PROCEEDINGS OF THE 24TH INTERNATIONAL CONFERENCE ON AUDITORY DISPLAY

SONIFICATION AS ADSR

ART | DESIGN | SCIENCE | RESEARCH

MICHIGAN TECHNOLOGICAL UNIVERSITY

JUNE 10-15, 2018

ISBN 0-9670904-5-8

[HTTP://ICAD2018.ICAD.ORG/](http://ICAD2018.ICAD.ORG/)

[HTTP://DOI.ORG/10.21785/ICAD2018.000](http://DOI.ORG/10.21785/ICAD2018.000)

TABLE OF CONTENTS

Preliminaries		6
Chair's Welcome		7
Conference Organizers		9
Reviewers		10
Sponsors		11
Keynote Speakers		12
Dr. Carryl Baldwin		12
Dr. Stefania Serafin		13
Program		14
Papers		18
EA = extended abstract, see extended abstract section for paper; IP = invited presentation, no accompanying paper		
Health Data		19
Andrea Aldana, Steffen Grautoff and Thomas Hermann	CardioSounds: A Portable System to Sonify ECG Rhythm Disturbances in Real-Time	20
Ryosuke Nakagawa, Hideki Kadone, Masaki Matsubara, Kenji Suzuki and Hiroko Terasawa	Realtime EMG Sonification System for Gait	EA
Daniel MacDonald, Thangam Natarajan, Richard Windeyer, Peter Coppin and David Steinman	Data-Driven Sonification of CFD Aneurysm Models	28
Designing Sonification		34
Jonathan Middleton, Jaakko Hakulinen, Katariina Tiitinen, Juho Hella, Tuuli Keskinen, Pertti Huuskonen, Juhani Linna, Markku Turunen, Roope Raisamo and Mounia Ziat	Sonification with Musical Characteristics: A Path Guided by User-Engagement	35
Takahiko Tsuchiya and Jason Freeman	A Study of Exploratory Analysis in Melodic Sonification with Structural and Durational Time Scales	42
Thomas Hermann	Wave Space Sonification	49
Kelly Snook, Tarik Barri, Joachim Goßmann, Jason Potts, Margaret Schedel and Hartmut Warm	Kepler Concordia: Designing an Immersive Modular Musical and Scientific Instrument Using Novel Blockchain and Sonification Technologies in XR	57
Joseph Ryan, Steven Landry, Jaclyn Barnes, Myoungsoon Jeon and Chung Hyuk Park	M.U.S.I.C. (Musical Utility Software for Interactive Creations): The Multipurpose Motion-Sound Mapping Platform	EA
Images, Graphs and Spaces		66
Robert Wheeler and David Worrall	Representing Twitter Users' Engagement by Sonification	EA

Shin'Ichiro Uno, Yasuo Suzuki, Takashi Watanabe, Miku Matsumoto and Yan Wang	Soundbased Image and Position Recognition System SIPReS	67
Niklas Rönnerberg and Jonas Löwgren	Photone: Exploring Modal Synergy in Photographic Images and Music	73
Braxton Boren and Andrea Genovese	Acoustics of Virtually Coupled Performance Spaces	80
Andrés Aparicio and Rodrigo Cádiz	Audiograph: Auditory Graphs for Google Sheets with Faust	EA
Sonification and Ambient Displays		87
Sara Lenzi and Francesca Gleria	Humanising Data through Sound: Res Extensae and a User-Centric Approach to Data Sonification	88
Marian Weger, Thomas Hermann and Robert Höldrich	Plausible Auditory Augmentation of Physical Interaction	97
Katharina Groß-Vogt, Marian Weger, Robert Höldrich, Thomas Hermann, Till Bovermann and Stefan Reichmann	Augmentation of an Institute's Kitchen: An Ambient Auditory Display of Electric Power Consumption	105
Patrick Cowden and Luke Dosiek	Auditory Displays of Electric Power Grids	113
Christopher Jette and James Buchholz	Fluor Sonescense: A Sonification of the Visualization of Brass Instrument Tones	121
Hyundai Motors Special Session		127
Lisa Frohmann, Marian Weger and Robert Höldrich	Recognizability and Perceived Urgency of Bicycle Bells	128
Tim Ziemer and Holger Schultheis	A Psychoacoustic Auditory Display for Navigation	136
Doon MacDonald and Tony Stockman	SoundTrAD, A Method and Tool for Prototyping Auditory Displays: Can We Apply It to an Autonomous Driving Scenario?	145
Justyna Maculewicz and Fredrik Hagman	Sonification Research and Trends at Volvo Cars	IP
Steven Landry and Myoungsoon Jeon	In-Vehicle Situation Awareness Using Sounds	IP
Maps, Location and Navigation		152
Ivica Bukvic and Gregory Earle	Reconsidering Human Capacity for Location-Aware Audio Pattern Recognition: A Case for Immersive Exocentric Sonification	153
Megen Brittell	Seeking a Reference Frame for Cartographic Sonification	160
Samuel Chabot, Wendy Lee, Rebecca Elder and Jonas Braasch	Using a Multimodal Immersive Environment to Investigate Perceptions in Augmented Virtual Reality Systems	168
Louis Commère, Florian Grond, Jean Rouat and François Côté	Sonification of 3D Point Clouds for Substitution of Vision by Audition for the Visually Impaired	EA
Brandon Biggs, Lena Yusim and Peter Coppin	The Audio Game Laboratory: Building Maps from Games	EA
AD Theory		176

David Worrall	Sonification: A Prehistory	177
Michael Quinton, Iain McGregor and David Benyon	Investigating Effective Methods of Designing Sonifications	183
Stuart Duncan Haffenden Cornejo	Towards Ecological and Embodied Design of Auditory Display	191
Kay Tislar, Zackery Duford, Brittany Nelson, Madeline Peabody and Myounghoon Jeon	Examining the Learnability of Auditory Displays: Music, Earcons, Spearcons, and Lyricons	197
Rehabilitation		203
Carolina Valli	Simpaty-Hand App, A Musical Instrument Designed for People with Limited Motor Skills	EA
Eric Vasey, Xingliang Li, Tejin Yoon and Myounghoon Jeon	A Preliminary Evaluation of a Smart Exercise Application with Wearable Sensors	EA
Ridwan Ahmed Khan, Myounghoon Jeon and Tejin Yoon	"Musical Exercise" for People with Visual Impairments: A Preliminary Study with the Blindfolded	204
Sonifying Science and Technology		212
Mark Ballora, Christopher Roman, Robert Pockalny and Karen Wishner	Sonification and Science Pedagogy: Preliminary Experiences and Assessments of Earth Science Data Presented in an Undergraduate General Education Course	213
Adrián García Riber	Planethesizer: Approaching Exoplanet Sonification	219
Michael Blandino	ToxSampler: Locative Sound Art Exploration of the Toxic Release Inventory	227
Robert Arbon, Alex Jones, Lars Bratholm, Tom Mitchell and David Glowacki	Sonifying Stochastic Walks on Biomolecular Energy Landscapes	232
Jaclyn Barnes, Erin Richie, Qiyue Lin, Myounghoon Jeon and Chung Hyuk Park	Emotive Voice Acceptance in Human-Robot Interaction	EA
Extended Abstracts		240
Ryosuke Nakagawa, Hideki Kadone, Masaki Matsubara, Kenji Suzuki and Hiroko Terasawa	Realtime EMG Sonification System for Gait	241
Joseph Ryan, Steven Landry, Jaclyn Barnes, Myounghoon Jeon and Chung Hyuk Park	M.U.S.I.C. (Musical Utility Software for Interactive Creations): The Multipurpose Motion-Sound Mapping Platform	245
Robert Wheeler and David Worrall	Representing Twitter Users' Engagement by Sonification	249
Andrés Aparicio and Rodrigo Cádiz	Audiograph: Auditory Graphs for Google Sheets with Faust	252
Louis Commère, Florian Grond, Jean Rouat and François Côté	Sonification of 3D Point Clouds for Substitution of Vision by Audition for the Visually Impaired	256

Brandon Biggs, Lena Yusim and Peter Coppin	The Audio Game Laboratory: Building Maps from Games	259
Carolina Valli	Simpaty-Hand App, A Musical Instrument Designed for People with Limited Motor Skills	264
Eric Vasey, Xingliang Li, Tejin Yoon and Myounghoon Jeon	A Preliminary Evaluation of a Smart Exercise Application with Wearable Sensors	267
Jaclyn Barnes, Erin Richie, Qiyue Lin, Myounghoon Jeon and Chung Hyuk Park	Emotive Voice Acceptance in Human-Robot Interaction	271
<i>Installations</i>		275
Niklas Rönnerberg and Jonas Löwgren	Photone	276
Jason Fick	Ceratonova Shasta	278
Michele Zaccagnini	Waveforms	280
Matteo Castiglioni	Offbeat News	281
<i>Concert</i>		282
Christopher Jette and Mark Broschinsky	Fluor Sonescence	283
Alessio Rossato	Temple of No Religion	285
Adrián García Riber	Planethesizer	288
Massimo Vito Avantaggiato	Atlas of Uncertainty	291
Christopher Plummer	Flow	IP
Daniel Grayvold	Snowfall	295

PRELIMINARIES

CHAIR'S WELCOME

It is our great pleasure to welcome you to the 24th International Conference on Auditory Display – ICAD 2018. The ICAD Conference is the unique venue that focuses on auditory displays and sonification showcasing the best research in the area, with participation from diverse communities of scholars, practitioners, artists, and designers, including computer science, acoustics, music, sound design, human-computer interaction, interaction design, medical and health care, engineering, psychology and social behavioral sciences.

The theme of this year's conference is "Sonification as ADSR". Auditory researchers are familiar with "ADSR" (Attack-Decay-Sustain-Release), which are basic components of sound. In ICAD 2018, we want to not only go back to the basics of sound but also explore and integrate various potentials of auditory displays and sonification as "ADSR" (Art-Design-Science-Research). Balancing the multifaceted aspects of sonification requires a multidisciplinary view, which the ICAD Conference offers. To support this intermingling of various areas, we solicited and reviewed all types of submissions, including papers, extended abstracts, concert pieces, demos and installations. Each submission was reviewed or curated by peers and an expert group of program committee members, who worked together with the program chairs to make the best program for each contribution type.

The program committee selected 31 full papers, 7 extended abstracts, 7 sonification concert pieces, and 4 installations from contributors worldwide, including Austria, Bangladesh, Canada, Chile, Finland, Germany, Italy, Japan, South Korea, Spain, Sweden, the UK and the US. The selected papers and conference proceedings are archived in Georgia Tech SMARTech repository.

This year, we have usual ICAD programs – single track paper presentation, poster presentation, sonification concert, installation, NSF-supported doctoral consortium (ThinkTank), workshops/tutorials, and jam session. In addition, we have the first presidents' panel discussion, the newly formatted diversity & inclusion initiative workshop, and the Hyundai Motors Company-supported "Futuristic In-Vehicle Auditory Interactions" session.

We also have the pleasure of presenting two inspiring keynote speakers from different backgrounds and expertise, who represent well the multidisciplinary nature of auditory display and sonification: Dr. Stefania Serafin, Professor in the Medialogy Section at Aalborg University in Copenhagen, and Dr. Carryl Baldwin, Associate Professor in the Human Factors and Applied Cognition Program at George Mason University in Virginia.

ICAD 2018 was made possible through the significant volunteer efforts of the organizing committee, program committee, student volunteers, reviewers and advisory committee. We are grateful for the keynote speakers, international reviewers, and financial sponsors for their support and participation. The conference is sponsored by Hyundai Motors Company, Michigan Tech Rozsa Center for Performing Arts, Michigan Tech Transportation Institute, Michigan Tech Institutional Equity and Inclusion, Center for Human-Centered Computing, Institute of Computing and

Cybersystems, Dept. of Visual and Performing Arts, Dept. of Cognitive and Learning Sciences and Dept. of Computer Science.

We are very excited to welcome the authors and performers of papers, extended abstracts, sonification concerts, and installations, as well as all other conference attendees. Your active participation and contribution make the conference a success each year and advance our community. We hope you will enjoy your time at ICAD2018 in Houghton, Michigan!

Myounghoon "Philart" Jeon
ICAD2018 General Chair
Michigan Tech, USA

CONFERENCE ORGANIZERS

General Chair Myounghoon Jeon (Philart)

Program Chairs Areti Andreopoulou

Jaclyn Barnes

Steven Landry

Workshop Chair Derek Brock

Music Chair Christopher Plummer

Installations Chair Josh Loar

Communications Chair Nicole Kelly

ThinkTank Chair Hiroko Terasawa

Accessibility Chair Katie Wolf

Local Chair Rachelle Gariepy

Advisory Committee Chair Matti Gröhn

Student Volunteer Chair Sangjin Ko

Advisory Board Mark Ballora

Bruce Walker

Tony Stockman

REVIEWERS

Areti Andreopoulou

Amit Barde

Jaclyn Barnes

Michael Barnhart

Megen Brittell

Derek Brock

Elliot Canfield-Dafilou

Samuel Chabot

Eugene Cherny

Teresa Connors

Allan Coop

Antonio D'Amato

Ziqi Fan

Thomas Gable

Nicolas Gold

Louise Harris

Laurie Heller

Thomas Hermann

Ridwan Ahmed Khan

Steven Landry

Marlene Mathew

Michael Nees

Arthur Pate

Camille Peres

Helen Petrie

Niklas Rönnerberg

Nina Schaffert

Brian Simpson

Kelly Snook

Takahiko Tsuchiya

Katharina Vogt

Richard C. Windeyer

Mike Winters

David Worrall

Jiajun Yang

Roberto Zanata

SPONSORS

Michigan Tech
Rozsa Center for
Performing Arts

Michigan Tech
Michigan Tech
Transportation
Institute (MTTI)

Michigan Tech
Visual and Performing Arts

Michigan Tech
Institutional Equity
and Inclusion

The Center for
Human-Centered
Computing
(HCC)

Michigan Tech
Cognitive and
Learning Sciences

KEYNOTE SPEAKERS

DR. CARRYL BALDWIN

Associate Professor, Human Factors and Applied Cognition Program, George Mason University

AUDITORY DISPLAYS TO FACILITATE ATTENTION MANAGEMENT IN HIGHLY AUTONOMOUS SYSTEMS

Two recent crashes in January 2018 involving Tesla's Model S, underscore the importance of the need for drivers to maintain awareness in semi-automated vehicles, even when the autopilot is engaged. Despite manufacturer's warnings and cautionary statements in owner's manuals, decades of research in vigilance indicates that this will be a challenge, if not impossible for most drivers. This talk will focus on our recent and on-going research developing novel methods of assisting the driver with

attention management in highly autonomous systems. Included in this discussion will be a discussion of the methods to develop and validate effective auditory collision avoidance alerts, driver state monitoring with low cost physiological sensors, and using specific types of music as a means assisting operators with maintaining sustained attention.

BIO

Dr. Baldwin teaches and conducts research in conjunction with the Human Factors and Applied Cognition program. Her primary research interests are in the area of Applied Auditory Cognition. She has over twenty years of experience investigating human factors issues in mental workload, surface and air transportation and cognitive aging. Much of her work involves the use of neurophysiological measures (i.e., EEG, ERP, EKG and eye tracking) to examine the effort expended when people perform multiple modality dual tasks as a function of changes in sensory or environmental condition or cognitive aspects of the task. Dr. Baldwin has an active line of externally funded research and is currently working in conjunction with sponsors such as the National Highway Traffic Safety Foundation, Federal Highway Administration, Northrup Grumman and Lockheed Martin. Previously she has successfully completed multiple projects for the National Institutes of Health, the National Science Foundation and both NASA Langley and NASA Ames.

DR. STEFANIA SERAFIN

Professor, Aalborg University Copenhagen

20 YEARS OF SONIC INTERACTIONS

Sonic interaction design is a fertile field of investigation at the intersection of sound and music computing and interaction design. For the past twenty years together with several collaborators, we have been designing sonic interactions for different applications, ranging from physics-based simulations of musical instruments and everyday objects, new interfaces for musical expression, cultural heritage, walking and rehabilitation, learning and training, as well as virtual and augmented

reality. In this talk I will present an overview of these activities reflecting on their impact and perspectives for the future.

BIO

Stefania Serafin is Professor with special responsibilities in sound in multimodal environments at Aalborg University in Copenhagen. She has been Associate Professor (2006-2012), and Assistant Professor (2003-2006) at Aalborg University Copenhagen. She received a Ph.D. in Computer Based Music Theory and Acoustics from Stanford University in 2004, and a Master in Acoustics, Computer Science and Signal Processing Applied to Music from IRCAM (Paris) in 1997. She has been visiting professor at the University of Virginia (2003), and visiting scholar at Stanford University (1999), Cambridge University (2002) and KTH Stockholm (2003). She is principal investigator for the EU funded project Natural Interactive Walking, and Danish delegate for the EU COST Action on Sonic Interaction Design. Her main research interests are sound models for interactive systems and multimodal interfaces and sonic interaction design.

PROGRAM

SUNDAY, JUNE 10TH

Time Place	Session
9:00 - 5:00 Forestry 144	Student ThinkTank

MONDAY, JUNE 11TH

Time Place	Session
9:00 - 9:30 Forestry Goo2	Welcoming Remarks/Introduction
9:30 - 10:30 Forestry Goo2	Health Data Chair: Matti Gröhn Andrea Aldana, Steffen Grautoff and Thomas Hermann, <i>CardioSounds: A Portable System to Sonify ECG Rhythm Disturbances in Real-Time</i> Ryosuke Nakagawa, Hideki Kadone, Masaki Matsubara, Kenji Suzuki and Hiroko Terasawa, <i>Realtime EMG Sonification System for Gait</i> Daniel MacDonald, Thangam Natarajan, Richard Windeyer, Peter Coppin and David Steinman, <i>Data-Driven Sonification of CFD Aneurysm Models</i>
10:30 - 11:00 Forestry Atrium	Break
11:00 - 12:00 Forestry Goo2	Keynote Dr. Stefania Serafin, <i>20 Years of Sonic Interactions</i>
12:00 - 1:30 Forestry Atrium	Lunch
1:30 - 3:00 Forestry Goo2	Designing Sonification Chair: Mark Ballora Jonathan Middleton, Jaakko Hakulinen, Katariina Tiitinen, Juho Hella, Tuuli Keskinen, Pertti Huuskonen, Juhani Linna, Markku Turunen, Roope Raisamo and Mounia Ziat, <i>Sonification with Musical Characteristics: A Path Guided by User-Engagement</i> Takahiko Tsuchiya and Jason Freeman, <i>A Study of Exploratory Analysis in Melodic Sonification with Structural and Durational Time Scales</i> Thomas Hermann, <i>Wave Space Sonification</i> Kelly Snook, Tarik Barri, Joachim Goßmann, Jason Potts, Margaret Schedel and Hartmut Warm, <i>Kepler Concordia: Designing an Immersive Modular Musical and Scientific Instrument Using Novel Blockchain and Sonification Technologies in XR</i> Joseph Ryan, Steven Landry, Jaclyn Barnes, Myounghoon Jeon and Chung Hyuk Park, <i>M.U.S.I.C. (Musical Utility Software for Interactive Creations): The Multipurpose Motion-Sound Mapping Platform</i>
3:00 - 3:30 Forestry Atrium	Break
3:30 - 5:00 Forestry Goo2	Images, Graphs and Spaces Chair: Derek Brock Robert Wheeler and David Worrall, <i>Representing Twitter Users' Engagement by Sonification</i> Shin'Ichiro Uno, Yasuo Suzuki, Takashi Watanabe, Miku Matsumoto and Yan Wang, <i>Soundbased Image and Position Recognition System SIPReS</i> Niklas Rönnberg and Jonas Löwgren, <i>Photone: Exploring Modal Synergy in Photographic Images and Music</i> Braxton Boren and Andrea Genovese, <i>Acoustics of Virtually Coupled Performance Spaces</i> Andrés Aparicio and Rodrigo Cádiz, <i>Audiograph: Auditory Graphs for Google Sheets with Faust</i>
6:00 - 8:00 Forestry Atrium	Welcome Reception

TUESDAY, JUNE 12TH

Time Place	Session
9:00 - 10:30 Forestry G002	Sonification and Ambient Displays Chair: Tim Ziemer Sara Lenzi and Francesca Gleria, <i>Humanising Data through Sound: Res Extensae and a User-Centric Approach to Data Sonification</i> Marian Weger, Thomas Hermann and Robert Höldrich, <i>Plausible Auditory Augmentation of Physical Interaction</i> Katharina Groß-Vogt, Marian Weger, Robert Höldrich, Thomas Hermann, Till Bovermann and Stefan Reichmann, <i>Augmentation of an Institute's Kitchen: An Ambient Auditory Display of Electric Power Consumption</i> Patrick Cowden and Luke Dosiek, <i>Auditory Displays of Electric Power Grids</i> Christopher Jette and James Buchholz, <i>Fluor Sonescense: A Sonification of the Visualization of Brass Instrument Tones</i>
10:30 - 11:00 Forestry Atrium	Break
11:00 - 12:00 Forestry G002	Keynote Dr. Carryl Baldwin, <i>Auditory Displays to Facilitate Attention Management in Highly Autonomous Systems</i>
12:00 - 1:30 Forestry Atrium	Lunch
1:30 - 3:00 Forestry G002	Hyundai Motors Special Session Chair: Michael Winters Lisa Frohmann, Marian Weger and Robert Höldrich, <i>Recognizability and Perceived Urgency of Bicycle Bells</i> Tim Ziemer and Holger Schultheis, <i>A Psychoacoustic Auditory Display for Navigation</i> Doon MacDonald and Tony Stockman, <i>SoundTrAD, A Method and Tool for Prototyping Auditory Displays: Can We Apply It to an Autonomous Driving Scenario?</i> Justyna Maculewicz and Fredrik Hagman, <i>Sonification Research and Trends at Volvo Cars</i> Steven Landry and Myoungsoon Jeon, <i>In-Vehicle Situation Awareness Using Sounds</i>
3:00 - 3:30 Forestry Atrium	Break
3:30 - 5:00 Forestry 144	Diversity and Inclusion Workshop
7:00 - 9:00 Franklin Square Inn	Jam Session

WEDNESDAY, JUNE 13TH

Time Place	Session
9:00 - 10:30 Various	Full Day Workshops Justyna Maculewicz, Fredrik Hagman and Myoungsoon Jeon, <i>Sonic Interaction in Intelligent Cars</i> (Forestry 108) Mike Winters and Prakriti Kaini with Bruce Walker, <i>Collaborative Design and Evaluation in Auditory Displays for Interactive Physics Simulations</i> (Forestry 144) Morning Tutorial Steven Landry, <i>Introduction to Pure Data: Beginner and Intermediate Techniques for Data Sonification</i> (Forestry 146)
10:30 - 11:00	Break

Forestry Atrium	
11:00 – 12:30 Various	Full Day Workshops and Morning Tutorial (cont.)
12:30 - 1:30 Forestry Atrium	Lunch
1:30 - 3:00 Various	Full Day Workshops (cont.) Afternoon Tutorial David Worrall, <i>Introduction to Data Sonification with Python and Csound</i> (Forestry 146)
3:00 - 3:30 Forestry Atrium	Break
3:30 - 5:00 Various	Full Day Workshops and Afternoon Tutorial (cont.)
6:00 - 7:00 Rosza Center	Sonification Concert

THURSDAY, JUNE 14TH

Time	Session
9:00 - 10:30 Forestry Goo2	Maps, Location and Navigation Chair: Thomas Hermann Ivica Bukvic and Gregory Earle, <i>Reconsidering Human Capacity for Location-Aware Audio Pattern Recognition: A Case for Immersive Exocentric Sonification</i> Megen Brittell, <i>Seeking a Reference Frame for Cartographic Sonification</i> Samuel Chabot, Wendy Lee, Rebecca Elder and Jonas Braasch, <i>Using a Multimodal Immersive Environment to Investigate Perceptions in Augmented Virtual Reality Systems</i> Louis Commère, Florian Grond, Jean Rouat and François Côté, <i>Sonification of 3D Point Clouds for Substitution of Vision by Audition for the Visually Impaired</i> Brandon Biggs, Lena Yusim and Peter Coppin, <i>The Audio Game Laboratory: Building Maps from Games</i>
10:30 - 11:00 Forestry Atrium	Break
11:00 - 12:30 Forestry Goo2	AD Theory Chair: Bruce Walker David Worrall, <i>Sonification: A Prehistory</i> Michael Quinton, Iain McGregor and David Benyon, <i>Investigating Effective Methods of Designing Sonifications</i> Stuart Duncan Haffenden Cornejo, <i>Towards Ecological, Embodied and User-Centric Design in Auditory Display</i> Kay Tislar, Zackery Duford, Brittany Nelson, Madeline Peabody and Myoungsoon Jeon, <i>Examining the Learnability of Auditory Displays: Music, Earcons, Spearcons, and Lyricons</i>
12:30 - 1:30 Forestry Atrium	Lunch
1:30 - 3:00 Forestry Goo2	Presidents' Panel: Jim Ballas, Matti Gröhn, Tony Stockman, Bruce Walker, and David Worrall
3:00 - 4:00 Forestry Atrium	Break/Posters
4:00 - 5:00 Forestry Goo2	Rehabilitation Chair: Hiroko Terasawa Carolina Valli, <i>Simpaty-Hand App, A Musical Instrument Designed for People with Limited Motor Skills</i> Eric Vasey, Xingliang Li, Tejin Yoon and Myoungsoon Jeon, <i>A Preliminary Evaluation of a Smart Exercise Application with Wearable Sensors</i> Ridwan Ahmed Khan, Myoungsoon Jeon and Tejin Yoon, <i>"Musical Exercise" for People with Visual Impairments: A Preliminary Study with the Blindfolded</i>
6:00 - 9:00	Banquet

FRIDAY, JUNE 15TH

Time	Session
9:00 - 10:30 Forestry G002	Sonifying Science and Technology Chair: Katie Wolf Mark Ballora, Christopher Roman, Robert Pockalny and Karen Wishner, <i>Sonification and Science Pedagogy: Preliminary Experiences and Assessments of Earth Science Data Presented in an Undergraduate General Education Course</i> Adrián García Riber, <i>Planethesizer: Approaching Exoplanet Sonification</i> Michael Blandino, <i>ToxSampler: Locative Sound Art Exploration of the Toxic Release Inventory</i> Robert Arbon, Alex Jones, Lars Bratholm, Tom Mitchell and David Glowacki, <i>Sonifying Stochastic Walks on Biomolecular Energy Landscapes</i> Jaclyn Barnes, Erin Richie, Qiyue Lin, Myounghoon Jeon and Chung Hyuk Park, <i>Emotive Voice Acceptance in Human-Robot Interaction</i>
10:30 - 11:00 Forestry Atrium	Break
11:00 - 11:30 Forestry G002	Open Mic
11:30 - 12:00 Forestry G002	Announcements

PAPERS

HEALTH DATA

CARDIOSOUNDS: A PORTABLE SYSTEM TO SONIFY ECG RHYTHM DISTURBANCES IN REAL-TIME

Andrea Lorena Aldana Blanco¹, Steffen Grautoff², Thomas Hermann¹

¹Ambient Intelligence Group, CITEC, Bielefeld University, Bielefeld, Germany

²Klinikum Herford, Emergency Department, Herford, Germany

aaldanablanco@techfak.uni-bielefeld.de

ABSTRACT

This paper is a continuation and extension of our previous work [1]. CardioSounds is a portable system that allows users to measure and sonify their electrocardiogram signal in real-time. The ECG signal is acquired using the hardware platform BITalino and subsequently analyzed and sonified using a Raspberry Pi. Users can control basic features from the system (start recording, stop recording) using their smartphone. The system is meant to be used for diagnostic and monitoring of cardiac pathologies, providing users with the possibility to monitor a signal without occupying their visual attention. In this paper, we introduce a novel method, *anticipatory mapping*, to sonify rhythm disturbances such as Atrial Fibrillation, Atrial flutter and Ventricular Fibrillation. Anticipatory mapping enhances perception of rhythmic details without disrupting the direct perception of the actual heart beat rhythm. We test the method on selected pathological data involving three of the most known rhythm disturbances. A preliminary perception test to assess aesthetics of the sonifications and its possible use in medical scenarios shows that the anticipatory mapping method is regarded as informative discerning healthy and pathological states, however there is no agreement about a preferred sonification type.

1. INTRODUCTION

Sonification of biosignals in medical scenarios is a common method to monitor and detect abnormalities in specific signals. A well-known example are pulse oximeters, medical devices to monitor the oxygen saturation levels, that produce a short sound in synchronization with the pulse rate. The starting frequency of the sound is defined by each company that manufactures the devices [2], however, when the oxygen saturation level decrease, regardless of the manufacturer, a sound of a lower frequency is triggered, calling for the immediate attention of clinicians.

In current times, besides the well-known sound of the pulse oximeter, sonification research focuses on an extended set of biosignals, among which is the electrocardiogram (ECG). One of the reasons for this is that the human hearing system has a great capability to recognize changes and patterns in the signal, therefore sonification appears to be a powerful tool for diagnosis and monitoring of multivariate time series such as the ECG [3] [4]. In fact, some ECG monitors already include a sonification that triggers a

short duration sound whenever there is a heartbeat, providing information about the heart rate. However, there are still many other parameters in the ECG signal that can be sonified to present a more complete view of the current state of the heart. Another reason for using sonification, is that monitoring through sounds allows users to focus their visual sense on a primary task while being aware of the changes in the monitored signal [5], and this is an interesting approach in situations where users need to perform a task while tracking changes in a signal.

Based on the previously explained observations, we have identified two cases in which ECG sonification could be applied: (i) *ECG auditory biofeedback when doing sports activities* and (ii) *auditory ECG monitoring to assist diagnosis during sport tests (e.g. Cardiac stress test)*. In the first case, sonification permits users to receive information about their heart rhythm when they are doing sport activities without blocking their visual sense. This is particularly important in activities where the users should focus their attention on the activity itself and avoid distractions from looking at a visual monitor (e.g. in jogging, biking). In the second case, clinicians can monitor the ECG signal from a patient while focusing their visual attention on other tasks such as examining the patient or looking at their clinical history.

In this paper, we first explain the ECG signal and particularly focus on the relevant background on arrhythmic patterns in ECG. We then introduce the *CardioSounds* system architecture which allows modular extension of other sonifications, recording of ECG data and interactive control of sonification parameters. We then introduce the *anticipatory mapping* method for sonifying arrhythmias as a concept and how to implement it in python/Supercollider. Subsequently we demonstrate the sonification at hand of selected healthy and pathologic ECG sonifications together with their clinical interpretation. Section 5 presents the results of the preliminary perception test where we evaluate the ability of anticipatory mapping to assist detection of changes in rhythmical organization. We therein also evaluate the participants' assessment of the aesthetic qualities and we consider compatibility with current clinical soundscapes by involving stakeholders from a German hospital. Discussion and conclusions summarize the paper.

2. ECG AND ARRHYTHMIAS

An electrocardiogram is a visual representation of the electrical activity of the heart measured as the time series of electric potential differences between two selected points on the skin (e.g. left arm and right foot). A standard clinical ECG usually has 12 leads using 10 electrodes, but details are not further relevant for the unfolding of the sonification presented here. Let's first understand a single

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

lead, e.g. Lead I (difference between left and right arm).

In each cardiac cycle a number of reference points can be spotted in the signal. Marking the beginning of the cycle, there is the P-wave, which represents regular atrial depolarization. Then, there is the so-called QRS complex representing ventricular activation; starting in the ventricular septum and finishing in the left ventricle. However, activation in the right and left ventricles happens almost simultaneously, for this reason the QRS complex is generally considered all together. Subsequently there is a time between depolarization and repolarization depicted by the ST segment and finally, there is the T wave that results from ventricular repolarization (see Fig. 1).

Figure 1: ECG standard reference points: P-wave, Q, R, S, J point, ST-segment, and T wave.

The reference points are the cues that clinicians use to determine if there are changes in specific parts of the heart, and if these changes correspond to a given pathology. For example, arrhythmias¹ can be detected by looking at the variations in the duration of the RR interval (see Fig. 1) and the ECG waves' changes.

There are several types of arrhythmias that can be the result of impulse formation and/or conduction problems [6][7]. Nevertheless, not all rhythm disturbances are considered pathological. For instance, sinus arrhythmia, is a variation of the RR intervals that can result from respiration activity [8]. On the contrary, there are disturbances that are considered serious medical conditions which could lead to other diseases (e.g. strokes) and life-threatening arrhythmias that require immediate medical attention. Among these there are three examples of important arrhythmias:

1. Atrial fibrillation: Next to the usually benign conditions (premature atrial or ventricular beats and sinus arrhythmia), atrial fibrillation (AFib) is the most common arrhythmia. AFib can lead to a higher prevalence of blood clotting in the heart which increases the risk of stroke especially if the patient is comorbid.² Therefore, most patients who suffer from atrial fibrillation have to take anticoagulants (blood thinner) on a regular basis for prevention. In this condition the atrium is not able to contract efficiently. This decreases the heart's ability to pump which can lead to quicker exhaustion of the patient. AFib is a disease which is often caused by fibrosis of the atrium, but can be due to other causes, too. In AFib there are multiple irregular electrical

impulses which lead to an irregular ventricular conduction. Instead of regular P-waves there are irregular fibrillatory (f) waves visible followed by irregular QRS complexes in AFib.

2. Atrial flutter: Atrial flutter (AFL) is also a common arrhythmia, but occurs less often than Afib. In AFL the rate of atrial beats is around 220 and 320 per minute. There is a delay of conducting these beats to the ventricles caused by the atrio-ventricular (AV) node. In AFL the AV node usually transfers the conduction in a regular pattern as in most cases it is 2:1 which is equivalent to blocking one after transferring one conduction. This leads to a ventricular rate of around 130 to 150 beats per minute. This rhythm disturbance also requires the intake of anticoagulants in most patients due to an increased risk of stroke. Additionally, it has an impact on the performance of the heart and the endurance of the patient.
3. Ventricular fibrillation: Ventricular fibrillation (VFib) is a life-threatening condition which requires immediate medical help for preventing sudden death. VFib is due to multiple chaotic electrical impulses in the ventricles which do not result in any regular contraction of the heart. In this condition the muscles of the heart are coming to a standstill and blood containing oxygen cannot circulate in the body. If VFib is present defibrillation (electric shock) is necessary without delay. This can immediately restore a regular heart rhythm such as a sinus rhythm. VFib can have multiple causes, but often occurs during the acute phase of myocardial infarction (MI)³. Due to lack of blood supply of the heart all kinds of rhythm disturbances can occur of which VFib is the most dangerous.

3. THE CARDIOSOUNDS SYSTEM

CardioSounds is meant to be used as a portable real-time ECG sonification system. Its structure can be decomposed into four modules: Input, processing, output and control, each with a hardware and software component. First, we explain the hardware components of the system:

(1) The input module uses a BITalino device⁴, a hardware platform that allows real-time data acquisition and wireless transmission of the recorded signals. (2) The processing module is done in a Raspberry Pi 3 Model B⁵, in which the signal is analyzed and sent to the audio engine. (3) The output module is a DIGITUS 7.1 USB Sound Box⁶ (portable sound card). Even though the Raspberry Pi has a headphone jack audio output, we use a small USB card in order to achieve better sound quality. Lastly, (4) the control module is a smartphone or tablet that allows the user to start and stop the signal acquisition and control basic settings such as the sonification type. A diagram of the system can be seen in Fig. 2.

On the following subsections, we explain how we process the signal in each module.

³Lack of oxygen supply in the heart due to a blocking of the coronaries, that in prolonged cases, could lead to death of the heart' tissue.

⁴<http://www.bitalino.com>

⁵<https://www.raspberrypi.org/products/raspberry-pi-3-model-b/>

⁶<https://www.digitus.info>

¹A rhythm disturbance in the heart

²The patient has one or more additional diseases

Figure 2: CardioSounds system architecture.

3.1. Input module: Acquiring the ECG signal

The input signal is recorded using BITalino. Currently we record one ECG lead (channel), however, it is also possible to extend to multivariate extensions by adding more ECG sensors to the system. The signal is sampled at 1000 Hz.

3.2. Processing module: R peaks detection and RR interval estimation

Once the signal is acquired, we remove unwanted information before extracting the features of our interest. First, we remove the DC component and then apply a low-pass filter with a *cutoff* frequency of 70 Hz. We do this in order to eliminate frequencies outside the accepted range for ECG diagnostics [9]. Additionally, we apply a band-stop filter to remove signal noise at 50 Hz. Then, we perform the R peaks detection by applying the method proposed by Worral et al. [10] within a time window of 200 ms. The reason for this window size is that the typical QRS duration for a healthy adult with a heart rate of 60 bpm is 100 ms. Once at least two consecutive R peaks are detected, we can estimate the RR interval Δt_{RR} . While filtering can be done in real-time at sample rate and thus only causes a latency of few milliseconds, the RR-peak detection necessarily introduces a latency of 100 ms in the creation of events to be used for sonification. However, as monitoring of ECG is not such a time-critical task, and rhythms can actually only be judged from a longer series of heart beats, the latency is irrelevant.

3.3. Output Module: Sonifying the signal

When we have extracted the ECG features of interest, we define which parameters will be mapped to the sound features. Then, this information is sent using Open Sound Control (OSC) messages to the audio engine developed in Supercollider. In the output module, users can select the sonification type and the playback volume.

The modular construction of the system allows us to use this module with offline data too and control the temporal resolution of the sonification. The advantage in this case is that clinicians can inspect the ECG data by playing the sonification at faster or slower speed. Therefore, it is possible to do a quick scan of longer data files, or focus on a specific part of the data by playing it slowly.

3.4. Control module: The CardioSounds app

In order to control basic features of the system such as start and stop the signal acquisition, we developed a mobile application along with *CardioSounds*. Using the app we provide a control that contributes to the portability of the system. Also, we want to take advantage of devices that people and clinicians already use in their everyday life in order to provide control of *CardioSounds* at lower costs.

The mobile app and the Raspberry Pi communicate through Open Sound Control (OSC) messages. Thus, the only condition is that they are in the same network. As the controls merely issue commands and parameter changes for sonification, latency is not a critical issue. Neither is bandwidth as currently only user actions such as start or stop are triggered. Later versions of *Cardiosounds* will enable to review ECG data on the smartphone, yet before data transmission can be implemented we need to make sure that necessary security measures and authentication is implemented. Hence so far, the visualization can only be done in a display connected to the Raspberry Pi.

Figure 3: CardioSounds Mobile App.

4. SONIFICATION DESIGN: ANTICIPATION MAPPING

The human auditory system is already highly capable of detecting rhythms and irregularities [11]. The Just Noticeable Difference (JND) for detecting time displacement in a metrical sequence is about 10 ms when the duration between sound events is shorter than 250 ms or around 5% when the time between events is longer

than this value [12]. Therein it might seem unnecessary to use any method at all (apart from playing sound events at the time of a heartbeat) and leave the assessment of rhythm to the user alone. So why use a method to enhance arrhythmia detection? First of all, if the interest is in subtle changes that fall below the JND, then rendering temporal changes more salient may help to assess them at greater accuracy. Secondly, in a situation of various sound sources competing for attention, perceivable yet small rhythmic variations may simply be overheard due to their limited saliency, particularly as habituation is at work to blend stationary sound streams more or less out. For these reasons, we here take a fresh look at ways to augment the standard pulse sonification but without interfering with the naturally audible rhythmic structure.

In essence the rhythm perception requires to be able to memorize a temporal pattern, to recognize parts of it in the stream of previously perceived events, and with increasing match between items of the memorized pattern and the heard pattern, an expectation for the timing of the subsequent events narrows down. If then the heard event doesn't agree with our expected beat onset, we witness a deviation which raises our attention as it contains more information (e.g. higher entropy) than the expected outcome. For the most simple type of rhythm a sequence of beats at equal time difference however, not much memory is required.

So, assuming that this anticipation and constant comparison of an anticipated onset with the observed onset happens already in the listener's auditory perception, would it be possible to connect to it and provide additional information about rhythmical deviations explicitly in the gap between the expected and the actual, whichever comes first?

This is exactly what we do by introducing the *Anticipation Mapping*, a parameter mapping sonification that adds a before-or-after shadow of the following event (here: heart beat) extending between the anticipated and actual time, whatever is first.

In case the anticipated beat comes first, we start a sound at that time to develop systematically in one attribute until the actual beat comes. If for instance, we choose pitch as the attribute, the resulting sound would up-chirp until the moment the sonification event of the real beat stops this before-shadowed component. In effect, we receive an explicit cue about the discrepancy in two variables: (i) an extra sound which temporally extends according to the difference between the anticipated and actual onset, and (ii) a pitch change which depends on the temporal difference. According to the choice of attributes and details of how to drive the parameter changes, we can fine tune the arrhythmia sonification to draw controlled attention to deviations larger than a defined threshold, for instance 5%.

In case that the actual beat comes before the anticipated beat, we could for example play a sound that extends precisely to the moment of the anticipated beat – in this case a shadow of the sound. This will also mark specifically where the beat should have been according to expectations gained from the sequence of past beats.

In the following we will detail out different strategies to assess the anticipation and three different alternatives how to realize the anticipation mapping using concrete synths and mappings.

4.1. Computation of the Anticipation signal

4.1.1. Constant Rate Anticipation

The most straightforward method to compute the anticipation signal (i.e. the landing point in time at which the next beat would be

expected) at hand of the observed previous data is by simply averaging their inter-beat intervals and adding the average to the last observed onset, in mathematical terms

$$t_{i+1} = t_i + \frac{1}{M} \sum_{k=0}^{M-1} |t_{i-k} - t_{i-k-1}| = t_i + \frac{t_i - t_{i-M}}{M} \quad (1)$$

This is a good estimate under the assumption of regular rhythm, i.e. that all heartbeat intervals should be equal. It would give poor anticipation potentials if the heart rhythm should contain some shuffling, i.e., every even heart beat is delayed from its regular rhythm. It also performs suboptimal if there is an acceleration or slowing down over time, which naturally occurs in human heart rhythms depending on cardiac stress. This motivates to consider Changing Tempo Anticipation which is introduced next.

4.1.2. Changing Tempo Anticipation

In this model, we compute the anticipation signal from applying a linear regression to the sequence of the past M inter-heartbeat intervals, i.e., to the data set $\{k, t_{i+k}\}_{k=-M, \dots, -1}$, fitting parameters a, b of our model $h(k) = a \cdot k + b$. The expected time is then simply

$$t_{i+1} = t_i + h(0).$$

Of course, more complex models are possible that consider rhythmical fine structure within the past beats, however these are not further considered at this stage of investigation. Figure 4 shows how the two forecast methods perform when $M = 20$.

4.2. Sonification designs

We propose three sonification designs based on the anticipation signal computation: (i) *Polarity-Pitch*, (ii) *ECGrains* and (iii) *Marimba*. All methods intend to magnify time changes between actual and expected onsets by mapping the anticipation signal to a set of auditory parameters.

4.2.1. Polarity-Pitch Sonification

The *Polarity-Pitch* sonification intends to magnify the time difference between the actual and expected R peaks, specifically, when this difference falls below the just noticeable difference. In order to do so, when an R peak is found, we compare the time difference to the expected onset (forecast), and trigger a sound with a duration equal to the previously mentioned time difference. By doing this, we have already an idea of how arrhythmic the signal is, however, despite of our high capacity to discern rhythm variations, our auditory system could miss subtle variations and clues about how the RR interval is increasing or decreasing. For this reason, we also take into account the order in which the actual and expected onset take place, and map this parameter to pitch of the sound event. If the actual onset happens after the expected onset, we trigger a sound of a given frequency, and for the opposite scenario, we trigger a sound an octave lower. As a result, a high-pitched sound corresponds to an increase of the heart rate, and a low frequency sound to a decrease. This basically reduces the details to a binary information $\text{sign}(t_{\text{actual}} - t_{\text{expected}})$. A more continuous variant could be called *Chirped Pitch* where the pitch deviation is set proportional to $t_{\text{actual}} - t_{\text{expected}}$, so that the size of pitch deviations becomes a continuous cue that informs about the amount of temporal mismatch.

Figure 4: Computation of the anticipation signal on surrogate data: (i) Real RR interval, (ii) Expected RR interval using averaging, (iii) Expected RR interval using linear regression.

Sonification example S1⁷ illustrates the Polarity-Pitch sonification for a dataset classified as healthy, while S2, S3 and S4, present the sonification for datasets containing Atrial fibrillation, Atrial flutter and Ventricular Fibrillation respectively.

4.2.2. *ECGrains sonification*

In noisy environments, such as an emergency room, perceived changes in duration or pitch from the previous method could not be enough to detect subtle variations. Also, there are other ECG features that can be taken into account when sonifying arrhythmias. For instance, as explained in section 2, pathologies such as atrial flutter and atrial fibrillation present a change in the morphology of the P-wave. In positive ECG leads (current flowing towards the positive electrode), atrial flutter presents a 2:1, 3:1 or 4:1 ventricular rate, meaning that during each heart beat there could be two, three or up to four P-waves present per QRS complex. Also, in atrial fibrillation the P-wave is absent. Therefore, in this sonification we look into other mappings that contribute to make arrhythmia features more salient and besides using the anticipated signal, we also consider if the P-wave is present in the signal and which is its duration.

In the *ECGrains* sonification we use a granular synthesis generator. Granular synthesis is based on using very short fragments of sound (called grains) that can be controlled and modified to form a new output. We can control parameters such as the number of fragments to be triggered, the duration of each fragment, among others. In our sonification the grains are sine waves, however any sound could be fragmented in the same way to provide a wide palette of synthesized sounds.

We propose a parameter-mapping sonification where we trigger a granulated sound whenever an R peak is found. The number of small fragments (grains) that compose the synthesized sound is determined based on the difference between the expected peak and the actual peak (anticipation signal). As a result, the lower the difference between actual and expected R peaks, the lower the

number of grains (1 grain is the minimum amount to be triggered with each R peak). This means that a healthy ECG signal, with no abrupt variations between actual versus expected onset, would produce a granulated signal containing only one grain whenever an R peak is found. However, a pathological signal would create a completely different rhythmical pattern because the resulting sound would contain more grains. The duration of the granulated sound event in each R peak, is bound to the duration of the last RR interval, regardless of the number of grains in it.

In order to estimate the presence and duration of the P-wave, we calculate the average amplitude in each heart beat and determine how many data points are above a given threshold during the part where the P-wave is expected to appear. We take as a reference Fig. 1 where in a healthy dataset we expect a positive P-wave marking the beginning of the ECG cycle.

The expected location and duration of the P-wave is predicted taking into consideration the typical ECG waves' duration for a heart rate of 60 Beats Per Minute (bpm)⁸ [9], however, in our calculations, we estimate these values based on the current heart rate $HR = \frac{60}{\Delta t_{RR}}$. Thus, we count how many data points exceed a given threshold within the previously defined "P-wave search" area, and use the resulting number as an ECG feature which we map to the frequency value of the resulting sound.

In the mapping, we set the range (min, max) of the source value (P-wave duration) to (0.0, expected duration of the P-wave according to heart rate) and the output (min, max) is set to a pitch corresponding to MIDI notes ranging from (80, 62). This means that an absent or uncommon P-wave morphology, in which the number of data points above the P-wave threshold is low or inexistent, will produce a high pitched sound. On the contrary, a healthy dataset produces a low pitched sound because the number of data points exceeding the threshold is large.

The motivation for selecting low frequencies values to represent healthy states and high frequencies for pathological states, is that in healthy conditions we want the resulting sound to be simi-

⁷See section 7 for supplementary material.

⁸P-wave width normal value = 100 ms (60 bpm)

lar to the heart heard through a stethoscope, however, if anything abnormal happens in the signal, we want to increase the sound frequency to call attention from the listener.

Sonification example S5 corresponds to the *ECGrains* sonification of a healthy dataset, while S6, S7, and S8 correspond to datasets with Atrial Fibrillation, Atrial Flutter and Ventricular Fibrillation respectively.

4.2.3. Marimba sonification

The *Marimba sonification* is based on the *ECGrains sonification*, however, instead of using sine waves as the grains in the synthesizer, we use audio samples of the musical instrument known as marimba. The motivation to do this is that at present, the sonifications used in clinical environments are mainly based on sine oscillators, but at the same time, clinicians often report that these sounds produce auditory fatigue. For this reason, we want to explore how other sounds are perceived in the same environment and if they could be used for long hours of monitoring a signal. In particular, we chose the marimba because it is a percussive sound with an envelope similar to the sound produced by the heart when heard through a stethoscope.

For this sonification, the number of small grains that compose the synthesized sound depend on the time difference between the expected peak and the actual peak that we calculate using the anticipation signal method. As a result, a dataset where there are not important rhythm disturbances, will produce a sonification that triggers one marimba sound when an R peak is detected. On the contrary, if the difference between actual and expected R peaks is higher, a larger number of grains will be triggered in each heart-beat. The frequency of the granulated sound in each heart cycle is set to either a C note or an E note. These two notes are alternated throughout the sonification.

Sonification example S9 corresponds to the *Marimba* sonification of a healthy dataset, while S10, S11, and S12 correspond to datasets with Atrial Fibrillation, Atrial Flutter and Ventricular Fibrillation respectively.

5. PRELIMINARY PERCEPTION TEST

We conducted a preliminary perception test involving six clinicians (one experienced cardiologist, one cardiologist in training, 1 anesthetist, 2 emergency physicians, and one nurse of intensive care medicine). The participants were asked to rate on a 7-point Likert scale, where 1 means *Not at all* and 7 means *Very much*, their agreement to five statements. The evaluated features were: (a) if the sonification was pleasant to listen to, (b) the suitability of the sonification to be heard for a longer period of time during a cardiac procedure or when working in the operation theater, (c) suitability of the sonification for everyday use in the medical emergency scenario: listening to it in the emergency room or emergency vehicle, (d) if the patients would not be annoyed, frightened or intimidated by the sound (e) if the sonification is informative (it gives an idea about the nature of the ECG data between healthy and arrhythmic states). At the end of the test the participants could add any comments they considered pertinent.

Fig. 5, shows the results of the preliminary perceptions test. Results are reported as ($\mu \pm \sigma$) in a scale from 1 to 7 (1 refers to *Not at all* and 7 to *Very much*). Pleasantness of the sonification, evaluated in question A, was rated 3.66 ± 1.36 for the *Polarity-Pitch sonification*, 4.83 ± 1.72 for the *ECGrains sonification* and

Figure 5: Results of the preliminary perception test (Means and Standard Error bars).

4.66 ± 1.03 for the *Marimba sonification*. Question 5, which evaluated if the information was informative was rated as 5.66 ± 1.86 for the *Polarity-Pitch sonification*, 5.5 ± 1.87 for the *ECGrains sonification* and 5.5 ± 1.37 for the *Marimba sonification*.

According to the results, the participants found the sonifications to be informative, which means that the sonifications could help to assess rhythm disturbances in the ECG. However, there is no agreement regarding the most appropriate sonification type. In all the aspects consulted there is a large variance, even though all participants are related to the medical field, their interaction with ECG data occurs in different environments (e.g., doctor's office, emergency room, operation theater) and therefore they could have diverse expectations about the sonifications. Nevertheless, regarding the sonification designs, the *Marimba sonification* presents the lowest variance across all questions, indicating that the participants agreed more about the rating of this sound.

Not all the participants included comments at the end of the test, but one of them mentioned that with the *Polarity-Pitch sonification* and the *Marimba sonification* it was easier to listen to the ECG information than with the *ECGrains sonification*. Another user commented that the sonifications draw too much attention, which could be interpreted as distracting.

After the preliminary perception test, we plan to conduct a study that evaluates ECG classification accuracy (e.g., healthy and pathological states) and assesses benefits in reducing just noticeable differences (JND) for detecting rhythm disturbances when compared under three feedback conditions (visual, auditory and visual+auditory).

6. DISCUSSION AND CONCLUSIONS

This work presents a system to acquire, process and sonify ECG signals in real-time. The system is aimed to assist users to monitor their heart without distractions from a visual display. In this paper we focused on the diagnostic and monitoring of arrhythmias, however the system and the sonification methods can be extended to other cardiac pathologies.

CardioSounds can be used in clinical environments to assist clinicians to monitor their patients' biosignals while focusing on a primary task such as performing a cardiac procedure, examination of the patient or revising their medical history. It can also be used by people interested in monitoring their heart while do-

ing sports, especially when there is evidence of previous abnormal heart activity. For instance, persons who suffered a myocardial infarction could present post MI arrhythmias that depending on their characteristics could become life-threatening rhythm disturbances. For this reason, it is important to monitor the appearance of these pathologies.

The current system setup supports the recording and sonification of one ECG lead. However, it is possible to have multivariate extensions by adding more ECG modules with their respective sensors to the BITalino platform. We consider important to include the recording of at least another lead to have a better overview of the heart's state. Also, given that the analysis process runs on a Raspberry Pi device, for the moment, visualization of the data relies on an external display, yet we plan to include this option as part of the *CardioSounds* app. So far the mobile application includes basic control over the system but it is in our plans to increase the number of commands that the user can use to interact with it. For example, graphical elements for controlling the sonification's parameters, playback rate, among others, are yet to be added to the application.

Regarding the sonifications, we propose three sonification designs that intend to magnify even subtle changes in consecutive RR intervals. The proposed methods are a continuation of our previously presented sonification methods [1] to magnify subtle differences in rhythm. In our current sonification design, we account for a wider range of parameters in the ECG signal and map them to a broader set of auditory parameters. The sonifications presented are based on the *Anticipatory Mapping* method where we compare the time difference between the actual onset of the detected R peak to the expected onset. We calculate the expected onset using forecasting methods such as averaging and linear regression. In this paper we present a basic approach to sonify changes in the P-wave morphology, however, future sonifications should rely on a better estimation method for the P-wave duration.

The preliminary qualitative perception test showed that the sonifications are considered informative when it comes to detecting rhythm disturbances in the ECG signal, however the opinions about the preferred sonification type or the suitability for everyday use in the medical context differ greatly from subject to subject. Given that all the participants work with ECG data in different scenarios (e.g., doctor's office, emergency room), it could be that the expectations they have about how the sonifications should sound or the information they should convey, differs. Nonetheless, in the upcoming study we need to include a larger sample size in order to reduce variance.

In the next stage of development of *CardioSounds*, we plan to extend the capabilities of the mobile application so that users can store and retrieve ECG recordings. This would allow users to modify the temporal resolution of the retrieved sonifications to listen to them at a faster or slower pace in order to do a quick scan of the data or focus on a specific part of it. We also intend to conduct a study where clinicians and non-clinician listeners can evaluate if the sonification contributes to the detection of heart rhythm disturbances when compared to traditional visual monitoring methods.

7. RESOURCES

Examples of the sonifications are provided in:
<http://dx.doi.org/10.4119/unibi/2919694>

8. ACKNOWLEDGMENT

This work has been supported by the German Academic Research Service (DAAD) and the Cluster of Excellence Cognitive Interaction Technology "CITEC" (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

9. REFERENCES

- [1] A. L. A. Blanco, S. Grautoff, and T. Hermann, "Cardiosounds: Real-time auditory assistance for supporting cardiac diagnostic and monitoring," in *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*, ser. AM '17. New York, NY, USA: ACM, 2017, pp. 45:1–45:4. [Online]. Available: <http://doi.acm.org/10.1145/3123514.3123542>
- [2] D. Chandra, M. J. Tessler, and J. Usher, "Audio spectrum and sound pressure levels vary between pulse oximeters," *Canadian Journal of Anesthesia*, vol. 53, no. 1, pp. 26–32, Jan 2006. [Online]. Available: <https://doi.org/10.1007/BF03021524>
- [3] H. Terasawa, Y. Morimoto, M. Matsubara, A. Sato, M. Ohara, and M. Kawarasaki, "Guiding auditory attention toward the subtle components in electrocardiography sonification," in *Proc. ICAD*. Georgia Institute of Technology, 2015.
- [4] J. N. Kather, T. Hermann, Y. Bukschat, T. Kramer, L. R. Schad, and F. G. Zillner, "Polyphonic sonification of electrocardiography signals for diagnosis of cardiac pathologies," *Scientific Reports*, vol. 7, p. 44549, 2017.
- [5] P. Vickers, "Sonification for process monitoring," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany: Logos Publishing House, 2011, ch. 18, pp. 455–491. [Online]. Available: <http://sonification.de/handbook/chapters/chapter18/>
- [6] F. Foxall and H. Blackburn, *Cardiac Arrhythmia Recognition: An Easy Learning Guide.*, ser. Easy Learning Guides From M and K. M and K Update Ltd, 2010. [Online]. Available: <http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=400965&site=ehost-live>
- [7] M. G. Tsipouras, D. I. Fotiadis, and D. Sideris, "Arrhythmia classification using the rr-interval duration signal," in *Computers in Cardiology*, Sept 2002, pp. 485–488.
- [8] J. Hirsch and B. Bishop, "Respiratory sinus arrhythmia in humans: How breathing pattern modulates heart rate," vol. 241, pp. H620–9, 11 1981.
- [9] G. D. Clifford, F. Azuaje, and P. Mcsharry, "Ecg statistics, noise, artifacts, and missing data," *Advanced Methods and Tools for ECG Data Analysis*, vol. 6, p. 18, 2006.
- [10] D. Worrall, B. Thoshkahna, and N. Degara, "Detecting components of an ecg signal for sonification," in *Proc. ICAD*. Georgia Institute of Technology, 2014.
- [11] B. Moore, *Psychoacoustics*. New York, NY: Springer New York, 2007, pp. 459–501. [Online]. Available: https://doi.org/10.1007/978-0-387-30425-0_13
- [12] A. Friberg and J. Sundberg, "Perception of justnoticeable time displacement of a tone presented in a metrical sequence

at different tempos,” *The Journal of the Acoustical Society of America*, vol. 94, no. 3, pp. 1859–1859, 1993. [Online]. Available: <https://doi.org/10.1121/1.407650>

DATA-DRIVEN SONIFICATION OF CFD ANEURYSM MODELS

Daniel E. MacDonald
University of Toronto
Biomedical Simulation Laboratory
Toronto, Canada
demacdo@mie.utoronto.ca

Thangam Natarajan
University of Toronto
Biomedical Simulation Laboratory
Toronto, Canada
thangam@mie.utoronto.ca

Richard C. Windeyer
University of Toronto
Centre for Drama, Theatre,
and Performance Studies
Toronto, Canada
rwindeyer@me.com

Peter Coppin
OCAD University
Perceptual Artifacts Lab
Toronto, Canada
pcoppin@faculty.ocadu.ca

David A. Steinman
University of Toronto
Biomedical Simulation Laboratory
Toronto, Canada
steinman@mie.utoronto.ca

ABSTRACT

A novel method is presented for inspecting and characterizing turbulent-like hemodynamic structures in intracranial cerebral aneurysms by sonification of data generated using Computational Fluid Dynamics (CFD). The intention of the current research is to intuitively communicate flow complexity by augmenting conventional flow visualizations with data-driven sound, thereby increasing the ease of interpretation of dense spatiotemporal data through multimodal presentation. The described implementation allows the user to listen to flow fluctuations thought to indicate turbulent-like blood flow patterns that are often visually difficult to discriminate in conventional flow visualizations.

1. INTRODUCTION

1.1. Aneurysms and sounds

A cerebral aneurysm is the localized bulging of a weakened artery wall within the brain. Often, no symptoms of the aneurysm are present in unruptured cerebral aneurysms. An estimated 2-5% of adults harbour an aneurysm, of which the risk of rupture per year is approximately 1%. Of ruptured aneurysms, 40% are fatal [1]. Rupture occurs when the stresses exerted on the artery wall exceed the strength of the wall.

Blood flow patterns (hemodynamics) are thought to contribute to the initiation, growth, and rupture of intracranial aneurysms [2], by stimulating degenerative remodelling (weakening) of the artery wall. Because the forces associated with these flows are difficult to measure non-invasively using current technologies, medical image-based CFD has been used to investigate these hemodynamic forces retroactively [2, 3, 4]. Using digitized patient-specific vasculature models extracted using 3D angiography techniques, blood flow is simulated using advanced numerical methods. The results of such simulations are typically visualized as part of the process of data analysis.

These previous investigations have revealed turbulent-like flow instabilities within aneurysms or the vessels that host them,

with implications for aneurysm rupture and initiation [2, 3]. Such instabilities are typically represented through conventional visualization paradigms (glyphs, streamlines, isosurfaces), and though these can provide tangible and intuitive representations of the hemodynamics, visual differences among unstable laminar, transitional, and turbulent flow regimes in complex vascular geometries are often indiscernible. In this context, a novel method originally inspired by Doppler ultrasound auscultation is presented for communicating the nature of the flow by augmenting conventional flow visualizations with data-driven sound (sonification).

Auscultation refers to the established medical technique of listening to the sounds of the body for evaluation purposes [5]. For example, clinicians are trained to identify turbulent blood flows by recognizing auditory properties of duration, quality, and timing during examination of bruits in the carotid artery [5]. Doppler auscultation has been used in preliminary stages to listen to valvular movements and blood flow velocity sounds during cardiac examination, resolving sounds undetected by stethoscope [6], though to our knowledge, this technique has not been demonstrated on intracranial aneurysms.

Intracranial saccular aneurysms have been observed to emit acoustic signals (bruits) as a result of fluctuating hemodynamic flows thought to be characterized as turbulent-like [7, 8, 9]. These vibrations are difficult to measure in vivo due to difficulty differentiating these bruits from other bodily sounds, and as such are not a reliable source of diagnostic information [8].

Inspired by these techniques, we present a method for sonifying numerically simulated flow of intracranial aneurysms for the purposes of identifying hemodynamic flow complexity during the process of flow visualization.

1.2. Conventional visualization of CFD data

Standard visualization techniques for representing multidimensional CFD datasets include static or animated views of isosurface rendering, glyphs, or pathline and streamline representations. Fig. 1 shows these conventional representations applied to our intracranial aneurysm CFD data set.

Particle pathline animations from CFD data provides a tangible representation of fluid flow and are often in good agreement with our intuitive understanding of structure and movement. Due to the large number of particles required to represent the spatiotemporal flow, the visual difference between laminar, transitional, and turbulent flow is difficult to discern - all may look chaotic, but each

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

This research was funded by the Canadian Natural Sciences and Engineering Research Council, the Social Sciences and Humanities Research Council, and the Heart & Stroke Foundation.

Figure 1: Examples of conventional flow visualizations created using ParaView. On the left, streamlines of velocity are augmented by glyphs, with coloured slice indicating internal flow velocity of the aneurysm sac. On the right, surfaces of Q-criterion indicate regions of unstable flow.

behaves in accordance with differing principles, and may have distinct consequences for artery wall remodelling.

Isosurfaces and thresholds of computed quantities, such as velocity, provide an uncluttered representation resembling discrete visual blobs, though such blobs may obscure or occlude regions of data. These visualizations may be more visually abstract than pathlines and are subject to further sculpting by the user. Isosurfaces of Q-criterion, a representation of vortex cores, are particularly useful for assessing location and intensity of turbulent-like flow [10]. There is nothing inherently visual about our CFD derived data - seed particles, surfaces, thresholds, and colours are ascribed by convention and experience. Conceptually, temporal fluctuations are also suitable for listening, and as mentioned previously, aneurysms are known to emit bruits as a result of velocity and pressure instabilities. From results of CFD, velocity or pressure measurements for each and every spatial location can be easily written as a WAV file for playback, however this method lacks control and ease of further manipulation.

Visually communicating spatiotemporal instabilities ignores the existing strengths of the human auditory system for resolving frequencies as compared to visual perception of spatial frequencies. For example, attempting to visually distinguish a difference of pitch between vibrating strings ignores the auditory ability to distinguish subtle pitch differences, of which Moore indicates can be as small as 0.2% [11].

This lack of visual acuity is not only a problem of aneurysm visualization, but flow visualization in general. By Nyquist's theorem, in order to produce a faithful digital reproduction of any signal, the sample rate of the digital medium (audio or video file) must be at least twice that of the highest frequency component of interest contained within the original signal. For example, given a video frame rate of 60 Hz, a fluctuating signal of at most 30 Hz can be reproduced. As the frequency of the fluctuating signal increases, a higher video frame rate would be required. Given the frequency range of interest for bruits observed in intracranial aneurysms (200-800 Hz) [8], uncommonly high frame rates (400-1600 Hz) or greatly slowed video would be required to communicate such instabilities, without even considering whether we may visually perceive these fluctuations.

1.3. Sonification of CFD Data

Sonification techniques can be broadly separated into five categories: audification, auditory icons, earcons, parameter mapping, and model-based sonification. Of these techniques, parameter mapping (the mapping of data dimensions to auditory dimensions)

is the most common technique [12], allowing continuous communication of information and greater flexibility of design compared to audification, auditory icons, and earcons. Mapping relationships must be considered very carefully in order to generate an informative sonification.

The generation of a parameter mapping sonification representation consists of three stages: choice of the mapping strategy (i.e. which data dimension maps to which auditory dimension), choice of polarity, and psychophysical scaling [13]. Within the choice of mapping strategy, the choice of data scale on which to operate (globally, or within local subsections) must also be considered with great care.

Several works have addressed the sonification of spatiotemporal flow data. Klein and Staadt [14] describe an implementation using wind sounds to indicate the velocity and vorticity at a local point within a three dimensional rectilinear vector field. By sampling a local volume representing the users headspace, vector properties manipulate a static sound sample using head-related transfer functions (HRTFs) to convey vector direction, magnitude, and vorticity to the user. Stereo speaker setups as well as headphones with head tracking systems were tested. The authors emphasize the need for an appropriately sized sample volume (headspace) for evaluating vorticity; too small, and the vectors are too similar, yielding low vorticity, while too high, and the sonification will always sound cacophonous.

Kasakevich et al [15] describe a multi-modal interface for use in a computational steering environment to augment 3D visualization using sonification for the purposes of interpreting velocity values and localizing vortex cores. Conceptually, it is similar to the work of Klein and Staadt, though this implementation is performed on an unstructured grid with a variable listening radius. Importantly, a usability analysis was presented along with the implementation. Multi-modal feedback was shown to improve localization of vortex centers when compared to visual-only feedback. Audio-visual feedback provided the greatest performance, followed by visual-only, with poor performance using audio-only feedback. Participants took less time and used shorter path lengths when using audio-visual feedback, though accuracy was unaffected. In evaluating polarity of auditory feedback in the audio-visual tests, accuracy improved for positive amplitude conditions, though path length remained unaffected.

Menelas [16] provides a technique for interactively assessing open-cavity flow phenomena using virtual reality (VR) techniques. With an emphasis on visual feedback, Menelas attempts to address the complexity related to the analysis of CFD data using visual, haptic, and audio feedback. User feedback was provided from four participants, two users versed in CFD and two computer scientists, though no feedback is relevant to the auditory domain.

Although the current research began independently of the work of Klein and Staadt, Kasakevich, and Menelas [14, 15, 16], it has converged upon a similar solution in which high-density data sonification is navigated within a three-dimensional auditory field with a focus on the communication of spatiotemporal data. The design of the current work benefited from Klein and Staadt's observations regarding sample volume sizing; a balance must be struck between providing a granular representation of a single point or a cacophonous representation of a large volume.

The motivation for the current work was to address the shortcomings of data visualization with respect to temporal fluctuations, whereas previous works sought to address the visual clutter associated with high-dimensional datasets. The works of Klein and

Stadt, Kasakevich, and Menelas directly communicate time domain values, such as velocity of flow, in order to reduce the strain on visual bandwidth by mapping magnitude of vectors directly to amplitude or filter cutoff [14, 15, 16]. For those previous works, this is suitable, allowing the user to directly query the data and receive auditory feedback proportional to the measured data. Rather than communicating vector magnitude, the current research seeks to sonify the temporal evolution of the vector, particularly the complexity of instability through time. For example, a region of high flow velocity that is stable in time may indicate a region of low complexity, while a slow-moving but unsteady flow may indicate a region of high complexity.

The current work seeks to decompose the time domain values (velocity) in the frequency domain in order to discriminate regions of unstable flow. The current method allows inspection of individual frequency bands independently. Although point vector values remain the basis for our current work, the frequency domain processing between these vector values and the audio output distinguishes the current research from previous works. By sonifying the spectrogram, the envelope shape (percussive or smooth nature) is communicated along with spectral content.

Additionally, the flow is pulsatile, indicating that greater consideration should be given to the evolution of the signal through each period, i.e., cardiac cycle. The temporal resolution of the data is therefore considered very important in order to perform meaningful manipulations in the frequency domain, and as such is finer than previous works. In combination with use of a fine, unstructured mesh, our datasets tend to be large and unwieldy, with number of points on the order of hundreds of thousands or millions, and temporal resolutions on the order of milliseconds.

2. METHOD

High-fidelity CFD data sets produced using cases from the open-source Aneurisk database [17] serve as the basis of this methodology. Within an interactive visualization environment (ParaView - an open-source scientific visualization tool [18]), the average power spectral density (PSD) of velocity-time traces is computed for a probed sample volume, defined by centre point and radius, in order to evaluate a local region of flow within the aneurysm sac. The probe radius is chosen based on two considerations: first, the real-time computation must support interactivity, and second, the probe should not be too spatially localized or too global. To define this balance of local and global volume, the probe diameter should not exceed the neck diameter of the aneurysm and should be small enough to separate coherent Q-criterion structures. The frequency content and time-varying power of the resulting spectrogram are communicated to a real-time audio server (SuperCollider sound server [19]) to control the frequency and loudness of a bank of audio-rate sound generators. The summation of these generators results in a harmonically rich signal representative of the probed location.

This on-the-fly implementation is contained (though not constrained) within the Python shell and intended to run alongside a synchronized animation. Sound networking protocol Open Sound Control (OSC) is used to communicate the spectrogram data between the live ParaView environment and audio server. The audio presentation of the signal is highly configurable while maintaining audio-rate synthesis, allowing the user to pitch, slow down, or threshold by frequency band in real-time, synchronized to the visualization.

Figure 2: Overview of the real-time sonification using ParaView and SuperCollider.

2.1. Mapping strategy, polarity, and psychological scaling

Of the various sonification techniques, the flexibility and continuous nature of parameter mapping was most influential in design. The structure of the instrument was designed such that any aneurysm dataset (or any collection of spatiotemporal data) may be used as input with little manipulation (only requiring a transpose of velocity vectors). As the user probes the data, varying audio outputs are generated local to the probe. As such, different frequency content and amplitude envelopes are readily generated for each new sample of the dataset.

A guiding philosophy for sonic aesthetics was to caricature the existing physiological phenomena by generating sound inspired by fluid sounds existing in nature, and not necessarily to recreate the full acoustic field tied to the flow. By taking advantage of our intuitive understanding of how certain sounds (wind, bubbles) map to velocity of fluid, the cognitive load is reduced. The basis of our parameter mapping was generated as simply as possible: the spectral content of velocity is mapped with auditory frequency values, while magnitude of spectral content is mapped with loudness. The parameters to be mapped are all derived from physical velocity values, and as such are not subject to ambiguity of order; the mapping is intended to resemble a simulation of underlying physical phenomena. As such, polarity choices were made to follow the logic of ecological perceptions. Use of spatial audio was considered to represent space of the mesh, but was ultimately found more useful for communicating the kinetic directionality of flow within the sample volume.

2.2. Implementation

Fig. 2 provides an overview of the method of sonification. In the following sections, the following aspects of the real-time method of sonification are described in detail:

- Probing the local volume for extraction of point velocity data
- Filtering velocity data by spectral power index (SPI) criteria
- Computing the PSD spectrogram from the filtered velocity data
- Passing the PSD data to SuperCollider to produce audio

Figure 3: Isolation of velocity fluctuations by high-pass filtering the pulsatile signal according to SPI criteria.

Data is stored in binary data frames and accessed interactively from the Paraview render view. SuperCollider is used to generate the sonification from the flow data.

2.3. Probing Data

Much like a stethoscope for the body, the user may interactively drag the probe (representing the sample volume) through the domain to listen to the local frequency spectrum. The chosen probe radius used is 0.25 mm, containing approximately 50 discrete points within the volume. A list of point identifiers is generated for a 20 uniformly distributed random points within this local volume. The velocity values for these points through time will serve as the basis for the sonification, to be discussed in subsequent sections. This sonification is intended to accompany a synchronized visualization with time duration chosen by the user. As such, this pipeline is updated once for each period of flow. The data has temporal resolution of 0.004s (2500 samples/second) and about 500 000 points.

2.4. Filtering Data

As shown in the Fig.3, the velocity content for each point is extracted and filtered by criteria of spectral power index (SPI) to isolate flow instabilities.

SPI is a frequency domain decomposition method introduced by Khan et al [20] as a means to quantifying the transient instabilities of turbulent-like flows without cycle-to-cycle variations. This decomposition separates the pulsatile carrier signal from the turbulent-like flow instabilities. By using a high pass filter to remove the 99% most energy-carrying frequencies (corresponding to a cutoff frequency of 25 Hz), flow instabilities can be isolated. SPI is defined as the ratio of energy contained within the flow instabilities as a fraction of all signal energy.

2.5. Computing the average spectrogram

The local PSD spectrogram for each point is calculated using a window length of 256 samples with 192 sample overlap, resulting in a smooth, temporally shifting view of the power content for each frequency bin. The average spectrogram for the sample volume is

Figure 4: Average power spectral density computed for N=20 sample points in radius 0.25 mm.

then computed (Fig.4). This matrix of power content corresponds to 128 frequency bins x 36 time bins.

2.6. Generation of sonification

A synthesizer template used for creating synth nodes (synthdef) was generated in the SuperCollider environment and stored on sc-synth (a real-time SuperCollider audio server [19]). Using sound networking protocol Open Sound Control (OSC), the power matrix is sent to scsynth.

The frequency content and time-varying power of the computed spectrogram are fed to a bank of sound generators. Each frequency bin of the spectrogram corresponds to the frequency of a generator. The time-varying power values of the spectrogram are used to modulate the loudness (amplitude) of each generator in time.

An array of synth nodes are generated, with each node procedurally assigned a range of frequency bins from the power matrix. The power matrix is broken into chunks by frequency, with each chunk containing five frequency bins to be generated per node. The power associated with each bin corresponds to the temporal loudness (amplitude envelope) of the sound as well as a bubble parameter, while the frequency value corresponds to the resonant frequency of a constant skirt gain filter (Ringz generator) with noise input. The bubble parameter controls the cutoff frequency and modulation depth of a resonant high-pass filter. By pseudo-randomly shifting the cutoff frequency of the high pass filter near the resonant frequency of the Ringz generator, and with a shift depth proportional to the temporal power, sound textures resembling bubbles are generated to indicate a higher level of flow complexity.

The summation of these generators results in a harmonically rich signal with frequency content varying in time. By tweaking the bubble parameter, the output sound varies between a harmonically clean sound reminiscent of Doppler auscultation, to windy sound textures, and more evocative ecological fluid flow sounds. By modulating this parameter with the amplitude envelope, the resulting signal sounds more chaotic during the systolic pulse of flow.

The apparent directionality of flow is communicated by rotating an ambisonic B-format sound field around the user. Given

the position of the aneurysm model onscreen, the directionality of flow relative to the user is determined and used to spatialize the sound; for flow visually moving left-to-right, the perceived sound field rotates clockwise around the user, and vice versa. The perceived rotation angle indicates the relative velocity orthogonal to the user.

2.7. User Interaction

The signal can be pitched up (maintaining relational frequency content) and slowed down while maintaining audio-rate synthesis, and can be pruned of low-amplitude frequency content or filtered in order to evaluate content by frequency band.

This on-the-fly generation is intended to accompany a conventional flow visualization in ParaView. All computed data (signal filtering and extraction of PSD) is computed within the Paraview python shell.

2.8. Plan for User Study

A preliminary user study is planned to test the accuracy and speed of discriminating a variety of complex flows, compared to using standard visualizations alone.

Conventional visualizations of intracranial aneurysms effectively communicate flow pattern, but do not effectively communicate flow quality (laminar, transitional, turbulent). By allowing users to discriminate the spectral quality of the flow by augmenting the visualization with sound, it is hypothesized that users will be able to better identify turbulent flow as compared to standard visualization alone.

Two training steps will be necessary. First, the correspondence between sonified spectral data and turbulent-like flow will be established in agreement with current auscultation literature, noting timings, pitch, and intensity [5]. Second, flow quality will be demonstrated using conventional flow visualizations, particularly pathline representations.

Three hypothesis have been generated regarding interpretation of unstable flow in numerically simulated aneurysms:

1. Users can accurately discriminate levels of turbulence through sound alone.
2. Users cannot accurately discriminate regions of turbulence by visuals alone.
3. Users will be able to best identify regions of turbulence using multimodal representation.

First, users will be asked to comparatively listen to sounds alone and rank which contains higher complexity or turbulence. User responses will be compared to known quantities of flow complexity. Second, the users will be asked to visually compare and rank-order regions of complex flow in a conventional visualization. Third, a conventional visualization will be shown to the user with sonified data for preselected regions. The user will be asked to identify which region contains more turbulence. Evaluation of the user study will occur in April-May 2018.

3. RESULTS AND DISCUSSION

An current example of the implemented method can be found at [21].

The present framework of sonification results in a highly configurable audio representation allowing on-the-fly exploration of

flow instabilities within CFD data. Frequency bands can be individually emphasized or extracted for closer inspection. Different sound textures, ranging from Doppler-inspired sounds to bubbly flow sounds, exist within the instrument framework. This sonification runs in real time next to a simplified conventional visualization on consumer grade hardware, allowing for flexible inspection of data.

Careful choice of mapping strategy, parameter polarity, and psychological scaling provide a range of sonic palettes for known conditions of flow within intracranial aneurysms. The prototype has been tested with five aneurysm datasets, each of similar temporal and spatial resolution but varying flow complexity. Further exposure to new datasets will confirm the existing method of psychological scaling, or suggest further refinement to accommodate a broader range of flow conditions.

The biological and clinical implications of high frequency content from a decomposed velocity signal are not the basis of this report; these effects are discussed in greater depth in CFD literature. The physiological presence of high frequency fluctuations, along with frequency-specific amplitude envelopes, are to be communicated to the user for interpretation. Other simplified strategies such as directly communicating Q-criterion to indicate turbulent flow were sonified, but by reducing the number of communicated parameters, the feedback system was found too simplistic for anything other than locating regions of instability. If it were as easy as presenting a prepared quantity, sonification would not be necessary; by design, this sonification attempts to maintain the nuance of the original signal for interpretation by a trained user. The objective of the current research is to present the spectral data using audio, and not to interpret the data.

The frequency resolution of the calculated PSD is limited by the temporal resolution of the data. By windowing the data, variation in time is gained at the expense of frequency resolution. This is a reasonable balance if the signals often approximate noise; broad temporal and spectral characteristics are communicated. From this balance, there are two potential disadvantages. First, the hard-set designation of frequency bands results in similar timbral properties between datasets. Secondly, and perhaps more importantly, if the signals have strong harmonic content, greater frequency resolution may be required to explore spectral relationships. Given the existing knowledge of spectral content in turbulent-like flow, the current chosen windowing parameters are acceptable.

By choosing the data for sonification based on known physiological parameters and carefully scaling, initial impressions appear to indicate good correspondence between data and sonification.

The methodological framework of the current research has been presented without formal user-study, and as such further evaluation of the design is the most crucial next step. In addition to evaluation of the hypotheses described in section 2.8, the effectiveness of the multidimensionality presented within the framework should be evaluated. From preliminary observation of the prototype, there is a possibility that the salience of each dimension within the sonification may not be perceptually balanced. For example, temporal characteristics of the sound may overshadow the harmonic content of the signal, which may indicate a need for finer resolution in harmonic content in order to discriminate different flows. Similarly, fluctuations persisting long in into the cardiac cycle may be perceived as more turbulent compared to short, intense bursts. Such an assessment may indicate that temporal envelope characteristics may require further refinement. Until the

framework is evaluated through user training and user study, the effectiveness of this sonification can be only incompletely assessed based on the design criteria of the authors.

A fundamental design limitation of the present framework lies in the ill-defined relationship between hemodynamic forces and degenerative remodelling of the aneurysm wall [2]. As such, harmonic and temporal components of the decomposed flow data are treated approximately equally to retain characteristics of the physical phenomena.

This work extends previous sonifications of multidimensional vector data by directly addressing the shortcomings of data visualization with respect to temporal instability. This insight resulted from in-depth knowledge of the physical and simulated hemodynamics specific to the current application of intracranial aneurysms.

The proposed method for interpreting cerebral flows is not limited by application, and may be useful for other time-series spatial datasets containing instabilities. A similar implementation has also been tested in the 3D creation suite Blender [22], and would presumably function in any other python-based environment.

4. CONCLUSIONS

A novel method for sonifying CFD data for the purpose of better communicating high frequency fluctuations within simulated intracranial aneurysmal flows is described. This frequency-domain method is intended to augment conventional visualization in order to ease the interpretation of dense spatiotemporal datasets by allowing the user to visually concentrate on one field while listening to another. This method for data inspection builds on established analysis of intracranial hemodynamics with a focus on real-time inspection inspired by the existing field of auscultation.

5. REFERENCES

- [1] "Brain aneurysm statistics and facts," Web Site, 2018. [Online]. Available: <https://www.bafound.org/about-brain-aneurysms/brain-aneurysm-basics/brain-aneurysm-statistics-and-facts/>
- [2] K. Valen-Sendstad, K. A. Mardal, and D. A. Steinman, "High-resolution CFD detects high-frequency velocity fluctuations in bifurcation, but not sidewall, aneurysms," *Journal of Biomechanics*, vol. 46, no. 2, pp. 402–407, 2013.
- [3] K. Valen-Sendstad, M. Piccinelli, and D. A. Steinman, "High-resolution computational fluid dynamics detects flow instabilities in the carotid siphon: Implications for aneurysm initiation and rupture?" *Journal of Biomechanics*, vol. 47, no. 12, pp. 3210–3216, 2014.
- [4] N. Varble, J. Xiang, N. Lin, E. Levy, and H. Meng, "Flow Instability Detected by High-Resolution Computational Fluid Dynamics in Fifty-Six Middle Cerebral Artery Aneurysms," *Journal of Biomechanical Engineering*, no. 6, p. 061009.
- [5] A. J. Taylor, *Learning Cardiac Auscultation: From Essentials to Expert Clinical Interpretation*. Springer.
- [6] M. J. Mc Loughlin and S. Mc Loughlin, "Cardiac auscultation: Preliminary findings of a pilot study using continuous Wave Doppler and comparison with classic auscultation," *International Journal of Cardiology*, no. 2, pp. 590–591.
- [7] L. N. Sekhar, M. Sun, D. Bonaddio, and R. J. Scialabassi, "Acoustic recordings from experimental saccular aneurysms in dogs," *Stroke*, no. 8, pp. 1215–21.
- [8] G. G. Ferguson, "Turbulence in human intracranial saccular aneurysms," *Journal of neurosurgery*, no. 5, pp. 485–97.
- [9] Y. Kurokawa, S. Abiko, and K. Watanabe, "Noninvasive Detection of Intracranial Vascular-Lesions By Recording Blood-Flow Sounds," *Stroke*, vol. 25, no. 2, pp. 397–402, 1994.
- [10] G. Haller, "An objective definition of a vortex," *Journal of Fluid Mechanics*, vol. 525, pp. 1–26, 2005.
- [11] B. C. J. Moore, "Frequency difference limens for short duration tones," *The Journal of the Acoustical Society of America*, no. 3, pp. 610–619.
- [12] G. Dubus and R. Bresin, "A systematic review of mapping strategies for the sonification of physical quantities," *PLoS ONE*, vol. 8, no. 12, 2013.
- [13] T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*, 2011.
- [14] E. Klein and O. G. Staadt, "Sonification of three-dimensional vector fields," in *Proceedings of the 12th SCS High Performance Computing Symposium (HPC 2004), part of the SCS Advanced Simulation Technologies Conference (ASTC'04)*, 2004.
- [15] M. Kasakevich, P. Boulanger, W. F. Bischof, and M. Garcia, "Augmentation of Visualisation Using Sonification : A Case Study in Computational Fluid Dynamics," *IPT-EGVE Symposium*, 2007.
- [16] B.-A. J. Menelas, "Interactive analysis of cavity-flows in a virtual environment," *Proceedings of the 28th Spring Conference on Computer Graphics*, pp. 31–38, 2012.
- [17] Aneurisk-Team, "AneuriskWeb project website, <http://ecm2.mathcs.emory.edu/aneuriskweb/>," Web Site, 2012. [Online]. Available: <http://ecm2.mathcs.emory.edu/aneuriskweb/>
- [18] U. Ayachit, "The ParaView Guide: A Parallel Visualization Application," *Kitware*.
- [19] J. McCartney, *SuperCollider*, 2018. [Online]. Available: <https://supercollider.github.io>
- [20] M. O. Khan, C. Chnafa, D. Gallo, F. Molinari, U. Morbiducci, D. A. Steinman, and K. Valen-Sendstad, "On the quantification and visualization of transient periodic instabilities in pulsatile flows," *Journal of Biomechanics*, vol. 52, pp. 179–182, 2017.
- [21] D. MacDonald, *Sonification of flow instabilities in intracranial aneurysms*, Biomedical Simulation Laboratory, University of Toronto, 2018. [Online]. Available: <https://www.youtube.com/watch?v=UmDvnPjnpV4>
- [22] Blender Online Community, *Blender - a 3D modelling and rendering package*, Blender Foundation, Blender Institute, Amsterdam, 2017. [Online]. Available: <http://www.blender.org>

DESIGNING SONIFICATION

SONIFICATION WITH MUSICAL CHARACTERISTICS: A PATH GUIDED BY USER ENGAGEMENT

*Jonathan Middleton¹, Jaakko Hakulinen², Katariina Tiitinen², Juho Hella², Tuuli Keskinen²,
Pertti Huuskonen², Juhani Linna², Markku Turunen², Mounia Ziat³ and Roope Raisamo²*

¹ Eastern Washington University, Department of Music, Cheney, WA 99203 USA, jmittleton@ewu.edu

² University of Tampere, Tampere Unit for Computer-Human Interaction (TAUCHI), Tampere,
FI-33014 Finland, {firstname.lastname}@uta.fi

³ Northern Michigan University, Dept. Psychological Science, Marquette, MI 49855 USA, mziat@nmu.edu

ABSTRACT

Sonification with musical characteristics can engage users, and this dynamic carries value as a mediator between data and human perception, analysis, and interpretation. A user engagement study has been designed to measure engagement levels from conditions within primarily melodic, rhythmic, and chordal contexts. This paper reports findings from the melodic portion of the study, and states the challenges of using musical characteristics in sonifications via the perspective of form and function – a long standing debate in Human-Computer Interaction. These results can guide the design of more complex sonifications of multivariable data suitable for real life use.

1. INTRODUCTION

Sonifications that rely on musical elements such as rhythm, pitch, and harmony are often reserved for creative purposes, yet musical characteristics can offer important contributions towards analysis [1]. When a variety of musical elements are integrated in auditory display for analytical contexts, such as process monitoring, researchers can capture a variety of features from multivariate data [2]. One concern, however, is that musical representations can create distractions and that music adds something more than is needed for an analysis. In this sense, music is seen as a more favorable component of form rather than function – an issue tied to the long-standing Human-Computer Interaction (HCI) debate of which comes first [3].

This paper offers the view that musical elements can facilitate analytical objectives from improved data perception, data analysis, and data interpretation through user engagement and enhanced experiences.

The use of musical elements and characteristics in sonification has been formally explored by members of the International Community for Auditory Display since the mid 1990's [4], [5], [6]. A summary of music-related research and software development in sonification can be obtained from Bearman and Brown [7]. Earlier attempts were made by Pollack and Flicks in 1954 [8], yet according to Paul Vickers [9] and other researchers, such as, Walker and Nees [10], the path for music in sonification remains uncertain or unclear. Vickers recently raised the significant question: “how should sonification designers who wish their work to be more ‘musical’ approach the task?” Vickers asserts “the path we ought to take is currently unanswered.” Walker and Nees [10] say that “questions of aesthetics and musicality remain open in the field of sonification.” The interest in incorporating musical characteristics in sonification can be summarized from the research of Brown et al. [11]: “the use of musical sounds has been recommended because of the ease with which musical sounds are perceived.”

The use of musical characteristics in sonification can serve multiple purposes. To start, musical expressions that relate to aesthetics can enhance the user experience, and this can, in turn, enhance the perception of data. Positive user experiences can translate to more time spent with the data analysis, and improving interpretations. Finally, the mapping possibilities to musical traits are numerous and this brings opportunities to associate multivariate data to different types of musical features and expressions.

The work reported in this paper was carried out in a project called Data-to-Music. It focused on the development of a software to map data to musical characteristics. The data came mostly from monitoring the conditions of buildings, machines, weather, and athletic experiences. What made the project unique, was the focus on the user experience with music.

In order to test the possibilities of musical characteristics in data representation and display, the research team designed and programmed the software called D2M (Figure 1). It enables the users to assign

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

Figure 1: Weather mapping to MIDI notes with the D2M software.

variables of time series data sets to “streams” which control different aspects of the generated audio. Each stream can be associated with a unique or shared instrument timbre and the data values can be set to control pitch, the complexity of rhythmic patterns, accent chords, articulation, or loudness of the instrument. The user can determine the duration of the generated music and the data are scaled accordingly. The data streams specify parameters of each generated note based on the given settings. The global structure of the generated music is the result of the data content.

Additional tools include the possibility to set thresholds based both on data values and time range to filter out some of the data. There are also preprocessing tools for linear and logarithmic scaling, as well as ways for inverting data values. The high degree of flexibility of the tool enabled the team to create auditory displays that sound composed, even though the results were prepared by an algorithm.

In this article, we first outline our sonification approach and then describe a user survey that seeks to measure to which extent the approach leads to greater levels of user engagement. The D2M software was used to turn weather data into musical pieces used in the survey.

2. FORM AND FUNCTION

The form and function debate in HCI is similar to the one found in sonification with music.

In the process of interpreting data, sound characteristics with musical qualities can contribute to enhanced levels of engagement or distract from the analysis [12]. The challenge is to make sure the aesthetic qualities of the auditory display do not compromise coherence. We submit that data-to-music sonification can not only engage users, but levels of engagement with coherence can be increased by careful combinations of musical characteristics.

A significant goal in HCI is the user experience, which is often tied to aesthetics. Dillon [13] traces the importance of aesthetics, as a basic motivator, back to A. Maslow (1954). Tractinsky et al. argue that aesthetic perceptions are highly correlated by individuals to interface usage; therefore, whether we intend to incorporate aesthetics by design into usage or not, aesthetics are most likely already tied to usage [3], [14].

In 2006 Andrew Dillon [13] introduced the term mediation to describe an essential mechanism whose role is to translate information between data and human. In describing this mechanism, he states that information technology is a mediator, among others, which “carries, stores, retrieves, and presents data”; and from these functions, information technology “can provide a physical instantiation of data or the mechanism for making visible or audible the data of interest.”

In this paper, we are guided by the research question: Can musical characteristics contribute to meaningful data perception, analysis and interpretation? In addressing this challenge, we reference two studies that focus on the significance of the audible experience

Figure 2: Cloud cover data being mapped into two independent streams of pitch and rhythmic values.

mentioned by Dillon and in particular the instantiation of data called sonification. We will then seek to prove that interpretive experiences with the aid of musical characteristics in the auditory display can offer analytical advantages.

In Human-Computer Interaction, auditory display has maintained an increasing role in offering new perspectives on data analysis and interpretation. These processes are undoubtedly complex. Dillon simplifies the process with a figure using two key terms called arcs (or paths), similar to Don Norman's interaction cycle [15]. There is the "arc of exploration" from human to data, and the "arc of interpretation" from data to human. Each path crosses the barrier of mediation (a central line) in a continuous circle. In other words, one leads to the other with the human as a starting and ending point.

We suggest that the human experience within the arcs of exploration and interpretation can be enhanced by aspects of aesthetics that come with musical characteristics. A step in this direction is not easy and can be demonstrated by a recent study in condition monitoring.

3. SONIFICATION AND CONDITION MONITORING

Condition monitoring with auditory display presents one of the most significant areas of exploration for sonification for practical usages. A recent study by Hildebrandt et al. [16] demonstrated the challenges of continuous monitoring and a successful method for real

time peripheral monitoring with sounds. Their main goal was to create a continuous soundscape based on sonification that does not distract users from their main task(s). In their study forest sounds were chosen for continuous soundscapes during peripheral process monitoring. The decision to use sounds from nature came after consulting the work of Gaver et al. [17] and Vickers et al. [18] to achieve "unobtrusiveness" and avoid a sense of "disturbance." One can imagine, in a practical sense, that sounds with musical characteristics would be more challenging to hear during extended "continuous" periods; however, given this assumption, in non-continuous environments, there may be opportunities for the inclusion of sonifications with musical characteristics.

While their study was successful, their conclusion asserts: "potentially more pleasing sound designs will be tested to test the suitability for long periods and effectiveness." They submit that these sounds could be based on longer looped samples "or even musical concepts." Although forest sounds are often pleasant over long periods, the level of engagement may be quite uniform. The researchers' motivation for more pleasing sound designs may be related to interests in creating a more engaging environment.

4. USER ENGAGEMENT

Multiple studies have shown that the task of defining and measuring engagement is quite complex, and many evaluations are connected to the field of education to

Figure 3: Melody Test flow with cloud cover data from three separate months shown as: Data 1, Data 2, and Data 3. Participants listened to each condition twice with different tasks before completing the questionnaire..

understand how students are engaged in learning [19], [20]. Engagement levels relative to musical interaction and loss of focus were recently explored by Tahiroglu et al. [21] where they monitored facial expressions of musicians and used the Positive and Negative Affect Schedule (PANAS) questionnaire to validate levels of perceived engagement.

In various studies on engagement, common descriptive terms emerge, such as motivation, persistence, and effort. Lutz Klauda and Guthrie [20] elegantly separate motivation as goal-oriented based on values and beliefs, while engagement refers to “behavioral displays of effort, time, and persistence in attaining desired outcomes.” Additional attributes from other studies include focused attention, curiosity, novelty, and challenge [22], [23], [24], [25].

The most seminal work on engagement evaluation appears to be by O’Brien and Toms [26]. The O’Brien and Toms study defines engagement in great detail by building on the premise that engagement is a process in three stages: point of engagement, sustained engagement, and disengagement, and as the process unfolds in time, there are multiple layers of experience, called threads. The categorization of threads is derived from the work of McCarthy and Wright [27].

We surmise that musical characteristics can

contribute to each thread: a) compositional (narrative), b) spatiotemporal, c) emotional, and d) sensual. While the O’Brien and Toms study is based on a framework for assessing levels of engagement in the context of visual display, our study adopts a similar framework to fit auditory/musical contexts.

5. USER SURVEY

The impact of musical characteristics in sonification is being measured by a user engagement survey that comprises three 40 minute tests, each anchored by either melody, rhythm, or accent chords (as opposed to harmonic progressions). Each test relies on fundamental areas of musical expression: pitch, rhythm, and timbre. As an example, the Melody Test uses these musical elements to express weather forecasting data for clouds (called oktas); and by virtue of the use of inversion the data can also express amounts of sunshine perceived by listeners. Note: This inversion technique is similar to one described by Walker and Kramer¹ [28].

The oktas data, on a scale of 0-8, were mapped to a one and half octave range with a tempo of one beat per second. Forty data points in a 24 hour period were displayed in forty seconds. A MIDI range of 43-60 was used as a destination span and all results adhered to a C major diatonic collection (Figure 1). The resulting MIDI

¹ Temperature trends were perceived as pitches ascending to represent a rise in temperature (with heat), or inverted so that ascending pitches represent cold, and although the latter seems counterintuitive, subjects perceived cold slightly better when pitches were high.

Figure 4: Average UES responses. The lines represent the three conditions, Sine, Timbre and Rhythm, and the related data points represent the average responses to specific items (x-axis). The items are ordered by the UES factors. Items that revealed statistically significant differences in Friedman's test ($p < 0.05$) are marked as follows: *) some differences, but none identified in the pairwise comparisons between the conditions, SR) differences between Sine and Rhythm, and ***) differences between both Sine and Timbre, and Sine and Rhythm.

files were rendered through software instruments on Logic Pro X and Ableton Live, and the resulting audio files were exported and converted to mp3 files for Web delivery with loudness normalized to negative 25 LUFS.

In our first test, listeners are presented with different experiences of three data sets in melodic form to determine if they hear clouds or sunshine. There are three conditions: Pitched sine waves with uniform rhythmic values with quarter notes (or crotchets); pitched timbre (marimba) with uniform rhythmic values with quarter notes; and pitched timbre (marimba) with a few basic rhythmic values in combinations using quarter notes and eighth notes (or quavers) (Figure 2). In the following, these three conditions are referred to as *Sine*, *Timbre* and *Rhythm*.

Each melodic experience is associated with: 1) tasks with questions (to give users time to experience the sounds), and 2) survey questions, at the end of a set of tasks for a given condition. The participants complete the same user engagement scale (UES) questionnaire on three occasions, with questions modelled after O'Brien and Toms [29] and answered on a 5-step scale from "strongly disagree" to "strongly agree." The participants complete tasks from three conditions (Figure 3) in two counterbalanced schemes of 6 (one scheme nested in another), hence participant 1 has the same experience as

participant 7.

Examples from our questionnaire below demonstrate how the Data-to-Music team adopted the original O'Brien and Toms [29] questions from a Web shopping experience for a musical context. "I was absorbed in my shopping task" transformed to "I was absorbed in my listening task." "I felt frustrated while visiting this shopping website" changed to "I felt frustrated while listening to these sounds." "This shopping website is attractive" became "These sounds were attractive." The transformation of questions was as close as we could make them.

In a few cases the transformations offered a challenge. For example, we modified the question "I liked the graphics and images used on this shopping website" to "I liked the beats and rhythms used in these sounds." The team is satisfied with the transformations based on a pre-existing engagement study and how the user experiences from our sonifications may relate to varying levels of engagement.

The complete set of modified questionnaire items can be seen in Table 1. We raffled the item order once and used the order (seen in Table 1) across all conditions and for all participants. Similarly as with O'Brien and Toms [29], we used reverse coding in the analysis for 8 out of 31 items for easier visual interpretation. Thus,

*Table 1: The UES statements modified to address the context of sonifications. Item number indicates the order in which the statements were presented to the participants. * indicates that the item-related responses were reverse-coded.*

#	Item
1	I was so involved in my listening task that I lost track of time.
2	The time I spent listening just slipped away.
3	My sound experience was rewarding.
4	I felt interested in my listening task.
5	During this sound experience I let myself go.
6	When I was listening, I lost track of the world around me.
7	These sounds appealed to my auditory senses.
8	I could not identify some of the things I needed to identify on these sounds.*
9	I liked the beats and rhythms used in these sounds.
10	If made available, I would continue to listen to these kinds of sounds out of curiosity.
11	I felt frustrated while listening to these sounds.*
12	The content of the sounds incited my curiosity.
13	I felt involved in this listening task.
14	Listening to these sounds was worthwhile.
15	I felt in control of my sound experience.
16	I found these sounds confusing to understand.*
17	This sound experience was fun.
18	I consider my sound experience a success.
19	These sounds were aesthetically appealing.
20	This sound experience did not work out the way I had expected.*
21	The sound layout of these sounds was auditory pleasing.
22	I felt annoyed while listening to these sounds.*
23	I was absorbed in my listening task.
24	These sounds were attractive.
25	I lost myself in this sound experience.
26	I blocked out things around me when I was listening to the sound data.
27	This sound experience was demanding.*
28	I felt discouraged while listening to these sounds.*
29	I would recommend listening to these kinds of sounds to my friends and family.
30	I was really drawn into my listening task.
31	Understanding these sounds was mentally taxing.*

item 11 can be considered as “I did not feel frustrated while listening to these sounds.”

The melodic portion of the study was conducted with 24 participants (11 female, 13 male) aged 18–50 years ($M=22.4$, $SD=7.5$). 17 of the participants had not heard data sonifications before the experiment. IRB exemption was provided by Eastern Washington University (human subjects protocol HS-5429) on December 1, 2017.

6. RESULTS AND CONCLUSION

Results from 24 participants in our validation study indicate how combinations of musical characteristics in melodic forms can engage listeners, as these relate to the sonification of sun and clouds. The subjective results on the UES items per condition, i.e., Sine, Timbre or Rhythm, are represented in Figure 4. The figure shows average values instead of medians for easier visual distinction. We ran Friedman’s tests to compare whether the ordinally scaled data would show differences between the conditions. The analysis revealed altogether 19 items with statistically significant ($p<0.05$) differences, and 8 of these findings showed such differences even between the conditions in pairwise comparisons.

Based on our results, engagement levels increase with melodies that are based on pitch with timbre combined, and those with pitch, timbre and rhythm combined. The findings show that sun and clouds data converted to only pitches with sine waves had lower engagement levels overall than those with timbre and rhythm (Figure 4). Average response values on a scale of 1–5 in the range of “strongly disagree” to “strongly agree” broadly demonstrate the importance of timbre in the six factors defined by O’Brien and Toms [29].

In “perceived usability” and “aesthetics” the differences are most pronounced. For the factor “perceived usability” averages are: sine waves 3.47; timbre and pitch 3.96; timbre pitch and rhythm 3.82. In the area of “aesthetics” averages are: sine waves 2.74; timbre and pitch 3.59; timbre pitch and rhythm 3.84. “Perceived usability” relates to the emotions experienced by users, and “aesthetics” to sensory appeal [29].

We see similar trends among some questions in the other factors: “focused attention,” “felt involvement,” “novelty,” and “endurability.” The trends suggest that timbre can significantly contribute towards engagement levels, even without the aid of moderately active rhythms. These results are from the Melody Test, which represents one third of our validation study. Tests with accent chords and active rhythms are on-going at this time.

The conducted study provides understanding on the relevance of timbre and rhythm to the engagement with musical sonifications. Both of these properties appear to

have a significant effect, in particular to the aesthetics of the auralization. Aside from focused attention, other measured aspects are affected as well.

While the evaluation was conducted with simple sonifications of single data variables to provide control to focus on individual factors on user experience, for real life use we believe that more complex sonifications are needed. Such sonifications can communicate more information and will likely have better durability and more novelty. Our D2M tool supports the creation of these complex musical combinations. The observations presented in this paper set a foundation to investigate these more elaborate musical representations of data.

7. ACKNOWLEDGMENT

Special thanks to Tekes, the Finnish agency for Innovation: decision 40296/14. This work was carried out in Finland in a three-year project funded by the agency and participating companies. The work was conducted at the University of Tampere within their unit for Human-Computer Interaction in collaboration with five companies with representatives seeking to hear their companies' data for analysis.

REFERENCES

1. Middleton J, Dowd D. Web-based algorithmic composition from extramusical resources. *Leonardo* 2008;41(2):128-135. doi:10.1162/leon.2008.41.2.128.
2. Hermann T, Hunt A, Neuhoff, J.G. (Eds) *The Sonification Handbook*. Logos Verlag, 2011.
3. Tractinsky, N., Katz, A.S., & Ikar, D., What is beautiful is usable. *Interacting with Computers*, 13, 2000, pp. 127-145.
4. Hermann, T., Hunt, A., Neuhoff, J. G. (Eds), *The Sonification Handbook*, Logos Verlag, 2011.
5. Sound and Geographic Visualization. In Visualization in Modern Cartography Mac Eachern and Taylor eds. New York: Pergamon, 149-166, 1994.
6. Flowers, Buhman, and Turnage. *Data sonification from the desktop: Should sound be part of standard data analysis software?*, in Proceedings of the 2nd ICAD 1996.
7. Bearman and Brown, *Who's Sonifying Data and How are they Doing it? A Comparison of ICAD and Other venues since 2009*. Proc. 18th ICAD Atlanta GA., 2012.
8. Pollack, I. & Ficks, L., Information of elementary multidimensional auditory displays. *The Journal of the Acoustical Society of America* 26/155, 1954.
9. Vickers, P., Sonification and Music, Music and Sonification. In M. Cobussen, V. Meelberg, & B. Truax (Eds.), *The Routledge Companion to Sounding Art* (pp. 135-144). Oxford: Routledge, 2017.
10. Bruce N. Walker and Michael A. Nees. Theory of Sonification (chap 2 p. 27) in *Sonification Handbook*.
11. Brown, L.M. & Brewster, Stephen & Ramloll, S.A. & Burton, R & Riedel, B. (2003). Design Guidelines for Audio Presentation of Graphs and Tables. Proc. ICAD, Boston, MA, USA, July 6-9, 2003.
12. Vickers, P., Hogg, B., & Worrall, D., Aesthetics of sonification: Taking the subject-position. In C. Wöllner (Ed.), *Body, Sound and Space in Music and Beyond: Multimodal Explorations*, 2017, pp. 89-109. Routledge.
13. Dillon, A., Information Interactions: Bridging disciplines in the creation of new technologies. In *Human-computer Interaction and Management Information Systems: Foundations*. D. F. Galletta & Z. Ping (Eds.) Oxford: Routledge, 2006.
14. Tractinsky, N., Aesthetics in Information Technology: Motivation and Future Research Directions. In D. F. Galletta & Z. Ping (Eds.), *Human-computer Interaction and Management Information Systems: Foundations*. Oxford: Routledge, 2006.
15. Norman, D., *The Design of Everyday Things*, MIT Press, 1988.
16. Hildebrandt, T., T. Hermann, and Rinderle-Ma, S., Continuous sonification enhances adequacy of interactions in peripheral process monitoring. *International Journal of Human-Computer Studies*, 95, 2016, pp. 54-65, Elsevier.
17. Gaver, W.W., Smith, R.B., O'Shea, T., Effective sounds in complex systems: the ARKOLA simulation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching Through Technology (CHI'91). ACM, New Orleans, 1991, pp. 85-90.
18. Vickers, P., Laing, C., Fairfax, T., July 2014. Sonification of a network's self-organized criticality. arXiv:1407.4705.
19. Appleton, J. J., Christenson, S. L., Kim, D., Reshley, A. L., Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology*, 44/5, 2006, pp. 427-445.
20. Lutz Klauda, S., & Guthrie, J. T., Comparing relations of motivation, engagement, and achievement among struggling and advance adolescent readers. *Read Writ*, 28(2), 2015, pp. 239-269.
21. Tahiroglu, K., Vasquez, J. C. Kildal, J., Facilitating the musician's engagement with new musical interfaces: Counteractions in music performance. *Computer Music Journal*, 41/2 2017, pp. 69-82.
22. Webster, J., & Ahuja, J.S., Enhancing the design of Web navigation systems: The influence of user disorientation on engagement and performance. *MIS Quarterly*, 30/3, 2006, pp. 661-678.
23. Webster, J., & Ho, H., Audience engagement in multimedia presentations. *The DATA BASE for Advances in Information Systems*, 28(2), 1997, pp. 63-77.
24. Jacques, R., Preece, J., and Carey, J. T., "Engagement as a Design Concept for Hypermedia, *Canadian Journal of Educational Communications*, 1995, pp. 49-59.
25. Skelly, T.C., Fries, K., Linnett, B., Nass, C., & Reeves, B. Seductive interfaces: Satisfying a mass audience. In C. Plaisant (Ed.), *Proceedings of the Conference on Human Factors in Computing Systems*, 1994, pp. 359-360.
26. O'Brien H. L., & Toms, E.G., What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6), 2008, pp. 938-955, Wiley.
27. McCarthy, J., & Wright, P., *Technology as Experience*. Cambridge, MA: MIT Press, 2004.
28. Walker, B.N. & Kramer, G. Mappings and Metaphors in Auditory Displays: An Experiment Assessment. *ACM Transactions on Applied Perception* 2/4, 2005, pp. 407-412.
29. O'Brien, H.L. & Toms, E.G., The Development and Evaluation of a Survey to Measure User Engagement: *Journal of the American Society for Information Science and Technology*, 61 (1), 2010, pp. 50-69. <http://dx.doi.org/10.1002/asi.21229>

A STUDY OF EXPLORATORY ANALYSIS IN MELODIC SONIFICATION WITH STRUCTURAL AND DURATIONAL TIME SCALES

Takahiko Tsuchiya and Jason Freeman

Georgia Institute of Technology,
Center for Music Technology,
840 McMillan St. Atlanta GA 30332, USA
{takahiko, jason.freeman}@gatech.edu

ABSTRACT

Melodic sonification is one of the most common methods of sonification: data modulates the pitch of an audio synthesizer over time. This simple sonification, however, still raises questions about how we listen to a melody and perceive the motions and patterns characterized by the underlying data. We argue that analytical listening to such melodies may focus on different ranges of the melody at different times and discover the pitch (and data) relationships gradually over time and after repeated listening. To examine such behaviors in real-time listening to a melodic sonification, we conducted a user study employing interactive time and pitch resolution controls for the user. The study also examines the relationships of these changing time and pitch resolutions to perceived musicality. The results indicate a stronger general relationship between the time progression and the use of time-resolution control to analyze data characteristics, while the pitch resolution controls tend to have more correlation with subjective perceptions of musicality.

1. INTRODUCTION

Consider asking someone to sing a melody based on one-dimensional, numeric, and sequential data (e.g., an EEG sensor input). Although there may be various mapping possibilities with the human voice, many people may choose to modulate the tone (pitch) of their voice by looking at the approximate value changes of the data sequence. Now, consider listening to such a pitch-modulated melody and trying to analyze the characteristics of the underlying data, such as recurring shapes that might be short, long, or both. Although these attributes might depend largely on the sonified data itself, are there any common patterns in how we explore and identify unknown data attributes over time?

First, we need to ask how our mind processes a newly-encountered melody in general. To understand the structure of a melody in real time, our mind does not seem to simply follow the instantaneous pitches from moment to moment. Instead, depending on the analytical goals, we may compare the current pitch to the immediate past, several seconds ago, to the beginning (as much as working memory helps), or possibly even to the upcoming (i.e., “anticipated” [1]) pitches to search for the characteristics of the melody. In other words, in an auditory analysis of data within a melody, our listening “focal range” over time may change dynamically.

This study examines our general exploratory behaviors when listening to and analyzing the data attributes in local and global time spans in a melodic sonification. Considering the obvious difficulty of measuring variable focal levels of listening, we have instead conducted an experiment featuring a generative system with dynamic resolution controls for the listener¹. With the control sliders, the user can physically emulate the varying focal ranges in real time by abstracting (e.g., quantizing) the data in time and pitch dimensions.

Our primary goal was to observe and statistically test how the listener explores and understands the structure of an unfamiliar melody, and how their analysis develops over time as the melody unfolds and repeats. However, the experimental design was also informed by two additional interests: to examine the relationships of the emulated resolution parameters to a subjectively-chosen musical balance, as the notion of time structures has a strong tie to musical aesthetics; and to evaluate the advantages of an exploratory real-time sonification system that allows the alternation of the system behavior.

We expect that this study contributes to the development of sonification techniques in many ways. It may, for example, inform the parameter-tuning process of a simple melodic sonification for optimal analytical experiences without expending additional mapping channels in the audio synthesis. It may also help creating a more engaging and intuitive sonification by taking a “balanced” resolution mapping for musical and analytical interests according to our test results. Lastly, by employing active control of parameters for users rather than a typical passive listening environment, we hope to present unique opportunities as well as challenges in analyzing a highly interactive sonification.

2. RELATED WORK

The perception and treatment of time structures in a melody involve various philosophical and pragmatic discussions. Particularly in the field of contemporary music, where digital audio synthesis has enabled “sub-symbolic” manipulations of sounds, the study of time structures in sound has been popular among many composers and theorists. Here, we briefly address their work and concepts most relevant to our experiment in two broad categories: ones for musical (qualitative) and cognitive perspectives and ones for more quantitative perspectives related to (functional) sonification, visualization, statistics, and digital signal processing. These opposing categories, however, often intersect with each other.

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

¹ <http://quantizationtest.herokuapp.com>
(User-experiment website, Retrieved: 3/13/2018)

2.1. Multiplicity of Musical Time

The common consensus in the study of musical time is that sounds are expressed and perceived not simply in the absolute “clock” time but in multiple ways. Desain and Honing, for instance, observe that a typical musical expression consists of, at minimum, discrete time intervals (e.g., meters) and continuous and “expressive” timing manipulations (e.g., “rubato” and “accelerando”), which poses a challenge in automatic transcriptions [2]. While these competing timing factors are both in the note- or phrase-level time durations, various theorists consider musical time to be hierarchical, even for a simple monophonic melody. Roads describes nine levels of time scale that contribute to the formation of musical piece, sections, rhythms, pitch, timbre, etc. for both perception and generation [3]. Here, different time scales imply different resolutions or levels of detail. The sensation of tone (pitch), for instance, happens at the “sound object” (i.e., note) and “microsonic” levels when there are enough repetitions of a microsonic pattern, while timbral effects are noticed when there are “micro” and “sample” level irregularities. The melodic patterns are recognized in the “meso” time scale, one level above the sound objects. How would these structural time scales or resolutions play a role when analyzing the data within a melodic sonification, such as value-by-value fluctuations, local and global peaks, and gradual shifts of the central point? How does the listener identify these scales in a new melody?

In addition to the hierarchical time scales, our experience of musical time is multiple while listening to and analyzing a melody, as expressed as “polyphony of viewpoints” by Jonathan Kramer [4]. He observes that when we listen, learn, and compare different parts of the melody, our mind does not simply follow the absolute time with a linear progression. Instead, we learn the structure of the melody in terms of the duration of auditory patterns and their various proportions (cf. time scales). As we listen to the piece / melody, Kramer observes, we acquire new information about the proportions. As such, the focal length of time may be constantly adjusted with two cognitive processes present: one following the durations in passing (from a still sounding past moment until the current moment), and the other experiencing and comparing the remembered durations in retrospect. While we may continuously learn the time structure of the melody as it progresses, he also argues the importance of repeated listening for a thorough analysis [5].

The notions of nonlinearity and multiplicity in musical time, therefore, have informed our user-experiment designs, where we attempt to capture the experience of absolute time vs. the duration relative to the total performance time of the melody (including repetitions), take the progress of time into consideration as the process of learning the periodic structures, and encourage the repeated listening experience where the user may learn the optimal focal ranges according to the current position (phase) of the data.

2.2. Measurable Contents in Absolute Time

The measurement of the listener’s response to multiple time scales is, however, a complex task. Pragmatic sonifications tend to rely on the uniformity of absolute time with a well-defined and steady clock. Here, instead of the varying qualities in time scales, the experience of time may be altered by the content or the data within the absolute time scale. This may involve various strategies in mapping and transforming data to sound, including the use of statistical abstraction techniques.

In this regard, the audification technique maps the individual data points to successive samples, with the assumption that our ability of listening to and processing the audio-rate high-density data is as efficient as the short-time Fourier transform, with the time and frequency-domain resolutions high enough to perceive the timbral details [6]. Such assumption may, however, become questionable for the perceptibility of sample-level movements, especially with data with irregular motions, even though the global characteristics might be effectively captured as a spectral / timbral impression.

To counter the uncertainty of the playback speed, many melodic sonifications such as the Sonification Sandbox [7] utilize the auditory “tick marks,” or a static metronome, to indicate the current time position of the playback in the data. This approach, however, does not necessarily capture the potential hierarchical structure in a melodic sequence, or might even introduce a false sense of periodicity. The mapping of data to the tempo can also be variable and arbitrary, potentially affecting the perception of the melodic structures.

Our experiment takes an alternative approach where we present the melody with a fixed tempo, but instead offer statistical abstraction tools to alter the mapping of data contents to a constant (absolute) time line.

Data visualization, as well as sonification, frequently employ statistical aggregation or abstraction techniques that effectively translate continuous data to discrete symbols. For example, the histogram of continuous data groups values into discrete “bins” with uniformly-spaced quantization steps [8]. Scaler or vector quantization are used ubiquitously for digitization [9], structural analysis and compression [10]. While various researchers in sonification warn against quantizing signals to create “music-like” effects [11], nonlinear transformations, including quantization, are an important tool for analyzing and handling both continuous and symbolic data, especially ones with discrete and hierarchical structures [12].

3. USER EXPERIMENT

In this interactive listening experiment, we collected anonymized user data from 20 Georgia Tech Music-Technology students who, according to the survey, had a wide range of listening and musical-performance skills from novice to professional.

3.1. Goals and Hypotheses

In this experiment, we asked subjects to perform analytical tasks utilizing the dynamic time and pitch resolution controls for aiding and measuring their structural analysis process. While we hoped to find general correlations between their “final” parameters and the optimal perception of certain data attributes, such results may be also dependent on the data source and, ultimately, the user’s personal preference. Instead, our greater interest was to find patterns in their exploratory behaviors on a newly-encountered sonification, measured over several different durational time scales. In other words, our study questions if a progress of time correlates the way they like to hear the melody in certain time / pitch resolutions for a particular analytical goal.

We have hypothesized that, taking Kramer’s argument of the listener gaining new information about melodic proportions through the progress of time and repetitive listening, there may be converging behaviors or common directionalities for both time- and pitch-resolution adjustments

over time. We also questioned if the total duration spent for a task has a correlation to their final values.

As for the use of the time / pitch resolution controls that emulate the structural analysis of listening, Roads' hierarchical time-scale theory suggested to us that creating an appropriate pitch resolution may be more important to capture the phrase-level and longer time-scale structures of the melody, such as the global contour, than for local details. On the other hand, the time-resolution control may have stronger relationships especially to the perception of local-level data attributes, such as small peaks and rapid fluctuations.

3.2. The System for Experiment

The recent introduction of the Web Audio standard¹ has enabled us to develop real-time audio synthesis for all major web browsers. Utilizing web audio, we have previously developed an application programming interface (API) called data-to-music (DTM) [13] that facilitates rapid prototyping and embedding of sonification into any web applications [14]. This online user study utilizes the DTM API to create a real-time user-customizable sonification.

3.3. The Dataset

Figure 1: The dataset for the experiment: the US Gas Price time series.

For every test, we used the US Gas Price dataset², single-dimensional data sampled at a fixed time interval from 1990 to 2003 with 695 data points in total (Figure 1). We mapped the gas-price time series to a monophonic oscillator voice using the square waveform, with a logarithmic frequency modulation (FM) that scales the data domain of [88.5, 169.3] (USD) to a MIDI-note range of [45, 110]. The data was played back for the duration of 17.5 seconds, with about 39.7 data points per second at the default time / pitch resolution. (Though not strongly intended, 40Hz is used typically for speech analysis and modeling as the minimum threshold for the perception of tone [3].) The amplitude of the melody stays constant. In the experiment, the user does not see the visualization of the data, but only hears the sonified result.

3.4. User Control of Resolutions

The test provides a minimal user interface consisting of the Play / Stop buttons, a playback-position indicator, a time-resolution control slider, and a pitch-resolution control slider. The resolution controls modify the melody being played in real time by altering the mapping of the underlying data.

¹ <https://webaudio.github.io/web-audio-api> (Retrieved: 3/13/2018)

The pitch-resolution control applies a uniform but variable-step-size scaler quantization onto the log-scaled (i.e., MIDI-note) frequency. As the user lowers the pitch resolution, for example, the output MIDI range is rounded by the factor up to 9.0.

The time-resolution control down-samples and interpolates the data. It first segments the data into a smaller number of blocks (with the minimum of 30), calculates the mean value for each block, then interpolates the result into the original length of the data (695). We assumed a linearly-interpolating characteristic between data points (as in a line graph) rather than a step interpolation (as in a bar graph) and rendered the pitch-modulated sound accordingly. The use of the linear interpolation is perhaps the most debatable design decision we took. This is because, when mixed with the pitch quantization in some tasks, the effects of step-interpolated data points would become rather indistinguishable from the pitch quantization.

When the time and pitch abstractions are combined (e.g., in tasks 0, 3, 6, and 7, described in detail in the following section), the pitch quantization is applied after the time abstraction (interpolation) process. Both time and pitch resolution controls are recorded in the normalized range of [0, 1], and set to 1 (max resolution and unaltered) as the default value for each task.

3.5. User Analytical Tasks

Table 1: Task types, numbers and enabled controls.

Type / Ctrl	Time Res.	Pitch Res.	Both
Practice	---	---	0
Details-Local	1	2	3
Local-Global	4	5	6
Musical	---	---	7

The main purpose of the experiment was, again, to observe the subject's exploration over time as they perform an analytical task on a melodic sonification. However, defining an analytical listening task is itself a challenge. We cannot, for instance, ask to identify a specific pattern in the melody without giving away the information about the data and biasing their exploratory behaviors. Instead, the tasks ask the user to look for a "balancing" point or a perceptual boundary between two high-level categories of data attributes that are most likely in different time spans.

We categorize the target attributes into the time spans of "details," "local," and "global" levels. The "details" level includes perceived attributes such as rapid fluctuations or continuities of individual data points. The "local" time span may include small peaks and periodic patterns, as well as the current central tendencies. The "global" attributes may be such as the overall contour of the data, the total value range, and general directions of the data sequence. The analytical tasks are grouped into two main stages: Tasks 1 through 3 ask the user to find a boundary between the "details" and "local" characteristics, and tasks 4 through 6 ask to identify the balancing point between "local" and "global" levels. Each task features different combinations of the time / pitch resolution controls (see Table 1).

The task 0 was presented as the practice stage in order for the user to get familiar with the behaviors of the control UIs. The task 7 is an extra stage that asks to find the most musically-balanced resolutions without the consideration for data analysis.

² <https://vincentarelbundock.github.io/Rdatasets/datasets.html> (File: gasprice.csv; Retrieved: 3/13/2018)

Figure 2: Time / pitch resolution control over the normalized time. The thickness of the line indicates the current phase of the data. The dotted lines are the 3rd order polynomial regression models. Note that some values are filtered out for unbiasing the regression models (see section 4.1).

3.6. Measurements and Abbreviations

The experiment records the user input of the time resolution (TR) and pitch resolution (PR) controls. These user inputs were polled every second with the current time stamp and playback position (phase) of the data. Not only the relationships between TR and PR, our analytical interests are their relationships to various scales of time, including the tasks {1, 2, ..., 7}, absolute time (AT), normalized time (NT), and data phase.

AT, measured in seconds, provides the information about the total duration the user spent until reaching the final value of the resolution parameters for each task. This may include repeated playback of the data. AT, however, cannot be used for directly comparing the user inputs over time.

NT scales the AT spent in each task to the range of [0, 1], where the TR or PR value at NT = 1 is the final value chosen for the task. With NT, we can compare the users' exploratory / adjusting behaviors over the duration of the entire listening.

The data phase, recorded in [0, 1] range, is the third time scale regarding the current playback position of the melody. It encodes the information about the data source itself, where we may potentially find exploratory trends corresponding to a particular region in the data.

In the following analyses, “TR or PR” is denoted as “TR/PR” as a shorthand. The final resolution values at NT = 1 are denoted as “fTR/fPR.”

4. ANALYSES

4.1. Preprocessing

Every user task starts with the default TR/PR set to 1 (i.e., no alternations to the data). This has resulted in artificially correlated tendencies near the starting point of each task. To reduce the biases, for TR/PR vs. NT, we excluded the data

points that were TR/PR > 0.999 (i.e., values most likely unchanged by the user) as well as NT < 0.1. Similarly, for the histograms and TR/PR vs. phase, we filtered out the data when TR/PR > 0.999 as the default-value bias. Since the phase cycle can repeat multiple times per task, we did not filter out their starting points where phase < 0.1.

4.2. Regression Models

The statistical analyses utilize linear or third-order polynomial (for the user data over the timeline that tend to fluctuate) regression models, computed with the ordinary-least-squares method in Tableau 10.5. In the model statistics, the R^2 describes the variance of the recorded data points from the prediction model, with higher values (~ 1.0) being a better fit to the model. The P-value shows the probability that the output is affected by random chance (i.e., the model coefficients are set zero), with a lower value (e.g., $p < 0.05$) indicating the statistical significance of the input variable. Many of the regressive models below use the time (e.g., NT or phase) as the input, and TR/PR as the predicted output.

4.3. Overall Results

In general, the regression models with any input (e.g., time) to TR/PR outputs show very low R^2 values, commonly below 0.1 (10%), suggesting that the trend lines cannot be used for a precise prediction of the “exploratory” analysis. We speculate the factors to be either the general random nature of user exploration, strong personal biases of perceived structures in the melody, or potentially the ineffectiveness of the task instructions (see section 4.4). On the other hand, some models have the p-value below 0.01, indicating the presence of common directionalities of two variables when compared to the case where the input coefficient is set to zero (i.e., a null hypothesis; see sections 4.5 and 4.6). The model directionalities are, however, generally moderate as seen in Figures 2 and 7.

4.4. Analyses with Aggregate Time

Before analyzing the time-evolving aspects of the subjects' data exploration, we examined the aggregate TR/PR for each task disregarding the time progress. The results generally suggest strong personal preferences of TR as the final values.

Figure 3: The histogram (20 bins) of the average TR/PR for each task. The frequencies were normalized by the total AT for each user.

We first looked for the regularities in TR/PR in each analytical task, which may indicate commonly-agreed structural resolutions when disregarding the time progress. Here, a normal or sharp distribution may indicate a value range for the optimal perception of structures. The results (Figure 3) show some trends in PR-only controls (tasks 2 and 5) and task 7 (“musical”). TR, however, show high spreads and little trends. Note the filtered values (see section 4.1) affecting the TR results in tasks such as 6 and 7 – indicating that multiple users preferred to mostly alter PR for those tasks.

Figure 4: The final TR vs. PR values for tasks 3, 6, and 7. The size of dot indicates the amount of time spent. The gray areas contain the 2nd and 3rd quartiles.

Are there any proportional patterns with fTR and fPR? The tasks 3 (details-local), 6 (local-global), and 7 (musical) take the combined controls of TR and PR. The comparisons (Figure 4) show almost no correlations between fTR/fPR. It shows the shifting median for fTR but not fPR. The stationary nature of fPR can be explained with regard to subjective musical preferences (see section 4.6).

Table 2: The linear regression model statistics between fTR/fPR (outputs) and task transitions (inputs).

Ctrl	TR				PR			
	1→4	3→6	3→7	6→7	2→5	3→6	3→7	6→7
R ²	0.052	0.022	0.122	0.036	0.001	0.008	0.036	0.003
P-val	0.156	0.356	0.026	0.237	0.844	0.575	0.234	0.736

Next, we used the linear regression to model the general directionalities of the final values between different tasks. While the aggregate fTR looked to have a linear trend (see Figure 4), the individual fTR values between tasks had few common directionalities. Table 2 shows that only fTR for the tasks 3→7 were statistically significant, implying some time-scale relationships between “details-local” and “musical,” but more general evidence of strong personal preferences.

Figure 5: The final TR/PR values vs. the total time spent for each task.

Lastly for the aggregate time analyses, we examined the effect of the total duration spent vs. the final values. All results, however, showed little to no correlation between these variables (Figure 5). The number of repetitions also did not uniformly affect the final values. These results may signify the widely-different personal preferences among the subjects for the use of TR/PR for analysis.

4.5. Analyses with Time Progress

While there appears to be little commonly-agreed fTR/fPR that may explain the melodic or data structure, are there any regularities in how the subject explores the data before arriving the final values? Although the user explorations have certainly complicated the analyses of the result, we observed several characteristic outcomes. In order to directly compare the activities among users, we scaled the recorded data into a normalized time line (NT), and also mapped to the current phase of the data.

4.5.1. Explorations over Normalized Time

Table 3: The 3rd-order polynomial regression model statistics for TR/PR (outputs) vs. NT (input).

TR	1	3	4	6	7
R ²	0.010	0.018	0.079	0.013	0.038
P-value	0.002	0.006	< 0.0001	0.024	< 0.0001

PR	2	3	5	6	7
R ²	0.005	0.007	0.026	0.003	0.242
P-value	0.231	0.225	0.001	0.609	< 0.0001

Figure 6: The TR/PR controls over the data phase. The circle size indicates the replay count.

In the normalized time line for each task, the regression analysis (Table 3; Figure 2) indicates that there are common directionalities in the user exploration affected by the time progression. Particularly, the subjects seem to utilize the TR control for exploring the data across all tasks. PR, on the other hand, only had correlations to time in the PR-only local-global structure analysis (5) and the musicality task (7). This suggests the limitation of PR for analyzing particular attributes, such as the sample- or local-level details.

Figure 7: The standard deviation of TR/NR over NT.

Would the time progression directly affect the choice of a commonly-accepted resolution value? Figure 7 shows the amount of convergence (standard deviation) toward a specific resolution value over time. The results, however, did not indicate such converging behaviors except for PR in tasks 2 and 7 and TR in task 4. This may suggest the stronger influence of individually-controlled TR on the local-global analysis while PR has more influence on the details-local analysis as well as the musical-balancing task.

4.5.2. Explorations over Data Phase

Do the subjects react similarly to any “points of interest” in data? To examine this, we compared their exploratory paths with regard to the phase of data, where, within the total duration of NT or AT, each user may repeat the playback and adjust the previously-set resolutions.

Table 4: The 3rd-order polynomial regression model statistics for TR/PR (outputs) vs. phase (input).

TR	1	3	4	6	7
R^2	0.022	0.006	0.005	0.006	0.006
P-value	< 0.0001	0.460	0.360	0.401	0.353

PR	2	3	5	6	7
R^2	< 0.001	0.023	0.007	0.018	0.029
P-value	0.928	0.037	0.324	0.060	0.001

The results (Figure 6, Table 4) are, to our surprise, almost contrary from the case of NT, that TR shows much less common directionalities than PR with regard to the phase. We consider that these high variance and random nature may indicate several possibilities:

- As the combined controls (tasks 3, 6, and 7) indicate the significance of phase vs. PR over TR, the PR is more useful in location-sensitive analysis.
- TR may be used more than PR for random-directional explorations. TR may be first used as the base line for PR adjustments in NT time scale, and PR is used for micro adjustments. As such, the results may be reflecting the “replay” factor of phase where each iteration of TR may start and progress from a much different base line.
- The TR control is not suitable for an “immediate” analysis sensitive to the phase. Or, this particular dataset or melodic configuration (e.g., speed) does not guide the TR analysis very well.

4.6. Analytical Balance vs. Musical Balance

Lastly, we observed the relationships between the final resolution values chosen for each task and subjective preferences of musicality using those controls.

TR analysis vs TR musical

PR analysis vs PR musical

Figure 8: fTR/fPR (inputs) vs. fTR/fPR in task 7 (output). The users are color coded.

Table 5: The linear regression model statistics for fTR/fPR (inputs) vs. fTR/fPR in task 7 (outputs).

Ctrl	TR				PR			
Tasks	1	3	4	6	2	3	5	6
R ²	<0.1	0.126	0.128	0.142	0.342	0.265	0.178	0.130
P-val	0.989	0.123	0.121	0.101	0.006	0.020	0.063	0.117

Figure 8 and Table 5 compare the analytical fTR/fPR with the musical fTR/fPR in task 7. Again, the final values are “static” and does not capture the time-varying explorations. However, as the analytical task moves from details-local to local-global, the TR seems to become more correlated whereas the PR slightly decorrelates. This may also confirm that the individual choices of fPR for analytical tasks are generally related to their musical preferences, thus stay stationary across tasks (see section 4.4). The fTR, in contrast, does not generally correlate with musical preferences, especially for details-local time-scale analysis.

5. DISCUSSIONS

In this study, we observed several contrasting patterns in how the listener explores and matches the TR/PR controls to the structures of the melody (data). In the task level, where we disregard the time progression, we saw strong correlations between the musical preferences and all pitch configurations, whereas the TR varied for both the perception of data and musical balances. Over the time line of data exploration, on one hand, the TR had much larger effects in all tasks until arriving the final configurations. This defied our hypothesis that TR/PR would affect separately on local and global structures. On the other hand, in repeated listening analyses over phase, the PR had more significance in the exploration. These suggest the durational time scales for listening (i.e., separate tasks, normalized time, and phase) may utilize different combinations of the resolution controls, but not as simple as time to the low-level and pitch to the high-level relationships.

The results also showed that the final and overall motions of the resolution parameters were highly divergent, perhaps reflecting the personal preferences for listening and analysis.

In this regard, it is important to note that, while the use of the time / pitch abstraction techniques may bring forward different time-scale structures in the melody to our attention, using a lower resolution physically filters out the detailed motions from the currently-produced sound. Thus, it should be encouraged to actively change and scan through different resolutions than assuming the final optimal value early in the exploration.

6. CONCLUSIONS

In this study, we examined the use of dynamic time and pitch resolution controls for listening and identifying various attributes of a melodic sonification. We argued the considerations in multiple hierarchical structures in a melody, and how to possibly measure them. Our experiment also emphasized the exploratory natures of listening and understanding a newly-encountered melody. Unlike a passive listening-and-response test, our interactive user test provided a real-time sonification with user customization and recording. This presented challenges in analyzing their exploratory approaches, which were highly variable and complex. However, the unique relationships that were found between different time scales and musical preferences may bring new design opportunities for creating engaging and dynamic melodic sonifications.

7. REFERENCES

- [1] D. B. Huron, *Sweet Anticipation: Music and the Psychology of Expectation*. MIT press, 2006.
- [2] P. Desain and H. Honing, “The Quantization of Musical Time: A Connectionist Approach,” *Comput. Music J.*, vol. 13, no. 3, pp. 56–66, 1989.
- [3] C. Roads, *Microsound*. Cambridge: MIT Press, 2001.
- [4] J. D. Kramer, “Multiple and Non-Linear Time in Beethoven’s Opus 135,” *Perspect. New Music*, vol. 11, no. 2, pp. 122–145, 1973.
- [5] J. D. Kramer, *The Time of Music: New Meanings, New Temporalities, New Listening Strategies*. 1988.
- [6] R. L. Alexander, J. A. Gilbert, E. Landi, M. Simoni, T. H. Zurbuchen, and D. A. Roberts, “Audification as a Diagnostic Tool for Exploratory Heliospheric Data Analysis,” Jun. 2011.
- [7] B. N. Walker and J. T. Cothran, “Sonification Sandbox: A Graphical Toolkit for Auditory Graphs,” in *Proc. ICAD*, 2003, vol. 3.
- [8] F. J. Gravetter and L.-A. B. Forzano, *Research Methods for the Behavioral Sciences*. Nelson Education, 2015.
- [9] P. R. Cook, “Sound Synthesis for Auditory Display,” *Sonification Handb. Logos Publ. House Berl.* 2011.
- [10] T. Painter and A. Spanias, “Perceptual Coding of Digital Audio,” *Proc. IEEE*, vol. 88, no. 4, Apr. 2000.
- [11] T. Bovermann, J. Rohrerhuber, and A. de Campo, *Laboratory Methods for Experimental Sonification*. Berlin: Logos Verlag, 2011.
- [12] R. Rowe, “Sub-Symbolic Processes,” in *Machine Musicianship*, MIT press, 2001, pp. 93–144.
- [13] T. Tsuchiya, J. Freeman, and L. W. Lerner, “Data-to-Music API: Real-Time Data-Agnostic Sonification with Musical Structure Models,” in *Proc. ICAD*, 2015.
- [14] R. M. Winters, T. Tsuchiya, L. W. Lerner, and J. Freeman, *Multi-Modal Web-Based Dashboards for Geo-Located Real-Time Monitoring*. Georgia Institute of Technology, 2016.

WAVE SPACE SONIFICATION

Thomas Hermann

Ambient Intelligence Group
Center of Excellence in Cognitive Interaction Technology (CITEC)
Bielefeld University, Bielefeld, Germany
thermann@techfak.uni-bielefeld.de

ABSTRACT

This paper introduces *Wave Space Sonification* (WSS), a novel class of sonification techniques for time- (or space-) indexed data. WSS doesn't fall into the classes of Audification, Parameter-Mapping Sonification or Model-based Sonification and thus constitutes a novel class of sonification techniques. It realizes a different link between data and their auditory representation, by scanning a scalar field – defined as wave space – along a data-driven trajectory. This allows both the highly controlled definition of the auditory representation for any area of interest, as well as subtle yet acoustically complex sound variations as the overall pattern changes. To illustrate Wave Space Sonification (WSS), we introduce three different WSS instances, (i) the *Static Canonical WSS*, (ii) *Data-driven Localized WSS* and (iii), *Granular Wave Space Sonification* (GWSS), and we demonstrate the different methods with sonification examples from various data domains. We discuss the technique and its relation to other sonification approaches and finally outline productive application areas.

1. INTRODUCTION

Sonification allows to experience data by listening, enabling listeners to understand variations in data as they manifest in systematic variations in auditory space [1]. It affords a systematic connection of the *data space* to the *sound signal space* which in turn, on listening, navigates the *auditory perceptual space* and ultimately leads to shifts in a *cognitive space of interpretations*, i.e., result in insights. The art of sonification design is to bring together knowledge of data, sound and perception to craft a suitable linkage between data- and sound signal space that allows meaningful patterns in data space to translate to distinguishable patterns in perceptual space and hence, give listeners insights about patterns in the data or awareness on their change, whatever the purpose of the auditory display might be.

Many sonification techniques have been introduced over the past decades, which illustrates how complex, rich and different a linkage can be achieved (s. icad.org and [2]). From the perspective of *structure*, these sonification techniques can roughly be organized into five different classes: (1) earcons, (2) auditory icons and parameterized auditory icons, (3) parameter-mapping sonification (PMSon), (4), audification, (5) model-based sonification (cf. corre-

Figure 1: Wave Space Sonification: data-driven trajectories probe a static or data-refined wave space scalar field.

sponding chapters in [2]). Of course, the classes are never sharply separated and many sonifications combine different sonification techniques into hybrid sonifications.

Focusing on time-indexed data, as they occur in all walks of life and data science, from biomedical data to motion analysis, to name a few, all the techniques have offers to make, yet the dominating methods are audification and parameter-mapping sonification. Let's review how in these approaches, data space and sound space are connected.

In audification, the relation is most intimate: the data speak for themselves, as they are used as instantaneous sound pressure level that constitutes a sound signal as function of time. Although many useful processing steps such as resampling, filtering, non-linear warping and time-stretching allow to refine the transition between data and sound space, data *become* a signal.

In parameter-mapping sonification, the relation is mediated: data features are used as input for a mapping functions which computes synthesis parameters for a sound synthesis. While being highly flexible and adaptable, the nature of the relation between data and signal is that data *control* (rather than *become*) the sound signal.

Are there any more different ways of linkage? Model-based sonification offered yet another way to connect the spaces: here the data are neither becoming a sound nor used as control, but the data are turned into (or parameterize) a sound-capable system (in other words: the instrument), while interaction with them via the model is left to the user. As, however, relations in the data systematically translate to properties of the 'data-informed instrument', it offers

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

a versatile and highly interaction-builtin way to explore structures in data.

We see that all these approaches exercise a qualitatively different paradigm how data and sound signals are connected. This paper introduces yet another, novel mechanism for this connection: the data *define the navigation of high-dimensional sound signal spaces* – resulting in a new cluster of methods which are in between parameter-mapping sonification and audification, yet different in structure from both, as will be shown below. This approach is primarily intended for time- (or space) indexed data, and thus we start with a focus on multivariate time series data in the following section. In turn, WSS will be formally introduced and then illustrated by few example instances demonstrated with sonification examples. This should provide ample orientation to discuss how WSS relates to audification, parameter-mapping sonification and model-based sonification in Sec. 5. The paper concludes with an outlook on future work.

2. MULTIVARIATE TIME SERIES IN STATE SPACE

A necessary requirement for WSS is that the data are indexed according to a scalar variable. The most frequent case for that is that of time-indexed data, i.e., time series. As most phenomena of interest deliver several measurements at each point in time, the more general class is that of multivariate time series. Examples are EEG, ECG, EMG, inertial sensor data, motion capturing data, climate models, to name a few. However, WSS will also work if the index is of other nature, e.g. a spatial variable, such as ice core or other geological probe properties varying over a spatial index. As time series are the more typical situation we anchor the mathematical treatment on an index t .

Let $\vec{x}(t)$ be a d -dimensional measurement of a system over time t . A discrete measurement process of n steps yields a dataset $X = \vec{x}(t_i)_{i=1\dots n}$. Often, but not necessarily, timestamps are equidistant, i.e., $t_i = i \cdot T = i/\nu$, where ν is the sampling rate in Hz and we can more easily write it as $\vec{x}_i$. A frequently used visualization of such data is by plotting component functions of time, as depicted in Fig. 2 (left). If we interpret measurement tuples as points in a vector space, we can consider the data to describe a trajectory in a d -dimensional space. Fig. 2 (right) depicts for illustration the linear projection onto a 3D-subspace spanned by 3 features.

This perspective stimulates interesting paths to deepen the understanding of the phenomenon. For instance, if we see orbits, this corresponds to oscillations in the time series; if we see convergence to a point, we understand a fix point attractor. If we see that trajectories cross each other, i.e., there are divergent trajectories from one and the same point, we understand that the representation is not a *state*, but merely a low-dimensional projection of the truly higher-dimensional situation. In other words, knowing a vector doesn't allow us to infer in what direction it will continue.

For that sake, methods are used to extend the vectors so that they are *states*. One of them is delay-embedding, which augments a time series $\vec{x}_i$ by its past version $\vec{x}_{i-k}$ delayed by k time units, into a now $(2 \cdot d)$ -dimensional vector. Figure 3 illustrates delay embedding using a 1-dimensional time series for two delays. While only low-dimensional projections (up to 3) are attractive for visualization, higher-dimensional embeddings can easily be created (and may be necessary to yield a *state*), yet visual inspection is clearly limited. Here lies an important selling point for WSS, as the audibility of the trajectories in the truly high-dimensional space is

Figure 2: Typical plot of a multivariate time series (top) and a projection of the corresponding trajectory on three selected axis (right). The example shows human EEG during an epileptic seizure.

possible without problems, as we will see, resp. hear.

We see that details of the dynamics manifests in properties of the trajectories, such as velocity and its change, curvature and its change. In modern medicine and systems biology there is the notion of ‘dynamical disease’, which argues that we fail to understand disease by merely looking at averages (e.g. hormone levels, body temperature, etc.) but instead disease manifests in their rhythmical orchestration and thus how values change in concert with each other. From that perspective we can take additional motivation to strive at visualizations and sonifications which make such coherences explicitly perceivable as shapes that depend on the dynamics [3].

While time series analysis is a very elaborated research area, the above-introduced basics should suffice to unfold Wave Space Sonification.

Figure 3: Delay-embedding of a one-dimensional time series (few cycles shown top) using delay $\tau = 16$ (left) and $\tau = 42$ (right), here showing human ECG.

3. THE FRAMEWORK OF WAVE SPACE SONIFICATION

Wave Space Sonification, as motivated in the introduction, offers a novel approach to connect data space with sound (signal) space, by means of a data-driven sampling of audio signals organized in a high-dimensional scalar sound field, called here *wave space*. So what is a wave space? We are familiar with the idea of audio signals stored in a vector (i.e., buffer or table) as sequence of samples, if we look at it as a one-dimensional sound space. Using a phasor at constant speed and moving along the dimension we reproduce the stored sound at a specific playback rate, associated with the phasor increment. WSS generalizes this process by defining the sonification to be the scanning of a sound space according to the trajectory that a time series forms in its vector space. Let's define it formally before we continue to specify useful fillings of wave space.

Let $\vec{x}(t)$ be a sampling of time-indexed data X , either as continuous function or time-discrete sampling, e.g. $t_i = i/\nu$. Let $\vec{v}(t) \in \mathbb{R}^d$ be a trajectory, for instance as directly used by taking the multivariate time series $\vec{x}(t)$ and/or augmented by selected appropriate delay embeddings. We obtain $\vec{v}(t) = E(X)(t)$ as an embedding of X into a d -dimensional space. Furthermore, let $V : \mathbb{R}^d \rightarrow \mathbb{R}$ be a scalar function which we call *wave space* function. Finally we need a warping or morphing function $M(t)$ which allows to control how exactly we move along the trajectory, which could be different from how the original time series moves.

With the above assertions, we define a *wave space sonification* (WSS) as the transformation of data into a sound signal via

$$s(t) = V(E(X)(M(t))) = V(\vec{v}(M(t))) \quad (1)$$

which is basically created by sampling the wave space function while moving along the embedded trajectory by means of a chosen time-advancement morphing.

WSS defines sonification as a scanning process by means of (a) a trajectory definition in wave space, and (b) a suitable definition of a wave space function, and (c) a suitable way of moving along the trajectory. The method renders interesting *auditory views* already for simple and static choices for E , V , and M . However, a particular benefit will arise from regimes where the wave space is filled or modified according to the trajectory itself, as will be demonstrated later on.

As to (a), the trajectory definition, Section 2 has already featured the most important ideas. However, it may be useful to additionally apply dimensionality reduction methods to limit the complexity of the space, e.g., by applying a principal component analysis (PCA) [4] to select only those dimensions in which the trajectory exhibits large variations. Often this (or other data mining methods for pre-processing) can remove irrelevant detail and facilitate focusing on key properties at the same time. The trajectory definition might furthermore involve some filtering to remove random processes, such as additive measurement noise, to get smoother paths through wave space, which in turn translates to clearer signals in auditory space.

As to (b), the definition of the wave space function can be achieved at hand of different approaches. Firstly, it can be explicitly specified by using an algebraic expression such as

$$V(x_1, \dots, x_d) = \sum_{i=1}^d h_i(x_i) \quad (2)$$

for given functions $h_i : \mathbb{R} \rightarrow \mathbb{R}$ (see examples in Sec. 4.2). Secondly, it can be defined by help of given sound signals $s_i[n]$, us-

ing an interpolation/resampling function to define associated real-valued functions h_i , for instance to be used in the linear combination in eq. (2) above. Thirdly, it is possible to anchor specific sound signals at user-selected locations with a defined orientation in wave space by using a mixture of signals approach, as explained next.

Assume we wish to anchor K localized wave fields $\phi_i()$ at given centers $\vec{\mu}_j$, $j = 1, \dots, K$ with given orientation defined by a transformation matrix $\Sigma_j = U_j D_j U_j^T$. We can then compute V by

$$V(\vec{v}) = \sum_{j=1}^K \phi_j((\vec{v} - \vec{\mu}_j)\Sigma_j(\vec{v} - \vec{\mu}_j)^T). \quad (3)$$

If the localized wave fields contribute only around the origin, for instance by using

$$\phi_j(\vec{v}) = s_j[v_1] \cdot \exp(\vec{v}_{2\dots d}^2/\sigma_j), \quad (4)$$

the resulting WSS will exhibit the selected timbre of $s_j[\cdot]$ whenever the trajectory passes close to location $\vec{\mu}_j$. Depending on the chosen morphing, the sample is activated at higher or lower sampling rate.

As to (c), the suitable way for moving along the trajectory, the first and most basic choice would be to let the trajectory itself determine the speed. The instantaneous velocity is then proportional to the distance between subsequent embedding points, and so velocity $\propto \vec{v}[n+1] - \vec{v}[n]$. It makes sense to interpolate the segmentwise velocity linearly to avoid unsteady changes while moving along. This way, faster trajectories naturally translate to up-sampled signals. Note that this would mean an equal number of samples in wave space for any segment $[\vec{v}_j, \vec{v}_{j+1}]$ in embedding.

In contrast, one could also define a morphing that moves along the trajectory but at *constant* velocity, which is called the natural parameterization (also: unit-speed or arc-length parameterization) in Differential Geometry of curves. We skip mathematical details, yet given a set of embedding points $\vec{v}_i, i = 1, \dots, N$ we could just sample each segment $[j, j+1]$ with a number of samples in wave space that is proportional to the length $\|\vec{v}_{j+1} - \vec{v}_j\|$.

Any interpolation in between these two extremes, and also any extrapolation can be achieved, and they can all be characterized by means of a morphing function $t' = M(t)$, which could simply be said to *warp time*.

Let's look at some practical realizations of WSS next, before we discuss the method further.

4. WAVE SPACE SONIFICATION INSTANCES

To better understand how WSS translates multivariate time series into sound, let's explore a few examples. Same as Model-based Sonification (MBS), as overarching framework, pre-structures many different sonification models which can be called instances of MBS, WSS can have different instances, according to the choice of embedding, wave space and morphing. We start with a most basic WSS instance, that might be called *canonical wave space sonification*, proceed with a *data-driven localized wave space* definition and close with GWSS, the *granular wave space sonification*. Sound examples are provided via the following DOI: [10.4119/unibi/2919709].

4.1. Canonical Wave Space Sonification

The most basic choice of constituents for WSS is to (a) use the data-driven trajectory itself for embedding, (b) define the wave space as a sum (or product) of mathematical functions, and (c) use a trajectory-controlled speed, i.e., a temporal morphing $M(t) = t$. A canonical choice for wave space is

$$V(x_1, \dots, x_d) = \sum_{i=1}^d \sin(2\pi\omega_i x_i). \quad (5)$$

If angular frequencies $\omega_i = \omega_0 \cdot i$ are chosen as multiples of a fundamental ω_0 , a straight line trajectory along the i -th axis will translate into a sine signal of frequency ω_i , i.e., a harmonic of frequency ω_0 .

Let's assume for example the multivariate time series $\vec{x}(t) = (\cos(2\pi\omega_0 t), \sin(2\pi\omega_0 t))$. The trajectory moves on a circle in wave space at constant speed. At $t = 0$, only x_2 changes as x_1 varies little around the maximum. Hence, the WSS produces a tone of frequency ω_2 . This frequency decreases as the sine's slope decreases while at the same time the frequency ω_1 of the x-component increases until $t = \pi/2$. This transformation between activation of the ω_2 and ω_1 sine continues and becomes the audible pattern as result of the sonification. Fig. 4 depicts the trajectory, the wave space and the sound signal. The sonification is provided as Sound Example S1¹.

Figure 4: Canonical Wave Space Sonification - the trajectory is a circle, the 3D plot depicts it as spiral using t as z -axis. Right side plots show the beginning of the sonification (first, below the spectrogram).

Things become auditorily more interesting when more complex timbres are used to define or fill the wave space. For instance using the same frequency along x_1 and x_2 , but only even/odd harmonics along x_1/x_2 . Then the trajectory manifests as timbre transition between two timbres. (sonification example S2).

The following sound examples S3.1-4 illustrate canonical WSS with the monthly number of sunspots dataset (measurements over 150 years which exhibits the 11-year solar cycle) [5]. The embedding of this one-dimensional time series into 2D-space is obtained by using a delay-embedding of $k = 60$, i.e. roughly 5 years. According to the choice of the rate parameter, which scales the velocity at which the trajectory is scanned, very different auditory views emerge, from sonifications that allow to understand

the overall evolution over time (e.g. how does the sunspot oscillation vary over centuries) (see sound examples S3.2 and S3.4) to auditory views that allow to understand individual orbits (e.g. fast/slow, small/large) and compare their inner shape and structure (see sound examples S3.1 and S3.3).

Figure 5: Wave Space Sonification of monthly sunspot data: top: time series of monthly sunspots and two delay-embedding of the filtered signal below, with $k=6$ (left) and $k=24$ (right). Listen to sound examples S3.1-4.

4.2. Static Sample-based Wave Space Sonification

Instead of using merely mathematical functions to define wave space, it is well possible and acoustically much more interesting to define wave space by means of available samples, i.e., recorded sound signals. The most straightforward way is to choose a sound file $s_i(t)$ for each dimension $i = 1, \dots, d$ of wave space and define the wave space function as

$$V(\vec{x}) = \frac{1}{d} \sum_{i=1}^d s_i(c_i \cdot x_i), \quad (6)$$

where c_i are scaling factors that allow to tune each dimension differently.

As a first sonification example, listen to the sunspot orbits where the x_1 dimension is filled with the spoken vowels 'a-e-i-o-u' (vowels as in bath-bear-bee-beau-boo) with a short gap in-between, and along the x_2 it is left silent. The sonification examples S4.1-2 (using the sunspots data set) and S4.3-4 (for ECG, realtime and /4 slowdown) illustrate that now much more understanding of the dynamics comes from the more structured layout. This WSS is called static as the wave space filling does not depend on time or on the data. Both extensions are possible and likely paths towards more useful sonifications. Of the many possible paths, the following section examines the manual adaptation of wave space to given data.

4.3. Data-Driven Localized Wave Space Definition

In many situations time series form a 1D-attractor, i.e., a limit orbit, around which deviations occur depending on initial conditions or external perturbations. For instance, the human ECG signal of a

¹All sound examples are provided with description on <https://doi.org/10.4119/unibi/2919709>

healthy subject will have some variations between individual heart beats but largely exhibit a structurally reproducible pattern, as depicted in Fig. 3. For understanding properties of the cycle in detail, it would be beneficial to craft *data-driven wave space functions* so that samples are centered at relevant locations and aligned to the dominant direction in the delay-embedding. For instance, the QRS-complex (the peak of the heart beat) manifests for the large slightly rotated D-shaped orbit in the lower left plot of Fig. 3, shown also in Fig. 6. A data-driven wave space definition could position a sound sample $w_1(t)$ on the horizontal part, and likewise $w_2(t)$ and $w_3(t)$ on the vertical part and along the main diagonal at the origin. Naturally the sonification would exhibit a sequence of w_1 , w_2 and w_3 sounds, depending on the idiosyncrasies of individual spikes.

Specifically, we suggest a mixture of k localized *wave space hyper-cylinders*², centered at locations $\vec{\mu}_k$, oriented along unit vectors $\hat{u}_k$ by

$$V(\vec{x}) = \sum_{k=1}^K g_k \cdot w_k(r_k \hat{u}_k^T (\vec{x} - \vec{\mu}_k)) \cdot K_\sigma(\|(\mathbf{1} - \hat{u}_k \hat{u}_k^T)(\vec{x} - \vec{\mu}_k)\|),$$

where g_k is the gain of $w_k(t)$, which is a sound sample, and $K_\sigma(y)$ is a window function that decays with increasing y such as $K_\sigma(y) = \exp(-y^2/2\sigma)$. The scalar factors r_k allow to stretch or squeeze the sound sample along $\hat{u}_k$.

The data-driven localized wave space sonification is of practical utility if the analysis focus is on deviations of orbits from a reference orbit, such as in monitoring situations, diagnosis or exploratory data analysis contexts.

Sound example S5.1 is a sonification of the ECG cycle depicted in Fig 6, using a sound sample of spoken letter 'r' and 's' for the outgoing (horizontal towards the right) and returning (vertical, downward) trajectory of the QRS-complex. A whistling sound is added to the origin with a 45° degree orientation, which is turned into a discernable pattern by the other parts of the ECG (u-wave, p-wave, etc.). The used samples were quickly recorded with the laptop microphone and are provided as sound example S5.0. Sound example S5.2 represents a sonification at 1/4 the rate, allowing to perceive the signal between the QRS-complexes in greater detail.

4.4. Granular Wave Space Sonification

The instances of WSS introduced above use the state trajectory directly to sample the Wave Space, and thus have a highly direct coupling between trajectory dynamics and sound signal. However, this comes at some disadvantages. On the one side, with the data-driven velocity mode, the changing trajectory speed causes strong (often too extreme) resampling so that the timbre cannot anymore be properly recognized, yet at least the timing is correct, i.e., sonification time is proportional to data time. On the other side, using the 'unit speed' natural parameterization or arc-length velocity mode, the extreme sampling problem is solved but then the temporal pattern is strongly distorted, causing short large spikes (such as the QRS-complex in the ECG data example) to be represented by much more sonification time than it proportionally takes within the cardiac cycle.

To overcome these limitations we finally introduce *Granular Wave Space Sonification*, which decouples the trajectory from the sound-collecting process to some degree. While moving along

²'hyper' because the definition is not only for 3D but any dimension

Figure 6: Localized Wave Space Sonification of ECG: the spoken sound 'r', 's' and a whistling sound at constant pitch are inscribed into wave space at the visible location and orientation. The whistling sound is high pitched and more quiet and thus less visible here unless zooming in.

the trajectory we regularly spawn 'particles' to probe wave space along the trajectory. These particles have a limited range and can propagate at their own velocity, independently of the moving trajectory. According to the morphing function (or velocity model) this spawning could happen equally often per second, or whenever a determined distance has been moved along the trajectory. The *Granular Wave Space Sonification* (GWSS) is then simply the superposition of all particles' individual wave space sonifications. However, as usual in granular synthesis, we also apply to each grain a temporal amplitude envelope window (hanning window) to reduce transient effects.

Compared to standard WSS we gain three control parameters: (i) the particle rate (in particles per second), the duration (in seconds) of the particles' life time, allowing the sound to become a mixture of slightly different particle views on the time series, and finally (iii) the propagation speed of particles in unit length per second, which can be adjusted between a narrow local spatially confined focus at slow speed and long-ranging particles, even up to half a cycle or more at large speed.

The following sound examples illustrate GWSS for the ECG data set used before. A sonification of the Lorenz attractor and EEG data will be made available as supplementary material and shown at the conference. Sound example S6.1 is a GWSS rendered with a particle rate of only 3 Hz, particles of 20 msec, using the static WSS wave space function. You can hear the well isolated grains and distinguish their timbre. Compare this to sonification S6.2., where the particle duration has been increased to 100 ms. Here the temporal contour in each grain becomes perceivable. In sonification S6.3 the short grains are kept but the particle rate is

increased to 20 Hz. Try to attend to the rhythm of the heart beat (0.5 Hz, as a 2x slow-down is used). This is much easier in sonification S6.4 where the particle duration is increased to 3 ms and the particle rate has been increased to 40 Hz so that grains blend and you can follow the 5 heart beats. If more details are needed without affecting the spectrum, the rate can now easily be changed, as for instance to a 6x slow-down in sonification S6.5.

Of course, the GWSS can also be used with any of the other proposed wave space fillings, such as the localized wave space samples. Examples S6.6 and S6.7 illustrate sonifications for a used spoken 'a-e-i-o-u' sound file laid out along the x_0 -axis.

5. DISCUSSION

The above presented WSS instances demonstrate that Wave Space Sonification is a structured yet still quite versatile and flexible framework to investigate time- (or space-) indexed data. Depending on the selection of the rate, different phenomena in given data can be explored – from differences in-between individual orbits to the overall structure of the dataset as a whole, including trends and changes in rhythmical organization. The possibility to define the wave space function at hand of user-selected timbres enables to rather easily create sonifications that meet expectations in timbral complexity or pleasantness. For instance, very precise but slightly annoying wave files which might be perfect for analysis and diagnosis could then easily be replaced by less precise but much more pleasant sounds, whatever is regarded as such, for using the system in a real-time monitoring scenario.

The number of parameters to control WSS is nicely limited, as too many parameters would only make it overly difficult to adjust the system to one's needs. They include the embedding, the rate, spatial scale, selection of morphing function, selection resp. definition of wave space. Depending on the latter, a number of additional parameters might be needed. For the GWSS, we furthermore have introduced particle duration, particle rate and particle speed, which all are self-explanatory when you understand the sonification process.

The chosen datasets for sonification examples have in common that their trajectories are repetitive with variations, i.e., they are quasi-periodic. This is a particularly useful situation as then the ear can easily understand the cycle, learn and memorize the basic pattern and subsequently focus on differences therein. However, non-stationary patterns would also make sense for WSS, for instance in the case of monitoring: assume that due to a trend or change of parameters the trajectory shifts to another area of wave space which is occupied by a completely different timbre – the listeners will very likely notice the salient change corresponding to such a non-periodic shift.

The second commonality in the datasets is that the trajectories are sufficiently smooth. This makes both the visualization easier to follow and the sonifications smoother. Smooth trajectories can usually be obtained for any data by applying a low-pass filter. However, it often makes also sense to apply a high-pass filter to remove DC-offsets or any slow waves that would manifest in delay embeddings as spatial translations of the orbits, so that practically a band-pass filter with user-selected low- and high-frequency cut-off is the best.

How WSS performs without this filtering, and in how far the filtering may remove relevant information from the underlying data depends on the data and remains to be learnt in long-term experience with – and further investigations of – WSS.

In the introduction, we advertised WSS to be a new class of sonification technique, on the same level as audification, model-based sonification and parameter mapping sonification, yet not well fitting into any of these. Thus, the relation of WSS to these techniques should be clarified more thoroughly.

5.1. WSS is not Audification

Audification is indeed a special case of Wave Space Sonification: If we take the wave space definition to be $V(\vec{x}) = V_{\text{aud}}(\vec{x}) = x_1$, then the WSS is exactly the audification of the feature represented as the first feature, i.e., x_1 . If the state space is defined by a delay-embedding, then x_1 is just directly the series of measurements.

This makes it clear that in turn WSS can not be a special case of audification. Audification, in its original form, doesn't generally include the application of arbitrary transfer functions apart from probably dynamic compression which is sometimes considered as signal conditioning [6]. In WSS, an audification with dynamic compression is likewise obtained by using $V(\vec{x}) = h(x_1)$ where $h(x)$ is a sigmoidal transfer function. More advanced approaches for signal conditioning in audification is the application of a phase vocoder or buffer granulator (similar to TGrains UGen in Supercollider3) to decouple time progression to some degree. This would also be a special case of WSS using $V_{\text{aud}}(\vec{x})$ with the GWSS introduced above.

Looking at audification as special case of WSS illustrates that WSS is a more general approach, and only in very special situations a WSS might appear to be an audification. The most important differences are that (i) typically not the data series alone but a data-orchestrated trajectory in data- (or state-) space is used for probing the wave space, that (ii) the wave space is a multi-dimensional signal and thus much more complex than a linear or smooth transfer function, and consequently (iii), that thus already slowly oscillating signals such as 1 Hz rhythms can deliver audible results – whereas audification would require significant temporal compression to shift oscillations to the audible frequency range which would in turn reduce the temporal resolution.

Audification and WSS share the problem that a speedup (i.e., resampling in audification resp. scaling of the embedding in WSS) can easily result in aliasing and cause misleading spectral reflections at the Nyquist frequency ν . As the wave space might usually be already filled with many frequencies (up to ν), ways to realize band-limited sonification would be a very useful and relevant focus of ongoing research.

Extensions of Audifications, such as *Augmented Audification* by Vogt and Höldrich [7] perform frequency shifting using a single-side-band modulation. This is an extension that shares with WSS the ability for flexible adjustment of the sonifications frequency range and at the same time and independently, to interactively zoom into temporal structures at any scale, yet it is achieved with a different method (multiplying vs. concatenating functions), so that the two techniques do not relate easily to each other.

5.2. WSS is not Parameter Mapping Sonification

In Parameter Mapping Sonification (PMSon), data are mapped to synthesis parameters of a sound synthesis algorithm, thus allowing to route different data features (i.e., components of a data vector) to different auditory qualities. If we would use a synth that taps into a wavetable and map our given time series data to the phasor (or tapping location) we would look at WSS from the perspective

of mapping. So, isn't WSS just a PMSon? The described PMSon is possible but w.r.t. the key idea of PMSon, it is a bit pathological: usually in PMSon we would define our parameters to systematically vary an auditory quality, e.g., pitch, level, duration, attack time, etc. This is not the case with mapping on a phasor.

Also, with the same argument we could explain an audification to be but a PMSon – which is obviously not helpful, as in terms of the central ideas these methods are clearly distinct.

Looking at WSS as a mapping doesn't do justice to its rather signal-near (to the level of sample-based control) nature of how data relates to the sonification – in parameter mapping this is usually much more indirect and usually driven by the wish to map a variable's change to likewise perceivable change of a perceptual quality.

Furthermore, WSS abstracts from the underlying data towards the definition of an embedding which is then subjected to a morphing before accessing the wave space, so we can't really say that the data index directly a wave table. Finally the WSS is non-parametric with regards to data and sound parameters: while in PMSon problems arise when there are more features than synthesis parameters – such as which of them to ignore – WSS is immediately applicable for data of any dimension.

5.3. WSS is not Model-based Sonification

WSS uses the idea of movement in space, which may remind to the idea of model space in Model-based Sonification. So can we regard WSS as a sonification model? The key idea in Model-based sonification is that the data *become* a dynamic system, capable of showing acoustic behavior. This is typically achieved by using the data to determine the setup of elements in the model, e.g. mass-spring systems, etc. An MBS is usually in a state of equilibrium and thus silent until excited by interaction, and this interaction is usually initiated by the data-exploring user. So, this is structurally quite different from WSS, where the data can more be said *to operate on or to drive* wave space than *to become* wave space. Thus WSS is clearly not a MBS.

In summary, we conclude that Wave Space Sonification is best regarded as a new class of sonification methods, conceptually closest to audification, and equally distant to parameter mapping sonification and model-based sonification, sharing some characteristics from the latter.

5.4. Related Sound Synthesis Sonification Methods

Key ingredients of WSS are the definition of wave space and the use of a trajectory that slowly taps into it and a scanning of wave space. We don't know of any other related work that uses this specific approach for sonification of multivariate time series. So we start with the more distant area of computer music.

With *Scanned synthesis*, Verplank et. al. [8] introduced a similar idea in computer music, however, more related to allowing users to interactively control timbre for musical expression. Their aim was to improve the human interface in controlling complex timbre using dynamical systems whose 'state' becomes the wave form that is scanned.

The Wave Space, if filled with sound files, is related to *Wave Terrain Synthesis* by Roads [9], which exercises a similar generalization from 1-dimensional wavetables to 2 or 3-dimensional wave fields, but again only for the sake of computer music applications.

Note that the Wave Space Sonification here is generally introduced as a scalar field of d dimensions, where d is the dimensionality of the multivariate data. The definition was tailored to match properties of time series in state space, which can be high-dimensional and visualization falls short above 3 dimensions. Here, Wave Space Sonification might help to give extra information, particularly when combined with interactive and dynamic visualization.

The idea of following a trajectory in data space for defining the sonification is for instance used in Principle Curve Sonification by Hermann et al. [10]. But being a MBS, here some other acoustic processes are needed to explain the audio signal.

Williamson and Murray-Smith introduced Sonification of probabilistic feedback using granular synthesis [11]. Here the idea is to facilitate the operation of a control space by using granular synthesis to experience the probabilities of likely goals, using display quickening. While the combination of using state space, particles and granular synthesis makes this approach to appear similar, it's main idea is interactive control space feedback.

Among the many papers on parameter mapping sonifications, one finds granular synthesis but not with much overlap to this definition of WSS here.

6. CONCLUSION

In summary, this paper has introduced Wave Space Sonification (WSS) as a new class of sonification techniques, starting from a formal definition, then presented concrete instances of WSS as new sonification techniques that apply the WSS logic to represent a number of test data sets from various areas. The sonification examples showed that very different aspects, ranging from detailed information of individual cycles to overarching structures such as trends manifest in sound. While the core concepts are outlined here, much research remains to be done to better handle issues such as aliasing in canonic WSS or providing guidelines how to best define Wave Space in case of higher-dimensional systems. This ongoing work will hopefully also provide some body of knowledge and experience how to shape wave spaces so that the resulting sounds are both informative and aesthetically pleasant.

As for applications, we see the highest potential in two areas: (i) in biomedical data contexts (analysis, differential diagnostics, monitoring), for instance in fields such as ECG and EEG, and (ii), in interaction scenarios such as in sports and rehabilitation, since in these domains we often have repetitive movement cycles that would naturally translate into orbits in wave space. Since WSS is realtime-capable, the sound can be perceived as immediate feedback during the activity, allowing users to interactively explore the interrelations between their movement execution and its consecutive Wave Space Sonification.

7. ACKNOWLEDGMENT

This research/work was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG). Thanks to A. L. Aldana Blanco for selecting and preparing the ECG data sample from PhysioNet. Thanks to the developers of python/numpy/sciopy/matplotlib/jupyter for their amazing tools.

8. REFERENCES

- [1] T. Hermann, "Taxonomy and Definitions for Sonification and Auditory Display," in *Proceedings of the 14th International Conference on Auditory Display (ICAD 2008)*, P. Susini and O. Warusfel, Eds. IRCAM, 2008.
- [2] T. Hermann, A. Hunt, and J. G. Neuhoff, Eds., *The Sonification Handbook*. Berlin, Germany: Logos Publishing House, 2011. [Online]. Available: <http://sonification.de/handbook>
- [3] G. Baier and T. Hermann, *Temporal Perspective from Auditory Perception*. World Scientific, 2008, pp. 348–363.
- [4] I. T. Jolliffe, *Principal Component Analysis*. Springer, New York, 1986.
- [5] B. WDC-SILSO, Royal Observatory of Belgium, "Sunspot data," <http://www.sidc.be/silso/datafiles>.
- [6] F. Dombois and G. Eckel, "Audification," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany: Logos Publishing House, 2011, ch. 12, pp. 301–324. [Online]. Available: <http://sonification.de/handbook/chapters/chapter12/>
- [7] K. Groß-Vogt and R. Höldrich, "Augmented audification," in *Proceedings ICAD 2015 (ICAD 2015)*, Graz (Österreich), 07 2015.
- [8] B. Verplank, M. Mathews, and R. Shaw, "Scanned synthesis," *The Journal of the Acoustical Society of America*, vol. 109, no. 5, pp. 2400–2400, 2001. [Online]. Available: <https://doi.org/10.1121/1.4744477>
- [9] C. Roads, *Computer Music Tutorial*. Cambridge, MA, USA: MIT Press, 1996, ch. Wave Terrain Synthesis, pp. 163–167.
- [10] T. Hermann, P. Meinicke, and H. Ritter, "Principal Curve Sonification," in *Proceedings of the International Conference on Auditory Display*, P. R. Cook, Ed. International Community for Auditory Display (ICAD), 2000, pp. 81–86.
- [11] J. Williamson and R. Murray-Smith, "Sonification of probabilistic feedback through granular synthesis." *IEEE Multimedia*, vol. 12, pp. 45–52, 2005. [Online]. Available: <https://doi.org/10.1109/MMUL.2005.37>

KEPLER CONCORDIA: DESIGNING AN IMMERSIVE MODULAR MUSICAL AND SCIENTIFIC INSTRUMENT USING NOVEL BLOCKCHAIN AND SONIFICATION TECHNOLOGIES IN XR

Kelly Snook, Tarik Barri, Joachim Gofmann, Jason Potts, Margaret Schedel, Hartmut Warm

University of Brighton
Fab Lab, EM13 Heavy Engineering, Lewes Road, Brighton, BN2 4GJ
K.Snook@brighton.ac.uk

ABSTRACT

This paper describes the first steps in the creation of a new scientific and musical instrument to be released in 2019 for the 400th anniversary of Johannes Kepler's *Harmonies of the World*, which laid out his three laws of planetary motion and launched the field of modern astronomy. Concordia is a musical instrument that is modularly extensible, with its first software and hardware modules and underlying framework under construction now. The instrument is being designed in an immersive extended-reality (XR) environment with scientifically accurate visualizations and data-transparent sonifications of planetary movements rooted in the musical and mathematical concepts of Johannes Kepler [1], extrapolated into visualizations by Hartmut Warm [2], and sonified. Principles of game design, data sonification/visualization optimization, and digital and analog music synthesis are used in the 3D presentation of information, the user interfaces (UX), and the controls of the instrument, with an optional DIY hardware "cockpit" interface. The instrument hardware and software are both designed to be modular and open source; Concordia can be played virtually without the DIY cockpit on a mobile platform, or users can build or customize their own interfaces, such as traditional keyboards, button grids, or gestural controllers with haptic feedback to interact with the system. It is designed to enable and reward practice and virtuosity through learning levels borrowed from game design, gradually building listening skills for decoding sonified information. The frameworks for uploading, verifying, and accessing the data; programming and verifying hardware and software module builds; tracking of instrument usage; and managing the instrument's economic ecosystem are being built using a combination of distributed computational technologies and peer-to-peer networks, including blockchain and the Interplanetary Filesystem (IPFS). Participants in Concordia fall into three general categories, listed here in decreasing degrees of agency: 1) Contributors; 2) Players; and 3) Observers. This paper lays out the broad structure of Concordia, describes progress on the first software module, and explores the creative, social, economic, and educational potential of Concordia as a new type of creative ecosystem.

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

1. INTRODUCTION AND HISTORICAL CONTEXT

2019 marks the four-hundredth anniversary of Kepler's publication *Harmonices Mundi* (Harmonies of the World), a treatise on geometry, astronomy and music [1]. The book, commissioned by the City of Linz and published in 1619, details his first three laws of planetary motion, as well as his process of discovering them. The three laws, now well known, were not initially accepted, even by contemporary astronomer, Galileo, because they conflicted with accepted theological world-views based on the geocentric model of Ptolemy:

1. *All planets move about the Sun in elliptical orbits, having the Sun as one of the foci*
2. *A radius vector joining any planet to the Sun sweeps out equal areas in equal lengths of time*
3. *The squares of the sidereal periods (of revolution) of the planets are directly proportional to the cubes of their mean distances from the Sun*

Four specific aspects of Kepler's approach were new or unique and bear mentioning. First, his writings make clear that he was primarily motivated by exposing and making accessible the harmonic order that he considered to be God's organizing principle [3]:

I feel carried away and possessed by an unutterable rapture over the divine spectacle of the heavenly harmony.

This, in itself, was not unique in the early 17th century. But his dogged insistence on understanding and publicly exploring deep spiritual principles, rather than supporting any of the rival Christian dogmas of his day, led to all of them rejecting him and he was left to seek spiritual satisfaction in astronomy [4]:

...You see how near I come to truth. As often as this happens, how can you doubt that I amply shed tears! For God knows, this *a priori* method [the scientific method developed by him and Galileo] serves to improve the study of the movements of the celestial bodies; we can put all our hopes into it if others cooperate who have observations at their disposal... I am eager to publish soon, not in my interest, dear teacher... I strive to publish in God's honor who wishes to be recognized from the book of nature. But the more others continue in these endeavors, the more I shall rejoice; I am not envious of anybody. This I pledged to God, this is my decision. I had the intention of becoming a theologian. For a long

time I was restless: but now see how God is, by my endeavors, also glorified in astronomy.

Rejection from his native Lutheran Church caused him deep personal grief and sorrow. He found solace from this sorrow in his new astronomy and mathematics:

If God fulfills these wishes, my mathematics will always be ready to offer delight... and give relief from sorrow either in the astronomical practices or in the contemplation of the heavenly works of the harmonies of the universe.

Second, in true Renaissance style, in addition to deep mathematical, astronomical, and theological knowledge, Kepler also possessed profound understanding of the most recent music theory of his day. At a critical juncture in his development, he was inspired by the work of Galileo's father, Vincenzo Galileo [5], who examined music theory from a broad historical and newly experimental perspective, with special attention paid to early Greek theory and tuning systems. While the Pythagorean idea of the "Music of the Spheres" was widespread and even long-embedded in Christian dogma, Kepler approached music as a tool for physical investigation of reality with unprecedented rigor and enthusiasm.

Third, Kepler, during his tenure as Imperial Mathematician to the Holy Roman Emperor Rudolph, had sole access to the most comprehensive and accurate data ever collected in the meticulously compiled astronomical observations of his predecessor, Tycho Brahe. His refusal to accept previously-ignored or unknown discrepancies between observed data and prevailing models, whether Ptolemaic or Copernican, eventually led to his breakthroughs in discovering the correct mathematics reflected in his three laws. No one before him had so rigorously pursued precise agreement between theory and observation, one of his principal contributions to the scientific method.

Finally, the fourth and perhaps most breathtaking and admirable aspect of Kepler's approach was his ardent prayer, granted several times over, to be relieved of any attachment to ideas that stood in the way of the truth he so tirelessly pursued. Most notably, when the unexpected mathematical truths of his three laws finally dawned on him, he was forced to let go of his cherished, physically and philosophically beautiful, early model of the solar system, consisting of nested platonic solids, which had launched his career and sustained his passionate inquiry for decades. His own discoveries undermined not only his contemporaries's worldviews, but his own favourite ideas. Coupled with this was his ability and willingness to extrapolate truths to the broader human condition:

While I struggle to bring forth this process into the light of human intellect by means of the elementary form customary with geometers... may He prevent the darkness of our mind from bringing forth in this work anything unworthy of His Majesty... Holy Father, keep us safe in the concord of our love for one another, that we may be one... just as through the sweetest bonds of harmonies Thou hast made all Thy works one... as Thou didst erect the heavens themselves out of harmonies.

In light of this brief examination of Kepler's approach, it is evident that modern science, by distilling his contributions to their most basic mathematics, does a disservice to his multidimensional, uniquely multidisciplinary pursuit of truth, which called on every possible human faculty and begs further research. Kepler, himself, issued a call to future musicians to investigate [1]:

...There is need for louder sound while I climb along the harmonic scale of the celestial movements to higher things where the true archetype of the fabric of the world is kept hidden. Follow after, ye modern musicians, and judge the thing according to your arts, which were unknown to antiquity.

Although it is unorthodox in modern times to supply such a detailed religious and historical account in a scientific paper, it is necessary to demonstrate the various ways that Concordia is a direct response to this call by Kepler. Far beyond a simple academic exercise in data sonification, it is a scientific and musical instrument designed explicitly to celebrate and make accessible Kepler's ideas as an immersive, visceral, educational experience. The creation of a modular, customizable musical instrument, as opposed to a musical performance or recorded piece of music or art, gives agency to each player to explore deeply their own truths. Building such an experience requires the application of modern astrophysical data, computational architectures and music technology to modern datasets, to create a comprehensive, scientifically accurate solar system model with abstract representations of geometric harmonies superimposed. Both the hardware and software for Concordia will be open source, developed in modules, and uploaded to the Concordia blockchain and other distributed computational infrastructures. The launch in 2019 will feature live streamed virtual-reality compositions by musicians that can be observed from anywhere in the world.

2. A NEW TYPE OF MUSICAL INSTRUMENT

Figure 1: The overlapping arenas of Concordia and some examples of other established instruments, tools, and games

Lying at the intersection of music, science, technology, gaming, and pedagogy (Figure 1), Concordia draws elements from all in equal measure. It resembles a game-like interactive, immersive installation in which single or multiple players navigate through one or more datasets. The instrument's visualization/sonification software and hardware are both modular, similar to the pedals and

effects of a live guitarist, the modules of a modular synthesizer, or custom objects/patches in MaxMSP or Pure Data. Concordia provides the computational and philosophical framework upon which modules can be built, played, and observed, and has three main elements:

1. **Interactive data sources** - players choose which physical phenomena they would like to explore and are presented with a wide array of parameters to vary, including time, selection of planets, moons, etc. (specific controls are designed by module creators). Concordia data and the algorithms comprising each module are stored in distributed, synchronized peer-to-peer networked and verified databases;
2. **Immersive audiovisuals, haptics** - Players explore via VR headset, AR glasses, or 360 projection of a dynamic solar system to scale and other geometric objects highlighting harmonic relationships. Spatialized sound is either rendered binaurally to headphones or realized by a multichannel loudspeaker array;
3. **Hardware controls** - Controls can be minimal, such as through a tablet or computer, or can expand to match the imagination of players through an optional array of traditional or non-traditional controls for interaction with the data and navigation through space and time.

Figure 2: Tarik Barri playing a composition in Versum

Figure 2 shows Tarik Barri interacting with his Versum [6] audiovisual composition environment inside a projection dome. The initial Concordia module prototype uses Versum to display visualizations and sonifications of planetary data. Whereas Versum allows a composer unlimited creative freedom in sound and visual design, all sounds and visuals in Concordia modules are generated by planetary data. The release version of Concordia will behave like a game. It will use rewards and levels of play to train players gradually in the new listening and control skills required to understand, explore, and eventually achieve virtuosity with the instrument. Its modular, open-source, DIY hardware designs and distributed architecture of software invite participation globally and reduce barriers present in traditional music creation and collaboration.

Participants in Concordia fall into three general categories, listed here in decreasing degrees of agency:

1. **Contributors** - those who design, program, create and upload new sonification or visualization modules, new DIY hardware control interface designs for others to download and build, or new compositions within existing modules. Contributors could also be people who facilitate new use cases for Concordia, such as presets or ways of playing Concordia that can be used in teaching curricula.
2. **Players** - those who use the instrument controls actively to choose which Concordia modules to play, to choose which data within given modules to play, and then to navigate, manipulate, and explore the data and their mappings into the 4D immersive world of sound, visuals, and haptics. Players have the option of inviting others to observe their live explorations, and/or recording their compositions back into Concordia for others to play and remix.
3. **Observers** - those who “ride along” with someone else who is playing live or with a recording inside Concordia. This category of participant can act as an audience or beneficiary of experiences others are creating.

These three categories of Concordia participants together form a community that functions as a creative economy and a collaborative ecosystem based on built-in, blockchain-based, community-defined smart contracts for tracking the value of contributions and use. This allows Concordia to serve as a socioeconomic experiment as it grows and evolves.

2.1. Astronomical Sonifications and Visualizations

There have been numerous sonifications of planetary movement and other astronomical data [7, 8, 9, 10, 11], and sonifications directly referencing Kepler's work, notably Laurie Spiegel's *Kepler's Harmony of the Worlds*, which was included on Voyager's Golden Record [12], and Rodger's and Ruff's Realization for the ear of Johannes Kepler's astronomical data from *Harmonices Mundi* [13]. These sonifications owe a conceptual debt to Pythagoras's Music of the Spheres, a philosophical conceit that sought to describe the movement of the sun, moon, and planets using harmonic proportions. The Music of the Spheres was not thought to be literally audible, but rather to be a harmonic, mathematical or religious concept; in later Christian reinterpretation of the Music of the Spheres, instrumental music was thought to exist solely to make audible the other two types of inaudible music: the music of the spheres and the music of the human spirit.

As discussed earlier, Kepler was highly influenced by the idea of the Music of the Spheres, and he sought a rational explanation of the movements of the planets, of which only six were known in his time. After exhausting, unsatisfied, sonifications of planetary orbital distances and other static parameters, he progressed to more abstract and complex relationships between angular velocities of the planets during particular times in their orbits, such as aphelion and perihelion. Over the centuries since, the solar system has inspired sonifications too numerous to adequately survey here. Many recent solar system sonifications translate direct planetary measurements such as position, size, density, composition, etc. into controls for the sound. Quinton et al. [14] discuss important issues that arise in the context of their research into effective solar system sonification methods, if they are to be used for anything other than mere entertainment.

Through sonification, abstract concepts like planetary movements, can be made more tangible and

comprehensible to the general public. Since many sonifications have been created as artistic pieces a certain degree of scepticism from scientists has made them wary of using sonification as a scientific tool. Barrass refers to the conflict between the more traditional scientific view with relation to data analysis and the new view, which embraces the advantages of the human auditory system and its cultural significance. There is a fine line between sonification as a means of scientific exploration and merely being perceived as a popular mass media marketing tool.

Considering the large amounts of data involved in astrophysics the use of sonification in this field can be explored in more detail. The difficulty arises when trying to find the right balance between artistic input and sonifications that are usable in scientific data analysis [14].

The solar system is a structure of significant spatial extent and complexity offering a seemingly infinite variety of possible perspectives. Kepler was interested in the relationships between the heavenly entities, as opposed to any individual planet considered alone, so it is these relationships that are the focus of the first Concordia modules. In this arena, Hartmut Warm has created a body of work [2, 15] that extrapolates and extends Kepler's ideas by representing them visually and examining the geometric archetypes intrinsic to dynamic planetary relationships.

Warm has created a computer application based on open hybrid observational and theoretical ephemerides models (e.g. [16]) that visualizes a wide variety of geometric, spirograph-like figures discussed in the text. Using his app, one can calculate and draw astronomical objects and events, such as relative positions, velocities, conjunctions, and oppositions. The first Concordia module prototype then uses the strings of numbers output from Warm's code to generate 3D visual and sonic objects in Versum for module creators to experiment with mappings to waveforms, envelope shapes, filters, colors, visual shapes, and time-based effects.

The approach of Concordia, mirroring Kepler's, is to minimize the imposition of pre-formed ideas or musical constructs as much as possible from the sonification algorithms, with the goal of allowing people to find mappings that naturally expose the symmetry, beauty, harmony and dissonance - the "musicality" - of the physics. For example, a Concordia contributor, as they program a new module, might favor, when choosing sonic mappings for the lines in Figure 3, mapping directly to frequency space rather than mapping to "pitch" via pre-selected notes in a pre-selected scale. In short, minimizing subjective aesthetic impositions from the mappings will enable maximum freedom to create subjective, aesthetically rich results in the exploratory playing of the instrument. Sophisticated, model-based sonification algorithms will be favored over direct audification, allowing for ever richer layering of information into the sounds being created. The simple, direct mappings of planetary information may be used as anchors for the subtler details that characterize the data, embracing rather than neglecting those particularly interesting imperfections or deviations from averages. For example, sound designers know that to make a truly interesting sound, things like Low Frequency Oscillation (LFO), filter cut-off frequency changes, Attack/Decay/Sustain/Release Envelope control, and thousands of other variables in synthesis are critical considerations. Eventually, any given combination of Concordia modules could employ

Figure 3: Screenshots of Hartmut Warm's *Keplerstern* program displaying linklines between Venus and Earth every 4 days over a period of 100 years.

thousands, if not hundreds of thousands, of algorithmic, model-based sonification and visualization mappings in a form that is designed to grow and evolve through open-source contributions to the code base. Composers will choose which of those relationships are pleasing to their ears and how to trigger or manipulate those in time. As Scaletti notes, Sonification $\neq$ Music [17]; the complexity and number of possible mappings will ensure that Concordia is an expressive musical instrument.

One primary consideration will be modularity and the flexibility of mapping algorithms, finding the optimum balance between the limitations of scientific accuracy and integrity versus the freedom of sonic exploration and expression. Another consideration in module design will center around deciding which parameters to map to the visual domain and which to map to the sonic or haptic domains. Concordia's hardware control interfaces and modules will be similarly flexible, with the ability to manipulate the data through a wide range of controllers, from simple switches

on a tablet to complex gestural control such as with motion capture or wearables like the Mi.Mu gloves [18].

2.2. AR/VR/XR - Visual and Sonic

Concordia will be built in extended reality (XR), a combination of Virtual Reality (VR) and Augmented Reality (AR), sometimes called “immersive computing.” There is not yet widespread agreement on the definitions of boundaries between AR and VR, so it is perhaps more useful to consider VR and AR as two points on a spectrum of the degree to which computer-generated imagery and audio is superimposed upon traditionally-perceived reality [19]. In the case of Concordia, most of the reality represented in the instrument must be computer-generated because its spatial and temporal scales are far outside normal human perception, as Kepler describes in *Harmonies of the World*:

But whose good will it be to have harmonies between the journeys, or who will perceive these harmonies? For there are two things which disclose to us harmonies in natural things: either light or sound: light apprehended through the eyes or hidden senses proportioned to the eyes, and sound through the ears. The mind seizes upon these forms and, whether by instinct or by astronomical or harmonic radioncination, discerns the concordant from the discordant. Now there are no sounds in the heavens, nor is the movement so turbulent that any noise is made by the rubbing against the ether. Light remains. If light has to teach these things about the planetary journeys, it will teach either the eyes or a sensorium analogous to the eyes and situated in a definite place; and it seems that sense-perception must be present there in order that light of itself may immediately teach. Therefore there will be sense-perception in the total world, namely in order that the movements of all the planets may be presented to sense-perceptions at the same time.

Thus, Concordia itself will present the celestial harmonies to the eyes, the ears, and the sense of physical immersion of the player so their “mind can seize upon the patterns,” using whatever technology is available to the player. Many modules will likely be designed for VR goggles, headphones, and a cockpit of hardware controls, but all will be required to be at least minimally playable on a simple tablet. In any format, Concordia will use both visualizations and sonifications of planetary data, but its primary function is to emphasize the auditory experience and compel musical explorations.

By sonifying information in addition to visualizing it, new pathways are created for mapping data structures that can exploit the human brain's natural expertise for decoding sound, perceiving patterns and correlations, and gaining understanding of the world from it. For people who are both sighted and hearing, using sonification and visualization together provides important corroboration of information between senses that are instinctively used in the world at large [20]. A great deal of time, trial, and error will be required to find elegant mapping algorithms to expose the natural harmonies and dissonances intrinsic in the system, and users will find their own way of playing the instrument, choosing the mappings that please them. Recent advances in technology enable us to combine new sonification concepts with powerful visual cues in

an immersive and interactive way that educates through direct, visceral, multi-sensory experience rather than traditional pedagogical methods.

In order to be useful as a musical instrument, Concordia needs to combine interactive affordances with a transparent mapping of scientific data into perceptual domains. In general, pitch (or fundamental frequency) as a musical mapping for information is avoided, in favor of more subtle and emergent sonic parameters such as timbre, envelope shaping, particularly in attack spectral content. Multichannel (3D) spatial sonic information will be correlated exactly with visual spatial information where possible, although additional sonic information may be available without being visible. In the cases where this is done, it is well described to the player.

2.3. Distributed Peer-to-peer Computing and Community

The rapidly advancing fields of distributed database computing, collaborative work platforms, media streaming, and blockchain technologies provide exciting and fertile ground for experimenting with large-scale data-driven creative ecosystems. Concordia fully embraces this experimental spirit [21]. While blockchains are most well known in the application of cryptocurrencies, a blockchain is simply a distributed ledger of records linked and secured using cryptography. Any kind of information or contract can be held on a blockchain, and it exists on the network as a continuously reconciled, shared database. Removing the need for trusted third parties such as banks or lawyers, or indeed any large corporations, blockchain technologies are rapidly transforming the socioeconomic landscape in ways that are only beginning to be understood and applied [22]. These technologies present enormous, as yet untapped, potential for large-scale creative collaboration, which can incorporate, as part of the basic infrastructure of a computational musical instrument, the economies and attributions formerly handled by a complex collection of third parties, such as scientific review boards, digital rights management organizations, record labels, and managers.

Blockchain, combined with other distributed database technologies, such as the Interplanetary File System (IPFS), are useful in four different aspects of Concordia implementation:

1. **Data storage and verification** - Automatically maintaining reliable, accurate, standardized datasets and software modules for the Concordia sonification/visualization modules
2. **Attribution and version control of software and hardware modules** - Automatically tracking and rewarding or charging contributors, builders, and remixers of hardware and software modules
3. **Attribution and source-data-matching for compositions** - Tracking composers, performers, and players and ensuring that the compositions are accessing correct versions of data and modules. Rewarding or charging players and observers.
4. **Management of live collaborative composition or performance data** - Building on existing partially-centralized, partially decentralized technologies such as those used for massively multiplayer online gaming for an optimized playing experience

Blockchain technology can be used to provide permanent (tamperproof, censorship resistant) records of the evolution of a

distributed instrument as it develops and changes, enabling searchable and verifiable metadata and outputs of the instruments configurations and discoveries that can be accessed and replayed, or reconstructed, by anyone from anywhere. That offers an open science approach to open music (or smart music, a la smart contracts): a world instrument as a global virtual machine.

Clearly, many aspects of the functioning of Concordia will benefit from the affordances of blockchain technologies, while many others would be unnecessarily bogged down by the inefficiencies introduced when requiring cryptography and verification. Where possible, other, more efficient, distributed architectures will be employed. However, blockchain is useful as a mechanism for handling Concordia's embedded economies and social contracts. A system of valuation will be assigned to the various ways of contributing to or benefitting from the Concordia instrument. However this value is defined - whether strictly financial or some other defined value - if that value can be traded, economies emerge. Players will "spend" a very small, fixed amount, to play the instrument or to download the code for modules. Once they create something like a composition or a new code or hardware controller design, they become Contributors. Contributors will earn fixed amounts of value by contributing in any of the ways defined and agreed by the community, such as

1. programming and uploading new sonification or visualization software modules;
2. composing/performing and uploading new pieces that are uploaded for others to experience; and
3. designing and uploading DIY hardware control modules for others to build.

For example, someone who composes a piece might earn enough value to request a new feature from a programmer. Conversely, to play the instrument without uploading anything does not earn value and requires the expenditure of value. Metrics for excellence and algorithms for rewarding excellence will be established such that the most excellent modules, performances, or compositions will earn added value.

As Concordia evolves and grows, and as additional scientific discoveries are made and new data becomes available, the original datasets upon which the instrument is based may change, requiring version tracking that links each performance, composition, and module with the correct versions of each other and the data itself. By careful version control and matching, no composition or performance becomes obsolete and a full history of the evolution of the instrument, and any compositions made with it over time, is accessible, allowing the instrument to be sustainable and all compositions to be easily re-performed [23]. It should be noted that this is a departure from many electronic music frameworks such as digital audio workstations (DAWs) or plug-ins, in which compositions or sessions must be bounced to stems to preserve the details of the performances so that when operating systems or compatibility changes, the sessions can still be heard.

Active exploration of broader socioeconomic and educational implications of this new type of instrument, particular in its distributed ledger aspects, is underway and might best be followed online (e.g. [24]), as progress in these fields outpaces traditional academic research publishing. From an economic perspective, Concordia appears to be the world's first distributed-ledger-based musical instrument, and because of that, it will also be the world's first musical instrument with its own economy. This makes Con-

cordia of interest not only to musicians, musicologists and technologists, but also to cultural economists and innovation economists.

Concordia reconceptualises an instrument from a private good or unit of capital to a distributed infrastructure (a protocol), and therefore enables a different economic model, namely instrument as a common pool resource. This works over the protocol and design of the instrument and its settings and other meta-data, and also to the data inputs and outputs. This then changes both the economic incentives at work, the business models that are effective, and the policy settings available. This new economic model also implies new policy models, based on institutional design of protocols and governance. Concordia also illustrates how the creative arts and creative economy can be a vanguard for other parts of the economy or society by early adoption of new technologies and experimental development of the institutions to support them.

2.4. Concordia in Education

A full discussion of the educational potential of Concordia, both as a direct element in curriculum design for primary, secondary, and post-secondary or higher education and as a networked installation in museums or institutions is beyond the scope of this paper but bears mentioning as an area being actively explored by the Concordia team.

3. MODULE 1 - PROTOTYPE

As a proof of concept, while the underlying computational and social/philosophical frameworks outlined above are being developed, work on a prototype of the first Concordia module has begun. Where possible, existing tools (e.g. the data visualization and sonification environments created by Warm and Barri) are being used for early mapping experiments. As a starting point, two datasets of planetary ephemeris have been generated using standard examples provided in Warm's app, *Signatures of the Celestial Spheres* [2]:

1. Lines linking Jupiter and Uranus, shown in Figure 4; and
2. Lines linking Venus and Earth drawn every 4 days for 100 years (as shown in Figure 5).

In this simplest of modules, the lines linking the planets form pentagonal and hexagonal figures, respectively, when drawn at the resolution generated in the examples. A MaxMSP patch was created to read in the data, add one dimension of input data control that could be mapped to an OSC controller, then output OSC commands to Barri's Versum audio-visual composition environment to draw the lines and spheres in 3D, apply mappings to audio and visual synthesis, and enable navigation through the data. The objects created in Versum comprise lines and spheres with a large number of time-variable visual and sonic parameters that can be mapped using data-generated OSC commands with an emphasis on visual simplicity and on sonification as the main carrier of geometric and relational information. As the lines between planets are drawn, their mathematical properties are sonified in a way that allows sonic patterns to be revealed gradually, just as the visual patterns are.

The one dimension of data input control chosen was the sampling frequency of the lines drawn over a given time, achieved in the max patch by creating a slider for the player to choose how many lines of the original input data file to skip between lines drawn. As shown in Figure 5.

Figure 4: Lines linking Jupiter and Uranus at regular intervals over time

In this simplest of experiments, as the players adjust the time interval between lines drawn and those lines are drawn inside Versum, they begin to see and hear changing symmetrical figures appearing in the data, from the beautiful dense 5-petal structure of the original set, to sparser 3-fold, 4-fold, 5-fold, 6-fold, 9-fold, 17-fold, etc symmetrical figures. This basic geometric property of varying symmetries may be obvious to mathematicians, but it causes great delight to a first-time explorer and when mapped well sonically, can offer hours of intrigue. Of course, this shifting symmetry pattern happens also in a similar way with the Jupiter-Uranus link lines when the interval between lines is changed; therefore, this parameter emerges as an interesting and fun potential “knob” for one of the first hardware controller modules.

The player's experience is a dynamic exploration of unfolding information as objects are made visible and audible in Versum's displays (see Figure 6). Perspective (location, camera-pointing) and zoom level within the data is under the control of the player, similar to an in-world avatar game, such as Second Life. In the full module, before starting the navigation or even during navigation, the player will set up their special universe, via an interface such as a mobile device or hardware cockpit control, by choosing from a whole host of variable parameters, such as which planet to include, which moons to include, over what time periods, and at what intervals.

Versum enables fluid experimentation with mappings to hundreds of sonic and visual parameters at any moment. Objects can be created, destroyed, and altered in real time either by changing the mappings from the data or by changing navigational perspective. A brief video of this initial proof-of-concept module can be found here: https://drive.google.com/open?id=1AWaM2C1uLV5m_lm32babRxyz1ttd5kWB Currently, navigation in the Versum prototype is manual and interactive using a 3D space navigator. Our version of Versum uses Supercollider for

Figure 5: Linklines of Venus and Earth over a period of 100 years with increasing intervals of time between lines, created by varying the number of lines skipped in the database from 0, representing a line drawn every 4 days (upper left) to every 195th line drawn, representing a line being drawn every 780 days (lower right).

audio synthesis for mappings from the data to audio, such as line length, line angle, regularity of line intersections, distance of intersections from the center of the figure, line start and end points, points where there are multiple intersections, etc. can, themselves, be changed in real time to explore the most compelling mappings for hearing geometric regularity.

4. CONCLUSION

Concordia is a musical instrument, inspired by and being built to celebrate the 400th anniversary of Johannes Kepler's 1619 tome *Harmonies of the World*. The instrument plays scientific data, and is therefore an instrument of investigation and discovery. This new type of instrument - an immersive VR protocol - will enable new types of music and music making. It is built to be massively parallel and collaborative, and to be iterative and cumulative, with plays and discoveries in one moment becoming part of the instrument at the next moment. This extra complexity in the instrument requires extra complexity in its institutional and technical infrastructure, employing new distributed database handling and coordination unimaginable to Kepler, such as blockchain and IPFS.

Still in the initial development stage, a simple proof-of-concept has been explored using some existing audiovisual tools. For this first module, data from Hartmut Warm's asynchronous, visual data exploration tool has been replicated and in the fully in-

Figure 6: Link lines between Jupiter and Uranus displayed visually and sonically inside Versum.

teractive Versum 3D audiovisual composition environment, such that all mapped sonic and visual properties are data-generated and navigable in real time. The first module is not fully implemented in Virtual Reality, but is an extension into Augmented Reality, allowing an experience of reality at scales not normally accessible to humans. Construction on initial frameworks for the distributed data, computational, and economic architectures has begun, onto which the prototype module will be uploaded for a small group of initial Concordia developers. While the process of creating Concordia will be documented on the blockchain throughout 2018, the four hundredth anniversary of *Harmonices Mundi* in 2019 will be the official release of Concordia as a functional instrument. Development can continue indefinitely, as contributors add more capabilities to the instrument and upload to the blockchain and distributed platforms, creating an ever expanding, continually updating sonification instrument ecosystem.

5. FUTURE WORK

Essentially, nearly all the work on Concordia is still in the future, but development is now occurring at a rapid pace to meet the 2019 anniversary launch deadline. Teams are working in parallel on the various broad areas, including the sonifications, visualizations, immersive media, distributed ledger technologies, education packages for schools and museums, and economics of this new instrument platform.

6. REFERENCES

- [1] J. Kepler, *Epitome of Copernican Astronomy and Harmonies of the World*. Prometheus Books, Publishers, 1995, google-Books-ID: gUJdcFmRUGMC.
- [2] H. Warm, *Signature of the celestial spheres: discovering order in the Solar System*. Forest Row: Sophia, 2010, oCLC: 756267079.
- [3] A. M. Duncan and E. Aiton, *Johannes Kepler: Mysterium Cosmographicum (The Secret of the Universe)*. Abaris Books, 1981.
- [4] J. Kepler and C. Baumgardt, "Johannes kepler: life and letters," *New York, Philosophical Library [1951]*, 1951.
- [5] V. Galilei, *Dialogue on ancient and modern music*. Yale University Press, 2003.
- [6] T. Barri, "Versum: audiovisual composing in 3d," in *Proceedings of The 15th International Conference on Auditory Display (ICAD-2009)*, Copenhagen, Denmark, 2009.
- [7] R. L. Alexander, J. A. Gilbert, E. Landi, M. Simoni, T. H. Zurbuchen, and D. A. Roberts, "Audification as diagnostic tool for exploratory heliospheric data analysis," in *The 17th International Conference on Auditory Display (ICAD-2011)*, Budapest, Hungary, June 2011. [Online]. Available: <https://smartech.gatech.edu/handle/1853/51574>
- [8] M. Ballora, "Sonification Strategies for the Film Rhythms of the Universe," in *The 20th International Conference on Auditory Display (ICAD-2014)*, New York, USA, June 2014. [Online]. Available: <https://smartech.gatech.edu/handle/1853/52075>
- [9] R. M. Candey, A. M. Schertenleib, D. Merced, and W. L., "Xsonify sonification tool for space physics," in *Proceedings of the 12th International Conference on Auditory Display*, London, UK, June 2006. [Online]. Available: <https://smartech.gatech.edu/handle/1853/50697>
- [10] N. Leonard, "Sonification: Celestial Data and Poetic Inquiry," in *The 16th International Conference on Auditory Display (ICAD-2010)*, Washington, D.C, USA, June 2010. [Online]. Available: <https://smartech.gatech.edu/handle/1853/50058>
- [11] B. J. Tomlinson, R. M. Winters, C. Latina, S. Bhat, M. Rane, and B. N. Walker, "Solar System Sonification: Exploring Earth and its Neighbors Through Sound," in *The 23rd International Conference on Auditory Display (ICAD 2017) June 20-23, 2017, Pennsylvania State University*, June 2017. [Online]. Available: <https://smartech.gatech.edu/handle/1853/58359>
- [12] S. Feisst, "18 Negotiating Nature and Music through Technology," in *Current Directions in Ecomusicology: Music, Culture, Nature*. Routledge, 2015.
- [13] W. Ruff, J. Rodgers, and J. Kepler, "The harmony of the world : a realization for the ear of Johannes Kepler's astronomical data from Harmonices Mundi 1619," 1979.
- [14] M. Quinton, I. McGregor, and D. Benyon, "Sonifying the solar system," in *The 22nd International Conference on Auditory Display (ICAD-2016)*. Canberra, Australia, 2016, pp. 28–35. [Online]. Available: <http://researchrepository.napier.ac.uk/id/eprint/10317>
- [15] "Keplerstern.com." [Online]. Available: <http://www.keplerstern.com/signature-of-the-celestial-spheres/program/>
- [16] "ssdl.jpl.nasa.gov." [Online]. Available: <https://ssdl.jpl.nasa.gov>
- [17] C. Scaletti, "Sonification Music," in *The Oxford Handbook of Algorithmic Music*. Oxford University Press, 2018.
- [18] "mi.mu." [Online]. Available: <https://mimugloves.com>
- [19] C. Bavor, "Virtual and Augmented Realities: Asking the right questions and traveling the path ahead," May 2017. [Online]. Available: <http://bit.ly/2pNfVra>

- [20] J. Gossmann, "Towards an Auditory Representation of Complexity," in *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland,, 2005*. [Online]. Available: <http://www.icad.org/Proceedings/2005/Gossman2005.pdf>
- [21] "ssrn.com." [Online]. Available: <http://ssrn.com/abstract=3174172>
- [22] S. D. Chris Berg and J. Potts, "Byzantine political economy," Oct. 2017. [Online]. Available: <http://bit.ly/2zPjoHg>
- [23] M. Schedel and E. DeMartelly, "Sustainability of Performing with Technology," in *Proceedings of the International Symposium on Electronic Art*, 2008.
- [24] "<https://medium.com/@jason.potts/>" [Online]. Available: <https://medium.com/@jason.potts/concordia-a-new-future-economy-of-music-7371f21e917e>

IMAGES, GRAPHS AND SPACES

SOUND-BASED IMAGE AND POSITON RECOGNITION SYSTEM ~ SIPReS ~*Shin'ichiro Uno, Yasuo Suzuki, Takashi Watanabe*

Nihon Fukushi University
 Faculty of Health Sciences
 Department of Human Care Engineering
 26-2 Higashi-Haemi-cho Handa,
 Aichi 475-0012, Japan
 uno@n-fukushi.ac.jp
 yasuo-s@n-fukushi.ac.jp
 wata-t@n-fukushi.ac.jp

Miku Matsumoto, Yan Wang

Shouei Denshi Kenkyusyo Ltd.,Co.
 26-145 Icchouda Kutsukake-cho Toyooka,
 Aichi 470-1101, Japan
 hirate@shouei.co.jp
 wang@shouei.co.jp

ABSTRACT

We developed software called SIPReS, which describes two-dimensional images with sound. With this system, visually-impaired people can tell the location of a certain point in an image just by hearing notes of frequency each assigned according to the brightness of the point a user touches on. It can run on Android smartphones and tablets. We conducted a small-scale experiment to see if a visually-impaired person can recognize images with SIPReS. In the experiment, the subject successfully recognized if there is an object or not. He also recognized the location information. The experiment suggests this application's potential as image recognition software.

1. INTRODUCTION**1.1. STEM and data sonification**

One of the important study themes in the assistive technology field is how to guarantee visually-impaired people's right to access information. Science Technology Engineering and Mathematics, or STEM, has been held to be important in recent years[1]. Emphasis is put on STEM also in the education field for disabled people[2]. Actually, there are more opportunities in the STEM field for disabled people to be independent and to exercise their abilities by making use of information technology. That is why efforts for better STEM education are being made in many areas such as in the education system, environment improvement, or software/ hardware development.

One of those efforts has been made in the study of conveying data through the medium of sound. Dubus and Bresin have surveyed 179 papers on sonification for the past 20 years, and made a database of them. Many of those papers are about sonification in the scientific fields. One recent example in the field of astronomy is sound models of the eight solar planets made by Michael Quinton. He reported visually impaired people can tell each planet's feature by sound[3].

Our project, the Astronomical Data Sonification Project, aims to develop methods with which space science data can be understood by visually-impaired people without depending on the visual information. The project is promoted by two institutions: Uno Laboratory under the Faculty of Health Sciences at Nihon Fukushi University and the Institute of Space and Astronautical Science(ISAS) of the Japan Aerospace Exploration Agency(JAXA). In this project, we explored possibilities of sonification of space science data. We already accomplished sonification of X-ray pulsar data [4] and sonification of Kp indices indicating the strength of geomagnetic disturbance[5]. In addition, we made prototype software called splot and sonified general one-dimensional histogram [6], [7]. After all these achievements, sonification method of two-dimensional data of astronomical bodies has been anxiously awaited.

1.2. Tactile Image Recognition Environment for Visually Impaired People

Various efforts have been made to develop methods for visually impaired people to recognize images.

Examples of tactile image representation are a) braille embossing printer, b) tactile images or three-dimensional copies made on capsule paper, and c) pin display, all of which are designed to give tactile information to readers.

A braille embossing printer can draw figures with raised dots just as an ordinary braille printer does when it prints letters. One of its advantages is that it can print letters along with images. So it is very useful for books such as expository books or braille books. One of its disadvantages is that printing requires time as it prints on paper. Another disadvantage is portability. It's too heavy to move it around. Figure 1 is an example of a chart made with a braille printer.

Tactile images or three-dimensional copies are made on special paper called capsule paper. It has a layer of capsules which expand with heat. An image is expressed by elevation of the paper. More specifically, the process is two-fold: making a copy with an ordinary copier and heating the paper with a dedicated machine. Then, a three-dimensional paper will be printed out. Compared to the other types of tactile image representation methods, it is easier to make three-dimensional diagrams on paper. Three-dimensional diagrams on paper are relatively easily made. The dedicated ma-

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

chine, however, has similar disadvantages as a braille printer. One of them is its size. It is bulky.

A pin display is a display to show what is on the computer display just like an ordinary braille display does. Figure 2 is an example of a pin display [8]. Its advantages are its easy connection with a personal computer and real-time display. On the other hand, the number of actuators is limited and its density is fixed. So the size of what can be shown on the display is limited and its granularity is coarse. In addition, it is too pricey, as it consists of special parts.

Figure 1: Example of an image printed with a braille embossing printer

Figure 2: Example of a pin display

1.3. Auditory-Visual Sensory Substitution (AVSS) System

Many attempts have been made so far to convert two-dimensional image information into sound. This type of conversion is called AVSS, or Auditory-Visual Sensory Substitution System.

“The vOICE” is the first general-purpose AVSS[9], on which various researches have been conducted[10, and references there in]. “The vOICE” firstly cuts image information into vertically long strips, then allocates lower-part information of each strip to low frequency, higher part to high frequency, and brightness of each pixel to sound volume, and finally generates composite tone. It passes along two-dimensional information to a user by playing the composite tone along with its automatic sweep at a constant rate in a horizontal direction. The user can grasp a non-uniform image or space by the sweep.

Another system called GUESS creates two-dimensional auditory space and passes along the information of a shape to a user by moving the location of sound[11]. GUESS system has succeeded in recognition of a circle or a triangle. In addition, it conveys the location information with the help of a device such as a tablet.

Another system called sonicphoto allocates information of vertical section to sound pitch and that of horizontal section to time. It represents image by sound along with automatic scanning at a constant rate. The program sonicphoto is available on the Internet[12].

Another system called iSonic developed at the University of Maryland [13] is an interactive system and enables users to understand a choropleth map linked with data in a table. With operation using arrow keys or a touchpad, users can understand data such as statistics linked to a certain point on the map[14]. Using iSonic, Delogu *et al.* compared recognition rates of sonified maps with a keyboard or a touch pad. They showed potential of iSonic with touchpad operation, though its serviceability depends on the type of a target map and statistical data[15].

2. DEVELOPMENT

In this study, we tried to develop image description software using sound. By describing image information with sound on a small device, we aimed to achieve image recognition method with alacrity and portability, which were difficult to achieve with conventional methods. We set our eyes on smartphones and tablets to achieve our goal, since those devices have been rapidly spread among people. System features requirements are as follows:

- It can run on a smartphone and also on a tablet.
- It can recognize the coordinate of the point a user touches on the panel, and strike a note according to the brightness of the point.
- It can run camera application on the device, and take a picture and import it.
- It can read image files such as jpg, png, and pdf.
- It can read screenshot images.

We gave it a name Sound-based Image and Position Recognition System, or SIPReS. And its target operating system is Android. The development platform is Eclipse version 4.6.2, development language is Java version 8, and the current version of SIPReS is 1.0.

2.1. Sonification

The system will strike a note of an assigned frequency according to the brightness of the point a user touches, and the note will continue as long as the user touches the point. Input images can be color or black and white. For black and white images we use the brightness value without any conversion. For color images, we convert RGB data to brightness information (Equation 1) [16].

$$y_n = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (1)$$

Next, we derive the frequency value of a note it will strike based on the brightness information. Firstly, we read the minimal and maximum value of the brightness of an image and assign them to $y_{\min}$ and $y_{\max}$ respectively. Then we set y_n as the brightness of the pixel the user touches. y_n 's frequency, f_n , is calculated using equation 2 below.

$$f_n = f_{\min} \times 2^{\left(\frac{(y_n - y_{\min})}{(y_{\max} - y_{\min})} \times \log_2 \frac{f_{\max}}{f_{\min}} \right)} \quad (2)$$

$f_{\min}$ and $f_{\max}$ are frequency values according to the brightness and can be set by users. Default value of $f_{\min}$ and $f_{\max}$ are 220 Hz and 1760 Hz respectively. y_n can be a) the brightness

value of a single pixel touched or b) the average brightness value of r layers of surrounding pixels from the point the user touches. The default value of r is 3, but it can be changed by users.

2.2. Application

After starting the application, a menu will come up for selection of an input source. A user can choose a file or take a picture for sonification here. According to the point the user touches, it strikes a note.

There is a menu icon in the upper right corner of the application. From the menu a user can choose an image or change settings of the application. It can also show the coordinate of the point touched, x and y , its brightness, y_n , and corresponding frequency, f_n , in the upper part. In Figure 3, we show the SIPReS's display.

Figure 3: SIPReS's display. An image of a coffee cup is shown on it. There is a blank space due to the difference between the image's aspect ratio and the number of pixels of the display, and/or the height of the menu bar.

3. EXPERIMENT

We carried out a small-scale experiment to see if SIPReS enables readers to recognize images. The aim of the experiment at this stage is just to see if the application is useful or not. So the number of subjects is only one this time and used simple images. The subject is a 31-year old man blind from birth. We will experiment with a number of subjects using more complicated images from next time onwards.

Devices used in the experiment are a tablet, Galaxy Note Pro and smartphone, Galaxy Note 5, both of which were made by Samsung Electronics Co., Ltd. We installed SIPReS ver 1.0 to both devices and ran it with default settings. The specification of both devices is shown in table 1. Since images for the experiment were horizontally long, we used Galaxy Note 5 turned sideways.

3.1. Simple Images

Experiment 1 is the simplest. We had the subject look for the center of a white circle against the black background and touch the point. The white circle is gradational with the whitest center. The size of the picture is 2560×1440 pixels, and the radius of the circle is 300 pixels. The center is located 30% lower from the top and 30% right from the left, or at the coordinate of (768, 432). In figure 4(a), we show the image used in the experiment.

The subject touched the tablet's display with his finger and based on the sound information gave us an oral answer on the location of the center, or how far, or the ratio on the whole surface,

Table 1: Device used for evaluation

	galaxy note pro	galaxy note 5
developer	samsung	samsung
display size (inch)	12.2	5.7
display resolution (pixel)	2560×1600	2560×1440
dimension (H/W/D, mm)	204/295.6/7.95	153.2/76.1/7.6
bezel (u/d/l/r, mm)	19/19/15/15	14/12/2/2
OS(android) version	5.0.2	7.0

The values of bezel are measured in the experiment. Other figures are extracted from the device catalogs.

the point is located from the upper left corner. When the subject said he found the point, we read the coordinates displayed on the SIPReS. Although we did not set any time limit, he reached the brightest point and gave an answer of a ratio of x and y in a matter of seconds.

Because of the frame of the device and the fringe parts of the application such as the menu bar, the center shows up at a little different location. As the frame size of a device is different from model to model and it is difficult to tactually tell the border between frame and display, we have decided not to offset this gap. The table 2 shows a) the location of the center in the picture, b) that in the whole device including its bezel, c) his answer on the ratio, and d) the real point he touched.

As the table 2 shows, the point he touched is almost correct with a margin of just 30 pixels or 2% of the size of the image. Though the point he touched is almost accurate, the margin of error of his oral answers is a little bigger. We should keep in mind that bigger margin of error is expected for location recognition due to the bezel of a device or the space for application's menu.

Table 2: The real location of the center of the circle and the subject's answer

	galaxy note pro	galaxy note 5
Entire tablet (include bezel)		
from left	93mm(31.5%)	45mm(29.4%)
from top	80mm(39.2%)	32mm(42.0%)
Coordinates in the display (scaled)		
x,y(pixel)	(741,417)	(486,273)
Subject's answer (pointing)		
x,y	(775,418)	(491,290)
residual	(+34,+1)	(+5,+17)
(impressions including bezel)		
from left (%)	30%	30%
from top (%)	30%	30%

3.2. Experiment 2

We use an image of a slant line for the experiment 2 to see if the subject can trace the line. The line is gradational like the circle we used in experiment 1. The center in the cross direction is the whitest and the width is 200 pixels. It starts from the lower left corner, passes the center of the image, and ends in the upper right

(a) finding circle

(b) tracing line

Figure 4: Sample images for experiment 1 and 2. The image size is both 2560×1440 pixels. In the image (a) there is a circle with 300-pixel radius at the point 30% lower from the top and 30% right from the left. In the image (b) there is a line passing the center of the image at a slant angle of 45 degrees.

corner at a slant angle of 45 degrees. Figure 4(b) is the image we used.

All he knew before the experiment is there would be an image of a line. Then in the experiment, he gave us an oral answer on the location of the start points and end points along with the direction. He gave us answers like “the start point is 20% right from the lower left edge,” and “the end point is 80% right from the upper left edge.” With the prior information, he successfully trailed points which struck high notes. He told us, the clearer the border against the background, the easier he finds the line.

3.3. Experiment 3

We also carried out an experiment on SIPReS’s another function, picture-taking function. One advantage of image recognition by sound is not limited to recognition of prepared images, but also of a shape in a picture taken on the spot. To make sure the function works as we intended, we experimented to see if the subject could tell the rough image of an object in a picture. We used a picture of a banana and an orange. Figure 5 is the picture we used.

After we presented the picture, the subject recognized there were two objects, one on the left was vertically long, and the other on the right was close to a circle. That said, he could not tell exactly what those objects were. After we told him there were a banana and an orange, he answered the left object was the banana. So if certain conditions are met, there is a possibility of providing shape information to the extent to which visually-impaired people can tell the difference between a round object and an elongated object.

Figure 5: These fruits were used for a picture to see if the subject can recognize the shape of objects in it with SIPReS. We put the fruits on a blue cloth since the clearer the contrast between the object and the background is, more recognizable.

4. DISCUSSION

This study indicates possibilities for visually-impaired people to roughly recognize images using SIPReS on a tablet or on a smartphone. We believe the possibilities of location recognition without depending on visual information but with the combinatory technology of two-dimensional location information on the touch panel and information of different tones is very meaningful.

Having said that, we have not collected enough amount of data for verification yet. So we need to conduct larger scale of proving tests and assess its effectivity. As to the types of images, we may need to use more complicated ones such as a line of a rectangle or a star, or a lump of shapes such as a function or a chart. In addition, we may need to ask more subjects to join the experiment.

The current version of SIPReS 1.0 does not support video sonification. If it does, it is expected that users would recognize the change of image by the change of sound with the finger staying at the same spot. This is a challenge we should address from now.

4.1. Positioning

As mentioned above, the subject successfully located the center of the circle in experiment 1. He firstly moved his finger just like scanning the whole area, then looked for the center after recognizing the sound difference. He told us it was easier on a bigger tablet than on a smaller smartphone.

The gradation of the circle turned out to be very helpful. The subject understood the higher tone means brighter spot. With SIPReS which strikes a tone in real time, he looked for points with a higher tone and finally spotted the point. By making use of this mechanism such as changing gradation pitch, visually-impaired people might be able to recognize more complicated images such as a pointed cone, a cone with a bigger bottom, or hemisphere.

As we saw in experiment 2, however, it is easier to reach the point with the highest tone if the border between the line and the background is clear. In a gradational image, the frequency of the tone changes even by a tiny location change. This leads us to think it is easier to trace a line with the same tone without gradation. Line recognition could possibly be utilized for map reading or guideboard.

4.2. Sound Parameters

The default frequency range of SIPReS is from 220Hz to 1760Hz, three octaves across the standard pitch of 440Hz. We chose the range because it sits in the middle of frequencies that can be heard by humans and is expected to be easier to distinguish.

Having said that, among those who have multiple difficulties such as auditory and visionary, there are people who cannot hear sound of specific frequency range. And the range they cannot hear varies from person to person. Therefore the output sound range is modifiable and can be set by users.

In the current version of SIPReS, output sound is formed by sinusoidal function. We had other sound form options such as triangle or box. However we deployed a simple sinusoidal function only, as our main aim, at least for now, is correspondence between the touched location and sound.

We think, however, allocation of multiple sound forms is effective to some users. Furthermore, there is a possibility to express subtle difference of colors by correlating each RGB parameter to specific sound form, which may lead to expression of more complicated information. Sound form does not have to be limited to mathematical formulas. There is also a possibility to express color and a shape with a chord, or even a discordance, by incorporating sound form of musical instruments.

4.3. Taking Picture on the Spot and Its Recognition

Experiment 3 indicates possibilities of recognition of a picture taken on the spot as long as it has a simple image on it. That said, the application's current ability is limited to recognition whether there is an object or not. We cannot tell how minutely people can recognize images. We need to start to think how we can evaluate it first. The subject made a suggestion that the application might be used when visually-impaired people look for something like a key they dropped inadvertently.

Let me talk about an example of image recognition requests here. The Astronomical Data Sonification Project is getting a request from a visually-impaired space enthusiast. He wants to join an astronomical observatory meeting and sense the waxing and waning of the moon. The moon is in stark contrast to the night sky, and the difference in shape between the crescent moon and the full moon is expected to be easily recognized. We take this as one challenge we should address.

4.4. Application Possibility of SIPReS

We developed SIPReS as one image recognition method for visually-impaired people. But it could be used for other purposes other than image recognition.

We tried to have SIPReS play music as one possible application. Figure 6 is a score of simple music. Each part's brightness is set so that its corresponding frequency can strike a tone of the musical scale. To clarify the maximum and the minimal there are two parts in the upper right corner showing 0% brightness and in the lower left corner showing 100% brightness. Simple music is played, while a user moves their finger in a set direction.

As we experimented, we found it is a little bit more difficult than expected to move the finger straight and at the same pace. Though it takes some getting used to and complicated music cannot be played, there is a possibility that anybody with or without visual disability can play music and that even those who cannot

read scores or have not played instruments can play music just by moving their finger on the tablet display.

Figure 6: Music score for SIPReS. By swiping from left to right at the same pace, one can hear songs played by SIPReS. By calculating brightness back from frequency, anybody can make music scores of a variety of songs. This example is the score of a song called Charmera, which was used by conventional ambulant vendors.

5. CONCLUSION

We developed SIPReS which describes a two-dimensional image with sound. SIPReS has possibilities that even a visually-impaired user can roughly recognize an image by hearing an assigned tone according to the brightness of the point a user touches. The subject, who is totally blind, successfully recognized whether there is an object or not along with the location information. The experiment indicates this software's practicability. We hope SIPReS can be applied to new methods of image recognition.

6. REFERENCES

- [1] J. J. Kuenzi, "Science, technology, engineering, and mathematics (stem) education: Background, federal policy, and legislative action," *Congressional Research Service Reports*, 2008.
- [2] J. Hwang and J. C. Taylor, "Stemming on stem: A stem education framework for students with disabilities," *Journal of Science Education for Students with Disabilities*, vol. 19 Iss.1 Article 4., pp. 39–49, 2016.
- [3] M. Quinton, I. McGregor, and D. Benyon, "Sonifying the solar system," in *Proc. of the 22nd Int. Conf. of Auditory Display (ICAD)*, 2016, pp. 28–35.
- [4] S. Uno, T. Kameyama, M. Horiata, N. Asano, K. Ebisawa, T. Tamura, I. Shinohara, O. Miyashita, A. Miura, K. Matsuzaki, H. Murakami, and F. Furusawa-Akimono, "The astronomical data sonification project: An overview of the project and its initial data output," *Journal of health sciences, Nihon Fukushi University*, vol. 10, pp. 1–9, March 2007.
- [5] A. Miura, S. Uno, T. Kimura, and K. Ebisawa, "Visualization and sonification of space science data -for education and public relation-," *Journal of Space Science Informatics Japan*, vol. 1, pp. 13–22, May 2012.

- [6] S. Uno, S. Sotoya, A. Miura, and K. Ebisawa, "The astronomical data sonification project 2: experimental production of sound-based data plot software splot," *Journal of health sciences, Nihon Fukushi University*, vol. 14, pp. 1–9, March 2011.
- [7] S. Uno, S. Sotoya, A. Miura, and K. Ebisaw, "Current status of the astronomical data sonification project," *Journal of Space Science Informatics Japan*, vol. 1, pp. 7–11, May 2012.
- [8] <https://www.kgs-jpn.co.jp/index.php>.
- [9] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Transactions on Biomedical Engineering*, vol. 39 Iss.2, pp. 112–121, 1992.
- [10] <https://www.seeingwithsound.com/literature.htm>.
- [11] H. M. Kamel, P. Roth, and R. R. Sinha, "Graphics and user's exploration via simple sonics (guess): Providing interrelational representation of objects in a non-visual environment," in *Proc. of the 2001 Int. Conf. on Auditory Display (ICAD)*, 2001, pp. 261–266.
- [12] <http://www.skytopia.com/software/sonicphoto>.
- [13] <http://www.cs.umd.edu/hcil/audiomap/>.
- [14] H. Zhao, C. Plaisant, B. Shneiderman, and J. Lazar, "Data sonification for users with visual impairment: A case study with georeferenced data," *ACM Transactions on Computer Human Interaction*, vol. 15 Article 4., 2008.
- [15] F. Delogu, M. O. Belardinelli, M. Palmiero, E. Pasqualotto, H. Zhao, C. Plaisant, and S. Federici, "Interactive sonification for blind people exploration of geo-referenced data: comparison between a keyboard-exploration and a haptic-exploration interfaces," vol. 7. Springer Berlin/Heidelberg, 2006, pp. 178–179.
- [16] "Itu-r recommendation bt.470-6, conventional television systems," 1998.

PHOTONE: EXPLORING MODAL SYNERGY IN PHOTOGRAPHIC IMAGES AND MUSIC

Niklas Rönnerberg

Linköping University
Media and Information Technology
SE-581 83 Linköping, Sweden
niklas.ronnerberg@liu.se

Jonas Löwgren

Linköping University
Media and Information Technology
SE-581 83 Linköping, Sweden
jonas.lowgren@liu.se

ABSTRACT

We present Photone, an interactive installation combining photographic images and musical sonification. An image is displayed, and a dynamic musical score is generated based on the overall color properties of the image and the color value of the pixel under the cursor. Hence, the music changes as the user moves the cursor. This simple approach turns out to have interesting experiential qualities in use. The composition of images and music invites the user to explore the combination of hues and textures, and musical sounds. We characterize the resulting experience in Photone as one of modal synergy where visual and auditory output combine holistically with the chosen interaction technique. This tentative finding is potentially relevant to further research in auditory displays and multimodal interaction.

1. INTRODUCTION

The starting point for us was the use of musical sonification to augment information visualization. With musical sounds we mean deliberately designed and composed sounds, based on a music-theoretical and aesthetic approach. Adding sound as a complementary modality to an image is related to the research area of sonification. One of the aims of sonification is to use sound to enhance and clarify visual representations of data, and to simplify the understanding of these [1, 2, 3, 4]. Sonification using musical sounds is interesting as the use of musical elements gives better control of the design of the sounds, and invokes potentially useful musical properties such as timbre, harmonics, melodies, rhythm, tempo, and amplitude (for more information about musical properties, see for example [5]).

Our previous applications of musical sonification include situation awareness in a monitoring/control task, where peripheral information on a radar screen for air traffic controllers was sonified to support the formation of a Gestalt of the airspace using musical sounds in different pitches and with different rhythms [6]. Sonification was also used for data exploration in scatter plots and parallel coordinates for discrimination of visual density levels using musical sounds of different pitch and timbre [7]. Furthermore, sonification was evaluated in relation to perception of color intensity [8] where three sonification conditions were used and interactively changed in relation to intensity levels in the visual repre-

sentations: cut-off frequency of a band-pass filter, amplitude level, and a combination of these two.

In all these applications, music served successfully as a complementary modality to enhance the performance of primarily visual tasks. The hierarchy between the visual and the auditory was clear. However, when we applied a similar approach to photographic images for hedonic experience in an interactive installation called Photone, we found the relationship between image and sound to be less clear-cut – the visual and the auditory appeared to be contributing equally to the multimodal experience.

Combining images and music is nothing new in itself. The history is full of excellent examples of composed music for images and motion pictures. There are also different approaches to visualizing music and for turning an image into music. In most cases, music is composed to images or generated from images based on their denotative and connotative meaning. Our artistic intention in Photone was another: By building the sonification upon pixel values of hue and brightness, that is, syntactic rather than semantic properties of an image, we aimed to cut through conventional ways of seeing to a more foundational level. The resulting experience of Photone cannot be understood as a simple sum of visual and auditory stimuli; our tentative analysis is that the fine-grained details of interaction form the glue between the visual and auditory modalities and the experience emerges as a synergistic whole. We propose the concept of *modal synergy* to capture this phenomenon. More on this below, but let us first try to convey a sense of the synergistic interaction experience we are talking about.

A short video demonstration of Photone to complement the following vignettes can be found here:
<https://vimeo.com/246964768>

2. VIGNETTES: USING PHOTONE

Consider a first example (see Figure 1). When locating the cursor in the dark center of the leftmost flower, the music is muffled and somewhat anticipatory in its low pitch and intensity. As you move the cursor slowly along one of the petals, the fine-grained textural changes invoke different notes forming a unique melody. Intensity and pitch increase with the brightness, but the music is still homogeneous in its harmony (the red harmony, as you are about to find out). Then, you cross the border between the petal and the orange background, and the music shifts to a different basic chord. The orange background has much less texture, which yields a more ambient and less articulated melodic quality. You move back and forth across the soft gradations and the more abrupt changes in the orange region, attentive to the differences, and perhaps try to find a rewarding rhythm by going back and forth across an intensity

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

Figure 1: Photo: Niklas Rönnberg.

shift. Moving on to exploring the blue, green and yellow areas of the image introduces new harmonics, giving you reason for further comparison and impromptu composition by going back and forth between the different color areas. This simple example already illustrates how image and music unfold together in an engaging interactive experience – the visual properties of the image suggest moves to explore the musical effects, and the musical qualities of the emerging composition suggest moves to exploit certain visual properties.

Figure 2: Photo: Niklas Rönnberg.

In the next example (see Figure 2), the lighted windows against the dark façade of the building on the left form distinct shifts in brightness, serving the same purpose here that the keys serve on a piano keyboard. The cursor is drawn there more or less involuntarily, and you move at different speeds along the rows of windows, concentrating on the resulting rhythm and its interplay with the ambient chords. From there, it is only a short step to playing around with the ripple textures on the surface of the stream, and the dynamic range between the lighted patches under the stone bridge and the dark pillars.

Going back and forth horizontally in the area above the glass (see Figure 3) provides a clear and engaging sense of how the red and the blue harmonics relate to each other, while each of them contain enough brightness variation to provide for interest-

Figure 3: Photo: Patric Ljung.

ing melodic possibilities in themselves.

The final example (see Figure 4) emphasizes the subtle combination of melodics and harmonics that lies in some images waiting to be discovered. The foliage has a fine-grained and sharp-edged texture yielding interesting melodic possibilities, and you may spend a long time there feeling how the autumn colors harmonize with each other. But at some point, you move the cursor from the foliage into the blue sky – and the triumphant quality of the resulting harmonic shift is quite remarkable and rewarding.

Figure 4: Photo: Rickard Englund.

What all these examples aim to show is how the interaction with photo and music has holistic qualities where it would be pointless to insist that one modality augments or supplements the other. What they also demonstrate is a specific way to look at images as collections of pixels with colors and brightness values, and the temporal trajectories formed by spatial movement across the surface of the image. We will return to these points in the concluding discussion, after a presentation of how Photone is designed and constructed.

3. DESIGN OF THE SONIFICATION

As previous studies (i.e. [6, 7, 8]) have suggested that amplitude, timbre, and pitch are useful musical elements to sonify visualiza-

tion of data, the composition in Photone was made with these elements as the basis. In the previous studies, data points were read as pixel values and the sonification was changed according to these values; the mapping between pixel values and musical elements was similar in Photone. The design and composition rationale for Photone was to compose the fundamental elements in a style inspired by electronic drone music, where the musical elements would be changed by the user's interaction and exploration.

An image in digital form consists of numerous small dots, pixels, where the color in each pixel is normally represented with three numbers, one each for the three color channels red, green, and blue (hence this technique for color reproduction is referred to as the RGB color space). Black is represented with no light (using the value 0) and white is represented with maximum light intensity (the value 255) in all three color channels simultaneously, and virtually any shade of color can be reproduced by adjusting the values in the different color channels [9].

The musical sonification in Photone starts with determining the predominant hue of a given image: red, green, blue, yellow, cyan, magenta, or white (see Figure 5). The musical expression differs in terms of harmonics and melodies between these hues. The intention is to create slightly different impressions of, for example, a whiter image compared to a greener one. The composition within each hue consists of six musical elements that are affected by the user interaction and the color intensity in the image (see Figure 6). The overall harmonic ambience is composed for each color channel (red, green, blue) to range from a simple two-tone chord to more complex chords, depending on the intensity and complexity of the color hues in the image. This creates an impression of a simpler and subtler harmonic ambience for darker and more basic colors, while brighter composite hues are sonified with more complex harmonic chords. Even if the harmonic and melodic components in Photone differ between a more red image and a more blue image, the overall musical expression is somewhat similar. Photone reads the pixel values in an image without regards to the depicted motive. The differences in musical expression lie rather in the user's interaction and exploration of the hues and textures in the images.

The three color channels (i.e. red, green, and blue) are related to different melodic components as well, creating different intervals and different melodic movements as the intensity in the color channels change. An increase in intensity generates a melody going upwards, and an intensity decrease yields a downwards-going melody. In areas with lower intensity, two low frequency bass tones are used to amplify the impression of darker shades of color. These bass tones are then attenuated as the intensity levels in the image increase to further enhance the feeling of different light levels in the image. A high-pitched light intensity chord with high frequency components is used to create an airy and high-intensity feeling in image areas with bright colors. For image areas containing pure white, a bell-like sound is played to accentuate the dazzling intensity of white. Similarly, for areas with very low intensity levels, a low frequency downwards-sweeping sound is played to emphasize the change in intensity from different shades of color to darkness.

4. IMPLEMENTATION

The implementation of the interactive sonification is done in SuperCollider 3.8.0, a programming environment for real-time audio synthesis [10, 11]. The RGB values of an input image are saved

Figure 5: In Photone, the prominent hue of color of an image is determined, and the harmonics and melodic components selected accordingly. First (counted from above) a mainly red image, second a mainly green image, and third a mainly blue image. The small recessed images show the predominant hues. Photos (from above): Niklas Rönnberg, Patric Ljung, Paula Žitinski Elías.

from Matlab R2016a as text files that are read by SuperCollider for adjustment of the sound according to the intensity levels of the different color channels.

The composition consists of six musical elements (see Figure 6). These elements are 1) the overall harmonic ambience, 2) melodic components, 3) bass tones, 4) a high light intensity chord, 5) a bell-like sound for pure white, and 6) a low frequency sweep

Figure 6: In Photone, each image is classified to a certain hue of color and the sonification is adjusted accordingly. The RGB color channels have different chords building up the ambient sound as well as different melodic components. The rest of the musical elements (the bass tones, the high chord, the bell-sound for white, and the synth sweep for black) are consistent within a certain main color shade, and are dependent on the intensity levels for each pixel. Photo: Paula Žitinski Elías.

for pure black. Of these musical elements, the overall harmonic ambience and the melodic components were composed for three color channels. Each image is determined to be mainly red, green, blue, yellow, cyan, magenta, or white and the composition used, in terms of harmonics and melodies, will differ between these colors.

The *harmonic ambience* consists of two-tone intervals multiplied over five octaves, creating a broad harmonic ambience with eleven tones for each color channel (see Figure 6). The red channel, for example, has the tones C2, Eb2, C3, Eb3, C4, Eb4, C5, Eb5, C6, Eb6, and C7. This creates a two-tone interval when the pixel under the cursor has light intensity only in one primary color channel (e.g. red, green, or blue), a four-tone interval when two color channels are used (e.g. red and green are blended together to create yellow, green and blue for cyan, or red and blue for magenta), and a six-tone interval when there is information in all three color channels simultaneously (e.g. white or bright shades of color). Each tone consists of seven triangle oscillators, where six oscillators are slightly detuned in relation to the fundamental frequency (-10, -7, -4, +4, +7, and +10 cents relative the fundamental frequency). The light intensity (0 to 255) of the individual color channel determines the sound level between 50% attenuation of the sound and no attenuation. If the intensity is 0 there is no sound for that color channel. The intensity for each color channel also controls the cut-off frequency of a second order band-pass filter (between 100 and 4000 Hz), through which the two-tone intervals

for each color channel pass. This results in a rich harmonic content ambient sound, where the harmonics varies in sound level and frequency components according to intensity levels in the three color channels independently of each other.

The *melodic components* consists of five tones (e.g. G5, C6, Eb6, G6, and C7 for the red channel), played one tone at a time, for each color channel (see Figure 6). The intensity level is divided into five steps and one of the tones is used accordingly. Each tone is built up with seven triangle waves in the same way as the harmonic ambience, but with a more pronounced attack and an attenuated sustain period. Also, similar to the harmonic ambience, the melodic components vary in amplitude level and in band-pass filter cut-off frequency. The *bass tones* consist of two tones (e.g. C2 and C3) (see Figure 6), that are only present when the overall intensity level is less than 200 of a total of 765, to fill up the ambient music when the harmonic ambience and the melodic components are attenuated. The *high light intensity chord* is composed with three tones (e.g. C7, Eb7, and G7), and played only when the overall light level is above 750. Both these instruments consist of three sawtooth oscillators each. The bass tones pass through a low-pass filter, while the high chord passes through a serial combination of a phaser, a flanger, and a resonant low-pass filter to create the impression of a more non-static changing sound. The *short bell-like sound* (e.g. the tone C8), used to sonify a white area in the image, consists of one triangle wave oscillator with a short

amplitude envelope. The *downwards sweeping sound* (e.g. C2 and C3), used to emphasize a dark region, consists of two sawtooth oscillators and a high resonance low-pass filter sweeping the cut-off frequency from high to low. Finally, all sounds are mixed together and passed through a reverb effect.

A simple screen saver is also implemented in the system, where a circle is moving in random directions across the screen and the sonification alters according to the pixel values under the center of the circle. The screen saver starts after 30 seconds of inactivity, and the photos are changed every fourth minute. The screen saver is used to draw attention to the Photone by displaying photos and creating an ambient sound.

5. OUTCOME AND IMPACT: NOT FOR EVERYONE

Photone was exhibited for a period of two weeks in a science center for the general public focusing on visualization and interactive experience. During this time, the science center had approximately 5000 visitors. Even if there is no exact count of the number of visitors exploring Photone, guides and staff at the science center estimate 100+ visitors. For these two weeks, the system was observed by guides and staff at the science center. Moreover, visitors had the ability to volunteer free-form text comments. There was no specific task the visitors were supposed to do in Photone. It was rather an open invitation to explore the cross-modal experience provided by Photone.

Our informal analysis of the data from the exhibition suggests that some visitors merely toyed with the interaction for a very brief period of time. At other occasions there were groups of visitors discussing the system while taking turns interacting with it, exploring the photos and the musical sounds. Then, there were sessions where visitors spent considerable time exploring the images and sounds, and the occasional comment indicating a visitor having a more or less transformative experience.

"First of all, [I] would like to say that the exhibition [i.e. Photone] was an interesting experience! Seems to be an exciting research area. On Saturday's visit, I expressed to my company that I would have liked to be able to sit for hours and explore this".

A tentative conclusion might be that Photone attracts only some of the visitors to a public exhibition space, but the visitors engaging more deeply in the interaction are drawn into a more or less captivating state of exploration similar to the vignettes outlined above. Our conjecture at this point is that engagement of this kind might correlate with musical aptitude and a reflective state of mind; more systematic studies are of course needed to validate any such speculations. It is also worth noting that Photone implements some of the established practices for facilitating exploration in a public space, such as the notion of progressive internal complexity and the creation of a frame for performative interaction, but not others such as engaging the full body and challenging situated norms [12]. This may also contribute to explaining the wide diversity in terms of visitors' engagement.

6. FUTURE DESIGN AND RESEARCH IDEAS

The tentative conclusions and feedback from users have given rise to some future design and research ideas. It would be interesting to further explore the interplay between colors, musical elements,

and the perception of emotions in a future version of Photone. Musical sounds are well adapted, at least on a more general level, to conveying meaning and emotions (see for example discussions in [13, 14]). Not only music affects emotions but also colors are associated with emotions [15], and there are correlations between the emotional associations of music and of colors [16]. On a more general level the interactive musical sonification used in Photone could adapt more to the prominent shade of color in each image. For example, it would be interesting to make a mainly red image sound distinctly different from a blue image, and with corresponding emotional impressions of the musical sounds as well as of the colors in the image. The harmonic ambience and the melodic components could be composed to instill the emotions in the user that are also present through the predominant color in the image. Apart from obvious aspects of musical sounds, such as major chords and melodies for more positive color schemes and minor chords and melodies for more negative colors, there are other interesting musical elements to consider. One example is melodic movements where an ascending pitch is generally perceived as more positive, while a descending pitch is perceived as more negative [17]. This could be utilized in the melodic components to further emphasize the impression of the overall color hue. Other musical elements that could be fruitful to explore is the harmonic timbre (harmony versus dissonance), as more complex harmonic timbres (i.e. a simple major or minor chord, such as C or Cb, versus a suspended or a diminished chord, such as Csus4 or Cdim) are more captivating for a listener compared to simpler harmonic timbres [18]. A truly dissonant combination of tones is also experienced as more unpleasant compared to harmonious chords [19]. Both the complexity of the harmonics in chords and the use of dissonant tones in a chord could be used to further link the musical sounds, the experience of color, and the emotional response.

The overall lightness of the image could also be used to create the impression that the musical sounds are even more connected to and derived from the image. For example, lighter and brighter colors are associated with higher pitched tones [20, 21, 22]. There might also be a relation between the experienced difference between high and low pitch tones, and the experienced difference between brighter and darker areas (see for example discussions in [23, 24, 25]). The overall brightness in the image could be used to select different ranges in the harmonic ambience with higher pitches for brighter images, and lower pitches for darker images. The use of a more soft or dull timbre will also be experienced as more negative compared to a brighter timbre [17]. In a similar way, the melodic components could be transposed in accordance with the brightness level in the image.

In the current version of Photone the amplitude and the timbre change (by changing the cut-off frequency of a band-pass filter) depending on the intensity levels in the three color channels. This affects each color channel separately, but in a future version of Photone the overall brightness could also affect the overall amplitude and frequency content in all of the musical sounds. Interesting to note is that a louder sound is more activating and engaging for the listener/user than a quieter sound [18]. The perception of loudness is also mapped to brightness, for example in a image, via the amplitude of a sound [26]. Consequently, a brighter image would have louder musical sounds with higher pitched harmonics compared to a generally darker image.

Further more, it would be fascinating to use image processing algorithms to detect patterns and shapes in the images. The pattern recognition could detect different levels of patterns, such as con-

trasts between and differences in hues of color and brightness. The level of patterns detected could then be used to adjust the tempo of some musical elements, for example the tempo of an arpeggio or of a percussive rhythmic sound, where a more cluttered image would yield musical elements with a faster tempo. The shape recognition could be used to detect different shapes or movements in the image, such as downwards going shapes that could lead to a descent in pitch when the user explores that shape, or the level of roughness in the image which in turn could be mapped to the cut-off frequency of a filter or the wave-shaping (e.g. the pulse width of a square wave sound) of the musical sounds leading to more harsh sounds when the roughness in an area of the image is high, and more smooth sounds when the roughness is low.

Finally, Photone could also be expanded to a completely immersive experience, with stereoscopic images and spatial sounds, using VR or AR techniques. It would also be possible to involve a tactile modality, in order to explore the potential of multimodal synergy.

We believe that these future design and research ideas would make Photone into an even more interesting interactive experience of photographic images and musical sounds. The next implementation of Photone will have logs of all mouse cursor movements with time stamps. This data will be used to identify interaction patterns and areas of interest. These data will be accompanied by data collected from an observational study, as well as interviews about the user experience. More systematic studies of the interactions in Photone, and about the interplay between music and image, would also lead to new and interesting insights in modal synergies.

7. DISCUSSION: TOWARDS MODAL SYNERGY

The work we report here is explorative and to some degree tentative. Still, it has already yielded insights of some significance to the field of auditory/multimodal display and, more generally, to interaction design.

Figure 7: Photone can be explored either with a computer mouse or a touch screen with quite different experiences of the interaction. Photo: Niklas Rönnberg.

We have experimented with Photone in two different hardware configurations, i.e., one with mouse and cursor input and one with

a vertical touchscreen (see Figure 7). It is quite striking how different the two conditions are with respect to the resulting experience of use. When using a mouse and cursor, the experiential state is one of prolonged and incisive exploration (looking closely, manipulating precisely). The touchscreen, on the other hand, becomes more of a way to sample the general musical timbre of different color and brightness areas. This difference is mainly due to the current implementation of the sonification algorithm, where the sound modulation is based on the exact pixel value under the cursor. Another part of the reason is that music unfolds over time, and moving the cursor slowly over a textured area in the image while listening attentively to changes in the music is simply more comfortable with a mouse that allows the hand to rest. This issue is solvable as such, of course, with a persistent finger-controlled pixel-level cursor offset from the point of touch, but the point is rather this: *the fine-grained, bodily details of the interaction must be considered part of the use experience together with the visual and auditory output modalities*. Working with a conventional vertical touchscreen, for example, would properly speaking motivate redesigning the sonification towards a more expressive interaction with large, sweeping movements and "broad strokes". Furthermore, the touchscreen adds an interesting feature to the interaction, which is the possibility to go from one pixel in a certain area of the image to another pixel in a different area. This enables, as mentioned above, sampling of different color and brightness values and the musical sounds derived from these, without the transitions between these areas. Thus, the touchscreen permits a nonlinear exploration, and to some degree facilitates the use of the image as a musical instrument when "playing the image".

This line of reasoning aims to show how image, sound and interaction are tightly integrated in Photone. In fact, we propose to see the three elements as aspects of the same emergent experience in use. Multimodality and multimodal interaction have been established concepts in interaction design for a long time, but it is notable that the concepts have not been unpacked systematically from an experiential point of view to any significant extent. Most accounts of multimodality seem to focus on how to use different input modalities (keyboard, mouse, touch, speech, etc.) in concert and how different output modalities supplement each other [27]. In that sense, the prevalent conceptual understanding of multimodality seems to be more or less on par with where we started in musical sonification: "How can sonification supplement an existing visual modality to enhance visual presentations and increase the users' task performance?".

Our current work aims to go beyond this supplementary approach to explore how the visual and auditory modalities come together and form an experiential whole through interaction. We have focused specifically on photographic image and musical sound, and it is worth noting that our approach to the images is rather deconstructive. Musical sonification of photographic images based on their denotative and connotative meanings ("this is a picture of a flower" and "this picture suggests tropical mystique", respectively) would essentially amount to the kind of impression management found in movie music composition, or more simply in the approach known as auditory icons (such as the sound of a bee buzzing around the flower and the sound of a gong and wind chimes for tropical mystique). In Photone, the sonification is driven by pixel values of hue and brightness, that is, syntactic rather than semantic properties of an image. Other syntactic image features that could be used in the same way include edges, shapes and spatial frequency measures, see discussion above. Our aim

was to cut through conventional ways of looking at a photographic image. At the risk of sounding pretentious, we would argue that the current combination of photographic images, sonification algorithms, and mouse-and-cursor input represents the potential for a more foundational way of seeing – similar to how a beginning portrait artist needs to unlearn ideas of noses, ears and facial expressions in order to see lines, shapes and hues. To us, this represents one example of how image, sound and interaction together form something more than a mere addition of supplementary modalities. We call this emergent phenomenon *modal synergy*, and we propose that further design-led inquiries in this direction may be a way to add to our constructive knowledge of auditory displays and multimodal interaction.

8. REFERENCES

- [1] T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*, 1st ed. Berlin, Germany: Logos Publishing House, 2011.
- [2] T. Pinch and K. Bijsterveld, *The Oxford Handbook of Sound Studies*. Oxford University Press, 2011.
- [3] K. Franinovic and S. Serafin, *Sonic Interaction Design*. Cambridge, MA, USA: MIT Press, 2013.
- [4] L. Philipsen and R. S. Kjærgaard, *The Aesthetics of Scientific Data Representation: More than Pretty Pictures*. Denmark, Europe: Routledge. Routledge Advances in Art and Visual Studies., 2018.
- [5] S. J. Deliège, I., *Perception and Cognition of Music*. Hove, East Sussex: Psychology Press Ltd., 1997.
- [6] N. Rönnerberg, J. Lundberg, and J. Löwgren, “Sonifying the periphery: Supporting the formation of gestalt in air traffic control,” in *The 5th Interactive Sonification Workshop*, ser. ISON-2016. Germany: CITEC, Bielefeld University, 2016, pp. 23–27.
- [7] N. Rönnerberg and J. Johansson, “Interactive sonification for visual dense data displays,” in *The 5th Interactive Sonification Workshop*, ser. ISON-2016. Germany: CITEC, Bielefeld University, 2016, pp. 63–67.
- [8] N. Rönnerberg, “Sonification enhances perception of color intensity,” in *Proceedings of IEEE VIS 2017, InfoVIS posters*, ser. IEEE VIS 2017, 2017.
- [9] C. Ware, *Information Visualization: Perception for Design*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2013.
- [10] J. McCartney, “Supercollider: A new real-time synthesis language,” in *The International Computer Music Conference*, ser. ICMC’96. New York, NY, USA: ACM, 1996, pp. 257–258.
- [11] —, “Rethinking the computer music language: Supercollider,” *Computer Music Journal*, vol. 26, no. 4, pp. 61–68, 2002.
- [12] M. Hobye, “Designing for homo explorans : open social play in performative frames,” Ph.D. dissertation, Faculty of Culture and Society Malmö University, Malmö, Sweden, 2014. [Online]. Available: <http://muep.mau.se/handle/2043/16510>
- [13] N. Rönnerberg and J. Löwgren, “The sound challenge to visualization design research,” in *Proceedings of EmoVis 2016, ACM IUI 2016 Workshop on Emotion and Visualization*, ser. EmoVis 2016. Sweden: Linköping Electronic Conference Proceedings, 2016, pp. 31–34.
- [14] F. J. L. L. Tsuchiya, T., “Data-to-music api: Real-time data-agnostic sonification with musical structure models,” in *Proc. 21st International Conference on Auditory Display (ICAD 2015)*, 2006, pp. 244–251.
- [15] A. N. Gilbert, A. J. Fridlund, and L. A. Lucchina, “The color of emotion: A metric for implicit color associations,” *Food Quality and Preference*, vol. 52, pp. 203–210, 2016.
- [16] S. E. Palmer, K. B. Schloss, Z. Xu, and L. R. Prado-Leon, “Musiccolor associations are mediated by emotion,” *PNAS*, vol. 110, pp. 203–210, 2013.
- [17] P. Juslin and P. Laukka, “Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening,” *Journal of New Music Research*, vol. 33, pp. 217–238, 2004.
- [18] S. A. Iakovides, V. T. Iliadou, V. T. Bizeli, S. G. Kaprinis, K. N. Fountoulakis, and G. S. Kaprinis, “Psychophysiology and psychoacoustics of music: Perception of complex sound in normal subjects and psychiatric patients,” *Annals of General Hospital Psychiatry*, vol. 3, pp. 1–4, 2004.
- [19] K. J. Pallesen, E. Brattico, C. Bailey, A. Korvenoja, J. Koivisto, A. Gjedde, and S. Carlson, “Emotion processing of major, minor, and dissonant chords: A functional magnetic resonance imaging study,” *Annals New York Academy of Sciences*, vol. 1060, pp. 450–453, 2005.
- [20] W. G. Collier and T. L. Hubbard, “Musical scales and brightness evaluations: Effects of pitch, direction, and scale mode,” *Musicae Scientiae*, vol. 8, pp. 151–173, 2004.
- [21] L. E. Marks, “On cross-modal similarity: Auditoryvisual interactions in speeded discrimination,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 13, pp. 384–394, 1987.
- [22] J. Ward, B. Huckstep, and E. Tsakanikos, “Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all?” *Cortex*, vol. 42, pp. 264–280, 2006.
- [23] J. Best, *Colour Design: Theories and Applications*, 2nd ed. Duxford, United Kingdom: Elsevier Ltd. : Woodhead Publishing, 2017.
- [24] R. Bresin, “What is the color of that music performance?” in *Proc. International Computer Music Conference ICMC 2005*. San Francisco, CA: International Computer Music Association, 2005, pp. 367–370.
- [25] S. E. Palmer, T. A. Langlois, and K. B. 1 Schloss, “Music-to-color associations of single-line piano melodies in non-synesthetes,” *Multisensory Research*, vol. 29, pp. 157–193, 2016.
- [26] R. W. Pridmore, “Music and color: Relations in the psychophysical perspective,” *Color Research & Application*, vol. 17, pp. 57–61, 1992.
- [27] M. A. Mathias Nordvall, “Perception, meaning and transmodal design,” in *Proceedings of Design Research Society 50th Anniversary Conference*, ser. DRS 2016, 2016, pp. 1–12.

ACOUSTICS OF VIRTUALLY COUPLED PERFORMANCE SPACES

Braxton Boren

American University
Audio Technology Program
4400 Massachusetts Avenue
Washington, DC 20016
boren@american.edu

Andrea Genovese

New York University
Music and Audio Research Laboratory (MARL)
35 W. 4th St.
New York, NY 10012
genovese@nyu.edu

ABSTRACT

Many different musical applications, including remote sonification, sound installation, augmented reality, and distributed/telematic music performance, make use of high speed Internet connections between different performance spaces. Most of the technical literature on this subject focuses on system latency, but there are also significant contributions from the acoustics of all rooms connected: specifically, smaller auxiliary rooms will tend to introduce spectral coloration, and the “main” larger volume will send more reverberation to the off-site performers. Measurements taken in two linked networked sites used in telematic performance show that both of these issues are present. Some improvements are suggested, including physical room alterations and equalization methods using signal processing.

1. INTRODUCTION

Many different musical applications, including remote sonification [1], auditory display [2], augmented reality [3], and distributed/telematic music performance [4, 5], make use of high speed Internet connections between different performance spaces. These share the goals of fostering greater scientific, personal, and musical interaction over distance by connecting scientists, educators, musicians, and audience members who would not normally be able to communicate in real time. These distributed sound interactions share the goal that such virtual connections should sound as perceptually naturally as possible [6], ideally indistinguishable from a comparable interaction within a single acoustic space, wherein two performers are only distributed across a few meters of a natural stage [7].

In all these areas, which we will group together under the umbrella term Networked Audio Platforms [8], the chief barrier to achieving ‘naturalness’ has been latency in the time required for a signal to travel from one participant to another, and as such it is also the chief problem addressed in the corresponding literature [9, 10, 11]. Bartlette and Bocko specify this as “external latency” as opposed to “internal” latency or timing variance due to differences between individual performers [9]. This external latency can be further subdivided into transmission delay (“propagation latency”) and any additional delay introduced by signal-

processing demands during transit (“system latency”) [7]. Finally, propagation latency may be either acoustic or electronic, depending on whether the signal is traveling between a sound source and a transducer, or between two transducers via the Internet.

In the past two decades, increases in computational power have reduced system latency for simple transmissions, making propagation latency the controlling variable for final system latency. Depending on the exact distances in question, common electronic propagation latencies are often on the order of tens of milliseconds [7]. Some have made the case that because there are non-networked (i.e. acoustic) sound environments in which such latencies also occur, that electronic path latencies of similar magnitude are satisfactory [9]. This reasoning requires that all acoustic sound sources are close-miked with no appreciable sound path between source and microphone, and likewise that all listeners are wearing headphones, with virtually no time delay between the transducer and the listener’s ear drums. In practice, both of these assumptions are untrue in various networked audio settings, meaning that the total propagation latency may be larger. In addition, it is often assumed that electronic signals travel at the speed of light, but this is also not practically achievable due to certain Internet infrastructure limitations [12]. Acceptable latency values for musical applications are variously reported as up to 30 ms [13], up to 50 ms [14], or higher [9], depending on how the criterion for acceptability is defined.

Despite this understandable historical focus on electronic propagation and system latency, Network Audio Platforms are also profoundly affected by the room acoustics of the networked spaces at either end (or more than two ends in many cases). In many cases the physical acoustics of the rooms at either end of the transmission line may affect the perceptibility of the delays due to transmission. Though individual acoustic effects have been addressed at times in the literature, there exists no single systematic treatment of the relationship between the acoustic characteristics of virtually linked spaces and the final audio system that results from their combination. This paper provides a background on the acoustic effects present in networked spaces, as well as the psychophysics of coupled acoustic volumes and musicians’ responses to different types of non-networked acoustic spaces. It will then present a unified framework for addressing systematic musical asymmetries that may arise from differences in the acoustic responses of networked performance rooms.

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

2. BACKGROUND

Though the literature on networked audio platforms focuses primarily on propagation latency, there are several references to room acoustics, albeit mostly in the context of addressing latency. Many researchers have noticed that at the present, it is impossible to remove all latency issues from transmission, so they propose a compromise between the real world and the effects of the transmission line. This ‘compromise’ strategy includes “using networked time delay in a musical fashion rather than constantly trying to counter it” and “combining the acoustics of physical spaces with the acoustics brought by the network itself” [10]. Chafe [12] designed a system to use feedback delays within a networked audio platform to trigger comb filters, which were later processed with an all-pass delay filter to create a single virtual reverberation on both ends of the network, thus using a real acoustic phenomenon to mask an audio problem. Gang et al. [11] showed that reverberation can mask some deleterious effects of transmission latency. While Braasch [7] does caution that it is important for networked performers to have a greater awareness of real-world sonic environments, he also stresses that “unique affordances exist for telematic music systems, which is a basic requirement for them to serve as a platform for new types of music.”

The label of ‘compromise’ above is not meant as a pejorative - indeed this approach is responsible for a great improvement in the smoothness and fluidity of networked audio performances in the past two decades. However, it is not advisable to let the necessary limitations of the present obscure our vision for the future: to take an example from another area of audio technology, for years headphone design focused on free-field or diffuse-field equalized headphones because these were intended to reproduce stereo audio mixed for loudspeakers [15]. But as the demand for high-quality binaural audio for virtual reality has increased, it is clear that more comprehensive inverse filtering will be necessary to remove the response of headphones entirely to provide a perceptually blank slate for dynamic binaural synthesis [16]. In the same way, a ‘compromise’ networked audio brings participants on either end into a virtual, neutral acoustic space which is not necessarily similar to the physical rooms on either end of the transmission. But a future system might offer either a sort of “acoustic gateway” between two sites, wherein one participant receives the same 3D sound field as a listener in physical room miles away, or an “acoustic holodeck” [17] where participants meet in a neutral virtual space, but one that is fully changeable and not dependent on the acoustic or audio characteristics of the physical performance spaces or their distance apart. Neither of these goals are attainable at present, and they are constrained not just by propagation latency but also by room acoustics. To better understand the impediments to improved networked audio platforms, it is important to understand the acoustic constraints imposed by the physical coupling of two separate volumes.

3. ACOUSTIC COUPLING EFFECTS

If a signal $x(n)$ is produced in a physical room and recorded at a microphone in the same room, $R1$, the resultant recorded signal $y(n)$ will be equal to the convolution

$$y(n) = x(n) * h_{R1}(n), \quad (1)$$

where $h_{R1}(n)$ is the room impulse response (RIR) representing the acoustic system from the sound source, through the room, and

into the microphone (thus accounting for any filtering characteristics of the microphone itself). The input sound $x(n)$ will experience the room decay present in the RIR $h_{R1}(n)$. Assuming the room is reasonably diffuse, the RIR will exhibit an exponential decay, which when plotted in dB will appear as a straight line. The slope of this line determines the reverberation time, T_{60} , of the room.

If we imagine a second adjacent room $R2$ with corresponding RIR h_{R2} and a longer reverberation time, and then create an opening or aperture of surface area S between $R1$ and $R2$, then this ‘open-window’ aperture appears, from the perspective of a single room, as an additional source of absorption such that each room’s total absorption values becomes

$$A_{iS} = A_i\alpha_i + S, \quad (2)$$

where A_i is the amount of surface area in the i th room, and α_i is the average absorption coefficient of the i th room [18].

However, the sound traveling through the aperture is not lost completely, as with an open window - instead it enters the second room, whose sound will likewise spill back into the first room. The phenomena described here is called acoustic coupling, and the system consisting of two coupled volumes will exhibit a decay curve that is a combination of the decay curves of both constituent volumes. If we now measure the impulse response from a source to a receiver, both in $R1$, with $R1$ partially coupled to a more reverberant $R2$ through an aperture, then the resulting impulse response will have a non-linear decay curve, made up first of the steeper decay rate of h_{R1} , while the later part of the curve will represent the more gradual decay rate of h_{R2} . As the curve transitions between these linear portions, there will be a bend which has a greater dB value than either of the individual rooms’ decay slopes, as shown in Fig. 1. This is known as the double-slope effect (DSE) or a double-slope decay (DSD) [19, 20].

Figure 1: Double-Slope Effect resulting from the coupling of two rooms with different decay rates.

The early and late portions of the decay are respectively given by

$$L_{early}(t) = -\left(\frac{60}{T_1}\right)(t) \quad (3)$$

and

$$L_{late}(t) = -\left(\frac{60}{T_2}\right)t - 10 \log_{10} \left(\frac{A_{1S}A_{2S}}{S^2} \right), \quad (4)$$

where T_i is the independent reverberation time of the i th room, and t is the time index [18]. Note that the second term in Equation 4 implies that impulse response of the coupled system is not equivalent to the convolution of the impulse responses of each room alone.

Coupling effects need not be restricted to only two connected rooms, as more adjacent volumes and apertures could produce a non-linear decay made up of three or more piecewise linear components, each corresponding to the decay rate of one of the individual rooms [18]. Because of the nature of the DSE, it is necessarily more complicated than a single room's decay and cannot be described by a single number as can a diffuse volume using T_{60} . Different parameters to measure the degree of coupling have been suggested, but all vary in effectiveness depending on the ratio of the linear slopes and the transition zone between them [20].

In the example above, if we instead put the source and receiver both in $R2$, the more reverberant of the two rooms, its (more gradual) decay rate will also serve as the initial decay rate, but because the sound will decay much more quickly in $R1$, there will be no DSE from the perspective of $R2$. This introduces an interesting asymmetry that will be important for later analysis: *dry rooms can hear reverberant rooms coupled to them, but reverberant rooms cannot hear dry rooms coupled to them* [18]. Reverberant volumes can act as a sort of time capsule, keeping sound around past when it would normally have decayed in the drier room, and letting that sound back into the dry room. For this reason, coupled volumes have been studied in the context of room acoustics, especially regarding dry spaces (like balconies or small chapels) connected to more reverberant spaces (such as concert halls or the nave of a cathedral) [21, 22]. However, the concept of coupling can easily be extended to two spaces which are not acoustically joined, but rather connected via networked audio.

4. COUPLING IN NETWORKED AUDIO PLATFORMS

The situation for Networked Audio Platforms has many distinct differences from the purely acoustic coupled volumes presented so far, but there are also many similarities. At the most basic level, there are still at least two different rooms which are hearing sound from each other, and these rooms likely have somewhat different acoustic properties. The range of different combinations is too large to address every possibility, so instead we restrict ourselves to a fairly simple setup which is used in many telematic/distributed performances. In this scenario, we assume there is a single 'main' volume containing multiple performers and an audience, and a second 'off-site' volume connected via networked audio, containing a single performer in another location who is joining in with the performance virtually. Just as before, the combined coupled system's impulse response will not be identical to the convolutions of the individual impulse responses, based on differences introduced not through an aperture but through the audio system connecting them.¹

Before considering the acoustic issues introduced, we must define some basic features of the two rooms. Again, individual

¹Note that in the case of virtual coupling, the amount of additional absorption from the first room's perspective due to an 'aperture' is insignificant, since the 'aperture' is just a microphone with little acoustic absorption. However the spillover effects from room to room proceed as with physical coupling, increased to some extent because of the amplification of the virtual signal received at the microphone.

situations may differ, but because of the presence of multiple performers plus an audience in the main site, we may usually assume this location is the larger of the two volumes. The off-site location need not be very big, and may be assumed to be roughly the size of a practice room or iso booth. Similarly, the off-site performer needs only an instrument, a computer, a microphone, and headphones (which are not strictly necessary but will be assumed for this example), while the performers at the main site, along with the audience, are receiving audio over loudspeakers.

4.1. Frequency Domain Effects

What acoustic issues will arise in such a system, irrespective of the propagation delay between sites? First let us consider the frequency domain: if the smaller room is a well-designed iso booth built to favorable dimensions and with no parallel surfaces, it may have a reasonably smooth frequency response due to the wide distribution of resonant modes through the audible spectrum. If instead it is a simple rectangular volume with flat plaster walls, it may exhibit significant modal resonances which will distort the timbre of the signal being sent to the main room.

Conversely, the larger main volume will have a low Schroeder frequency, meaning that most if not all the reflections within the room can be considered diffuse and evaluated chiefly in the time domain rather than by considering individual room modes [23]. This means there are likely to be no large resonant peaks or notches in the main room's frequency response, either from the perspective of the audience or the off-site performer. However, the sound coming from off-site may be significantly affected, depending on the geometry of the smaller room. These frequency-dependent effects, however, can be easily compensated for via regularized inversion in the frequency domain, a technique that is widely used to equalize headphones, loudspeakers, and rooms [24, 25]. Even though the larger volume is likely to have a smoother spectrum, the reverse may also be applied to equalize the main performers' audio before reaching the off-site room.

This ability to get near-perfect removal of frequency effects means that the off-site performer may also be removed from their own room and put into the main volume, at least as far as the frequency domain is concerned. Since the off-site performer is listening through headphones, they will hear both their own instrument filtered through the off-site room, as well as their delayed version which has been equalized. If EQ is being applied in both directions, theoretically this delayed version could be free of coloration due to either of the rooms, within the constraints determined by regularization. One caveat is that such equalization assumes that the frequency response of both rooms has been measured in advance, which is not always the case.

4.2. Time Domain Effects

In the time domain, the situation is more complex: the smaller site will usually have a very small T_{60} value, unless it is extremely reflective. The larger main site will likely have a longer reverberation time, especially if it is a site intended for live music performance. Because of that, there are various sound paths that need to be considered: one-way paths, from one site to the other, and round-trip² paths, from one site to the other, filtered through the second site,

²In some distributed performances [26] audio is designed to flow only one-way and the return path is disabled, in which case the round-trip analysis will not apply.

and back to the sound's origin. These different paths are shown in Table 1. In addition, for reference we will sometimes consider the single-room impulse responses for either space alone.

Table 1: Sound paths

One-Way	Round-Trip
Off-Site to Main Room	Off-Site Back to Off-Site
Main Room to Off-Site	Main Room Back to Main Room

In general, the discussion of these four main paths involves the amount of reflection and reverberation added by sound traveling through both rooms. This merits the question of whether blind dereverberation may be applied in the time domain as easily as regularized inversion in the frequency domain. In principle dereverberation may be achieved similarly to frequency equalization, but since RIRs are orders of magnitude longer than their frequency domain equivalent, dereverberation is much more computationally complex. Since different source positions determine slightly different RIRs, many dereverberation techniques focus on a blind determination of the best inverse filter for the total audio signal received, a problem which is still unsolved. Exact inversion may actually degrade the input signal, whereas algorithmic processing adds significantly to system latency [27]. In addition, many existing techniques focus only on improving signal-to-noise ratio for speech intelligibility, which is applicable to some (e.g. teleconferences) but not all (e.g. distributed music performances) Networked Audio Platforms [28, 29]. While multiple input channels of audio can improve the output filter [30, 31], this also increases computational load. In the future, multiple audio channels may be desirable both for increased dereverberation and for improved immersion for performers, whose preferences include spatial as well as temporal placement of sound reflections [32]. While dereverberation certainly poses an important area of future exploration, at present it is not possible to cheaply and quickly remove the reverberation from a signal sent from one room to another, implying that these rooms are in a sense coupled together similarly to adjacent acoustic volumes connected by an aperture. Keeping this in mind, we will apply the theoretical framework developed in Section 3 to the virtual coupling in our imagined musical performance.

4.2.1. Off-Site to Main Room

Since the signal from the off-site performer receives a much lower T_{60} value than the main room, it emerges from a loudspeaker in the main room dry, and then experiences the longer reverberation of the larger volume. In this sense it is similar to the instruments of the performers in the main room, whose signal emerges dry – either from acoustic instruments or from loudspeaker-based virtual instruments – and is convolved with the RIR of the main volume. While some early reflections will be mixed with the off-site performer's direct sound, these will be lost in the larger mix of reflections in the main volume and will not be perceptible to the audience.

4.2.2. Main Room to Off-Site

The one-way path for sound originating in the main room and traveling off-site involves a larger reverberation time in the larger volume, which will then be heard in the signal received off-site. Assuming (as we are) that dereverberation is not a practical option,

this creates an asymmetry in the one-way sound paths: the off-site performer hears the added decay of the main room in addition to their own, but those in the main room do not hear the time decay of the off-site room.

4.2.3. Off-Site Back to Off-Site

From the perspective of the off-site room, the performer's own sound is transmitted through the smaller room, through the network, into the main room, which then acquires the reverberation of the larger space before being fed back into the network and reaching the off-site performer. This is a feedback echo, which is used in the 'compromise' strategy described above to create a single, shared reverberation time between both spaces [12]. If the virtualized T_{60} is greater than that of the main room, then the virtual reverberation will indeed be heard equally by performers on both ends. If, however, the main room's T_{60} is greater, then this will form a DSD similar to Fig. 1. The early decay time (EDT) may still be perceived as that of the shorter decay rate [18], but the late field reverberance will likely be perceived as the longer T_{60} value of the main volume, depending on the ratio between the two decay rates. This suggests that the minimum T_{60} of the virtually coupled system is constrained by the greatest T_{60} of any of the rooms being joined, i.e. $T_{system} \geq T_{main}$ in our example. In addition, this affects the stable range of achievable decay rates, constraining the types and tempos of music that may be realized in a reverberant acoustic system.

4.2.4. Main Room Back to Main Room

From the perspective of the main room, the performers' signal acquires some reverberation in the larger volume before being sent off-site. Since the off-site performer is wearing headphones, this signal will not be fed back, and thus there is no significant round-trip path in this case. However, the performers will still hear the T_{60} of the main room acoustically (more strongly than the off-site performer). If a virtualized reverberation is imposed on the network, then it will not be heard separately unless its T_{60} is greater than that of the main volume, which will also cause a DSD.³ Thus the acoustics of the larger volume become a controlling factor for the performance of both the on-site and off-site musicians.

5. MEASUREMENT RESULTS

In order to explore and validate the theoretical foundation under the given assumptions, the case scenario illustrated in section 4 was measured in the time and frequency domains. Two candidate locations known for having been previously used for distributed music performances were chosen within the NYU Steinhardt Education building. The Frederick Loewe Theatre (294 seats, dimensions approximately 17x21x6 m, volume approximately 2180 m³) was selected as the main site, while the Research Laboratory on the 6th floor of the same building (rectangular, hemi-anechoic, dimensions 4.5x3.7x2.5 m) was selected as the smaller off-site location for remote performance.

In both rooms, a Genelec speaker was placed one meter away from a C617 Josephson omnidirectional microphone. Three 2-second exponential sine-sweep measurement signals (ranging

³It is also possible that a performer very close to a loudspeaker could perceive the virtualized EDT, but we are principally addressing the room as a whole.

from 20Hz to 20kHz) as described in [33] were controlled from a central studio location able to patch to both environments. The sine-sweeps were first measured independently in both rooms, and then through “one-way paths” in both directions, meaning that a sine-sweep reproduced in the main room was captured in the main room, routed to the off-site room and measured, and vice versa. Finally, round-trip paths were measured from a source in one room, through the other room, and back to the first - this was done by playing the “one-way” sweep recording from the opposite room’s speaker and recording its response in the originating room.

5.1. Frequency domain

Despite the claim that inverse filtering of room effects could remove frequency coloration from off-site volumes, in practice real-time filters to implement this equalization will add somewhat to the propagation latency between the two sites, and are therefore often not used in telematic performance. However, the measurements in fig. 2 illustrate the frequency responses of the two measured rooms, plus that of a source in the off-site room and measured at a microphone in the main room. What comes primarily to attention are the strong notches captured within the off-site room, probably caused by the resonant modes typical of rectangular spaces. The same notches are also captured in the sound transmitted from the off-site room to the main room. This suggests that the incoming transmissions to the main room present timbre coloration introduced by the acoustical character of the off-site room. In contrast, the response of the main room alone indicates a more diffuse field and less dominance of any particular resonances.

Figure 2: Frequency spectrum resulting from the coupling of a small offsite space with a larger main room, smoothed according to [34]. The red line represents the off-site room’s transfer function alone, the blue line is the spectrum of a source in the off-site room measured from a microphone in the main room, while the dashed black line represents the spectrum of the main room alone.

Based on these significant notches of more than 10 dB, some equalization seems advisable for use in musical performance. A simplified inverse IIR filter or short FIR filter could be used to correct for the biggest spectral coloration without large increases in system latency. In addition, some networked audio applications using spatial audio may already require real-time convolution for implementing binaural filters on transmitted audio – in this case, the addition of a single static inverse filter does not require extra FFT operations, but only an additional vector multiplication in the frequency domain, which is not a large computational load

increase. In addition the convolution of a binaural filter with the static room inverse filter could also be pre-computed, which would further reduce latency at the expense of additional system memory.

5.2. Time Domain

The backwards-integrated Schroeder decay curves were calculated by octave band for the off-site room alone, the off-site room to the main room, and the round-trip path from the off-site room back to itself. The results for the octave band centered on 1000 Hz are shown in fig. 3. The results show an expected agreement with the Double-Slope Effect: as expected the T_{60} value for the small off-site room is approximately 0.1 s. At the receiving end, the sound is played back through a loudspeaker and mixes with the hall’s reflections. Given the slower decay rate of the concert hall, both the one-way and round-trip RIRs show a distinct bend indicative of coupling between the faster decay of the off-site room and the slower decay of the main volume. In the late field, the captured sound presents now approximately the same decay characteristics as the one reproduced in the main room. The off-site performer will thus be subjected to both the resonant modes of the smaller space, and the reverberation effects of the larger hall.

Figure 3: Decay curves measured for coupled rooms. The red shows the decay in the off-site room alone, the blue curve is the signal decay from a source in the off-site room measured at a microphone in the main room, and the black curve the resulting round-trip loop from a source in the off-site room, through the main volume, and measured back in the off-site room.

6. CONCLUSION

Due to the nature of acoustic coupling and DSD, rooms which are virtually linked will also transfer their acoustic parameters to each other to a certain extent. In the frequency domain, it is relatively simple to remove the worst modal resonances of small spaces’ audio signals. But in the time domain, it is currently not possible to adequately remove reverberation from a larger space’s signal, and applying existing dereverberation techniques adds to system latency considerably. This poses significant tempo constraints on the performers, who will tend to play slower in the presence of greater reverberation [35, 36]. Networked audio often tends to feature slow tempos, and reverberation is thought to help mask feedback delay, so this is not always a problem. However, it puts a hard constraint on the overall tempo space that may be achieved via telematic performance. To use a historical analogy, it was not

a problem to perform slow monophonic chants in highly reverberant churches, but the development of faster, more complex music necessitated drier spaces with greater acoustic clarity [37].

On top of the very real constraints imposed on distributed music performance by propagation delay and feedback echoes, the acoustic properties of any large space used during the performance are a permanent constraint on the composite Networked Audio Platform. This can of course be addressed by adding sufficient absorption to any large volumes to create extremely low T_{60} values. Though performing in a very dry space can also cause performers to lower their tempo [38], additional virtual reverberation can always be added in.

This suggests a problem, in that many distributed performances' main (i.e. audience-containing) room is often intended for acoustic performance and thus has a T_{60} between 1-2 seconds at mid-frequency. One possibility for distributed performance is to simply remove the large room, sending out an Internet live stream to an audience that is equally distributed, free to listen wherever they are. However, this also removes any communal aspect to the experience of the music, and would diminish the experience of attending a performance rather than simply consuming it like any other Internet video. A better solution is to always place the audience for a Networked Audio performance in a dry space, such as those intended for rock music or electroacoustic performance. This removes the need for dereverberation while ensuring that, whatever virtual acoustic is painted on the audio transmissions, all performers are starting from a blank slate.

Aside from the discussed case scenario, a further question of telematic presence has yet to be explored. Whenever the goal is to create a virtual common performance space between distributed musicians, the reproduction systems at both ends will influence the decision on whether incoming signals should be either filtered or dereverberated, and whether performers should be shielded by their own real signals (for example by using closed headphones) or let the virtual sound mix equally with the real sound. Performance style, latency, reproduction systems and acoustics are ultimately all interacting factors that affect the quality of performance, sometimes asymmetrically at each connected node.

7. ACKNOWLEDGMENTS

The authors would like to acknowledge Michael Ikonomidis for his invaluable technical help in connecting the measured rooms and patching the signals through the studios. Thanks also to Tom Beyer and Dr. Agnieszka Roginska for their advice.

8. REFERENCES

- [1] K. E. Wolf and R. Oda, "MalLo March: A Live Sonified Performance With User Interaction," in *Proceedings of the 22nd International Conference on Auditory Display - ICAD 2016*, Canberra, Australia, 2016.
- [2] M. Musick, J. Turner, and T. H. Park, "INTERACTIVE AUDITORY DISPLAY OF URBAN SPATIO-ACOUSTICS," in *Proceedings of the 20th International Conference on Auditory Display (ICAD)*, New York, NY, 2014.
- [3] J.-M. Jot and K.-S. Lee, "Augmented Reality Headphone Environment Rendering," in *Proceedings of the Conference on Audio for Virtual and Augmented Reality*, Los Angeles, CA, 2016.
- [4] J. Lazzaro and J. Wawrzynek, "A Case for Network Musical Performance," in *Proceedings of 11th International Workshop on Network and Operation Systems Support for Digital Audio and Video*, Port Jefferson, NY, 2001.
- [5] J. Braasch, B. Woodstrup, P. Oliveros, and C. Chafe, "Tele-Colonization," in *International Conference on Auditory Display*, Montreal, Quebec, 2007.
- [6] A. A. Sawchuk, E. Chew, R. Zimmermann, C. Papadopoulos, and C. Kyriakakis, "From remote media immersion to distributed immersive performance," *2003 ACM SIGMM Workshop on Experiential Telepresence, ETP '03*, pp. 110–120, 2003.
- [7] J. Braasch, "The Telematic Music System: Affordances for a New Instrument to Shape the Music of Tomorrow," *Contemporary Music Review*, vol. 28, no. 4-5, pp. 421–432, 2009.
- [8] R. Mills, "Dislocated Sound : A Survey of Improvisation in Networked Audio Platforms," in *Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010)*, Sydney, Australia, 2010, pp. 186–191.
- [9] C. Bartlette and M. Bocko, "Effect of Network Latency on Interactive Musical Performance," *Music Perception*, vol. 24, no. 1, pp. 49–62, 2006.
- [10] J.-P. Caceres and A. B. Renaud, "Playing the Network : the Use of Time Delays As Musical Devices," *Proceedings International Computer Music Conference*, pp. 24–29, 2008.
- [11] R. Gang, S. H. Shivaswamy, S. Roessner, A. Rao, D. Headlam, and M. Bocko, "Convention Paper 8688 Audio Latency Masking in Music," in *Proceedings of the 133rd Audio Engineering Society Convention*, San Francisco, CA, 2012.
- [12] C. Chafe, "Distributed internet reverberation for audio collaboration," *24th AES International Conference on Multichannel Audio, The New Reality*, pp. 1–6, 2003. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=12302>
- [13] N. Schuett, "The effects of latency on ensemble performance," Master's Thesis, Stanford University, 2002.
- [14] E. Chew, R. Zimmermann, A. A. Sawchuk, C. Kyriakakis, C. Papadopoulos, A. R. J. François, G. Kim, A. Rizzo, and A. Volk, "Musical Interaction at a Distance: Distributed Immersive Performance," in *Proceedings of the MusicNetwork Fourth Open Workshop on Integration of Music in Multimedia Applications*, Barcelona, Spain, 2004.
- [15] H. Moller, C. Jensen, D. Hammershoi, and M. Sorensen, "Design Criteria for Headphones," *Journal of the Audio Engineering Society*, vol. 43, no. 4, pp. 218–232, 1995.
- [16] B. B. Boren, M. Geronazzo, F. Brinkmann, and E. Choueiri, "Coloration Metrics for Headphone Equalization," in *The 21st International Conference on Auditory Display (ICAD-2015)*, Graz, Austria, 2015.
- [17] S. Marks, J. E. Estevez, and A. M. Connor, "Towards the Holodeck: Fully Immersive Virtual Reality Visualisation of Scientific and Engineering Data," *IVCNZ '14 Proceedings of the 29th International Conference on Image and Vision Computing New Zealand*, pp. 42–47, 2014.
- [18] D. T. Bradley and L. M. Wang, "The effects of simple coupled volume geometry on the objective and subjective results from nonexponential decay," *The Journal of the Acoustical Society of America*, vol. 118, no. 3, p. 1480, 2005.

- [19] M. Ermann and M. Johnson, "Exposure and materiality of the secondary room and its impact on the impulse response of coupled-volume concert halls," *Journal of Sound and Vibration*, vol. 284, no. 3-5, pp. 915–931, 2005.
- [20] D. T. Bradley and L. M. Wang, "Quantifying the Double Slope Effect in Coupled Volume Room Systems," *Journal of Building Acoustics*, vol. 16, no. 2, pp. 105–123, 2009.
- [21] C. F. Eyring, "Reverberation Time Measurements in Coupled Rooms," *Journal of the Acoustical Society of America*, vol. 3, pp. 181–206, 1931.
- [22] J. S. Anderson and M. Bratos-Anderson, "Acoustic coupling effects in St Paul's Cathedral, London," *Journal of Sound and Vibration*, vol. 236, no. 2, pp. 209–225, 2000.
- [23] H. Nélisse and J. Nicolas, "Characterization of a diffuse field in a reverberant room," *The Journal of the Acoustical Society of America*, vol. 101, no. 6, pp. 3517–3524, 1997. [Online]. Available: <http://asa.scitation.org/doi/10.1121/1.418313>
- [24] S. G. Norcross, M. Bouchard, and G. A. Soulodre, "Inverse Filtering Design Using a Minimal-Phase Target Function from Regularization," in *Proceedings of the 121st Audio Engineering Society Convention*, San Francisco, CA, 2006.
- [25] F. E. Toole, *Sound Reproduction*. Oxford, UK: Focal Press, 2008.
- [26] R. Rowe and N. Rolnick, "The Technophobe and the Madman: An Internet-2 Distributed Musical," in *Proc. of the International Computer Music Conference (ICMC)*, Miami, FL, 2004.
- [27] P. A. Naylor and N. D. Gaubitch, "Speech dereverberation," in *Proceedings of the 2005 International Workshop on Acoustic Echo and Noise Control (IWAENC 2005)*, Eindhoven, Netherlands, 2005.
- [28] T. Nakatani, M. Miyoshi, and K. Kinoshita, "One Microphone Blind Dereverberation Based on Quasi-periodicity of Speech Signals," *Advances in Neural Information Processing Systems 16*, pp. 1417–1424, 2004.
- [29] H. Kameoka, T. Nakatani, and T. Yoshioka, "ROBUST SPEECH DEREVERBERATION BASED ON NON-NEGATIVITY AND SPARSE NATURE OF SPEECH SPECTROGRAMS," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2009*, Taipei, Taiwan, 2009, pp. 45–48.
- [30] M. Miyoshi and Y. Kaneda, "Inverse Filtering of Room Acoustics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 145–152, 1988.
- [31] M. Triki and D. T. M. Slock, "BLIND DEREVERBERATION OF A SINGLE SOURCE BASED ON MULTICHANNEL LINEAR PREDICTION," in *Proceedings of the 2005 International Workshop on Acoustic Echo and Noise Control (IWAENC 2005)*, Eindhoven, Netherlands, 2005, pp. 173–176.
- [32] S. V. Amengual, M. Kob, and T. Lokki, "Investigations on stage acoustic preferences of solo trumpet players using virtual acoustics," in *Proceedings of the 14th Sound and Music Computing Conference (SMC)*, Espoo, Finland, 2017, pp. 167–174.
- [33] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.
- [34] J. G. Tylka, B. B. Boren, and E. Y. Choueiri, "A Generalized Method for Fractional-Octave Smoothing of Transfer Functions that Preserves Log-Frequency Symmetry," *Journal of the Audio Engineering Society*, vol. 65, no. 3, pp. 239–245, 2017.
- [35] Fischinger Timo and K. Frieler, "Influence of Virtual Room Acoustics on Choir Singing," *Psychomusicology: Music, Mind, and Brain*, vol. 25, no. 3, pp. 208–218, 2015.
- [36] Z. Schärer Kalkandjiev and S. Weinzierl, "The Influence of Room Acoustics on Solo Music Performance: An Experimental Study," *Psychomusicology: Music, Mind, and Brain*, vol. 99, no. 3, pp. 433–441, 2015.
- [37] B. B. Boren, M. Longair, and R. Orlowski, "Acoustic Simulation of Renaissance Venetian Churches," *Acoustics in Practice*, vol. 1, no. 2, pp. 17–28, 2013.
- [38] K. Kato, K. Ueno, and K. Kawai, "Effect of Room Acoustics on Musicians' Performance. Part II: Audio Analysis of the Variations in Performed Sound Signals," *Acta Acustica United With Acustica*, vol. 101, pp. 743–759, 2015.

SONIFICATION AND AMBIENT DISPLAYS

HUMANISING DATA THROUGH SOUND: RES EXTENSAE AND A USER-CENTRIC APPROACH TO DATA SONIFICATION

Sara Lenzi

Density Design Lab, Design Department,
Politecnico di Milano
Via G. Candiani, 72
2058 Milano, Italy
sara.lenzi@polimi.it

Francesca Gleria

Servizio Supporto alla Direzione Generale e ICT,
Provincia Autonoma di Trento
Piazza Dante, 15
38122, Trento, Italy
francesca.gleria@provincia.tn.it

ABSTRACT

In this paper, starting from a case study (the mixed-media data sonification installation *Res Extensae*), we discuss a number of assumptions on the efficacy of sound as a means to represent and communicate numerical data. The discussion is supported by the results of a questionnaire aimed at validating our assumptions and conducted with fifteen of the participants to the experience. At the same time, we have the ambition to contribute to a wider debate on the value of data sonification. We introduce the first stage of a research on sonification as a design-driven, user-centred and multi-modal experience, in that closer to data design practices rather than to traditional composition and computer music. We describe the usage of physical objects to help users to put sounds and data into a wider context, improving the user experience and facilitating the comprehension and retention of the meaning of data.

1. INTRODUCTION

Res Extensae is a data sonification art installation premiered in Trento, Italy, in September 2017, and sponsored by the European Union (EU) program INTERREG Europe OSIRIS. The goal of the project was that of raising awareness, through art, on the (sometimes unexpected) possibilities of re-usage of open data from the public administration, in the framework of the EU vision of a Digital Single Market and uniformed Open Data policy within the Union.¹

The ambition of the project matched well the author's belief, sustained, as we will see shortly, by a current debate among the experts in the emerging necessity of humanising the approach to data design in order to counterbalance data overload in the untrained, non-expert user. In the case here presented, the use of sonification as the main means of representing the chosen data sets aims to investigate the role of the auditory sensory dimension in engaging with structured, quantitative information at the emotional level. In an effort of experimenting with different modalities to measure the efficacy of sound both alone and combined with other modes of interaction, *Res Extensae* was designed as a multi-modal installation where sound elements are combined with found

objects and displayed in the exhibition space as a full rounded, physical *in situ* experience.

This is part of a long-term investigation on the potentialities and possibilities of a user-centred, multi-modal framework for data design.

1.1. Communicating Public Data in the European Union: Opening Silos

Starting from the year 2011, The European Union Commission has issued a series of Communications (a policy document with no mandatory authority) defining the 2020 strategy for a sustainable growth path inside the Union, focussing (among other goals) on the untapped potential of the re-use of public data. Public data are defined by the Commission as *all the information that public bodies in the European Union produce, collect or pay for* [1]. Behind the public advocacy of data re-use lies a wider and stronger need of fostering data culture inside the public sector. In an age where public administrations in Europe are facing several challenges [2], from structural and organisational changes to budget reductions and a decreasing political legitimacy, the EU policy makers openly advocate for the dissemination and re-use of public data with the final goals of increasing its economic value, foster innovation, improve transparency, and engage citizens. Digital data are, in fact, a rather unique type of raw material, one that is not consumed by usage. Quite the opposite, the more data are used, inside and outside the borders of their original domain, the more they become valuable [3].

While the main economic effort of the EU and its Members Countries is focussed on helping local public administration in the transition to the Digital Single Market [4] (harmonising data collection, storage and accessibility) the accent of the public discourse [5] goes on the need to break "data silos" crossing borders inside and outside the public administration. Public administrations are encouraged to learn to speak different languages to increase data value awareness among the public. In a user-centric vision of the data life cycle, there is a strong need to engage citizens in order to give back to them the data they contributed to create².

1.2. Humanising Data

In recent years, as an attitude that can be defined "data culture" is progressively emerging, the practice of translating large sets of discreet (mostly numerical) information into a

¹ Samples of the project can be listened to online at <https://soundcloud.com/user-467084838/sets/res-extensae-a-data>

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

² Paraphrasing <http://uciad.info/ub/2012/02/uciad-the-second-phase/index.html> (Last access 10th May 2017)

more efficient means of understanding has reached momentum. Data are, though, much more than an objective series of numerical values. Confronted with a data set, in the effort to make sense of it we will gather a wider picture which includes broader associations between the data and other relevant background knowledge we may have including “any previous knowledge and expectations” which will inevitably condition the way we connect to the data. [6]

Data Design practices are today at the centre of a cross—disciplinary public debate and emerging activism, involving mostly designers, artists, journalists, data philanthropists. In December 2016, Google News Lab launched an extensive data visualisation project to “tell stories with Google Data” [7]. With the words of Simon Rogers, Google News Lab data editor, the goal of the project is to “communicate ideas” [8] giving back to users their own data within a context, in a statement that echoes the European Union vision of a circular data economy with citizens at the centre.

The data journalist Jacob Harris shared a similar perspective in 2015 [9] when he openly advocated for an “empathic design” in data representation, able to engage the readers at the human level. Inside academia, in 2007 Cawthon and Vande Moere run an experiment to investigate the relationship between aesthetic beauty and usability in data visualisation. They explicitly quoted Donald Norman [10] in theorising that the affection developed by users towards an aesthetically beautiful data visualisation encourages them to think creatively in order to solve any problem or in general to explore more efficiently and effectively the information presented [11].

Given the huge quantity of data now out in the public debate, mainstream data design disciplines (such as data visualisation) are seeking new ways of representing data to engage the public at a more visceral [10] (or “human”) level in an effort of raising awareness on the importance of data for the current and future society. Recent examples of this trend include “data cuisine” [12], “data pottery” [13] and “data sculpture” [14]. Among this emerging trans-disciplinary practice, data visualisation experts are timidly starting to recognise the potentialities of sound and music as an effective way of communicating emotional connotations of information, through qualities such as tempo, volume, or timbre [15].

1.3. Sonification in a multi-modal context

Traditionally, Data Sonification has been understood as a tool for the representation of scientific data through computer generated sounds [16]. Despite being an academic subject of research since the 1990s, “While visual rendering (visualisation) has become an everyday affair, the representation of data or information by sound remains much less common.” [16] Sonification is still, disappointingly, a relatively isolated field let aside the occasional pseudo-sensational article on “eerie sound of Saturn”³.

A number of hypothesis can be made why sound remains an under-represented, under-used means of conveying information, while there seem to be no theoretical aversion in the design community to that sound “when used properly, is both emotionally satisfying and informationally rich” [10]. Factors such as absence of a friendly audio technology both

during the usage of sonification [17] and during the presentation of results to the peer community [18]; prevalence of visual language in the Western cultural landscape over the centuries [16]; arbitrary mapping of data to sound parameters without a “natural” connection between data and sound [19]; and lack of education in interpreting auditory stimuli among scientists [17], are among the most quoted.

More in general, sonification has been traditionally presented to the public with a lack of a real life, physical context that the acousmatic experience (as Pierre Schaeffer defines it, an experience based on “Sounds one hears without seeing their originating cause” [20]) usually entails. This aspect, together with the lack of education in integrating abstract, non-speech sounds into our everyday experience (something that happens on a daily basis with non-physical, digital visual elements [21]) might be another reason why sonification has not managed to become a relevant tool in information communication. The lack of focus on the listening experience (in favour of a focus on the sonification process), underestimates the “wider sonic space” [6] that nevertheless the recipient of the sonification (be her a scientist, a researcher, a visitor of an art exhibition) will intuitively refer to, to decode the relationship between sonic parameters and data. We propose to turn outside the world of sound programming and computer music, in search of other fields where sound is successfully used as embodied in a multi-dimensional, cross-modal context and widely accepted as a valuable, sometimes even necessary element of the user experience.

In the history of sonification, the Geiger-counter is possibly the most quoted [22] and the most successful example of using sound to monitor and alert users, and the one stuck into the collective memory. But more recently multi-modal design research has shown a renewed interest in leveraging sound specificities, drawn mainly from psychology and psychoacoustics, cognitive sciences and neurosciences [15], to design more efficient user experiences in order to increase the information bandwidth and avoid information overload in Virtual Reality (VR) highly complex environments [17]. Maybe surprisingly, in these contexts not only functionality and efficiency but also the aesthetic of sound appears as an increasingly important characteristic to make auditory feedbacks more relevant and easier to interpret [37]. VR training and research environments in scientific disciplines could highly benefit of positional audio and dynamic soundscapes synthesis common to video games to create an immersive environment. Immersive soundscapes could be used by a user to constantly monitor the overall status of a system, while freeing the visual channel for other, simultaneous precise tasks [17].

Cinematography is another field where sound is used, quite fruitfully, to efficiently convey relevant (sometimes fundamental) information on the development of the narrative in the context of a multi-modal experience. Despite cinema not being, undeniably, a data design process, it is nonetheless a field where sound is widely accepted as an extremely successful and necessary means of communicating information, and therefore we believe it deserves our attention in the context of the present work.

1.4. Three converging perspectives

The project that we will be describing in detail in the coming sections is, to sum it up, the resultant of three converging perspectives on data:

³ <https://www.sciencealert.com/nasa-has-recorded-the-first-sounds-of-the-eerie-void-inside-saturn-s-rings>

- We need to engage citizens on the importance of public data and the intrinsic value of the re-use of data produced by the public administration.
- There is a growing quest for a design approach able to foster a more “human” relationship with data representation for a better user experience.
- Sound is widely recognised in a number of disciplines, ranging from art to interface and product design through branding, as an efficient means of communicating information and emotion at the same time. Sonification based on acousmatic sounds is still sometimes met with a skeptical mind by the scientific community [23]. We will be looking at overcoming this aspect integrating sound into a full rounded experience in order to test the potentiality of a multi-modal, contextualised, embodied approach to data sonification.

2. RES EXTENSAE: A MULTI-MODAL SONIFICATION EXPERIENCE

INTERREG Europe OSIRIS is a European Union project dedicated to Open Social Innovation policies driven by co-creative Regional Innovation ecosystems [24]. Started in 2016, OSIRIS was born in the frame of the EU policies and activities aimed at facing increasing social challenges in the Union, as outlined in the introduction to this paper. Seven public bodies from seven different regions in Europe (Vasterbotten, Sweden ; Ostrobothnia, Finland; Provincia Autonoma di Trento, Italy; Drenthe Province, Netherlands; Computer Institute and Press Diophantus, Greece; Abo Akademi University, Finland; Fundao Municipality, Portugal; Regional Development of Presov Region, Slovakia) share good practices to foster innovation in the field of employment, education, social inclusion and climate/energy with the objective of implementing cross-borders smart solutions and disseminating improvements.

The goal of the program is to increase the efficacy of local policies for a better use of the EU structural funds. In particular, the OSIRIS project aims to do that within the framework of a co-creative approach sharing best practices among the partners. In September 2017, an OSIRIS meeting was held in the partner region of Provincia Autonoma di Trento, an autonomous region in the north of Italy. The three days of activities, titled “Data Driven Innovation and Territorial Growth” focussed on the discussion about data (mostly open local governments data) presenting ways to enable data driven economies and open social innovation and facilitating networking between public administrations as data provider and private companies (as potential data buyers).

2.1. Background

In the context of the meeting, OSIRIS commissioned to xxx an artistic work with the objective of raising awareness on the re-usability of open data sets from the public administration among its member regions and their related citizen networks. The project, which took the shape of a multi-modal, mixed-media installation, was experienced by the public in the outdoor courtyard of Palazzo Roccabruna, in the Italian City of Trento. This section will describe the decisions taken by the author during the conceptualisation of the work, later called “Res Extensae”. The title is an homage to the philosopher of the XVI century Rene Descartes and his well-known investigation of the relationship between body and mind (*res extensa* and *res cogitans*, respectively). We are today facing a

new dualism, between an extended body of digital data and our human-scaled minds. On one side, quantitative information is continuously collected that tracks and describes our everyday life and the world around us. On the other side, this extended body is coupled by our limited cognitive capacities that, at the other extreme of the Descartes dualism, both collectively and intimately, try to grasp the extension and attribute meaning to these large collections of numbers. The artistic experience is seen, in this case, as a means to overcome our limits and make numerical data closer to the human scale.

2.2. Choice of Data Sets

The data sets were freely chosen by the author from the online open repository of the Autonomous Province of Trento, the Trentino Open Data web portal. Several considerations were made in order to choose, mostly related to the unique characteristics of the location where the installation was to be performed, in the historical centre of the city of Trento. The Trentino region is, in fact, cut in the middle by the second most important river of Italy, the River Adige. The River shaped the historical town in her early days, before its course being deviated in the nineteenth century by the Austrian Empire’s engineers.

Nowadays, the river is constantly monitored with a system of sensors which collect, every fifteen minutes, inputs on the river height (hydrometer), flow rate, rain rate (hygrometer) and temperature. The measurement is taken at each station along the course of the river in the Trentino territory, and released for open use from the portal in the form of a spreadsheet.

Figure 1: The River Adige sensor network as it appears on the open data web portal.

For the purpose of the installation, the author chose the centric station of the San Lorenzo bridge (in Trento City centre) and two independent data sets: flow rate and hydrometer. The sets covered the measurements of five days, at a rate of one measurement every fifteen minutes. The reasons for limiting the data to five days of history are to be found in the sound design choices and are explained in the following paragraphs.

2.3. Sonification as a User – Centred Design Process

Taking the risk of being over-reductive and for the sole purpose of this paper, we identified two main approaches to data sonification, both inside the academia and in the art practice world. From one side, we have what we will call a generative approach: sounds are synthesized algorithmically by purposely designed software; the translation from numerical data to sound is based on pre-determined parameters which associate certain values of the data set to certain elements of the sound material (amplitude, frequency, timbre, morphology). The objective of such projects is the

scientific translation of quantitative information into an auditory experience, be it for analytical exploration or synthetic representation of data. The resulting soundtrack is usually streamed either offline or online and listened to via a headphone set, with notable exceptions (for instance, the Sonification Festival where projects were performed live in Berlin in late 2017). The second approach, though still generative by nature, has the objective of obtaining musical compositions from data sets, often performed with physical instruments in a performative setting. The process involved in this second case will be that of composing with the support of a traditional, written score to be handed out to the performer (for instance, the composition “The sound of climate change from the Amazon to the Arctic” by D. Crawford and S. St George for string quartet).

We propose here a third approach that we call design-driven and user-centred [25]. The accent of the sonification process is thus on the design of a sound experience that has at the core the user herself and the context in which she will listen to the sounds. We believe that sound is always heard, or listened to, in combination with other sensory modalities [26] and their respective inputs. Far from being a limitation, this is an opportunity to use sound to design meaningful, integrated and rich user experiences. In a traditional perspective that sees academia reflecting mainly on utility and functionality as main criteria to evaluate data sonification, the design perspective focuses on (paraphrasing G. Judelman) an aesthetic, though rigorous approach to the representation of numerical information [27]. We define here “aesthetic” not as a romanticised, subjective sense of beauty but as a design rationale conceptually analogous to that of traditional product design, architecture, or art [28].

2.3.1. Choice of sound material

In “SonifEye: Sonification of Visual Information using Physical Modeling Sound Synthesis” [23], Roodaki and colleagues purposely design the sounds elements to resemble sounds generated by natural phenomena, thus leading to an intuitive, therefore less tiring and stressful, processing of the auditory information. This attribute of sound is used to avoid alarm fatigue (a well-documented problem where professionals exposed to too many monitoring systems end up ignoring them in order to avoid perceptual stress, with clear decreased safety consequences [29]) and allow the visual channel to focus on other tasks, specifically high precision medical tasks conducted in a Virtual Reality training environment (such is the case in the SonifEye project). The aesthetic component of real time physical modelling sonification plays an important role leveraging sounds attribute that limit the cognitive effort needed to understand the underlying meaning of abstract sound material (and the information it carries) correctly.

In his “Audio-Vision: Sound at Cinema” [30], Michel Chion illustrates, through multiple examples derived from the history of the moving image, how sound is able to radically modify our perception of a movie narrative influencing the sense of time and space (synch point, spatial magnetization), overcoming the physical barriers of the screen (off-screen sound), and adding to the materiality of the image contributing to conveying tactile information such as speed and weight (materializing sound indices).

Cognitive sciences have added, in recent years, several elements that can be taken into account while designing sounds in a multi-modal setting. Prof. Charles Spence and his Cross -

Modal Interaction Lab at Oxford University have been investigating how the feeling of taste, colours, and sizes is influenced or can be stimulated by specific characteristics of auditory inputs such as pitch, rhythm, timbre. For example, people seem to consistently match high-pitched sounds with small, bright objects located high up in space [31] or low-pitched sounds with bitter taste [32]. Cross modal correspondences involving the auditory channel and other sensory modalities have by now been proven by a large body of research [31] and can be integrated in a design-driven approach to sonification.

All these elements were taken into account for the design of the sound material of the sonification process, in particular for the design of texture and morphology of sounds. Paraphrasing Michel Chion [30] we call this approach embodied sonification, i.e., an approach that focuses on the relationship between sounds and the listening experience, mediated by elements borrowed from other sensory modalities, like sight (objects, images, moving images); smell (food); touch (objects, food); taste (food). In this perspective, the design of each sound element is the result of a research on texture, timbre, and in general the matter of sound in order to suggest or amplify such a relationship.

We cannot not mention the work of Stephen Roddy on Embodied Sonification [33] which utterly resonates with the approach we have taken for Res Extensae. We too, with Roddy, believe that “Embodied cognition is defined by how it answers the age-old mind-body problem, the question of how the human mind relates to physical matter” and that this approach can help to bridge the “Cartesian divide” offering a “truly cohesive account of auditory meaning-making” [6].

An investigation of the cognitive and philosophical framework of embodied cognition exceeds the scope of this paper but it sits at the background of every decision we took owing much to the work of the Chilean philosopher Francisco Varela and his “The embodied mind” [34].

2.4. Building the context. Found objects and narrative

The quest for a data humanism [35], an approach to data design that goes beyond a merely technical relationship with the data representation and towards the story behind data itself, calls for posing the accent of data representation on the larger context. In the current perspective of sonification, where, as already noted, acousmatic sounds are mainly predominant, physical objects can play the role of mediators [36], reinforcing the narrative and embodying the acousmatic sounds into a *res extensa*. The subjective, often intimate relationship of an individual with the materiality and physicality of objects is part of what “authenticity can be defined by - the subjective experience through which the physicality of an object can elicit experiential, meaningful, and affective responses” [21].

In the Found Object Project, an extensive study on our relationship to objects we happen to find by various means by Paul M. Camic, personal attachment to everyday object is described as the resultant of a cognitive process involving “the interaction of aesthetic, cognitive, emotive, mnemonic, ecological, and creative factors” [37]. As Donald Norman puts it, though, our attachment is not to the thing, it is to the relationship, to the meaning and feelings the thing represents [10].

The use of acousmatic sounds deprives the user from such a meaningful relationship. We argue that the lack of the

physical element negatively affects the appreciation of the context and the interpretation of the auditory element as part of a narrative, making it more difficult for the listener to engage with data in a meaningful way. For Res Extensae, we decided to include in the listening experience a few found objects (five, each representing one day of data), with the goal of evaluating the impact of the physical element on the overall user experience and on the comprehension of the context and its narrative.

2.4.1. Reinforcing the narrative: objects as sonic boards

As we already mentioned, the way sound is broadcasted and therefore experienced by the final user is sometimes quoted as a disincentive to the affirmation of a sonification field [16]. For this reason, and to reinforce the relationship with the found objects as a background canvas for locating sounds in a context, we decided to use audio transducers instead of traditional speakers. Transducers were placed inside (or under) each object so that the sound material would come out of the object itself, inviting the users to get closer to it and facilitate what Csikszentmihalyi calls a “flow experience” [38].

3. WORK PROCESS

3.1. Sonification process

The initial sound material was synthesised using proprietary software by Native Instruments® (NI). Having to deal with two different data sets (flow rate and data from hydrometer at a given sensors station) we identified two different sounds connected to data by two different rules, and we approached these two aspects differently in terms of work process. Being, as we will see, the whole process manual, we were obliged to select a human scale set of data, limiting it to five days in the life of the river, with one data input every fifteen minutes, for each set (flow rate and hydrometer). It is obviously a fundamental limitation in the sonification process and we will get back to this issue in the conclusions of this paper.

3.1.1. Hydrometer

For the data set relative to the hydrometer, which measures the river height at a given point, identification of the mapping rules came first. In particular, we chose one of the most common sonification association [19], i.e., the representation of height values as pitch values, but we applied this criterion loosely giving priority to aesthetic rules. In other words, we twisted the concept of a direct translation of height numerical values into frequency numerical values, in order to obtain an increased variety of texture and richness of the sound samples.

The need for an increased variety emerged while analysing the data set. Undeniably, measurements taken within fifteen minutes are, under normal circumstances, most likely to be very similar, giving as a result a monotonous sound experience. We therefore applied a mixed parametrisation, splitting the numerical value in two parts and using the resulting parts to program two distinct synthesizers. The sound

sample associated with each data was a resultant of this combination, applied to a customised patch of the software Absynth® 5 by NI. For example, the starting height of 102 cm on Day 1, time 00:15 was translated into a sound combining an Oscillator (A) at a frequency of 10,000 Hz and an Oscillator (C) at a frequency of 2,000 Hz. A third Oscillator (B) was kept by default at the frequency of - 2.000 Hz. The choice of timbre took into consideration the intrinsic melodic nature of a soundtrack resulting from the association between a numerical value and a frequency.

In the context of a design approach, we proceeded manually recording sound by sound (one sound for each single numerical data), before moving to the editing software (Steinberg Nuendo® 5) for montage and post-production. This approach conceptually followed the idea of using sonification to convey analytical information to the listener.

3.1.2. Flow Rate

For the flow rate, we experimented with the sound texture trying to keep a “natural” connection with the data itself [19] with the intention of triggering an intuitive apprehension of the context of origin of the data represented [17]. In this sense, this second approach conceptually followed the idea of using sonification to convey synthetic information (a bird’s eye view) on the overall status of the river [23]. The initial sound material was synthesised using NI Reaktor® 6 looking for a texture that could aurally mimic the feeling of an underwater current. The mapping of this complex sound element to the single elements of the data set was done manually, loosely associating each numerical value to the position of the cursor on the sample while recording the whole process in real time. The idea being that the abrupt changes in the sample formant provoked by this process would have been intuitively perceived as changes in the river current’s energy. The entire data set was performed, manually scrolling the cursor along the timeline. The performance was recorded in real time and subsequently imported into Steinberg Nuendo® 5 for post-production.

Figure 2: Real time performance and recording process of the flow rate data set.

3.2. Choice of objects

Each object was hand-picked in the area of Trento, and chosen among those objects of the local tradition which used to or still do contain water, or liquids. The idea behind, quite simply,

was connecting data, sounds and objects within the same narrative, focussing around the same topic, the river. We expected this process to enhance the memorability of the experience and to facilitate the intuitive appreciation of the data sets through the embodiment of sounds in the physical dimension of the objects.

Figure 3: One of the found objects with sound transducer used for the project.

3.3. The installation

The exhibition space consisted in an outdoor area of circa fifty square meters. The objects were placed on supports (one support per object) and located in the space for guests to encounter them on their path. As mentioned, each object also functioned as the sound source thanks to the audio transducers. Each object was labeled with the data set it played and the day it referred to, in a way that reminded that of natural science museums' display of items, thus encouraging the public to get near the object for close observation (in this case, for close listening). The event took place at night in a low-light setting. Each object had its own LED source of light to amplify, once again, the focus on the sound source.

Figure 4: One of the objects and sound stations during the installation.

4. TESTING WITH USERS

Six months after the event, a ten questions survey was emailed to thirty selected participants. The selection was made based on several factors including country of residence (in order to have the highest possible variety of cultures) and possibility of reach (there was not an official list of participants except those belonging to the delegations of the OSIRIS project or personally invited by the author). Of the 30 surveys sent out, 15 came back for inclusion in this paper. The response rate was therefore of 50%. The questionnaire was easily accessible via [SurveyMonkey.com](https://www.surveymonkey.com). The following paragraphs highlight the most relevant aspects of questions and answers, for further discussion. Question 10 is excluded from this report because of its vague formulation that led to unclear/unusable results; Question 9 asked to quote three keywords describing the work: the result, though very interesting for the author, is not deemed to be relevant for the present discussion.

4.1. Efficacy of sonification

The first question addressed the main objective of *Res Extensae* and asked to participants to express with a Yes/No answer whether they felt sound could be an effective means of communicating the value and the potentialities of public administration data sets. 86.67% of the users answered positively. This group subsequently rated 62% (Question 2) the efficacy of sound as a means of communicating the value of data, on a scale from 1 (lowest value) to 100 (highest value).

Figure 5: Question 1, results.

4.2. Influence of the objects on the experience

More relevantly for the present work, the association between sounds and physical objects during the experience was clearly perceived as a relevant added value to the sound experience by the totality of the respondents.

Figure 6: Question 4, results.

The meaning of “added value” was subsequently defined as: suggesting a context for the data sets (60%), creating an emotional connection with the sound (53.55%), building a narrative, attracting attention, and enhancing memorability (46.67%); amplifying the sound experience and creating an emotional connection with the data set (26.67%).

Figure 7: Question 4, results.

Interestingly, the capability of sound to engage the listeners seems to lose more than half of its power when sonification is experienced without the object. Asked to re-listen to the sonification through headphones, without the physical presence of the object, the engagement was given a rate of 48 on a scale of 100, where 100 was the engagement of the sonification with the object.

4.3. Sound as a humanizing factor in data representation

Asked whether, after the experience, they would consider sound as a humanising factor in the representation of quantitative data set, i.e., able to make numerical data less abstract and closer to our everyday experience, the majority of users answered positively (71.43%).

Figure 8: Question 7, results.

The “humanising” power of sound was given an overall value of 59 on a scale of 1 (lowest value) to 100 (highest value).

4.4. Targets of sonification

Question 3 asked the participants to select target groups which they thought could most benefit from the use of sound to represent data. Media was the most relevant group according to the participants (66.67%), followed by researchers and scientists (60%), students (53.33%) and citizens (46.67%). “Other” (26.67%) included a suggested target group defined by its relationship with the arts (suggestions included art institution, cultural institutions, artists).

Figure 9: Question 3, results.

4.5. Discussion

For most, if not all, of the participants the event described in this work was the first encounter with the field of data sonification. It is therefore particularly encouraging to note that the majority of the participants considered sound as a valuable means to convey the meaning of the data set. “Valuable” is here intentionally used as a broad term meant to indicate a positive feeling towards the importance of communicating public administration data to a wider public which was, as explained earlier in this work, one of the main interests of the project’s stakeholders.

Quite surprisingly, given the artistic nature of the work, the respondents indicated “Researchers and Scientists” as one of the categories which would most benefit, in their opinion, of the usage of sound to represent data.

There is no doubt that the presence of physical objects facilitated the overall experience conveying, as we hoped, an increased sense of context to the data sets and creating an emotional connection with it. Enhancing memorability, building a narrative and attracting attention are other aspects of embodied sonification which seem to be promising. The objects seem to function as catalysts for an increased understanding and improved relationship with the sounds representing the data sets. It seems reasonable to hypothesize that physical objects are able to balance the acousmatic relationship grounding it in a physical context.

5. CONCLUSIONS AND FUTURE DEVELOPMENTS

This paper presents a seminal work in a design-driven, user-centred approach to data sonification. In particular, we introduce the usage of physical objects in association with sounds, where sound represents the actual data sets and objects help the users to put sounds and data into a wider

context, improving the user experience and facilitating the comprehension and retention of the meaning of data. This multi-modal approach to the design of data experiences seems, despite the small number of respondents to the testing phase, promising. We also believe that this approach, if validated, could be of interest for the development of Virtual Reality environments for data representation with an embodied approach to sonification.

Further development will investigate the role of sound material which is “naturally” connected to the meaning of the data they represent. Experiments will be carried out with different sound matter including speech and soundscape, and leveraging the potentialities of physical synthesis.

5.1. Limits of the present approach

The manual approach to data sonification that this project has presented, justified by the author’s interest in design practices, presents an evident limitation in the amount of sounds that can be created and therefore the amount of data that can be analysed and represented. Further developments will investigate how to automatize the approach, including in real-time sonification.

6. ACKNOWLEDGMENT

Sara Lenzi wishes to thank the INTERREG Europe OSIRIS program, Trentino Open Data and the Provincia Autonoma di Trento, Servizio Supporto alla Direzione Generale e ICT (Cristiana Pretto) for taking the risk to sponsor an-out-of-the-box project; the Trento Chamber of Commerce for hosting the installation in the beautiful space of Palazzo Roccabruna; Mr Bruno Lorengo of the Ufficio Dighe - Servizio Prevenzione Rischio for his help in organising the data sets on a human scale; Prof. Paolo Ceccarelli from Density Design Lab, Politecnico di Milano (Italy) and Prof. Lonce Wyse, National University of Singapore, for their precious feedback on this paper.

7. REFERENCES

- [1] Communication from the Commission of the European Parliament, “Open data. An engine for innovation, growth and transparent governance”, *COM (2011) 882 Final*, Brussels, Belgium, 2011.
- [2] É. Bassot and W. Hiller, “Key Policy Challenges for the EU in 2017 EPRS”, *European Parliamentary Research Service*, 2016.
- [3] W. Carrara, W. San Chan, S. Fischer and E. van Steenberg, *Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources*, European Union, 2015.
- [4] Communication from the Commission of the European Parliament, “A Digital Single Market for Europe *COM(2015) 192 Final*, Brussels, Belgium, 2015.
- [5] <https://www.dati.gov.it/content/disponibile-registrazione-webinar-ruolo-caratteristiche-portale-europeo-dei-dati> (Last access 15th Feb 2018).
- [6] S. Roddy and D. Furlong, “Sonification listening: an empirical embodied approach”, in *Proc of the 21st International Conference on Auditory Display (ICAD)*, pp. 181-187, Graz, Austria, 2015.
- [7] Data Visualisation Project, *Google Trends*, available at https://trends.google.com/trends/story/US_cu_6fXtAFIB_AABWdM_en (Last access 10th May 2017).
- [8] S. Rogers, *Data needs empathy to make it real*, Google News Lab, 2017. Available at <https://medium.com/google-news-lab/data-needs-empathy-to-make-it-real-c9ca0f0aea6d> (Last access 10th May 2017).
- [9] J. Harris, *Connecting with the dots*, 2015. [online] <https://source.opennews.org/articles/connecting-dots/> (Last access: 10th May 2017).
- [10] D. Norman, *Emotional Design: Why We Love (or Hate) Everyday Things*, New York, U.S.: Basic Books, 2005.
- [11] N. Cawthon and A. Vande Moere, “The Effect of Aesthetic on the Usability of Data Visualisation”, in *Proc of the 11th International Conference on Information Visualisation*, Zurich, Switzerland, 2007.
- [12] <http://data-cuisine.net/>
- [13] <http://alicethudt.de/>
- [14] <https://www.adriensegal.com/>
- [15] N. Rönnerberg, J. Löwgrén, “The Sound Challenge to Visualization Design Research”, in *Proc of EmoVis, ACM IUI Workshop on Emotion and Visualization*, Sonoma, CA, USA, 2016.
- [16] C. Dayé and A. de Campo, “Sounds sequential: sonification in the social sciences”, *Interdisciplinary Science Reviews*, Vol. 31, N. 4, 2006.
- [17] B. Rau, F. Friess, M. Krone, C. Müller and T. Ertl, “Enhancing visualization of molecular simulations using sonification”, in *Proc of 1st International Workshop on Virtual and Augmented Reality for Molecular Science (VARMS@IEEEVR)*, 2015.
- [18] A. de Campo, C. Dayé, C. Frauenberger, K. Vogt, A. Wallisch and G. Eckel, “Sonification as an interdisciplinary work in progress”, in *Proceedings of the 12th International Conference on Auditory Display*, London, UK, 2006.
- [19] C. Scaletti, A.B. Craig, “Using sound to extract meaning from complex data”, in *SPIE Vol. 1459 Proc of Electronic Imaging '91*, San Jose, California, U.S., 1991.
- [20] P. Schaffer, *Traité des objets musicaux*, Paris, France: Le Seuil, 1966.
- [21] Varnalis-Weigle, Anastasia S. “A Comparative Study of User Experience between Physical Objects and their Digital Surrogates”, *Journal of Contemporary Archival Studies*, Vol. 3, Article 3, 2016.
- [22] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. H. Flowers *Sonification Report: Status of the Field and Research Agenda*, 2010.
- [23] H. Roodaki, N. Navab, A. Eslami, C. Stapleton, and N. Navab, “SonifEye: Sonification of Visual Information using Physical Modelling Sound Synthesis”, *IEEE Transactions on Viz and Computer Graphics*, vol. 23 No. 11, Nov 2017.
- [24] <https://www.interregeurope.eu/osiris/>
- [25] D. A. Norman, S. W. Draper *User Centered System Design: New Perspectives on Human-Computer Interaction*, Boca Raton, Florida, U.S.: CRC Press, 1986.
- [26] P.M. Lindborg, *Sound perception and design in multimodal environments*, Doctoral Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, Dec 2015.
- [27] G. Judelman, “Aesthetic and Inspiration for Visualization Design: Bridging the Gap between Art and Science”, in *Proc the Eighth International Conference on Information Visualisation (IV'04)*, London, UK, 2004.

- [28] A. Calvera, “Nulla ethica sine aesthetica: should aestheticism still be stigmatized?”, *Utopia. The Radical Designer*, Dec 2016.
- [29] S. Sendelbach and M. Funk, “Alarm fatigue: a patient safety concern”, in *AACN Advanced Critical Care*, 24(4):378-86, Oct 2013.
- [30] M. Chion, *Audiovision: Sound on Screen*, New York, U.S.: Columbia University Press, 1994.
- [31] C. Spence, “Crossmodal correspondences: A tutorial review”, *Attention Perception Psychophysics* 73:971–995, May 2011.
- [32] A. S. Crisinel, C. Spence, “As bitter as a trombone: synesthetic correspondences in non-synesthetes between tastes/flavors and musical notes.”, *Attention Perception Psychophysics* 72(7):1994-2002, Oct 2010.
- [33] S. Roddy, *Embodied Sonification*, Doctor of Philosophy Thesis, University of Dublin, Ireland: Trinity College, 2015.
- [34] F. J. Varela, E. Thompson and E. Rosch, *The Embodied Mind. Cognitive Science and Human Experience*, Cambridge, Massachusetts, U.S.: The MIT Press, 1993.
- [35] G. Lupi, “We reached peak infographics. Are you ready for what comes next?”, 2017. [online] Available at: <http://www.printmag.com/information-design/data-humanism-future-of-data-visualization/> (Last access: 10th May 2017).
- [36] S. Sato, M. Takahashi and M. Matsuo, “Giving Context to Sounds through Mediation of Physical Objects”, in *Proc of UbiComp*, Zurich, Switzerland, Sept 2013.
- [37] P. M. Camic, “From Trashed to Treasured: A Grounded Theory Analysis of the Found Object”, *Psychology of Aesthetics, Creativity, and the Arts*, Vol. 4, No. 2, pp. 81–92, Oct 2010.
- [38] M. Csikszentmihalyi *Flow: The psychology of optimal experience*, New York, U.S.: Harper & Row, 1990.

PLAUSIBLE AUDITORY AUGMENTATION OF PHYSICAL INTERACTION

Marian Weger¹, Thomas Hermann², Robert Höldrich¹

¹ IEM, University of Music and Performing Arts, Graz, Austria

² Ambient Intelligence Group, CITEC, Bielefeld University, Bielefeld, Germany
weger@iem.at

ABSTRACT

Interactions with physical objects usually evoke sounds, i.e., auditory feedback that depends on the interacting objects (e.g., table, hand, or pencil) and interaction type (e.g., tapping or scratching). The continuous real-time adaptation of sound during interaction enables the manipulation/refinement of perceived characteristics (size, material) of physical objects. Furthermore, when controlled by unrelated external data, the resulting ambient sonifications can keep users aware of changing data. This article introduces the concept of *plausibility* to the topic of auditory augmentations of physical interactions, aiming at providing an experimentation platform for investigating surface-based physical interactions, understanding relevant acoustic cues, redefining these via auditory augmentation / blended sonification and particularly to empirically measure the plausibility limits of such auditory augmentations. Besides conceptual contributions along the trade-off between plausibility and usability, a practical experimentation system is introduced, together with a very first qualitative pilot study.

1. INTRODUCTION

As we live in a physical world, we interact with other physical objects: the floor we are walking on, objects we take in our hands, or a table where such objects may be positioned. Any of such physical interactions usually evoke specific sounds, depending on the physical properties of the involved objects, such as material, shape, or size, and also depending on the type of interaction, i.e., excitation (scratching, hitting, etc.). Physical objects in this context refer to objects with spatial location [1]. The resulting sounds being a reaction to a preceding (physical) action are referred to as auditory feedback, representing a major part of our everyday acoustic environment which we use either consciously or unconsciously while pursuing our daily activities.

According to activity theory, these activities can be separated into a number of goal-oriented conscious actions, which in turn consist of chains of unconscious operations [2, 3]. One possible action using auditory feedback may be to shake a box in order to get information on its contents from the resulting sound. Similarly, our operations are influenced by auditory feedback, for example, through footstep sounds which reveal important information on the material of the ground and thus subconsciously influence our walking style [4, 5, 6, 7].

In this context, we define *augmented auditory feedback* as the artificially modified sonic reaction to physical interaction. This type of augmentation could be exploited in various forms: First, the signal-to-noise-ratio may be improved in order to help communication of (a) the involved objects' physical properties such as material or spatial dimensions, and (b) the interaction type such as tapping or scratching. Both may facilitate specific activities. Second, specific physical properties of a physical object could be modified perceptually, in order to induce a certain change in user behavior. Apart from the aforementioned footstep sounds, successful behavioral change through sound has been shown for hand tapping [8, 9], as well as by for grasping [10, 11]. Finally, augmented auditory feedback creates a new communication channel that can also be used for sonification of data that are completely unrelated to both object and interaction, e.g., for continuous monitoring as a secondary task. The sonification is thus naturally and seamlessly fitted into the everyday acoustic environment.

In the context of this article, we explore the exploitability of augmented auditory feedback as communication channel. In information theory, a major characteristic of a given channel is its channel capacity, i.e., the highest rate at which information can be reliably transmitted. In our case, it is limited by at least two factors: the *plausibility* and the *usability* of augmented auditory feedback. Both will be discussed in detail in Sec. 2.

For a specific physical interaction, we make four assumptions:

- (1) There exists a manifold of sounds which serve as *plausible auditory feedback*. Its borders define the *plausibility range*.
- (2) There exists a manifold of sounds which serve as *usable auditory feedback*, i.e., the sounds are helpful to perform the specific actions. Its borders define the *usability range*.
- (3) The manifolds of plausible and usable sounds overlap. We define this overlap region as the manifold of *alternative auditory feedbacks*.
- (4) It is possible to discriminate between different alternative auditory feedbacks.

If all these assumptions are met (compare Fig. 1), we conclude that it is possible to convey additional information through (more or less subtle) sound changes within the intersection of plausible and usable auditory feedbacks.

Note that usability refers to the original physical interaction alone. Adding additional information can make an interaction more useful for the users and eventually let users even actively interact to query information that they regard as useful. This aspect of usability (or the additionally received information) is not meant here.

The above argumentation leads us directly to the already established concept of *auditory augmentation* [12]. In addition, we

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

Figure 1: Sets of plausible and usable variants of auditory feedback for a specific physical interaction. Points *a* and *b* represent two discriminable but still plausible and usable sounds.

contribute a more systematic design approach to the no less important original definition. According to [12], auditory augmentation is “a paradigm to vary the objects’ sonic characteristics such that their original sonic response appears as augmented by an artificial sound that encodes information about external data. [...] All this manipulation does not affect the sound’s original purpose.”

In this article, we test the method of auditory augmentation through the concept of plausible and usable auditory feedback and provide an exemplary implementation as proof of concept. What follows in Sec. 2 is an investigation of the above assumptions concerning plausible and usable auditory feedback. After a review of related sonification literature in Sec. 3, we present an experimental platform for an auditory augmentation system in Sec. 4, together with a pilot study with naive users.

2. PLAUSIBLE AND USABLE AUGMENTED AUDITORY FEEDBACK

In this section, we test the four assumptions formulated in the introduction in order to find rules how to meet them.

2.1. Plausible auditory feedback

*“I’m not concerned with plausibility;
that’s the easiest part of it, so why bother?”*
— Alfred Hitchcock [13, p. 99]

Sensory feedback is generally considered to be plausible if it is “conceptually consistent with what is known to have occurred in the past” [14]. In particular, “a highly plausible scenario is one that fits prior knowledge well: with many different sources of corroboration, without complexity of explanation, and with minimal conjecture” [14].

With increasing perceptual dominance of auditory augmentation, the resulting auditory feedback influences user perception, emotion, and behavior [9]. For example, auditory cues have been shown to influence the haptic perception of virtual textures presented by the Phantom Omni haptic device [15]. Likewise, perception of material properties (e.g., hard/soft, rough/smooth) has been shown to be strongly influenced by auditory cues [16]. As the perceptual plausibility depends on the congruency between different modalities such as haptic, visual, or auditory information, it has no meaning for the unimodal case of auditory feedback alone. Perceptual congruency and therefore plausibility is high if the information of the different modalities combined, i.e., the combination of different stimuli, matches the pattern we learned through natural interactions with our physical environment. It is therefore hypothesized that perceptual plausibility increases with increasing congruency (agreement) between cues (information) from different sensory modalities (information channels).

It must be considered that people are already accustomed to manipulated visual feedback, but generally have less experience (or none at all) in augmented auditory or tactile feedback: It is

common knowledge that a physical object’s inner structure can be concealed, e.g., through painted surfaces. This supports the assumption that there exist at least several interchangeably plausible visual representations of a physical object.

We argue that if an auditory augmentation alters the perception of only such physical properties that are hidden behind the surface finishing (e.g., lacquer or laminate), i.e., physical properties that are not conveyed through vision or haptics, then the auditory augmentation cannot lead to incongruent sensory information. The augmented physical properties, however, still need to remain physically viable in order to reduce the complexity of explanation. These physical properties concerning the inner structure of physical objects may include material category, density, hollowness, and spatial volume, as well as boundary condition (e.g., free or clamped) and coupling to other physical objects.

Individual exploitability of the mentioned physical properties for sonification purposes can be derived from a short literature review on auditory perception:

Material identification. Listening experiments showed that material identification of impacted plates is limited to gross density categories (steel/glass vs. wood/plexiglass), while in-category identification is strongly influenced by interpretations based on the everyday acoustic environment [17]. In particular, glass and wood were associated with smaller objects than metal and plastic.

Volume of cavities. It has been shown that volumes of cavities can be directly perceived through sound [18].

Hollowness. Participants of a listening experiment could auditorily distinguish between hollow and solid bars; however, they tended to base their judgments on simple differences in frequency, which accompany size variations of a resonant cavity [19].

Size. As a general rule when judging the size of a sound-producing object, people tend to rely on simple sound attributes such as loudness, duration, and base frequency [20, 21, 22].

Boundary condition. Concerning stiffness variations at different boundary conditions of vibrating plates, it has been shown that listeners used both pitch and loudness for discrimination [23].

Concerning other physical properties, the interested reader is referred to further literature on auditory perception of physical objects, as reviewed in [24, pp. 1–16] or [25, pp. 31–38].

Summing up, all of the above material properties have a perceivable effect on auditory feedback. In consequence, their modulation through auditory augmentation is supposed to provide a reliable communication channel. The resulting augmented auditory feedback is assumed to stay within the plausibility range if augmented in a physically meaningful and feasible way, i.e., if the illusory physical properties can be explained without effort.

Note that augmented auditory feedback that is far from physical feasibility, but rather represents a cartoonification of the original auditory feedback, might be perceived as even more plausible or realistic than the original sound [26]. However, as we seek for general applications, the more restricted approach of physically informed auditory augmentation is followed here.

2.2. Usable auditory feedback

According to ISO 9241-210 [27], usability is defined as the “extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”.

Auditory feedback generally has a positive effect on task performance as well as on motor learning [28]. Sigrist et al. summa-

alized the potential and challenges of augmented feedback in general in the context of motor learning, and provided design criteria to successful visual, auditory, haptic, and multisensory feedback [28]. However, a positive effect of auditory feedback observed in isolation may completely vanish in the presence of feedback in other sensory modalities. For example, auditory feedback has been shown to improve the typing performance on a computer keyboard, if no additional haptic feedback was present [29], concluding that in this case, haptic feedback had stronger influence than auditory feedback.

In any case, what makes auditory feedback usable is the information it carries about the performed action and the physical objects it affects. All information which is encoded in the auditory feedback but not needed for a specific task, physical object, and context, is therefore assumed to be irrelevant for usability. We define *relevant physical properties* as the properties of a physical object which influence its usability. Likewise, *irrelevant physical properties* are the properties of a physical object which do not affect its usability.

Irrelevant physical properties are therefore considered as possible candidates for usability-independent auditory augmentation. Note that auditory augmentation of irrelevant physical properties may mask the perception of relevant physical properties – either auditorily, or through informational masking effects, e.g., by adding additional disturbance or stress. Furthermore, the relevance of specific physical properties of one physical object diverge for different actions and context.

As an example, we examine a relatively simple category of physical objects: an even, horizontal, rigid, and stationary surface. Such a surface usually appears on tables, cupboards, bookshelves, etc. Due to its affordances [30, pp. 10–13], it is primarily used for putting things on top, moving these things around, and manipulating them (writing, cooking, etc.), but also manual interaction is possible (hitting, scratching, tapping, etc.).

For the observed surface, we consider three relevant physical properties: Hardness influences how we put fragile things on top, or how we interact with our hands. Roughness influences how objects or fingers can be moved. Sturdiness is relevant, as a fragile table might break while positioning heavy things. We thereby do not want to modulate these but rather leave them unchanged in order to preserve usability.

Similarly, physical properties that we consider as irrelevant include spatial volume, hollowness, underlying material category, and boundary conditions – if not in conflict with relevant physical properties. These relate mainly to the non-visible part of the surface, under the visible texture layer. In consequence, their perception may be securely altered through auditory augmentation within the plausibility range.

Note that not all physical objects incorporate a specified purpose or intended use. However, every physical object has affordances. The goal of usable auditory augmentation is to preserve these in the best possible way while adding new affordances such as exploratory data analysis through manual interaction.

3. RELATED SONIFICATION WORK

With Wetterreim [12], Bovermann et al. already presented a proof-of-concept for the use of auditory augmentation in a peripheral monitoring situation where realtime weather data modulated the auditory feedback of a computer keyboard. Their general implementation processes an audio signal from a contact microphone

attached to a physical object by a digital filter whose parameters are controlled by external data. The modified signal is rendered by loudspeakers and blends with the object's original sound, providing an additional layer between a user's action and a physical object's auditory feedback.

The concept of auditory augmentation is strongly connected to *blended sonifications* which “blend into the users' environment without confronting users with any explicitly perceived technology” and thus provide ambient sonification channels [31]. This general framework contains auditory augmentation as an important special case (where user-caused audio signals affect the auditory display via data-dependent filtering). Yet more generally, in Blended Sonification, physical environment, digital environment, and user can each provide both audio and data components which affect the auditory display via filtering and/or additive auditory elements, bound together by the shared claim that the formed auditory gestalt is still perceived as coherent auditory events.

Another interesting yet more remotely relevant approach for the topic of this article was presented by Barrass [32], who embedded interactive sonifications into found objects, making these serve as a physical medium for exploration of the otherwise imperceptible and abstract data.

Interactive control of in-home ambient sonification systems through small and unobtrusive sensors has been discussed by [33]. The presented prototypes were able to convey information through sound, triggered by simple everyday actions.

Specifically connected to the affordances of surfaces is the sonification of pen strokes presented in [34]. Müller et al. used a simple physical model of a cylindrical resonator (pen) excited by a mainly stochastic process which describes the surface texture. They discriminate between three auditory surface textures: The micro surface texture consists of fine-grained single peaks of stochastic distribution which cannot be identified individually. The meso surface texture is rougher, resulting in isolated perceptible peaks. The macro surface texture relates to visible patterns, similar to a chessboard, and describes larger-scale and regular structures. The authors argue that “a maximum of naturalism of the sound feedback can be mainly achieved by providing a feedback to the interaction with low latency and high precision.” [34, p. 3].

4. THE MONDRIAN-TABLE

In order to explore the potential as well as the constraints of plausible and usable auditory augmentation, we developed an experimentation platform which adheres to the horizontal, even, rigid, and stationary surface, as exemplified in Sec. 2.2. For even further simplification, we restrict it to rectangular shape. The experimentation platform is intended to offer two different modes of operation. On the one hand, it should enable calm and unobtrusive blended sonification outside the focus of attention while performing daily activities that affect it (e.g., writing or positioning of other physical objects). On the other hand, it should serve as an interface for exploratory data analysis through manual interaction in the form of tapping, scratching, etc.

At the same time, irrelevant of the performed action, this experimentation platform is designed to offer plausible and usable augmented auditory feedback. As derived in Sec. 2.2, possible usability-independent physical properties include spatial volume (thickness of the plate), material category (wood, metal, etc.), and boundary conditions (free, clamped, simply supported, etc.). All these can be referred to the inner structure of the physical object

Figure 2: Concept of the Mondrian Table, inspired by Piet Mondrian’s “Composition II in Red, Blue, and Yellow” (1930).

and thus may be augmented in a plausible way, as described in Sec. 2.1.

Even on an ordinary dining table, the tabletop can be composed of several elements of different materiality. Such structures can be simulated through a space-dependent auditory augmentation. Our experimentation platform might therefore be perfectly suitable for sonification of a time-invariant configuration of geometric structures, displayed as regions with different perceived physical properties. This relates to a macro surface texture [34]. The underlying data can be a map, a technical drawing, or a painting similar to Piet Mondrian’s “Composition” series. We therefore call this experimentation platform the Mondrian Table (see Fig.2).

The Mondrian Table ambiently sonifies geometric structures as plausible and usable changes of the surface material. By placing a sheet of paper on top of it, those structures can be traced back with a pencil and thus transferred into the visual domain.

In contrast to the original approach of auditory augmentation, we insert a physical sound synthesis model between data and sonification, in the hope to achieve naturally plausible auditory feedback. The data does not directly produce sound, but is mapped to properties of the intermediate physical model (e.g., material category or shape). This model serves as a filter, altering the original sounds emerging from the interaction between user and physical object. Note that this is *not* Model-Based Sonification (MBS) [35, p.399–427], but merely using physical models for perceptually more plausible intertwined coherences. However, the rendering pipeline makes our prototype as well perfectly suited for future interactions with MBS.

Tünnermann et al. [31] developed a standard for visualizing the audio- and dataflow of blended sonifications in a simple way. Such a blended sonification diagram usually involves the three factors User, Physical Environment, and Digital Environment as main sources of Data (D) and/or Audio (A). Connections of these sources to the Auditory Display as a sink visualize causal contributions to either a Filtered (F) or an Added (A) output. While filtered sonifications are “sonifications that stay very close to the original sound”, superimposed “sound samples or synthesized sound fall in the ‘added’ category” [31, p. 4].

The blended sonification diagram of the Mondrian Table (Fig.3) shows that our system produces a filtered sonification which is dependent on audio from the user in order to sonify data from the digital environment. The auditory display additionally depends on properties of the interaction (spatial location) as well as on properties of the physical object (original physical properties). We interpret the physical environment as something that is out of reach and not influenced by the user, and thus attribute properties of both interaction and involved physical object as data of the user.

Figure 3: Blended sonification diagram of the Mondrian Table.

4.1. Development of an experimentation platform

The technical implementation of our experimentation platform is similar to the augmented keyboard described in [12], replacing the keyboard by a rectangular plate with location tracking system. For the first prototype (see Fig. 4), we decided for a stylus-based graphic tablet which is basically a combination of both plate and tracking system. An advantage of the graphic tablet (compared to resistive sensitive surfaces) is, that the pen coordinates are already tracked while the pencil hovers contact-free over the surface, so that for instance filter parameters can be adjusted before any physical interaction happens. This is less critical for continuous interactions such as painting or scratching, but more so for tapping where the filter should have been parameterized *before* the pencil-paper interaction delivers the source signal for filtering. We use a Wacom Intuos 5 touch M which offers an active area of 224 by 140 mm – approximately the size of US Half Letter or A5 format, just enough for drawing simple sketches or writing small texts.

As we know that spatial congruency between sensory modalities has an impact on plausibility [36], sound playback needs to be somehow spatialized in a meaningful way. Horizontally, a landscape oriented sheet of size A5 within reach for writing comfortably covers an angle of approximately 21° . This is well below the usual 60° for simple stereophonic panning strategies with two loudspeakers. Considering a perceptual localization blur of $\pm 3.6^\circ$ for white noise-bursts in the horizontal plane [37, p. 41], a stereophonic setup is supposed to be sufficient for the given task. Due to the small size of the sonification interface, vertical spatialization is regarded as irrelevant, taking into account the stronger perceptual localization blur in vertical angle ($\pm 4^\circ$ for white noise, [37, p. 44]) as well as in distance perception ($\pm 25\%$ at 1.1 m for impulses, [37, p. 47]). Contact microphones and loudspeakers are therefore placed vertically centered on the left and right side of the graphic tablet to provide natural panning in the case that both signal paths (left and right) are processed individually.

The latency of augmented auditory feedback needs to lie below a certain threshold in order to be successfully and plausibly combined with visual or haptic information, as also emphasized by Müller et al. [34] who referred to naturalism in this context. Typical maximum thresholds for synchronous perception are at 25 ms for auditory-tactile delay [38] and 18 to 24 ms [39, 40] for auditory-visual delay. However, for trained users such as musicians, a latency below 10 ms is recommended [38].

As hardware platform, we therefore decided for the Beagle-Bone Black Rev. C with Bela audio cape, which is designed for sub-millisecond-latency audio and sensor processing [41] and targeting specifically for digital musical instruments [42]. Two low-cost piezo-electric contact microphones are installed on the left and right side underneath the graphic tablet’s active area. These

Figure 4: The Mondrian-Graphic-Tablet. The positions of the contact microphones underneath are marked through a red ‘M’.

are connected to the Bela cape through a FET buffer preamp (see [43] for schematics), driven by a 9 V battery. Two miniature loudspeakers are connected directly to the on-board class-D amplifiers of the Bela, and are placed besides the tablet.

For sound synthesis, in contrast to the sonification of pen strokes described in [34], we directly use the input signal from contact microphones as excitation signal for filter-based modal synthesis as described, e.g., in [44, pp. 46–48]. The physical model is intended to synthesize the normal modes (transversal waves) of a two-dimensional plate and to ‘apply’ these to the physical surface. The plate is seen as a linear and time-invariant system; its impulse response can be decomposed into exponentially decaying pure sine-waves [45, p. 12]:

$$x(t) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} A e^{-\alpha t} \sin(2\pi f_d t + \phi), \quad (1)$$

with mode indices m and n referring to the number of nodal lines in the two dimensions, respectively. Each mode is described by four factors: a frequency f_d , an attenuation rate α for the exponential decay, an amplitude A , and a starting phase angle ϕ , all dependent on mode indices m and n . In our simplified model, we assume amplitude and phase primarily influenced by the excitation signal and thus decided to ignore these in the first prototype.

As an example, we analyze an undamped, isotropic, and rectangular thin plate which is simply supported (hinged) along the edges. Its natural frequencies f_0 are computed as follows [45, p. 80]:

$$f_0 = 0.453 h \sqrt{\frac{E}{\rho(1-\nu^2)}} \left[\left(\frac{m+1}{l_x} \right)^2 + \left(\frac{n+1}{l_y} \right)^2 \right]. \quad (2)$$

In Eq. 2, the geometry of the plate is described by width l_x , height l_y , and thickness h . The material of the plate is expressed by Young’s modulus E , Poisson ratio ν , and density ρ .

When damping is considered, the actual resonant frequencies f_d , i.e., the frequencies of the exponentially decaying sinusoids in Eq. 1, are a bit lower than the undamped natural frequencies f_0 ,

Figure 5: Example image, generated with the Mondrian Generator [48], serving as test data.

depending on the individual damping ratio ξ [45, p. 11]:

$$f_d = f_0 \sqrt{1 - \xi^2}, \quad \text{with} \quad \xi = \frac{\alpha}{2\pi f_0}. \quad (3)$$

The damping of a plate depends on several factors including geometry, material, and boundary condition, and cannot be easily predicted. We therefore employ a simple but perceptually-tuned damping model that is parametrized by global damping α_G and frequency-relative damping α_R [46]:

$$\alpha = e^{\alpha_G + \alpha_R \cdot 2\pi f_0}. \quad (4)$$

The resonant frequencies further depend on the boundary conditions of the plate edges. We describe the four boundary conditions through B_{x0}^i and B_{x1}^i for the left and right edge, and B_{y0}^i and B_{y1}^i for the top and bottom edge, respectively, where 0 means free, 1 means simply supported, and 2 means clamped. Solutions for all combinations of these are derived in [47].

In our simplified physical model, we neglect the resonant behavior of the graphic tablet and thus interpret the input signal from a contact microphone as a pure excitation signal induced by the stylus or pencil. This excitation signal is then filtered through our physical model to form the augmented output signal. Technically, this model is represented through a parallel filter-bank of resonant bandpass-filters which are tuned to the frequencies and attenuation rates of the simulated plate’s normal modes. The implementation is done in SuperCollider (SC), based on the DynKlank object which adds any number of resonances to an input signal, taking vectors of undamped natural frequencies f_0 , amplitudes A (set constant to 1), and 60 dB decay times $t_{60} = \ln(1000)\tau$ (in seconds, with time-constant $\tau = 1/\alpha$) as arguments. The input signal is additionally cleaned from unwanted low-frequency noise through a high-pass filter set below the lowest mode frequency; the output signal is soft-limited.

Only the audio signal path is processed on the Bela running an SC server, while the mapping between data, physical model, and filter parameters is done on a separate computer running the SC language. Both computer and Bela communicate via network. The graphic tablet, connected via USB to the computer, is accessed as HID device.

In order to simplify the process of defining regions of different physical properties on the tablet already during development, a so-called Mondrian Generator [48] was used. This simple program generates compositions in red, blue, and yellow, “in the style of Piet Mondrian”. The result is a raster graphics; an example is shown in Fig. 5.

Each of the three colors red, blue, and yellow represents a reference model M^i with specific physical properties, with i being

Table 1: Physical properties of the reference models.

	M^1	M^2	M^3
material category	metal	ceramics	glass
E (GPa)	180	360	70
ρ (kg m ⁻³)	7740	3800	2600
ν	0.305	0.22	0.22
α_G	0.6	1.55	2.0
α_R ($\times 10^{-4}$)	2	1.75	1.5

the index of the model. Each model M^i is represented by its parameter vector:

$$\mathbf{p}^i = [B_{x0}^i, B_{x1}^i, B_{y0}^i, B_{y1}^i, E^i, \rho^i, \nu^i, \alpha_G^i, \alpha_R^i, l_x^i, l_y^i, h^i].$$

The three reference models are rendered in parallel with pre-set parameter vectors $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$, respectively. Together they form the set of possible auditory augmentations. The color directly controls the input gains g_{in}^1 , g_{in}^2 , and g_{in}^3 of the three models, respectively (red: $g_{in}^1=1$, $g_{in}^2=g_{in}^3=0$; blue: $g_{in}^2=1$, $g_{in}^1=g_{in}^3=0$, yellow: $g_{in}^3=1$, $g_{in}^1=g_{in}^2=0$). White leads to a level-preserving mix of the three models. Black means that all input gains are set zero; black lines between the different regions are thus interpreted as grooves filled with sound-absorbing material.

4.2. Informal evaluation

For evaluation of the prototype system, we defined three reference models with plate dimensions l_x and l_y set to match the active area of the graphic tablet. All three models were set to a constant thickness h of 3 mm, with the boundary conditions for all edges set free, i.e., simulating plates that were freely hovering in the air. The material properties, as listed in Tab. 1, were set to match the material categories metal, ceramics, and glass, respectively.

Due to limited computing power on the BeagleBone, only few resonances can be processed. However, the number of vibrational modes in the relevant audible range, i.e., below 16 kHz, is anyway small for the synthesized plates, especially due to the small spatial dimensions. In particular, these were 13 modes for metal (starting at 1.5 kHz), 6 modes for ceramics (starting at 3.0 kHz), and 13 modes for glass (starting at 1.6 kHz). For each model, only the 6 lowest out of 100 calculated modes have been used.

For informal evaluation, an unknown random image was created by the Mondrian Generator and sonified through the Mondrian Table by using the pre-set reference models. The graphic tablet was covered by a blank sheet of paper. The given task was to trace back the sonified structures with the stylus, and to draw a visual representation of the auditorily perceived structure. For this task, the stylus was paired with a real pen in order to be able to paint contours on the paper while exploring borders of a region of equal material properties. Such a sketch can be seen in Fig. 4.

An example video, recorded from first-person perspective with Soundman OKM binaural microphones, can be found in [49] (listening with headphones is recommended). It illustrates some tapping and scratching interactions with the pen at various locations of different data-driven (resp. location-dependent) augmentations.

The evaluation of the experimental platform provided us with valuable information on further development. In accordance with related studies, latency seems to be a major factor for plausible and successful auditory augmentation. During the evaluation, a round-trip latency of 25 ms was disturbing, while 14 ms was sufficient for the illusion of realistic auditory feedback when watching another

Figure 6: The Auditory Coloring Book.

person interacting (auditory-visual condition), but still felt unnatural for the interacting person (auditory-visual-haptic condition). With measured round-trip latency of about 1.5 ms, obtained with the Bela, the augmentation felt completely plausible and blended well with the direct sound of the original auditory feedback, despite spatially incongruent playback through loudspeakers. The synthesized materials felt realistic and successfully created the impression of a different materiality of the augmented physical surface. The sonified geometric shapes could be easily detected without effort while interacting in a completely natural way.

4.3. The Auditory Coloring Book

In order to get an impression on how unbiased and naive users might interact with such an augmented surface, we installed a modified version of the above-described system at an open house event of the university: The Auditory Coloring Book (see Fig. 6) is an interactive sound installation which auditorily displays regions of different color through augmented surface material of a real table. In contrast to the first prototype, this installation uses a real table equipped with two AKG C411 contact microphones underneath the surface. Two miniature loudspeakers are placed on top. Tracking of the pen is realized in Processing through the Microsoft Kinect v2 sensor, covering an A2-sized sheet of paper, fixated on the table.

During the event, volunteers were asked to trace back the regions of different materiality through natural interaction either with fingers or a ball-pen, both producing sufficient auditory feedback to drive the auditory augmentation. The ball-pen additionally provides an intuitive way for switching between “exploration” and “drawing” mode through (de)activation of the tip.

Due to the lower resolution of the Kinect-based tracking compared to the graphic tablet, the fine-grained “Mondrian-style” structures were regarded as too difficult to detect in the envisaged context. Instead, users were asked to detect three different shapes, a rectangle, a triangle, and a circle, and to draw them on the paper sheet. The same materials as in the informal evaluation were mapped to the three shapes: rectangle to metal, triangle to ceramics, and circle to glass. The spatial arrangement of the shapes randomly differed between participants. The shapes could overlap, i.e., partially mask others. An example-image with the correct proportions was shown to the participants. Most of the time, the room was crowded with people, inducing a relatively noisy and sometimes disturbing environment.

20 random visitors were documented through their individual drawings. Four of these drawings are shown in Fig. 7, with the auditorily displayed arrangement drawn as an overlay. An analysis of

Figure 7: Drawings from participants 3 (top left), 5 (top right), 8 (bottom left), and 19 (bottom right), overlain by the corresponding “correct answer”, respectively.

these drawings through visual judgment of the investigator reveals that 55 % of the participants correctly identified the approximate location of all three regions. 40 % additionally assigned all correct shapes, while 60 % roughly hit the correct size. 95 % confidence intervals, assuming binomial distribution, are 32–77 %, 19–64 %, and 36–81 %, respectively.

Results indicate that the task was very hard to accomplish in the given context. However, the experiment clearly reveals 4 different strategies for task completion (compare Fig. 7): (1) continuous oscillatory drawing with enabled pen, somehow tracing back the border regions (top left), (2) tracing back the borders through systematic tapping, marking positions of different sound and connecting these (top right), (3) random tapping interaction to quickly find extreme values for efficient identification of the shapes (bottom left), and (4) continuous scanning of the surface and marking material changes (bottom right).

Yet interestingly, strategy (3) was only used by children and led to quick and reliable results for shape-identification; however, always underestimating the size. The other strategies concentrated on exact border positions, and took much longer (up to about 15 min), but gained similar hit-rate for shape-identification. The relatively low overall performance surely comes from low resolution and instability of the tracking system.

5. CONCLUSIONS AND OUTLOOK

We have introduced the research question of plausibility of auditory augmentations of physical interactions, with the particular interest in shedding light on the available bandwidth for sonification using auditory augmentation, understood as the number of discriminable alternative auditory feedbacks, so that the sound is still plausible and useful for the underlying physical interaction. This is basic research, and thus requires as a first step to clarify concepts, and to develop a research method to investigate the emerging questions.

With the developed experimentation platform, we have succeeded in creating settings that allow both a high-fidelity control of discrete and continuous physical interactions, and at the same

time to manipulate the auditory feedback from subtle to gross. This platform is intended to form the basis for conducting empirical studies in order to investigate limits of plausibility of auditory augmentations and how blended sonifications affect usability of the original auditory feedback.

The informal evaluation helped to develop the experimental platform as well as to design further experiments with it. The qualitative user study gave an insight on behavior and strategies of naive users of the system, which is of great importance for future designs. Especially time-varying auditory augmentations are supposed to work well for real-time sonification; however, presumingly at higher risk of inducing implausibility.

A series of follow-up studies will focus on the discovery of just-noticeable differences in auditory augmentation. A marker-based motion-capture system will be used in order to enable the auditory augmentation of arbitrary flat surfaces such as existing tables or plates of different material with high-precision tracking.

6. ACKNOWLEDGMENT

This research was supported by the Cluster of Excellence Cognitive Interaction Technology ‘CITEC’ (EXC 277) at Bielefeld University, funded by the German Research Foundation (DFG).

7. REFERENCES

- [1] N. Markosian, “What are physical objects?” *Philosophical and Phenomenological Research*, pp. 375–395, 2000.
- [2] V. Kaptelinin, “Context and consciousness: Activity theory and human-computer interaction,” B. A. Nardi, Ed. Cambridge, MA: MIT Press, 1995, ch. Activity Theory: Implications for Human-computer Interaction, pp. 53–59.
- [3] K. Kuutti, “Context and consciousness,” B. A. Nardi, Ed. Cambridge, MA, USA: Massachusetts Institute of Technology, 1995, ch. Activity Theory As a Potential Framework for Human-computer Interaction Research, pp. 17–44.
- [4] R. Bresin, A. de Witt, S. Papetti, M. Civolani, and F. Fontana, “Expressive sonification of footsteps sounds,” in *ISon workshop*. KTH, 2010, pp. 51–54.
- [5] J. Maculewicz, C. Erkut, and S. Serafin, “An investigation on the influence of soundscapes and footstep sounds in affecting preferred walking pace,” in *ICAD*, 2015.
- [6] S. Papetti and F. Fontana, “Effects of audio-tactile floor augmentation on perception and action during walking: Preliminary results,” in *SMC*, 2012, pp. 17–22.
- [7] L. Turchet and R. Bresin, “Effects of interactive sonification on emotionally expressive walking styles,” *IEEE Trans. on Affective Computing*, vol. 6, no. 2, pp. 152–164, 2015.
- [8] E. Furfaro, F. Bevilacqua, A. Tajadura Jimenez, *et al.*, “Sonification of surface tapping: Influences on behaviour, emotion and surface perception.” *ISon Workshop*, 2013.
- [9] E. Furfaro, F. Bevilacqua, N. Berthouze, and A. Tajadura-Jimenez, “Sonification of virtual and real surface tapping: evaluation of behavior changes, surface perception and emotional indices,” *IEEE MultiMedia*, 2015.
- [10] U. Castiello, B. L. Giordano, C. Begliomini, C. Ansuini, and M. Grassi, “When ears drive hands: the influence of contact sound on reaching to grasp,” *PloS one*, vol. 5, no. 8, 2010.

- [11] A. Sedda, S. Monaco, G. Bottini, and M. A. Goodale, "Integration of visual and auditory information for hand actions: preliminary evidence for the contribution of natural sounds to grasping," *Experimental brain research*, vol. 209, no. 3, pp. 365–374, 2011.
- [12] T. Bovermann, R. Tünnermann, and T. Hermann, *Auditory Augmentation*, ser. Premier reference source, K. Curran, Ed. Information Science Reference, 2010.
- [13] F. Truffaut, A. Hitchcock, and H. G. Scott, *Hitchcock*, revised ed. Simon and Schuster, 1985.
- [14] L. Connell and M. T. Keane, "A model of plausibility," *Cognitive Science*, vol. 30, no. 1, pp. 95–120, 2006.
- [15] S. Serafin, H. P. Kjaer, C. Taylor, and F. Avanzini, "Audio-haptic physically based simulation and perception of contact textures," in *ICAD*, 2007.
- [16] R. Martín, J. Iseringhausen, M. Weinmann, and M. B. Hullin, "Multimodal perception of material properties," in *ACM SIGGRAPH on applied perception*, 2015, pp. 33–40.
- [17] B. L. Giordano and S. McAdams, "Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates," *JASA*, vol. 119, no. 2, pp. 1171–1181, 2006.
- [18] D. Rocchesso and L. Ottaviani, "Can one hear the volume of a shape?" in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 115–118.
- [19] R. A. Lutfi, "Auditory detection of hollowness," *JASA*, vol. 110, no. 2, pp. 1010–1019, 2001.
- [20] R. Mason, T. Brookes, and F. Rumsey, "Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli," *JASA*, vol. 117, no. 3, pp. 1337–1350, 2005.
- [21] D. R. Perrott and T. N. Buell, "Judgments of sound volume: Effects of signal duration, level, and interaural characteristics on the perceived extensity of broadband noise," *JASA*, vol. 72, no. 5, pp. 1413–1417, 1982.
- [22] D. Cabrera and S. Tilley, "Parameters for auditory display of height and size," in *ICAD*, 2003.
- [23] C. Marquis-Favre and J. Faure, "Auditory evaluation of sounds radiated from a vibrating plate with various viscoelastic boundary conditions," *Acta Acustica united with Acustica*, vol. 94, no. 3, pp. 419–432, 2008.
- [24] D. Rocchesso and F. Fontana, Eds., *The Sounding Object*. PHASAR, 2003. [Online]. Available: http://www.soundobject.org/SObBook/SObBook_JUL03.pdf
- [25] W. A. Yost, R. R. Fay, and A. N. Popper, *Auditory perception of sound sources*. Springer, 2007, vol. 29.
- [26] R. Wages, S. M. Grünvogel, and B. Grützmacher, "How realistic is realism? considerations on the aesthetics of computer games," in *International Conference on Entertainment Computing*. Springer, 2004, pp. 216–225.
- [27] ISO, "9241-210: 2010. ergonomics of human system interaction-part 210: Human-centred design for interactive systems," 2009.
- [28] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review," *Psychonomic bulletin & review*, vol. 20, no. 1, pp. 21–53, 2013.
- [29] Z. Ma, D. Edge, L. Findlater, and H. Z. Tan, "Haptic keyclick feedback improves typing speed and reduces typing errors on a flat keyboard," in *World Haptics*, 6 2015, pp. 220–227.
- [30] D. Norman, *The design of everyday things*. Basic, 2002.
- [31] R. Tünnermann, J. Hammerschmidt, and T. Hermann, "Blended sonification – sonification for casual information interaction," in *ICAD*, 2013.
- [32] S. Barrass and T. Barrass, "Embedding sonifications in things," in *ICAD*, 2013.
- [33] S. Ferguson, "Sonifying every day: Activating everyday interactions for ambient sonification systems," in *ICAD*, 2013.
- [34] C. Müller-Tomfelde and T. Münch, "Modeling and sonifying pen strokes on surfaces," in *DAFx*, 2001.
- [35] T. Hermann, A. Hunt, and J. G. Neuhoff, *The sonification handbook*. Logos, 2011.
- [36] Y.-H. Yang and S.-L. Yeh, "Unmasking the dichoptic mask by sound: spatial congruency matters," *Experimental brain research*, vol. 232, no. 4, pp. 1109–1116, 2014.
- [37] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, revised ed. MIT Press, 1997.
- [38] M. E. Altinsoy, "The quality of auditory-tactile virtual environments," *JAES*, vol. 60, no. 1/2, pp. 38–46, 2012.
- [39] B. D. Adelstein, D. R. Begault, M. R. Anderson, and E. M. Wenzel, "Sensitivity to haptic-audio asynchrony," in *ACM ICMI*, 2003, pp. 73–76.
- [40] D. J. Levitin, K. MacLean, M. Mathews, L. Chu, E. Jensen, and D. M. Dubois, "The perception of cross-modal simultaneity (or "the greenwich observatory problem" revisited)," in *AIP Conference*, vol. 517, no. 1, 2000, pp. 323–329.
- [41] A. McPherson and V. Zappi, "An environment for submillisecond-latency audio and sensor processing on beaglebone black," in *AES Convention*, 2015.
- [42] V. Zappi and A. McPherson, "Design and use of a hackable digital instrument," in *Proc. Live Interfaces*, 2014.
- [43] J. D. Tillman. (2001, 5) A discrete fet guitar preamp. [Online]. Available: <http://www.till.com/articles/GuitarPreamp/>
- [44] P. R. Cook, *Real sound synthesis for interactive applications*. CRC Press, 2002.
- [45] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. Springer, 1998.
- [46] M. Aramaki, M. Besson, R. Kronland-Martinet, and S. Ystad, "Controlling the perceived material in an impact sound synthesizer," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 301–314, 2011.
- [47] G. Warburton, "The vibration of rectangular plates," *Proceedings of the Institution of Mechanical Engineers*, vol. 168, no. 1, pp. 371–384, 1954.
- [48] J. Fan. Mondrian generator. [Online]. Available: <https://github.com/JEFworks/mondrian-generator>
- [49] M. Weger. (2018, 2) Supplementary material. [Online]. Available: <https://phaidra.kug.ac.at/detail/object/o:69732>

AUGMENTATION OF AN INSTITUTE'S KITCHEN: AN AMBIENT AUDITORY DISPLAY OF ELECTRIC POWER CONSUMPTION

Katharina Groß-Vogt¹, Marian Weger¹, Robert Höldrich¹,
Thomas Hermann², Till Bovermann³, Stefan Reichmann⁴

¹ IEM - Institute of Electronic Music and Acoustics,
University of Music and Performing Arts Graz, Austria,
² Ambient Intelligence Group, CITEC, Bielefeld University, Germany,
³ University of Applied Arts Vienna, Austria,
⁴ Center for Social Research, University of Graz, Austria.
vogt@iem.at

ABSTRACT

Efficient feedback on energy consumption is regarded as one step towards a more sustainable lifestyle. Sonification is very apt to convey such information continuously in an ambient and effective way. This paper presents a pilot system for sonifying the electric power consumption of an institute's kitchen. The reverberation of the kitchen is changed depending on the actual consumption and its difference to a weekly baseline. If the actual consumption is low, it is mapped to a plausible kitchen reverberation. If it is high compared to the baseline, the reverberation becomes unnatural. Evaluating the system gave insights on perceptibility and acceptance of auditory augmentation in a semi-home context.

1. BACKGROUND

Electricity consumption is a key issue in striving towards sustainability. Households in the EU are consuming 29% of final electrical energy, excluding energy used for transport [1]. Household devices are becoming more efficient, but at the same time are increasing in number and functionality. Therefore, the overall in-home electricity use is rising and solutions to a more careful use of electricity are needed. By 2020, smart meters (intelligent, digital electric meters) are predicted to be implemented in approximately 72% of European households [2]. In addition, consumer electronic devices are gaining more and more networking functionalities – by 2020, the Internet of Things (IoT) in the EU-28 is expected to number more than 6 billion devices [3]. This opens up to the possibility to access real-time information for home monitoring.

Numerous studies on smart meters in recent years have established the fact that feedback has measurable effects on behavior [4]. Typical forms of feedback are smart meters with in-home displays or web portals, informative billing, or similar [5]. Some current approaches visualize energy consumption metaphorically as flowing water [6], or even use a real indoor fountain for ambient sonification [7]. Schwartz et al. relate to energy as an imperceptible entity [8]. Therefore, design strategies have to make it

perceptible in our everyday lives in order to allow for an emotional connection. Furthermore, the rationalistic paradigm is criticized in the design of standard systems for energy feedback – people often have difficulties apprehending the information provided by feedback systems because appropriate *contextualization* is missing.

The aim of our ongoing research is to take advantage of auditory display to support energy awareness in an ambient in-home setting. The pilot study presented in this paper is a first step towards a sonification system for in-home electrical power consumption (henceforth abbreviated as 'power'). As relevant factors for efficient feedback, both perceptibility and contextualization are being explored.

2. AUGMENTED AUDIFICATION AND RELATED WORK

The sonification of the pilot system described in this paper is based on changing the properties of the kitchen's acoustics by adding artificial reverberation. A large amount of literature deals with related approaches from environmental sociology, HCI, sonification, and sound and music computing. A short overview will be given in the following paragraphs.

The concept of auditory augmentation has been proposed by Bovermann et al. [9] as "*building blocks supporting the design of data representation tools, which unobtrusively alter the auditory characteristics of structure born sounds*". The concept has been discussed to be extended to the more general blended sonification, which "*describes the process of manipulating physical interaction sounds or environmental sounds in such a way that the resulting sound carries additional information of interest while the formed auditory gestalt is still perceived as coherent auditory event*." In the *knock'knock* blended sonification [10], reverberation of knocking the door was used to communicate whether somebody is present behind the office door. In this approach, the time span after the person has left the room is mapped to reverberation: an empty room sounds more reverberant than a stuffed one. The authors of *knock'knock* address the importance of a calm display. Auditory augmentation and blended sonification only work with active input, and are, therefore, calm in the sense of calm computing.

Kilander and Lönnquist presented the concept of WISP, or Weakly Intrusive Ambient Soundscape [11]. In a ubiquitous

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

(a)

(b)

Figure 1: (a) Kitchen of the Institute of Electronic Music and Acoustics, IEM, Graz. (b) Wall plugs measuring and sending data on electric power consumption. Photos: Till Bovermann.

service environment, individual notifications have been presented with a sound associated to each user. Playback volume and reverberation have been used to convey three levels of intensity of the notification. High intensity was mapped to a dry, loud sound, while low intensity was saturated by reverberation, giving the impression of a far sound. The polarity of the mapping is opposite to our pilot study (see below) and it might be interesting to compare the two approaches to each other in a future study.

A similar project to WISP was realized by Brazil and Fernström [12] who explored a basic system for an ambient auditory display of a group. The presence of colleagues was sonified as soundscape of individual sounds, each time a person came or left the workplace. The proposed system has similarities to our pilot study in taking the work surrounding for a field test, but focusses on auditory icons for creating a soundscape instead of a calm approach.

Parthy et al. [13] also used reverberation for ambient data communication. They conducted a small study where they found that the reverberant decay time (RDT) has to increase by approximately 60% or decrease by approximately 30% in order to be clearly perceived. These findings are in approximate accordance with We-

ber's law which states that perceived change in a stimulus is proportional to its size. Parthy et al.'s study was conducted with music as input signal and with adapted (identical) RMS amplitude for various RDT values. These factors might play a role in the way that the RDT has to be changed so drastically in order to be perceived. We assume in our approach that the sensitivity is higher when (1) the amplitude is used as additional parameter of reverberation and (2) the input signals are more transient than with musical input.

The works cited above reflect on the goals of ambient displays, e.g., *awareness and lightweight interactions* [12], to "convey an intuitive sense of any graspable process" rather than a direct information display [11], or a system that is "not intended to give a precise representation of the data value" [13]. Our pilot system follows the same idea, exploring perceptibility of un/-natural reverberation as a sonification parameter.

3. THE AUGMENTED KITCHEN

This paper presents a pilot study in a surrounding that facilitates both easy maintainance and constitutes a valid field test for an in-home setting: our institute's kitchen. An institute's kitchen is a specific surrounding, neither private nor really public. Furthermore, the users of the kitchen do not pay for the electric power consumption, as would be the case in an in-home setting. Therefore, the purpose of the study can only be to raise *awareness* of power consumption. No large behavioural changes are to be expected.

The kitchen is a 16-square-meters room equipped with a table for 8 people and a kitchen unit including

- (A) a dishwasher,
- (B) a coffee maker,
- (C) a water kettle,
- (D) a microwave oven and
- (E) a fridge.

The *power* of the above mentioned five appliances has been monitored and sonified, as discussed in the next sections. The kitchen unit also comprises an electrical stove, but this has been unplugged permanently for security reasons (as it has never been used and the water kettle is placed on it). Furthermore, electrical light was neglected, i.e. the last permanent appliance in the kitchen that drains power. Pictures of the kitchen and the wall plugs measuring the power are shown in Fig. 1. The kitchen is used regularly by 15 members of staff, all of whom have their workplace in the same building as the kitchen, and by external teachers, students and visitors of the institute. It is also used as a small conference room in case all other rooms are occupied.

Our sonification design is based on virtual acoustics. Acoustic properties of a room can be changed by adding reverberation. In the simplest case this means recording the environmental sound with a microphone, applying filters and playing it back over a loudspeaker in real-time. Care has to be taken to avoid electro-acoustic feedback loops. Such a systems of virtual acoustics is used frequently at our institute to showcase the power of audio to diverse audiences in our main performance and lecture hall, the CUBE [14]. Therefore, all our kitchen users, i.e. the test subjects for our pilot study, are familiar with it. The sonification design will be discussed in Sec. 3.3.

A current field of research deals with the automatic analysis of *power* data in order to infer which appliances in a household are

switched on and how much energy is used by them (so-called Non-Intrusive Load Monitoring or NILM) [15]. However, NILM cannot be applied for data from standard smart meter devices without further information on the specificities of the plugged appliances. Alternatively, a wireless sensor network could be used to track the individual *power* of appliances. With the recent advancements towards low-energy sensors [16], this represents a viable alternative in terms of energy efficiency and yields the most precise results. In the long run, we aim at developing a system based on smart meter data. For developing our pilot system, we utilize individual measurement devices in order to comprehend our data better. The technical set-up of our pilot study consists of (1) five wall plugs of the intelligent home FIBARO system [17] sending data on *power* once a second over z-wave; (2) an AKG CK92 Blue-Line omnidirectional microphone, an M-Audio MobilePre audio interface, and two Genelec 8020CPM loudspeakers; and (3) a laptop running PureData and SuperCollider for audio processing and DOMOTICZ [18], a software for smart home measuring devices receiving the data from the plugs.

3.1. Electric power consumption in the kitchen

Data of the kitchen appliances listed above have been recorded and analyzed to design the pilot system. The data provide rich patterns containing both information on *power* cycles typical of each appliance and on usage patterns typical of the kitchen users. The measured data are discontinuous due to the sampling of the measurement and the power draining behavior of the appliances. Thus, smoothing is needed and was implemented within the data preprocessing described in Sec. 3.2.

The *power* cycles of the appliances vary widely. Making coffee takes roughly one minute, during which the maker grinds coffee, boils the pre-heated water, and pours the coffee. The water kettle takes about two to three minutes depending on the amount of water being heated, and follows a step-function either being off or on. Furthermore, some appliances are directly connected to the users' interaction while others follow their own rhythm. Schwartz et al. [8] differentiate between an embodied service which stems from an activity personally performed by the user, while a background service is part of the domestic environment, e.g., the fridge with a regular cycle of cooling. The fridge's *power* cycle is hardly disturbed by user interventions, like, e.g., opening the door. The dish washer is an embodied service with a weekly, irregular cycle of usage. It is turned on two to three times a week, runs approximately for two hours, and drains most energy when it is hardly heard at all – i.e., when it is heating up water. Finally, the microwave is hardly used at all. All appliances furthermore drain power in standby, e.g., over night. One week of *power* of these different appliances is shown in Fig. 4(a).

The technical cycles of the appliances are overlaid by the pattern of the kitchen users. Fig. 4(b) shows the weekly average of three weeks' data. The presence of the institute's staff can be indirectly measured mainly by the coffee consumption showing a clear peak on Tuesday noon – where the institute's *jour fixe* takes place. Towards the weekend, the kitchen is less frequently used but the frequency of use is not reflected in the overall maximum of *power*. For instance, single runnings of the dishwasher dominate the overall *power*. We have considered ethical issues but concluded that sufficient anonymity is provided by not referring the data to our personal knowledge about the individual kitchen users' preferences.

3.2. Data processing

The two patterns of *power* described above are reflected in considerations on how efficient feedback can be given. On the one hand, successful feedback typically comprises appliance-specific data, offers interaction and presents information in an appealing and understandable way [19, 5, 20]. Feedback should be specific for a certain task and given in real time, e.g., continuously. Ambient ecological-feedback systems are promoted in the HCI community, and are mostly used for their motivational effects [8]. On the other hand, information has to be *contextualized* for the users, rather than displaying mere physical entities, as described by the same authors.

In this pilot study we explored in what ways the two levels of information can be conveyed using continuous auditory feedback. We decided to create a self-adapting system that both conveys direct feedback on the actual *power* and relates this information to the typical weekly user pattern. The algorithm of data preprocessing consists of an initialization and four iterative steps:

0. Initializing baseline. The system is initialized with data of one average week of overall *power*, which has been gathered during the three weeks' monitoring period, see Fig. 4(b). This baseline contains $7d \cdot 24h \cdot 60m \cdot 60s = 604800$ samples for each second of the week.
1. Smoothing the raw *power*. Most appliances exhibit discontinuous *power* patterns by nature of their on/off state, see Fig. 4(a). Therefore, the raw measurement data is smoothed with a leaky integrator that retains a fraction a of the past output and adds $1 - a$ of the actual input. (This is equivalent to low-pass filtering the data with a 1-pole IIR filter, or to exponential smoothing, see, e.g., [21].) Formulated for a series of digital samples, the filtered output at time k , $y(k)$, is calculated from the input at time k , $x(k)$, by

$$y(k) = a \cdot y(k-1) + (1-a) \cdot x(k) \quad (1)$$

where the smoothing constant a relates the filter time constant τ and the sampling interval T (here, 1 s),

$$a = e^{-\frac{T}{\tau}}. \quad (2)$$

A suitable value for τ has been empirically adjusted in order to attain a good balance between sufficient smoothing and adequate temporal resolution for real-time feedback. In the first experiment, we have set $\tau = 60 \text{ min}$, which has led to a clear weekly *power* pattern shown in Fig. 4(b) but to poor results for the real-time response in the first evaluation (Sec. 4.1). Therefore, we have reduced it to $\tau = 15 \text{ min}$. (After $5 \cdot \tau$, past data are decayed below 1% of their initial value.

2. Baseline update. The baseline is constantly being updated using another leaky integrator following Eq. 1 with $a = 1/3$. This means that the actual smoothed *power* is iteratively averaged with the value of the past weeks with a weight of one third.
3. *Power* difference from baseline. For each second, the smoothed *power* is subtracted from the weekly baseline, resulting in the pattern shown in Fig. 4(c) for a randomly chosen week. The function is positive (resp. negative) if the smoothed (i.e., typical) consumption is above (resp. below) the baseline.

4. Clipping outliers. The power difference is clipped along the 95 % quantile of the baseline data in order to eliminate statistical outliers. The resulting value serves as driving signal for the sonification.

The steps 1., 2., and 3. are updated with new measurement data every second. Step 4. is updated once a week, as no large variation of the 95 % quantile of the baseline is expected.

3.3. Sonification design

Our pilot system is based on the mapping decision that a more reverberant room represents more *power*. Furthermore, the sound design is qualitatively different in two reverberation regions that are confined by three presets. The system has been implemented in SuperCollider, and in the final set-up we have used the plugin *JPVerb* [22]. The main differing settings of this algorithm for the presets are given in brackets below: *t60* is the reverberation time in seconds, i.e. the time the signal takes to drop by 60 dB; *damp* is in the range of [0, 1] where 1 means strong damping for high frequencies; *size* varies between [0.5, 5] and gives the impression of a smaller/ larger space; finally, the reverberation time for the high frequency band has been slightly modified, with the parameter *highcut* for the crossover frequency, and the multiplying factor *highx* for the reverberation time of the high band:

Preset 0 - Zero: no augmentation/ no added reverberation (parameter settings as preset 1 but with *amplitude* = 0.)

Preset 1 - Kitchen: additional reverberation which is so natural or plausible that only quick changes from or to this preset are perceived (*t60* = 0.8, *damp* = 0.0, *size* = 1.1, *highx* = 0.3, *highcut* = 1500)

Preset 2 - Church: the reverberation clearly exceeds plausible kitchen room acoustics, both in reverberation time and level (*t60* = 3, *damp* = 0.6, *size* = 2, *highx* = 0.5, *highcut* = 1000)

We experimented with more artificial sounds, especially applying a pitch shift of up to 25 Hz on the incoming microphone signal. The resulting ringing echo was not well accepted in the first round of experiments and, therefore, dropped in the final version.

The driving signal of the sonification is mapped to reverberation in the way that preset 0, 1, and 2 coincide with lower threshold, zero-level and upper threshold in Fig. 4(c). In between, the levels of the parameters of the reverberation are interpolated linearly. Thus, the sonification is running and changing the room acoustics nearly all of the time (except for the 5% statistical outliers of very low *power* difference, see Fig. 4(b)). In the lower region, where the smoothed actual *power* is lower than the baseline, the reverberation is changed very naturally. Only sudden and/or large changes in consumption will be perceived as change of room acoustics. In the upper region, however, the filtered actual *power* is above the baseline level and the reverberation quickly becomes unnatural. Our hypothesis was that the presets are tuned in a way that the change from just below to just above zero-level is obvious.

This sonification design allows to convey two different patterns. The first pattern is the *change* of the actual smoothed *power*. It is conveyed by a change in reverberation times and levels, and reacts immediately with the appliance itself (due to the smoothing it reaches the maximum at approximately 5τ in Eq. 1 or a shorter period if the initial value has been clipped). The second pattern is the level of the actual *power* compared to the baseline *power*. It

allows to find out if the actual *power* is high or low as compared to the same weekday and time of the last weeks' average.

A demo video with binaural audio recording of the pilot system can be found on [23].

4. EVALUATION OF THE PILOT STUDY

The pilot study has aimed at assessing two factors, i.e. *perceptibility* and *contextualization*, and has been realized in two rounds of experiments. A first round of evaluation has centered on the question of perceptibility of the system. Furthermore, as the feedback is continuous and persistent, it needs to be accepted by the users. Following the paradigm of auditory augmentation, the auditory environment must not be changed in a way that the communication of the kitchen users is disturbed. Still, the feedback must not be too subtle in order to still being perceived. The second round of evaluation has questioned the contextualization of the system, specifically if and how well information on energy consumption is conveyed and if the energy awareness of the users is influenced.

The test subjects are our colleagues, i.e. the users of the institute's kitchen. These are 3 females and 12 males all of which have their workplace in the same building as the kitchen is. On the one hand, this group of subjects does not reflect a representative part of the population. On the other hand, they provide cumulated expertise as experienced listeners with a background in audio – engineering, music, and/ or sound design.

4.1. Evaluation part I

The first round of experiment ran for ten days during the end of the the winter term 2017, i.e. at a time when the kitchen is frequently used. The staff members had been informed beforehand that an experiment would be conducted in the kitchen involving the measurement of electrical power consumption. No further information had been given.

Following this first test period, a questionnaire was distributed to collect impressions of the system. Eleven subjects filled in the questionnaire. It contained standardized questions on kitchen use and open questions on the system. Results showed that the participants use the kitchen on a regular basis (7 people use the kitchen 3-5 times a week, 3 people use it 1-2 times, and only one less often). An overview of statements drawn from the questionnaire is shown in Fig. 2. The answers can be ordered along two continua: one is the level of perception/ control – from no audio-related perception to interactive control of the audio system. The other axis shows the range of negative and positive valence.

Asked whether they noticed anything during the test period in the kitchen only two participants denied. Thus, the majority of participants perceived the feedback. However, the results also show that the feedback was sometimes not perceived, perhaps occasionally due to the visiting time of the users. When the remaining nine participants were asked if they had noticed anything *on all occasions*, only three agreed, of which two were not sure about it. These findings show that the sonification reached its goal of staying at least partially below the threshold of perception (for cases when the actual *power* is below the baseline). The nine participants who have noticed something during the test period all correctly identified either resonant audio phenomena, feedback, or a changed soundscape. Without being asked explicitly, three subjects reported that they felt able to interact over speech or making

Figure 2: Qualitative statements by the subjects labeled s1 to s11 collected during the first experiment. The statements are ordered according to their level of perception/control of and valence towards the system.

other noises with the sound; one participant reported that his behavior had no influence.

The questionnaire contained only open questions. Most participants answered neutrally, but three complained about the sound. Two even turned the system off during a joint meeting (using the emergency switch of the loudspeaker). By contrast, two other participants gave positive answers about the ‘subtle feedback’. They said that the kitchen was perceived more ‘lively’. Three subjects asked specifically how the audio feedback is related to electricity consumption; none had worked out a hypothesis on the mapping.

The qualitative results above have led to modifications of the system, mainly the calibration of the parameters and the implementation of a better algorithm for reverberation. The smoothing constant of the actual *power* was reduced in order to give feedback more promptly. The preset 2 (Church) was tuned in order to reach a more pleasing sound quality even when the system is running at the maximum, i.e. at the upper threshold of the power difference.

4.2. Evaluation part II

The second round of evaluation aimed at assessing the user’s possibility to contextualize the presented auditory information. Furthermore, the levels of control/perception and valence found in the first experiment have been explored more systematically, and completed with the dimension of arousal following research on affective reactions [24]. This evaluation has been conducted in the form of a diary study. The participants have been introduced to the system, both in terms of data preprocessing and the presets of the sound feedback. They have been requested to write a diary entry each time they would use the kitchen for a period of 2 two full weeks during the semester. The diaries have been mini-questionnaires asking for:

- Participant’s name
- Date and time
- Duration and reason of stay in the kitchen

(a)

(b)

Figure 3: (a) 5-scale self-assessment manikin (SAM) for the dimensions valence, arousal, and control, measuring the affective reaction. The pictograms are taken from [25]. (b) For analysis the 5 states of the above diagram were given the numbers -2, 1, 0, 1, and 2, where 0 marks the neutral manikin in the middle. The results for the SAM rating show a wide spread of affective responses in the diary entries.

- Whether the soundscape has attracted any attention:
 - A description of the situation
 - A 5-scale self-assessment manikin (SAM) for the dimensions control, valence, and arousal, measuring the affective reaction of a users towards the sound system, see Fig. 3(a). The SAM technique has been established for quick and non-verbal testing and has been successfully used to evaluate sound, e.g., [24].
 - Any other reactions.

The collected diary entries have been analyzed both quantitatively and qualitatively. In total, 14 users took part in the survey. Six users filled in their diary entries on a regular basis, with 13,6 entries per person. Eight more users only handed in 1-3 entries each. Not all of them could be taken into account, because the system has occasionally been turned off by the users and was only re-started with some delay. The ratio between entries where people reported perceiving something and not perceiving anything was about 50%. Astonishingly, these data show no correlation with the driving value of the sonification. The hypothesis, that the sonifi-

cation is clearly detected above the threshold of 0 is not supported by the data. This could be due to statistical reasons, with a sample size of approximately 60 entries that could be taken into account for the statistical analysis. Another explanation might be deduced from the qualitative answers. The diary entries showed large inter-personal differences. While some people appear to be very sensitive towards the mapping, describing the reverberation as "strong" or "reverberation of a church" even for values just above 0, others have not remarked anything until much higher values. As a conclusion on these results, the calibration of the mapping has to be thoroughly tested with the participants, and the participants have to become acquainted with the mapping in a longer learning phase.

Qualitative analysis of the SAM rating supports the findings above. Different kitchen users had very different affective reactions towards the system. Some being positive in general, some negative, and some varied their rating from entry to entry. The bar chart in Fig. 3(b) gives an overview of the heterogeneous ratings. For the dimension of pleasure the mean value is not significantly different to 0 ($mean = 0.13$ with confidence interval of $+/- 0.36$, calculated for a significance level of $\alpha = 5\%$ and $n = 60$ samples), i.e., positive and negative attitudes towards the system are in balance. Arousal has a $mean$ of $0.70 +/- 0.28$, thus the system has an arousing factor. In the case of dominance, the mean value is $-0.66 +/- 0.31$, thus subjects feel to be slightly passive towards the system.

5. CONCLUSION AND OUTLOOK

This paper describes a system of auditory augmentation that conveys information on electric power consumption of an institute's kitchen. Data collected from five kitchen appliances contain two patterns: one stemming from the technical cycle of each appliance, and a second one from the interaction of the kitchen users. The data have been pre-processed in order to give information on both levels. The smoothed actual power consumption relative to a baseline controls the level of reverberation within two sound presets: a 'typical' kitchen use leads to a typical kitchen reverberation, while a particularly high power consumption leads to the reverberation of a church.

The results of the evaluation indicate that the state of the system was not perceived saliently. One problem of the approach of auditory augmentation in the kitchen is that the stationary noise of running appliances interferes with the perceptibility of the reverberation. For instance, the water kettle or dishwasher sounds are broadband noise signals that mask the perception of reverberation. By contrast, reverberation is much better perceivable if the source sound is short and transient. These kinds of sounds stem from the interaction of the users, e.g., of footsteps or when placing a cup on the countertop. Thus, the ambient auditory display is more perceptible when excited by users' sounds than by the sounds of appliances. One solution could be to adjust the reverberation level dynamically to the level of the soundscape.

In general, sustainable HCI systems which aim at encouraging users to save energy face the conflict that they themselves utilize resources. For a prototype design, this can be accepted. In an optimized product, the net energy balance has to be weighed.

We consider a few design changes for a next, possibly in-home experiment. It might be useful to separate embodied services from background services. In our system, this would mean using a different smoothing constant for the fridge. Its regular cycles could be smoothed out, as feedback on whether the fridge drains power

or not is not relevant to the users.

Further evaluation is envisaged, e.g., conducting an experiment in the kitchen with participants placed in the room for a few hours, giving them a task, and checking if feedback on the power is still being perceived at the background of their attention. In a next step, the system should be tested in a real in-home setting.

The presented system is deliberately simple and neutral, i.e. completely data-driven and not judgmental with respect to an absolute level of consumption. On the contrary, it is self-adaptive, constantly updating the weekly baseline. A shortcoming of the system is its real-time nature, giving feedback about now and how this time and date relates to the respective previous ones. Times in which the kitchen are not or hardly used, e.g., at night, are not taken into account. Turning appliances off during the night is possibly the largest factor for saving energy at our institute's kitchen. In a future version of the installation this fact should be included in the sonification design.

6. ACKNOWLEDGEMENT

This research was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

7. REFERENCES

- [1] European Environment Agency, "Final energy consumption of electricity by sector." [Online]. Available: <http://www.eea.europa.eu/data-and-maps/indicators/final-energy-consumption-by-sector-9/assessment-1>
- [2] European Commission, "Benchmarking smart metering deployment in the eu-27 with a focus on electricity," *REPORT FROM THE COMMISSION*, 2014.
- [3] European Commission, Directorate-General of Communications Networks, Content & Technology, "Definition of a research and innovation policy leveraging cloud computing and iot combination, final report." [Online]. Available: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=9472
- [4] R. Hierzinger, M. Albu, H. van Elburg, A. Scott, A. Łazicki, L. Penttinen, F. Puente, and H. Sæle, "European smart metering landscape report, smartregions deliverable 2.1," vol. 24, p. 2013, 2012.
- [5] S. Darby, "The effectiveness of feedback on energy consumption: a review for DEFRA of the literature on metering, billing and direct displays." Environmental Change Institute, University of Oxford, Tech. Rep., 2006. [Online]. Available: <http://www.eci.ox.ac.uk/research/energy/electric-metering.php>
- [6] Y. Ishihara, M. Ishihara, F. Hirayama, and K. Yasukawa, *Domestic Electricity Consumption Visualized as Flowing Tap Water to Raise the Feeling of Waste*. Cham: Springer International Publishing, 2015, pp. 546–550. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-21383-5_91
- [7] A. Harman, H. Dimitrov, R. Ma, S. Whitehouse, Y. Li, P. Worgan, T. Omirou, and A. Roudaut, "Notifall: Ambient sonification system using water," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA

- '16. ACM, 2016, pp. 2667–2672. [Online]. Available: <http://doi.acm.org/10.1145/2851581.2892443>
- [8] T. Schwartz, G. Stevens, L. Ramirez, and V. Wulf, "Uncovering practices of making energy consumption accountable: A phenomenological inquiry," *ACM Trans. Comput.-Hum. Interact.*, vol. 20, no. 2, pp. 12:1–12:30, May 2013. [Online]. Available: <http://doi.acm.org/10.1145/2463579.2463583>
- [9] T. Bovermann, R. Tünnermann, and T. Hermann, "Auditory Augmentation," *International Journal on Ambient Computing and Intelligence (IJACI)*, vol. 2, no. 2, pp. 27–41, 2010.
- [10] R. Tünnermann, J. Hammerschmidt, and T. Hermann, "Blended sonification –sonification for casual information interaction," in *Proceedings of the International Conference on Auditory Display (ICAD)*, 7 2013.
- [11] F. Kilander and P. Lönnqvist, "A whisper in the woods - an ambient soundscape for peripheral awareness of remote processes," *Proceedings of the International Conference on Auditory Display (ICAD)*, 2002.
- [12] E. Brazil and M. Fernström, "Investigating ambient auditory information systems," *Proc. of the International Conference on Auditory Display (ICAD)*, 2007.
- [13] A. Parthy, C. Jin, and A. van Schaik, "Reverberation for ambient data communication," *Proceedings of the International Conference on Auditory Display (ICAD)*, 2004.
- [14] W. Ritsch, T. Musil, A. Sontacchi, J. Zmölnig, and F. Zotter, "The cubemixer a performance-, mixing- and masteringtool," in *Proceedings of the LAC 2008 (Linux Audio Conference 2008)*, F. Barknecht and M. Rumori, Eds.
- [15] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 1, pp. 76–84, 2 2011.
- [16] J. Rabaey, J. Ammer, J. L. da Silva, and D. Patel, "Picoradio: Ad-hoc wireless networking of ubiquitous low-energy sensor/monitor nodes," in *Proceedings IEEE Computer Society Workshop on VLSI 2000. System Design for a System-on-Chip Era*, 2000, pp. 9–12.
- [17] Fibaro, "Wall plug," 2017. [Online]. Available: <http://www.fibaro.com/uk/the-fibaro-system/wall-plug>
- [18] Domoticz. (2017) Home automation system. [Online]. Available: <https://domoticz.com/>
- [19] W. Abrahamse, L. Steg, C. Vlek, and T. Rothengatter, "A review of intervention studies aimed at household energy conservation," *Journal of Environmental Psychology*, vol. 25, no. 3, pp. 273 – 291, 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S027249440500054X>
- [20] C. Fischer, "Influencing electricity consumption via consumer feedback. a review of experience," in *European Council for an Energy Efficient Economy (ECEEE) (ed.):Energy Efficiency: Just do it!ECEEE 2007 Summer Study Proceedings*, 2007, pp. 1873–1884.
- [21] Performity LLC, "Exponential filter," 2018. [Online]. Available: <http://gregstanleyandassociates.com/whitepapers/FaultDiagnosis/Filtering/Exponential-Filter/exponential-filter.htm>
- [22] GitHub, "Extension plugins for the supercollider3 audio synthesis server." [Online]. Available: <https://github.com/supercollider/sc3-plugins>
- [23] Anonymized, "Supplementary material (link to be exchanged by data-doi for camera-ready version)," 2 2018. [Online]. Available: <https://goo.gl/NiWs7j>
- [24] F. Stevens, D. Murphy, and S. Smith, "Emotion and soundscape preference rating: using semantic differential pairs and the self-assessment manikin," in *Sound and Music Computing conference, Hamburg*, vol. 2016, 2016.
- [25] A. P. Soares, A. P. Pinheiro, A. Costa, C. S. Frade, M. Comesaña, and R. Pureza, "Affective auditory stimuli: Adaptation of the international affective digitized sounds (iads-2) for european portuguese," *Behavior Research Methods*, vol. 45, no. 4, pp. 1168–1181, Dec 2013. [Online]. Available: <https://doi.org/10.3758/s13428-012-0310-1>

Figure 4: (a) One week of electric power consumption during the second evaluation study, 12-18 March 2018. The data have been smoothed for this plot with $\tau = 1 \text{ min}$ to show the actual, short-time power consumption. The two largest peaks are due to runnings of the dishwasher on Tuesday and Wednesday of that week. During weekdays there are many smaller peaks due to coffemaker and the water kettle. Finally, the small periodicity that can be observed during the night is due to the cooling rhythm of the fridge. It continues throughout the week but is masked (in this plot) by higher consumption during the day. (b) Average weekly *power* derived from three weeks' data of a different period, in Autumn 2017. The dotted line shows the 95 % quantile that separates 5% of outliers of peak current. The filter time constant τ for generating these plots is one hour. (c) Power difference of the actual *power* shown in (a), smoothed with $\tau = 15 \text{ min}$, minus the weekly baseline shown in (b). The data are clipped at the lower and upper thresholds that correspond to the 95% quantile of the baseline data. This clipped *power* difference is the driving signal for our sonification approach.

AUDITORY DISPLAYS OF ELECTRIC POWER GRIDS

Patrick Cowden, Luke Dosiek

Union College
Department of Electrical, Computer,
and Biomedical Engineering
Schenectady, NY 12308
patrickjcowden@gmail.com
dosiekl@union.edu

ABSTRACT

This paper presents auditory displays of power grid voltage. Due to the constantly changing energy demands experienced by a power system, the voltage varies slightly about nominal, e.g., 120 ± 2 V at 60 ± 0.04 Hz. These variations are small enough that any audible effects, such as transformer hum, appear to have constant volume and pitch. Here, an audification technique is derived that amplifies the voltage variations and shifts the nominal frequency from 60 Hz to a common musical note. Sonification techniques are presented that map the voltage magnitude and frequency to MIDI velocity and pitch, and create a sampler trigger from frequency deviation. Several examples, including audio samples, are given under a variety of power system conditions. These results culminate in a multi-instrument track generated from the sonification of time-synchronized geographically widespread power grid measurements. In addition, an inexpensive Arduino-based device is detailed that allows for real-time sonification of wall outlet voltage.

1. INTRODUCTION

AC electric power grids are highly regulated such that under ambient operating conditions the magnitude and frequency of the voltages throughout the system remain virtually unchanged from their nominal values, e.g., 120 V and 60 Hz at a residential wall outlet in the United States. If one were to measure the voltage, however, one would notice that in reality there are small variations about nominal as seen in Fig. 1, which displays the residential mains-level voltage magnitude and frequency of the US Eastern Interconnect power system over a period of 1 hour.

This variability about nominal is due to the aggregate effect of small and seemingly random changes in electrical load caused by events such as lights being switched on and off. Indeed, the random component in the measured system outputs is well-modeled as white noise that has been filtered by the power grid system dynamics [1, 2]. Due to the fact that system dynamics information is embedded in these measured ambient data, they may be analyzed as predictors of overall power grid health and stability, e.g., [2–6]. The data may also be used to detect transient events

Figure 1: Example voltage magnitude and frequency measured from the United States Eastern Interconnect power system.

such as ringdown-type responses to system disturbances or rogue oscillations due to, e.g., damaged controllers [7, 8].

Despite the information-rich signals available in the measured power system outputs, the grid is so tightly controlled that the AC waveforms appear to have constant amplitude and frequency when viewed by the naked eye. Similarly, the audible effects of the AC signals appear to have constant parameters, e.g., the transformer hum that is often heard near substations. With some signal processing, however, one can amplify the ultra low-level variations in the AC waveform so that they are perceivable to the human ear.

This would allow for the use of audio displays to augment existing wide-area situational awareness tools currently used by power grid operators. These include the numerical methods discussed in the aforementioned references, as well as simple visual displays of system parameters such as the spectra of the frequency variations. Indeed, it has been shown in, e.g., [9] that the human ear can be more successful than the eye in perceiving certain spectral features of a signal. Additionally, there is evidence that auditory displays may be used in situational awareness applications; the authors of [10] utilize the sonification of network traffic to detect anomalies such as attempted attacks by hackers.

To the best of the authors' knowledge, the only prior work

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

on generating auditory displays for power systems has been conducted by the authors of [11–14]. They created a sonification tool for converting a measure of the energy consumed by household appliances into sounds, e.g., bird songs, that could be incorporated into the ambient soundscape of a person's home life. One of the goals of that project was to increase the awareness of the general public of the characteristics of their home energy use.

This paper focuses on the development of auditory display techniques for power system AC voltage measurements. These signals include a random component, and can be obtained from the multichannel geographically-widespread time-synchronized measurements typically used for wide-area situational awareness by power grid operators. Previous work outside of the power systems realm has demonstrated the feasibility of such a scheme. For example, the authors of [15, 16] developed methods of audification and musification, respectfully, of multichannel geographically-widespread seismograph data. The author of [17] describes the use of random processes, not unlike the random variations in power system voltage, in the composition of *Achorripsis*, a musical piece created by Iannis Xenakis in 1957. The auditory display techniques explored in this paper are applied to both simulated and measured voltage data that include both ambient and oscillatory disturbance conditions. The resulting audio waveforms range from the basic audification of a single AC voltage signal to a multi-instrument MIDI-based musification.

This paper is organized as follows. In Section 2, an overview of power system small-signal dynamics is given that includes measurement and preprocessing techniques. Sections 3 and 4 detail the audification and sonification techniques used, and Sections 5 and 6 provide audification and sonification results, including audio samples. Section 7 describes a device that was developed for the real-time sonification of mains voltage, and finally Section 8 provides concluding remarks and a discussion of ongoing and future work.

2. POWER SYSTEM DYNAMICS AND MEASUREMENTS

The power system dynamics under consideration here are of extremely low frequency. As seen in Fig. 1, the signals are dominated by a very slow moving process that is due to the speed governor located on each synchronous generator connected to the grid. Supported by the total inertia of turbines and generators, these dynamics are in the sub-0.1 Hz range and have large time constants typical of massive mechanical systems. The speed governors work to keep the grid in synchronization, meaning that for a given window in time, the sub-0.1 Hz variations in system frequency will be nearly identical throughout the system. The voltage magnitude is similarly controlled, but is in general altered by power transmission and distribution components according to linear circuit theory.

Electromechanical dynamics, ranging from 0.1 to 2.0 Hz, cause signal fluctuations that do vary from location to location. They are caused by turbine dynamics interacting with transmission line dynamics and manifest as groups of generators with speeds that swing with or against other groups of generators. They have low energy compared to the speed governor dynamics, seen in Fig. 1 as the small noisy process superimposed on the low-frequency trend.

Finally, note in Fig. 1 the occasional brief drops in voltage magnitude that are not present in the frequency signal. These are caused by the inrush currents of motors starting up near the measurement device. These brief spikes in current cause a corresponding drop in the voltage magnitude on all lines connected to the

same circuit breaker due to the resistive losses in the distribution wires. They do not affect the frequency since they cannot change the rotational inertia of the synchronized generators on the entire power grid. In order to observe similar transients in frequency, a large system-level event is required such as a transmission line failure.

In order to understand how one measures the instantaneous voltage magnitude and frequency discussed so far, one must examine the mathematical expression for the instantaneous voltage:

$$v(t) = \sqrt{2}V(t) \cos(2\pi f_o t + \theta(t)) \quad (1)$$

where $V(t)$ is the root mean squared (RMS) magnitude of the voltage, f_o is the nominal system frequency, and $\theta(t)$ is the instantaneous phase angle, which may be expressed as

$$\theta(t) = \theta_o + 2\pi \int_0^t \Delta f(\tau) d\tau \quad (2)$$

where θ_o is the initial phase angle of the sinusoid and $\Delta f(t)$ is the deviation in frequency from nominal such that the instantaneous frequency $f(t)$ is

$$f(t) = \frac{1}{2\pi} \frac{d}{dt} (2\pi f_o t + \theta(t)) = f_o + \Delta f(t) \quad (3)$$

In phasor notation, the instantaneous voltage is a two-element complex number comprised of the magnitude and phase

$$\bar{V}(t) = V(t)e^{j\theta(t)} = V(t)/\angle\theta(t) \quad (4)$$

where j is the imaginary number. Note that in Fig. 1, $V(t)$ is shown in the upper plot, while $f(t)$ is in the lower.

Estimation of the voltage parameters begins with the sampling of the raw voltage waveform followed by a filtering operation that attenuates harmonics, measurement noise, or other interference, allowing the acquired signal to closely resemble (1). Then any one of several signal processing approaches may be taken to obtain $V(t)$ and $f(t)$. They range from the simple, e.g., finding $V(t)$ from a direct calculation of the root-mean-squared value of a period of $v(t)$ and finding $f(t)$ through zero-crossing estimation; to the complex, e.g., FM demodulators, recursive maximum likelihood estimators, etc. See [18] for several examples.

Alternatively, phasor measurement units (PMU) are commercial devices that provide signals known as synchrophasors (named due to their GPS-time synchronized nature) that are estimates of the phasor quantity $\bar{V}(t)$ in (4). The voltage magnitude $V(t)$ is readily available as the synchrophasor magnitude, and $f(t)$ may be found by adding nominal system frequency f_o to the numerical derivative of the synchrophasor angle $\theta(t)$, which is $\Delta f(t)$. Note that in this paper, synchrophasor data was used with the audification and sonification techniques outlined in the following two sections.

3. AUDIFICATION METHOD

Because the instantaneous power system voltage is already a sinusoid, a simple audification process may be implemented similar to what was developed for seismograph data in [16]. Input to the audification algorithm are the voltage magnitude $V(t)$ and frequency $f(t)$. They are each normalized as

$$V_n(t) = \frac{V(t) - \min(V(t))}{\max(V(t)) - \min(V(t))} \in [0, 1] \quad (5)$$

$$f_n(t) = 2 \frac{f(t) - \min(f(t))}{\max(f(t)) - \min(f(t))} - 1 \in [-1, 1] \quad (6)$$

Noting that the expression in (3) implies that the normalized frequency is equivalent to the normalized frequency deviation, the normalized instantaneous phase $\theta_n(t)$ is found by numerically integrating $\Delta f_n(t) = f_n(t)$ and setting initial phase $\theta_o = 0$. Note this normalization is conducted with reference to the min/max values of the particular measurements under investigation such that maximum values of $V(t)$ and $f(t)$ get mapped to 1 and the minimum values get mapped to 0 and -1, respectively. Other normalization algorithms could be used here, such as normalizing with respect to “typical” min/max values under normal power system operating conditions. The exploration of these alternatives is a subject of ongoing work.

To create the audio signal, $V_n(t)$ and $\theta_n(t)$ are upsampled using linear interpolation to 44.1 kHz from, e.g., the 30 Hz reporting rate of a PMU. Then, audio signal $a(t)$ is formed as

$$a(t) = V_n(t) \cos(2\pi f_a + \alpha \theta_n(t)) \quad (7)$$

with user-tunable parameters f_a , which defines the “middle” note (e.g., 261.63 Hz for middle C), and α , which determines maximum excursion in pitch from f_a .

Once constructed, $a(t)$ provides a simple auditory display for the nonstationarity of practical power grid voltage.

4. SONIFICATION METHODS

Three different sonification techniques were implemented using the parameter mapping techniques from [19]. The first involves mapping the voltage magnitude $V(t)$ and frequency $f(t)$ to MIDI note velocity and pitch, respectively. The second technique consists of using a threshold detector on frequency fluctuation $\Delta f(t)$ to create a MIDI-based sampler trigger. Lastly, the third technique derives the MIDI note length from the sampling rate of $V(t)$ and $f(t)$. The data processing for each of the sonification algorithms is implemented in Python, and MIDI files are exported using the MIDI-Util Python library [20].

4.1. MIDI Note Velocity and Pitch Mapping

For this method, samples of voltage magnitude $V(t)$ are converted to an array of MIDI velocities and samples of voltage frequency deviation $f(t)$ are converted to an array of MIDI pitches. To create the MIDI velocity values, the normalized voltage magnitude used in the audification process, $V_n(t)$, is scaled by 127 to span between 0 and 127 (since there are 128 possible MIDI velocity values) and rounded to the nearest integer.

Creating MIDI pitch values is slightly more complicated as the user may specify both the musical key and the number of octaves above/below the middle note that are to be spanned by the mapping of $\Delta f_n(t)$. First, the range $[-1, 1]$ is divided into N equally sized bins corresponding to the total number of notes included in the desired number of octaves. Then, the user-selected key is used to determine which notes (and MIDI pitch values) correspond to each bin. Note that if the user specifies an invalid key value, the algorithm defaults to C major. Finally, each value of $\Delta f_n(t)$ is assigned a MIDI pitch based on which bin it belongs to.

For example, if one desired sonification in the key of C major with 1 octave above/below C4, then 2 octaves with 7 notes each plus the middle C note would require 15 bins in the range $[-1, 1]$ as seen in Table 1. So if three consecutive samples of

Table 1: Example MIDI Pitch Mapping in C Major with 1 octave above and below middle.

Bin	Δf_n Range	Note	MIDI Pitch
1	$[-1, -0.8667)$	C3	48
2	$[-0.8667, -0.7333)$	D3	50
$\vdots$	$\vdots$	$\vdots$	$\vdots$
8	$[-0.0667, 0.0667)$	C4	60
$\vdots$	$\vdots$	$\vdots$	$\vdots$
14	$[0.7333, 0.8667)$	B5	71
15	$[0.8667, 1]$	C5	72

$\Delta f_n(t)$ were $[-0.91, -0.99, 0.81]$, then the MIDI pitches would be $[48, 48, 71]$.

4.2. Sampler Triggering

This method analyzes the absolute value of the normalized voltage frequency deviation $\Delta f_n(t)$ to create an array of MIDI velocity values that can be mapped to sampler triggering. A threshold $\gamma \in [0, 1]$ is applied such that the MIDI velocity M_v is

$$M_v(t) = \begin{cases} 0 & \text{if } |\Delta f_n(t)| \leq \gamma \\ 127 & \text{if } |\Delta f_n(t)| > \gamma \end{cases} \quad (8)$$

To finalize the sampler triggering, the MIDI pitch is fixed at 60, the value used by MIDI samplers to play back the original audio sample. This may then be opened with an appropriate sampling device which will play a particular sound, e.g., a snare drum hit, each time the power system frequency $f(t)$ deviates far enough from nominal.

4.3. MIDI Note Length Assignment

If real-time sonification is not necessary, one may simply assign a constant note length, e.g. an eighth note, to all data samples. Or, one may randomly assign standard note lengths such as sixteenth, eighth, or quarter notes. Depending on the tempo of the musical piece and the sampling rate of the voltage measurement, this will in general result in an audio recording that is of a different duration than the power system voltage recording. For example, eighth notes at 120 bpm occupy 0.25 seconds each. If constant eighth note lengths are applied to 10 seconds of voltage data recorded at 60 samples per second, the resultant audio file will be comprised of 600 eighth notes and last 150 seconds.

When real-time sonification is desired, the MIDI note length may be set in decimal quarter note format based on the sampling rate of the streaming voltage measurements. Given sampling rate f_s in samples per second and musical tempo T in beats per minute, the MIDI note length M_l in beats per sample is found as

$$M_l = \frac{T}{60f_s} \quad (9)$$

For the above example, this would result in an M_l of 0.0333 beats per sample. When compared to using constant eighth notes, which would correspond to 0.5 beats per sample, the notes would be

played at a much higher speed, but the audio and voltage signals would have identical time durations, thus enabling real-time operation.

5. AUDIFICATION RESULTS

This section presents the results of the previously described audification process applied to both measured and simulated power grid voltage under ambient and oscillatory conditions.

5.1. Ambient Conditions

Thirty seconds of voltage magnitude and frequency were measured from a wall outlet at Union College in Schenectady, NY, as seen in Fig. 2, where one of the aforementioned motor startup effects can be seen in the voltage magnitude. The audification process described in Section 3 was applied with $f_a = 261.63$ Hz and $\alpha = 75$. In the resulting audio track, which can be found at <https://soundcloud.com/pcowden/ambient-data-audification> the fluctuations of voltage magnitude can be heard as changes in volume, while the frequency fluctuations can be heard as the slowly changing pitch of the tone.

5.2. Oscillatory Conditions

One minute of voltage magnitude and frequency was simulated in MATLAB using the miniWECC, a reduced-order linearized model of the Western Electricity Coordinating Council (WECC) power system [21]. A forced oscillation scenario was created by injecting a 0.35 Hz square wave as real power on a particular generator at the 20-second mark. Forced oscillations are deterministic periodic signals that appear in otherwise ambient data due to a periodic system input such as a poorly tuned controller, broken equipment, or a cyclic load. See [22] for a detailed discussion of power system forced oscillations. The simulated data is shown in Fig. 3.

The audification process was applied with $f_a = 261.63$ Hz and $\alpha = 175$, resulting in the audio file found at <https://soundcloud.com/pcowden/forced-oscillation>.

The power system forced oscillation is clearly heard as periodic fluctuations in volume and pitch.

6. SONIFICATION RESULTS

This section presents the results of the previously described sonification process applied to measured power grid voltage under ambient and transient conditions. It also includes a multi-instrument composition derived from multiple voltage measurements. Note that the reFX Nexus VST synthesizer plugin was used in each of the examples below.

6.1. Ambient Conditions

Sonification was performed on the ambient data from Fig. 2. MIDI note velocity and pitch mapping were used in the key of Bb major with 2 octaves above and below the middle Bb note. The sampling rate of the voltage data was 4 samples per second, and the tempo of the audio was 120 bpm. The MIDI note lengths were first set to be all eighth notes, and then were chosen randomly among quarter, eighth, and sixteenth notes. The MIDI file was played by the

Figure 2: Ambient voltage data measured from wall outlet at Union College in Schenectady, NY.

Figure 3: Ambient voltage data simulated from the miniWECC system with 0.35 Hz forced oscillation starting at 20 seconds.

synthesizer using the “UK Pikes 1 Pluck” synth, creating the following audio pieces:

Constant note length:

<https://soundcloud.com/pcowden/power-grid-ambient-data-sonification>

Random note length:

<https://soundcloud.com/pcowden/power-grid-ambient-data-sonification-bb-major-randomized-note-length>

In the resulting tracks, the fluctuations of voltage magnitude can be heard as changes in note volumes while fluctuations of the voltage frequency can be heard as varying note pitches. While both tracks provided an auditory display of the power system voltage, the version with the random note lengths provided a more melodic experience.

Figure 4: Voltage data measured from the WECC system during a transient ringdown event.

6.2. Transient Ringdown Response

A stable power system responds to a large system disturbance through a damped natural oscillation process referred to in the power community as a “ringdown response.” The frequencies and dampings of the ringdown are defined by the natural oscillatory modes of the power system. If the system disturbance is large enough, the ringdown signal will dominate, providing an extremely high SNR signal that engineers may use to estimate the system modes with great accuracy.

An example of one such ringdown response that was measured in the WECC system is given in Fig. 4. The voltage magnitude is given in per-unit (p.u.) in order to mask the location of the measurement due to confidentiality agreements.

For the sonification of this data, the MIDI velocity and pitch were mapped from the voltage data in the key of F major spanning 4 octaves above/below the middle note. As seen in Fig. 5, the synthesizer display of the MIDI values closely resemble the power grid measurements of Fig. 4. To simulate a real-time environment, the MIDI note lengths were set using (9) with a tempo of 150 bpm and a power grid data sampling rate of 10 samples per second, resulting in a MIDI note length of 0.25 in decimal quarter notes, which represents a sixteenth note. The resultant MIDI file was processed by the synthesizer using the “Waterhouse Pikes Pluck” synth to create the following audio file:

<https://soundcloud.com/pcowden/transient-ring-down-response-real-time-f-major>

6.3. Multichannel Ambient Data (Musification)

Time-aligned synchrophasor data from six geographically widespread locations in the WECC power grid were sonified to create a multi-instrument composition. The voltage magnitude (in per-unit) and the frequency are shown in Fig. 6. The locations of measurements are withheld and the voltage magnitude shown in the accompanying figure was normalized before converting to per-unit in order to honor confidentiality agreements.

The sampling rate for the synchrophasors was 60 samples per second, and 640 samples from each measurement location were

Figure 5: Synthesizer display of MIDI note velocity and pitch mapped from the voltage magnitude and frequency of Fig 4.

Figure 6: Voltage data measured from the WECC system during ambient conditions used in the multichannel sonification example.

analyzed, resulting in 10.667 seconds of power grid data. The musical piece was created at 120 bpm and one sampling period of the voltage data was assigned to be a sixteenth note, or 0.125 seconds, resulting in an 80-second musical composition.

The first measurement location was sonified with velocity and pitch mapping to provide a lead melody in the key of A major spanning 2 octaves above/below the middle note. The MIDI note lengths were randomly chosen as either 1, 2, 3, or 4 sixteenth notes. In order to preserve the time-aligned nature of the syn-

chrophasor data, if a note longer than 1 sixteenth note was chosen, the sonification algorithm would skip over the appropriate number of samples. For example, if the first note length was 3 sixteenth notes, then the first sample of voltage data would be sonified, and the next two samples would be skipped. The “Waterhouse Pikes Pluck” synth was applied to the resultant MIDI data.

The next measurement location was chosen to play chords with the same key and octave span as the melody. Every 16th sample of the voltage data was sonified, beginning with the 1st sample, and assigned a MIDI note length corresponding to a whole note. The “Andromeda Pad” chord synth was applied to the MIDI data.

The third measurement location was sonified using sampler triggering with a threshold of 0.15 to trigger a hi-hat sample. In order to create a rhythm common to dance music, where the hi-hat is only heard on even eighth notes (the “and” of each beat), every 4th sample of voltage data was sonified, beginning with the 3rd sample, and assigned a MIDI note length of a quarter note.

The fourth measurement location was sonified using sampler triggering with a threshold of 0.1 to trigger a snare drum sample. In order to create the common rhythm of snare drums only being hit on the back beats, every 8th sample of voltage data was sonified, beginning with the 5th sample, and assigned a half note MIDI note length.

The fifth measurement location was sonified using sampler triggering with a threshold of 0.275 to trigger a crash cymbal sample. Every 8th sample of voltage data was sonified, beginning with the 7th sample, and assigned a half note MIDI note length. This placed the cymbal crashes on the eighth note (the “and”) after a snare hit.

The last measurement location was sonified using sampler triggering with a threshold of 0.02 to trigger a kick drum sample. Every 4th sample was sonified, beginning with the 1st sample, and assigned a MIDI note length corresponding to a quarter note. This timing coupled with the low trigger threshold created a nearly constant quarter note kick drum feel typical to this genre of music.

The composition, which can be heard at

<https://soundcloud.com/pcowden/multi-channel-power-grid-musification-pure> reflects the synchronized nature of AC power grid dynamics in an auditory display that is truly musical in nature. In this track, every sound has been controlled by power grid data.

Transitional audio samples were manually added and the melody and chord tracks were processed with a kick drum side-chain compression effect to create the track that can be heard at

<https://soundcloud.com/pcowden/multi-channel-power-grid-musification-edited>

This provides an example of how the sonified power grid data could be incorporated into the regular work flow of an artist. The resulting composition is a piece of music that contains elements under both human and power grid control.

7. REAL-TIME SONIFICATION DEVICE

A sonification device was designed that allows for the conversion of wall-outlet AC voltage to a MIDI signal in real time. As seen in Fig. 7, the device consists of five main components. The first is a center-tapped 5:1 transformer that takes the 120-VAC voltage from the wall socket connection and steps it down to a safe level. The center tap is connected to the wall socket earth ground, providing three output voltages: 24-VAC line-to-line, and a pair of 12-VAC line-to-ground voltages that are anti-phase with each other.

Figure 7: Flow chart of real time sonification device.

One of the 12-VAC transformer outputs is connected to the second component of the device which scales, offsets, and filters the voltage for input to the Arduino. First, a voltage divider is used to reduce the signal to approximately 1 VAC. Then an op-amp circuit is used to add approximately 2 VDC to the signal. Finally, the signal is passed through a first-order RC low-pass filter with cutoff of approximately 61 Hz to remove harmonics and high frequency noise. All components are referenced to the transformer center tap. The result is a clean sinusoid with DC offset such that it satisfies the required 0 to 3.3-V range of the Arduino analog input pin.

The Arduino microcontroller serves as the main component of the sonification system. An Arduino Zero was used for its high precision and accuracy with regards to analog-to-digital conversion and subsequent signal processing. It completes all of the signal measurement, processing, parameter-mapping sonification, and MIDI signal creation via software.

First, it estimates voltage magnitude and frequency from the sampled analog voltage input. The software operates on windows of six periods of 60-Hz. For each data window, the mean is subtracted to remove the DC offset added by the op-amp circuit. Then the voltage magnitude is found by calculating the root mean square (RMS) of the zero-mean signal. The frequency is estimated through a zero-crossing detection technique. The signal is scanned for changes in sign, and the actual zero crossing time t_{zc} is estimated through linear interpolation

$$t_{zc} = t_1 - v_1 \frac{t_2 - t_1}{v_2 - v_1} \quad (10)$$

where t_1 and v_1 refer to the time and voltage values of the sample just before a change in sign, and t_2 and v_2 refer to the values just after the sign change. Given that there are six cycles of 60-Hz

voltage per analysis window, there will be twelve zero crossings, collected in vector $\underline{t}_{zc}$. The frequency estimate for the window is derived from an averaging process that is robust to timing errors and small sampling delays. It first averages frequencies estimated from odd- and even-numbered zero crossings separately, then takes the mean of those averages as the final frequency estimate:

$$\begin{aligned}\hat{f}_1 &= \frac{1}{5} \sum_{i=1}^5 \frac{1}{t_{zc}[2i+1] - t_{zc}[2i-1]} \\ \hat{f}_2 &= \frac{1}{5} \sum_{i=1}^5 \frac{1}{t_{zc}[2i+2] - t_{zc}[2i]} \\ \hat{f} &= \frac{1}{2} \hat{f}_1 + \hat{f}_2\end{aligned}\quad (11)$$

where $t_{zc}[i]$ is the i^{th} element of $\underline{t}_{zc}$.

With voltage magnitude and frequency estimated, the previously discussed sonification techniques are implemented. Prior to uploading the software to the Arduino, the user selects the sonification parameters, i.e., musical key, number of octaves, triggering threshold, tempo, and note length technique. The software implements the sonification and converts to a binary MIDI signal that is sent to a digital output pin.

The MIDI notes created by the software are passed through one of the Arduino digital output pins to a female MIDI connector following the MIDI OUT electrical specifications for 3.3-V signaling [23]. Any synthesizer may then connect to the device via the female MIDI connector to initiate musical playback.

Note that both the Arduino and the op-amp require DC power supplies. These were created by the fifth and final component of the device, which first converts the 24-VAC transformer output to a DC signal through a full-wave bridge rectifier with smoothing capacitors. This DC signal is then fed to voltage regulators that provide the ± 5 -VDC required by the op-amp and the 9-VDC required by the Arduino. Once again, all components were referenced to the transformer center tap.

Whether a student in the classroom setting, a musician looking to incorporate the environment into their compositions, or a casual observer, it is intended that the development of the Arduino-based sonification device will provide a simple and inexpensive way of creating auditory displays of the power grid voltage nearest them.

8. CONCLUSIONS

The methods developed for this research successfully generated interesting auditory representations of power grid data. The audification and sonification algorithms are easy to use and can be applied to measured power grid voltage from virtually anywhere. For those that are unaware that the power grid voltage parameters are not always 120-V at 60-Hz, these auditory displays allow for a simple comprehension of what is really happening in a power system. Furthermore, since the fluctuations in voltage are due to the physical properties of the power grid responding to the sum of society's ever-changing energy demands, these auditory displays allow one to experience the very heartbeat of an incredibly complex system that is only ever given a second thought when the power goes out.

Ranging from the other-worldly theramin-sounding audification tracks, to the synthesizer-based sonification pieces, it has been shown that in addition to being informative these auditory displays

are rather entertaining. The multi-channel musification composition sounds like a dance track that was entirely human-produced, when in fact the power grid was playing the instruments. It is an entertaining composition that speaks to the significance of auditory displays, and shows how the intersection of music and power engineering can be used to spark interest in both areas.

An important area of future work is the consideration of these approaches as a tool for actual power system operators to monitor the overall health of the system, or to detect events such as forced oscillations or ringdown responses. Currently this is achieved with visual indicators, e.g., dials, graphs, or raw numbers, that are the results of traditional numerical analysis. A study with grid operators should be conducted in which they are presented with auditory displays of typical grid parameters in order to determine practical usefulness.

Another area of future work is the exploration of how the various audification and sonification algorithm parameters should be selected. For the results in this paper, they were tuned until the authors found an audio track that was pleasing. Consultation with music theorists and surveys with listeners should be conducted to determine which settings make for a more globally pleasing sound.

Lastly, it must be noted that the multi-channel musification tracks are rather prototypical in nature. The six power grid measurement locations chosen for sonification were done so based solely on the availability of the data, and they were randomly assigned to instruments. This created a false hierarchy of importance among the power system measurement channels since the listener will tend to assign greater importance to the melody track as compared to, e.g., the hi-hat. Furthermore, the particular instruments used in the synthesizer were chosen simply based on the musical tastes of the authors, and did not relate in any physical way to the nature of the power grid.

The next step in improving upon the power grid musification process is to address these issues. Power system measurement location may be ordered by importance based upon grid topology. For example, measurements made closest to major generation stations or load centers could be assigned higher "importance" than those obtained from the middle of a long transmission corridor if one associates importance with energy sources and sinks. Or, one could assign "importance" of a measurement location based upon its level criticality to the overall reliability of the power grid.

One could also explore the fact that synchrophasor measurements typically include current phasors in addition to the voltage phasors, implying that one can easily calculate both real and reactive power flow quantities. Thus, a scheme could be developed that assigns instrumentation based on signal type. For example, since the voltage magnitude and frequency are the quantities that are most tightly controlled by the grid, they could be used to provide the rhythm section, while the more wildly varying current and/or power values could control the melodic instruments.

Finally, a study into integrating properties of the power grid into the selection of the instruments used in the synthesizer should be conducted. Perhaps there is a musical genre better suited to represent power grid dynamics than the electronic dance music presented here. The use of audio samples such as the bird songs used in [11–14] should also be considered.

9. ACKNOWLEDGMENT

The authors would like to thank Professors Palmyra Catravas and Nick Webb at Union College and Dr. Ryan McGee of Life Orange

LLC for their insightful discussions throughout this project.

10. REFERENCES

- [1] P. Kundur, N. J. Balu, and M. G. Lauby, *Power system stability and control*. McGraw-hill New York, 1994, vol. 7.
- [2] J. W. Pierre, D. J. Trudnowski, and M. K. Donnelly, "Initial results in electromechanical mode identification from ambient data," *IEEE Transactions on Power Systems*, vol. 12, no. 3, pp. 1245–1251, 1997.
- [3] D. J. Trudnowski, J. W. Pierre, N. Zhou, J. F. Hauer, and M. Parashar, "Performance of three mode-meter block-processing algorithms for automated dynamic stability assessment," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 680–690, 2008.
- [4] N. Zhou, D. J. Trudnowski, J. W. Pierre, and W. A. Mittelstadt, "Electromechanical mode online estimation using regularized robust RLS methods," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1670–1680, 2008.
- [5] L. Dosiek, J. W. Pierre, and J. Follum, "A recursive maximum likelihood estimator for the online estimation of electromechanical modes with error bounds," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 441–451, 2013.
- [6] S. N. Sarmadi and V. Venkatasubramanian, "Electromechanical mode estimation using recursive adaptive stochastic subspace identification," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 349–358, 2014.
- [7] J. Follum and J. W. Pierre, "Detection of periodic forced oscillations in power systems," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2423–2433, 2016.
- [8] Y. Song, W. Wang, Z. Zhang, H. Qi, and Y. Liu, "Multiple event detection and recognition for large-scale power systems through cluster-based sparse coding," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4199–4210, 2017.
- [9] R. L. Alexander, J. A. Gilbert, E. Landi, M. Simoni, T. H. Zurbuchen, and D. A. Roberts, "Audification as a diagnostic tool for exploratory heliospheric data analysis," in *17th International Conference on Auditory Display (ICAD-2011)*, June 2011.
- [10] P. Vickers, C. Laing, M. Debashi, and T. Fairfax, "Sonification aesthetics and listening for network situation awareness," in *Conference on Sonification of Health and Environmental Data (SoniHED)*, September 2014.
- [11] D. Lockton, F. Bowden, C. Brass, and R. Gheerawo, "Bird-watching: exploring sonification of home electricity use with birdsong," in *Conference on Sonification of Health and Environmental Data (SoniHED)*, September 2014.
- [12] —, "Powerchord: Towards ambient appliance-level electricity use feedback through real-time sonification," in *8th International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2014)*, December 2014.
- [13] D. Lockton, F. Bowden, and C. Matthews, "Powerchord: Exploring ambient audio feedback on energy use," in *Living Labs: Design and Assessment of Sustainable Living*, D. Keyson, O. Guerra-Santin, and D. Lockton, Eds. Berlin: Springer, 2017, ch. 23, pp. 297–308.
- [14] S. Clune and D. Lockton, "Design for behaviour change and sustainability," in *Design for Behaviour Change: Theories and practices of designing for change*, K. Niedderer, S. Clune, and G. Ludden, Eds. London: Routledge, 2017.
- [15] F. Dombois, "Using audification in planetary seismology," in *7th International Conference on Auditory Display (ICAD-2001)*, July 2001.
- [16] R. McGee and D. Rogers, "Musification of seismic data," in *22nd International Conference on Auditory Display (ICAD-2016)*, June 2016.
- [17] E. Childs, "Achorripsis: a sonification of probability distributions," in *8th International Conference on Auditory Display (ICAD-2002)*, July 2002.
- [18] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. ii. algorithms and applications," *Proceedings of the IEEE*, vol. 80, no. 4, pp. 540–568, 1992.
- [19] F. Grond and J. Berger, "Parameter mapping sonification," *The sonification handbook*, pp. 363–397, 2011.
- [20] M. C. Wirt. (2016) MIDIUtil - MIDIUtil 1.1.1 documentation. [Online]. Available: <http://midiutil.readthedocs.io/en/latest/>
- [21] D. Trudnowski, D. Kosterev, and J. Undrill, "PDCI damping control analysis for the western north american power system," in *2013 IEEE Power Energy Society General Meeting*, July 2013, pp. 1–5.
- [22] J. D. Follum, F. K. Tuffner, L. A. Dosiek, and J. W. Pierre, "Power system oscillatory behaviors: Sources, characteristics, & analyses," Pacific Northwest National Lab.(PNNL), Richland, WA (United States), Tech. Rep., 2017.
- [23] "MIDI 1.0 electrical specification update," MMA Technical Standards Board/AMEI MIDI Committee," CA-033, 2014.

FLUOR SONSESCENCE: A SONIFICATION OF THE VISUALIZATION OF BRASS INSTRUMENT TONES

Christopher Jette

Center for Computer Research in Music
and Acoustics (CCRMA)
Stanford University
Palo Alto, CA, USA
jette@ccrma.stanford.edu

James H. J. Buchholz

Department of Mechanical and Industrial
Engineering, and IIHR – Hydrosience and Engineering
The University of Iowa
Iowa City, IA, USA
james-h-buchholz@uiowa.edu

ABSTRACT

This paper is a discussion of the composition *Fluor Sonescence*, which combines trombone, electronics and video. The trombone and electronics are a mediated sonification of the video component. The video is a high framerate capture of the air motions produced by sound emanating from a brass instrument. This video material is translated into sound and serves as the final video component. The paper begins with a description of the data collection process and an overview of the compositional components. This is followed by a detailed description of the composition of the three components of *Fluor Sonescence*, while a discussion of the technical and aesthetic concerns is interwoven throughout. There is a discussion of the relationship of *Fluor Sonescence* to earlier works of the composer and the capture method for source material. The paper is an overview of a specific sonification project that is part of a larger trajectory of work. Please see <https://vimeo.com/255790972/> to hear and view *Fluor Sonescence*.

1. PROJECT OVERVIEW

Fluor Sonescence explores the creation of a piece of concert music for trombone, electronics and video using the visualization of air masses from a brass instrument as source material. Smoke is placed into the air column of a brass instrument during the production of a tone. As the smoke exits the bell of the instrument a laser illuminates a plane of the cloud and a high frame-rate camera captures the visual data (see figure 1). This visual material is then translated into sound and combined with the video. The process of constructing this work is a series of translations from sounds to visuals and then back from visuals to sounds. The resulting artistic artifact is a musical composition for trombone, electronics and video projection. The interpenetration of the translated material provides temporal and gestural continuity. The translations are mediated and not meant to strictly synthesize across media but rather to balance compositional intuition and expressive data. The basic musical material is generated by translating image to sound and these musical phrases are edited to create a musical composition. This approach of working with recorded data in a studio

Figure 1: A still from the video, illustrating gray and white smoke in a black frame

setting in order to create a finished composition is modeled on the approach of Musique Concrete composers working with recordings on magnetic tape in a studio environment. With Musique Concrete, the transformations of sound resulting from manipulations in playback speed, tape direction and tape splicing define a series of constraints. For composers working exclusively with tape, the source material and the operations present a set of constraints which gives rise to a range of sounds. The transformations in *Fluor Sonescence* are less extreme, with an adjustable samplerate during the reordering process and minor temporal expansions and compressions. Like the constraints of tape music, in *Fluor Sonescence* the sonification process defines the sonic palette and the organization of the material gives rise to a particular acoustic landscape.

The constraints that determine the construction of *Fluor Sonescence* are defined by the types of translations (discussed in Section 3) and these translations occur off-line (ie. not in real

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

time). The larger goal of this project is the creation of a real time instrument for controlling and producing both auditory and visual elements. The first step in this larger plan is the work *Fluor Sonescence*, where the composer both develops the techniques that will be incorporated in the forthcoming instrument and develops what will turn into a formalized translation process. This is also an opportunity to investigate the sonic and visual palette of working with this material. *Fluor Sonescence* serves as a first investigation of the sonic and visual material that both extends a compositional trajectory and combines research in the aesthetics of mixing sonic and visual material.

2. COMPOSITIONAL COMPONENTS

The construction process begins with the production of sound in a brass instrument and the introduction of smoke into the air column. The smoke makes the wavefront visible, revealing the undulations of the various forces involved. A cross section of the smoke is then recorded to video and this serves as the starting point for the sonic composition. The location of the whitest portion of the video is used to determine which pitch is used in the trombone part (see Section 2.2). The vertical axis of the video plane is subdivided and a pitch is associated with each region (more in Section 3.1). The array of available pitches changes as the piece traverses different formal sections. The trombone pitch is drawn from an irrational overtone series and the remaining overtones are rendered as sine tones, each of which is associated with a vertical division of the screen (see Section 3.2 & 3.3). The amplitude of the sine tones (ie. electronic material) is controlled by the brightness (amount of white) in their respective portion of the video matrix. This translation process yielded the raw material for *Fluor Sonescence*. Through manual interventions of amplitude envelopes and by introducing rhythmic subdivisions, the electronic and trombone material were each sculpted to follow the perceived dramatic contour of the video. The end result is a composition where the sound of a brass instrument produces video, which is used to generate a score for trombone and accompanying electronic material.

The audience experiences visual and sonic material that is generally tightly coupled. By shaping dynamics, the insertion of silences and varying the cross domain mappings the composer aims to create a dramatic formal structure while still adhering to the translations from one domain to another. The process of composing the final work is a blending of the video and sound material to construct a compelling flow of material. In performance, this presentation of both visual and sonic material creates a situation where the audience listens to the sound and interprets the sound as being causally related. This *causal listening* described by Chion([1]), places the visual information and the the multiple sound streams in a tightly coupled relationship. This relationship is reinforced because during composition the overall scaffolding of the sound components were designed ahead of time, and the movement of the video material provides the pitch and amplitude information.

2.1. Visualizing Air Motions

Visualizations were created by holding steady tones, scales, and other note variations on a horn while injecting aerosol into the brass tubing 0.5 meters before the bell. This allowed the visualization of the steady and unsteady jets of air issuing from the bell

Figure 2: Camera and laser sheet orientations used to capture images of smoke emanating from the horn. a) parallel orientation; b) perpendicular orientation.

as the notes were played, using smoke emerging from the bell of the horn to render the air visible.

Theatrical smoke was pumped deep into the the horn using a plastic tube of inner diameter 6.4 mm and outer diameter 9.5 mm. The tube slides over a barbed rigid nylon tube that is seated in a hole in the horn. Silicone is used to insure no leakage and the nylon tube does not protrude into the air column. The smoke accumulated in the horn during rests, resulting in intense bursts of high image intensity when notes were initiated, and lighter, lower-contrast patterns during held notes or musical phrases. The smoke patterns were imaged using an IDT NX-4 high-speed camera with a Nikon 35 mm f/2 lens, and recorded at various framerates (ranging from 2HZ to 5000HZ). Two camera orientations were used, consisting of viewing perspectives into the bell and perpendicular to the axis of the bell. The smoke was illuminated by a Laser-glow LRS-0532 diode-pumped solid-state laser, providing continuous illumination of green light (532 nm) with an output power of approximately 6 W. A lens mounted on the laser aperture spread the beam into an expanding sheet of light approximately 1 mm in thickness. Images of the flow patterns were acquired in the illuminated plane with the sheet either oriented along the axis of the horn, such that the flow progressed within the plane of the sheet, or perpendicular to the axis of the horn, such that the flow passed through the sheet, as illustrated in Figure 2. Images were also acquired using white light produced by an LED lamp and softbox modifier; however, they were not used here because the contrast was reduced, and deemed less visually appealing and resulted in a weaker audio translation.

The result was the acquisition of somewhat complex flow patterns. The initiation of sound produced a burst of air from the bell in the form of a *starting jet* that is well-known to organize into a propagating vortex ring [2]. A laser sheet cutting through the axis of the vortex ring would render it as two counter-rotating vortices. A steady tone or sequence of notes would tend to produce a nominally more parallel flow, but highly susceptible to perturbations due to the musician's unsteady breath, intermittent separation and reattachment of the flow from the surface of the curved, expanding bell of the horn, and natural instabilities in the jet due to the presence of the straining field in the emerging wind. Due to the topological constraints of the vorticity field, which generally require the vortices to ultimately form closed loops [3], the evolution of perturbations typically resulted in reorganization of the flow into vortex rings of all sizes, such that the vortex pair pattern

is ubiquitous in the visualizations. Imaging with the laser sheet perpendicular to the horn axis resulted in much more erratic fluctuations in intensity as the smoke-marked flow structures passing through the sheet had only a very short residence time in the light.

2.2. Previous Sonification Work

This composition extends work developed by the first author in previous compositions where recordings are translated to a score and real time sonification of a dancer or an instrumental soloist is a compositional element. The translation of recordings into a score is described in *Textual and Sonic Feedback Loops: Simultaneous conversations as a collaborative process in cmetq* [4] and emphasizes the importance of the composer interacting with the translation algorithm by hand in the studio environment. In translating *Fluor Sonescence*, the primary data does not come from an audio recording, but rather video. The composer extends the minimal high level control of the translation procedure of earlier work to new software for *Fluor Sonescence*. This new software employs the same detailed attention and rehearsal of performative gestures before creating an ideal translation across domains. This practice of translation of recordings is also explored in *#chronicled: Two realtime concert reviews sonified* [5] where recordings from a local newspaper facility serve as source material for sonification. In this case the sonification results from the performance of typing actions on a cell phone.

The sonification of dancers also informs this new work. The dance sonifications began with the work *SoundLines* and uses lines of the video matrix as an audio waveforms while exploring different different means of triggering sound. To increase variety, the waveform is smoothed to create a different timbre and multiple lines are employed for polyphony. The composition *utuquq* uses this sonification of the live video feed as a component of a piano and electronics work where one hand is dedicated to playing the piano and the other to playing the kinect or video camera. In *In Vitro Oink* ([6] a Wii-Remote is used to sonify movement, a practice which is extended in *Psionic* where A Muse EEG Headband is used for movement data sonification and also synthesis based on the EEG signal data. This use of the movement vocabulary of both an instrumentalist and a dancer is an opportunity to develop a unique means of translating data. The performer and the composer work together to discover and map meaningful data. These data mappings explore both tightly coupled and less direct relationships between movement and sound. These various translations of movement and biometric data into sound inform the techniques and software used in the translations of *Fluor Sonescence*.

With *Fluor Sonescence* the sonification moves beyond the performer and ties the sonification directly to a physical aspect of the instrument. The intention here is similar to that of sonifying a performer, to compose with meaningful information that is inherited from the performance. With this composition we extend the aesthetic palette to include visual material that is the result of sound production, creating an extended perceptual experience from the existing data. The format of the presentation draws from the tradition of presenting silent film with live instrumental accompaniment. The black and white film and the smoke echoes the candle projectors of early film. This presentation blends aspects of the history of cinema with a minimalist electro-acoustic setup in the proscenium stage.

STAGES	RESULT	IMPLICATION
Sound production with brass instrument & smoke in air column	Smoke makes sound visible	The amplitude & note choice of player determines visual material
Laser illuminates a cross section & this is recorded to video	A particular plane of the sound is visible & serves as the representation for a 3d space	The camera settings determine what is visible (zoom level & frame-rate)
Video pixels are translated to pitch [trombone & electronics]	Common notation material & array of frequency and amplitude values	Outcome determined by: overtone series chosen, subdivision of visual space and translation rate
Raw material from translation is sculpted by hand	A composed score and a composed electronic score	Human intuition (aesthetic & logistical concerns) dominate this shaping process

Figure 3: The results and implications of the four stages of the compositional process.

3. COMPOSING

In his discussion of synchresis, Choin notes

”Let us note that in the cinema, causal listening is constantly manipulated by the audiovisual contract itself, especially through the phenomenon of synchresis. Most of the time we are dealing not with the real initial causes of the sounds, but causes that the film makes us believe in.” [1].

This highlights the primacy of visual stimuli and influenced the decision in creating *Fluor Sonescence* to begin with visual media for a sound work. The visual material serves as a starting point in order that it will define the types of motion in the sound part. The following will describe how the composition is organized, followed by an examination of the compositional process. Figure 3 provides an overview of this discussion drawing out the results and implications of this approach.

To compose this work, the first step was the creation of video material from the large collection of still images of the capture process. After concatenating the stills into and determining a suitable speed for playback of the gray-scale material it was important to balance the black and white across all video to insure similar contrast in all sections of material. With this basic visual processing in place the act of composing the sonic and visual flow of the piece was undertaken. Moving back and forth between the development of the sonification software and the processing of the video at this stage enabled the constraints of each process to inform the other. This workflow also enables the composer to shift the focal point of concentration between aesthetic and technical issues. The continuous shifting makes for an enjoyable working process and allows work on one aspect to advance while contemplation and reflection on another occurs.

The composition of the sonic aspects of *Fluor Sonescence* are derived from data in the video material. Greater detail about the technical aspects of each translation will be discussed in Sections 3.2, 3.3 and 3.4. In more general terms, the video is black, gray and white, where white smoke is introduced into a black frame. From the perspective of a computer vision system the greater the density of smoke results in a cluster of white pixels and whitish-gray pixels. In this way, the sonification calls attention to the boundaries between opposing forces in the distribution of energy, where there are ridges of relative low pressure between moving air masses.

The formal structure of the piece is articulated with changes to rate of playback in the video and the movement from one col-

lection of pitch material to the next. The pitch collections are held constant for each of the four sections while playback rate is constant within each video clip. Hence, on the phrase level, *Fluor Sonescence* is a series of different vignettes where the pacing and pitch emphasized by each clip is constant while there are changes across adjacent clips. The four sections of *Fluor Sonescence* are framed by three pauses which are articulated by silence and a black screen. The material within the four sections was first organized by intuitively grouping together video material and then organizing those video clips into a time line. This first arrangement of material is therefore primarily informed by visual concerns. The video was then used to generate pitch material for the trombone and fixed electronics.

Each of the four sections uses a different collection of pitch material. This pitch material is constructed by generating a geometrically expanding array of thirty values that results in a similar distribution of pitches below, within and above the trombone range. Consider the array A containing 30 frequency values, indexed from $i = 1$ to $i = 30$.

$$A = [78, 78 \times 1.145, 78 \times 1.145^2, \dots, 78 \times 1.145^{29}] \quad (1)$$

This particular array of thirty values is used in the first section of *Fluor Sonescence*. The spectrally rich, monophonic trombone is nested between a collection of sinetones articulating an irrational overtone series. Using the given pitch collection, the translation of the video to sound produces a section that has a particular distribution of sonic and visual-movement energy.

The video determines which tones are heard. Moments of fusion and separation occur within the electronics as well as across the trombone and electronics as a result of pitches coming in and out of audibility. Each video clip is joined to the resulting trombone and electronic material and then these audio/visual clips are placed into the initial time line. The organization of the clips as a result of purely visual flow, is then reviewed and altered so the section makes sense both visually and sonically. This technique is an intuitive collage process, where the material is considered for what it does in both domains and how that contributes to the emergent form. This strategy of organizing the material first with visual flow and then with visual and sonic flow is a deliberate step to counteract a personal bias for organizing a work with an emphasis on the sonic components. *Fluor Sonescence* is an audio visual composition and it was important to design a working process that would result in a balanced treatment of material based on my own personal bias for organizing sound first. Each new composition is an opportunity to both extend a line of inquiry (in this case the translation of image to sound) and to further develop my own abilities and sensitivities as a person who composes sound, image and movement in time.

3.1. Compositional Process

Using software that leverages earlier work discussed in Section 2.2, *Fluor Sonescence* is composed in three layers; the trombone part, the fixed electronics and live electronics. These components are used in conjunction and opposition to articulate the markers of formal structures and the aesthetic elements that define the work. The end result of this composition presents the audience with a video and sonic composition, so the workflow emphasized working with visual alone, sound alone and then the interplay of these two.

Fluor Sonescence is constructed in layers where structural phrases and elements unfold on multiple timescales, both within single layers and across multiple layers. This compositional approach is informed by the multiscale analysis of landscape described by Hay et al. [7]. They describe a multiscale framework in order to capture the complexity of processes and patterns observed across a landscape. Presenting multiple layers: trombone, fixed electronics, live electronics and video projection, *Fluor Sonescence* strives to present a rich perceptual environment of sonic and visual information for the audience. This work is not a literal translation of any one ecological system, but rather draws from the conceptual model. This model serves as an ideal during points of critical reflection in the compositional process. By unfolding processes both within and across multiple layers, gestures create a network of interrelated material that imitates the patterns of a natural landscape (or soundscape for that matter). The result of a rapidly advancing wave of smoke on the video screen can be heard in the dissipation of a sound mass into a deep reverberant space. Similarly, a series of rapidly ascending pitches in the trombone are interrupted by a sharply articulated electronic mass, which in turn is replaced with a sure and triumphant spectrally distorted trombone tone. It is through these series of interactions that *Fluor Sonescence* seeks to represent the deep web of interactions found in the natural world. As Curtis Roads notes,

Human beings respond intuitively to context-dependent cognitive impressions that are difficult to formalize, like wit, irony, tension, surprise, virtuosity, humor, and clever twists and transitions. [8]

It is hoped that the interplay of perceptual information presented in a multiscale framework within *Fluor Sonescence* gives rise to this range of cognitive impressions. During the composition of this work, this quote served as a touch point when considering the effectiveness of material.

3.2. Trombone

To create the trombone score, optical flow analysis of the video's center of mass is used to determine both pitch and amplitude of the notes. The *cv.jit.centroid* object implemented by Jean-Marc Pelletier [9] provides the points at which the sum of all pixel values are equal on both the top and bottom, and the left and right sides. The vertical centroid is used to determine which of eleven pitches are played by the trombone. The video, which is used to generate sonic material, presents the audience with black and white image of smoke (see Section 4). As described in Section 2.2 the video is created in order to visually represent the sound emanating from the bell of a brass instrument. Human attention is drawn to the whitest portion of a mostly black image which means the visual attention is focused on the apex of the white portion. As described in Section 3 this perceptual focal point is used to set the pitch and amplitude of the trombone part. As Grond and Hermann point out,

the tight relation between action and perception is important for ... engagement with the sound: [if] the sound is clearly anchored to a physical cause. [10]

In *Fluor Sonescence*, the visual focal point is tightly coupled to the moving line of solo instrument in order to create a correlation between physical movement in the video and sonic movement in the trombone.

The pitchwise translation of centroid to pitch polls the video space in order to determine when the centroid has moved. While

Figure 4: A still from the video with a red hatchmark indicating the position of the centroid.

an accurate translation necessitates a polling rate aligned with the framerate, the ultimate goal in this case is the production of a musical line that is both playable and aesthetically satisfying. To achieve a polling rate that excludes extraneous information while retaining relevant detail, the strategy of manual intervention was implemented. In creating the translations, the update rate is manually set by the mouse's X position. To produce the ideal take, the video clip was watched several times and various translations attempted. This enables the composer to both memorize the visual trajectory and attempt to best represent it. The multiple *rehearsals* enable the composer to hear what pitch material the video is resulting in and to practice until a suitable take or takes are achieved. Working in short segments like this also affords the opportunity to emphasize different rhythmic or pitch figures as well as an opportunity to study the relationship between the visual and sonic material. The end result is a simple compositional tool that enables the composer to both study and craft material in a tightly coupled feedback loop between action and result. This simple tool provides an intuitive gestural means for sculpting material.

3.3. Fixed Electronics

Where the trombone is generated from a single centroid in the Y plane, the fixed electronics uses blob detection to identify multiple areas of intensity. This approach leverages the multiplicity of visual objects in the video matrix as material for control of individual components of a soundmass. Like the trombone, the vertical position of a blob determines which of the available sinetones is activated and the size of the blob detected determines the amplitude of the sinetone. In some cases, the threshold for detection of similar blob quantities is used as a criteria in determining which section a clip fits in.

To create distinct characters the timbre of the four sections is unique. In section one and three pure sine tones are the base element, while two and four use impulse trains with narrow band resonators to achieve percussively varied fixed tones. The four sections of *Fluor Sonescence* have unique sonic characters which are the result of the timbre of the sonified data, different impulse responses for convolution reverb and unique live processing of the

trombone.

In Section 3 the calculations used to derive the pitch material for each of the four sections is discussed. The range of the trombone constrains which pitches are used for the instrumental part and are employed in the fixed electronics. The electronics are an undulating cluster of pure tones that fuse into a single sonic mass. The perception of this mass varies as the relative amplitudes change and as the trombone plays different pitches. The combination of sonic elements into a mass is motivated by considering the myriad of droplets and the currents of air which combine these drops to form clouds.

The presence of a solo performer with a spectrally rich monophonic instrument stands in contrast to the additive synthesis of the fixed electronics. The fixed electronics provide a billowing mass of sound information for the listener reflecting the shifts and movements of the source imagery. This compositional choice reflects the perceptual role the soundscape (fixed electronics), in contrast to sound objects (trombone), plays in the perception of the listener. As Walk and Nees note,

Although the soundscape may not require a particular response at any given time, it provides ongoing information about a situation to the listener. [11]

The trombone's solo line traces the most visually vibrant component of the video material while the fixed electronics extends and surrounds this apex of perceptual information with an undulating resonance of the larger scene.

3.4. Live Electronics

Fluor Sonescence is a composition created in the studio environment where the composer has the ability to rework and craft every microsecond of the fixed elements of the piece. This is balanced with material that is presented live in each performance. The trombone sound production and live processing of the trombone sound must leverage the abilities of the performer to manage sound in real time in reaction to the acoustics of the space and the atmosphere of the performance/audience. The interface for managing the performance is the trombone and the goal of the composer is to create a score that provides ample information and is suitably open to interpretation.

Fluor Sonescence provides the performer with a fixed notation score and a maxpatch that includes a levelmeter, timers that indicate the bar, beat and the amount of elapsed time. The trombonist is able to perform the trombone part and blend with the electronics the way they would perform with another instrumentalist. In order to extend the sound of the trombone and create sonic continuity across layers, the trombone sound is presented in the speakers in addition to live processing. The live processing algorithms are triggered by the elapsed time counter. From performance to performance, the settings do not change, providing a predictable interaction for the performer. The proximity of the trombone to the microphone and the micro-adjustments of pitch and amplitude of the performer serve to shape the sound of not only the trombone in performance but also the live electronics. In this way the electronic components of the live material, which are close in timbre at times to the fixed electronics, become imbued with elements of the live performance. Because *Fluor Sonescence* blends multiple layers together, the electronics inherit aspects of this performative liveness.

The live processing uses a range of means to manipulate the sound of the trombone with filters, spectral manipulation, feed-

back, waveshaping and granulation. The audio is used both as a control signal and as audio stream in the live processing. The control envelopes for these algorithms are both composed and derived from live amplitude data. Using amplitude as control information echoes the translations of the composition process. This serves to obscure the distinction between live and fixed elements and enables the performer to influence the shape of the piece through the trombone as interface. Indeed, the trombonist does not need to be concerned with this process, but rather the composer must be concerned with making certain the mapping of the data is scaled in a manner that produces material that is suitable to the aesthetic of the piece and that is perceptually salient.

4. CONCLUSION

Fluor Sonescence is an interesting point in the formalization of a sonification approach. This piece continues a line of inquiry and looks forward to a new method of generating material as well as an audio visual controller. The piece is an opportunity to deeply consider the propagation of sound from a brass instrument. Often sonification and visualizations are used to present a large amount of data. In the case of *Fluor Sonescence* the compositional process is an opportunity to consider that data and explore it as a result of the images and sounds. This compositional process and the investigation of known patterns of air mass flow has reinforced a commitment to an aesthetic focused on considering the richness of subtle motions. The audience feedback on *Fluor Sonescence* has been very positive and many have reported that they sense both continuity and friction across the visual and sonic domains. This project will move forward with a further formalization of the sonification strategies and the development of a mobile instrument based on this research.

5. REFERENCES

- [1] M. Chion, *Audio - Vision, Sound on Screen*. London, UK: Columbia University Press, 1990.
- [2] M. Gharib, E. Rambod, and K. Shariff, "A universal time scale for vortex ring formation," *JFLUM*, vol. 360, pp. 121–140, 1998.
- [3] H. Lamb, *Hydrodynamics*. Cambridge, UK: Cambridge University Press, 1932.
- [4] C. Jette and N. Krueger, "Textual and sonic feedback loops: Simultaneous conversations as a collaborative process in cmetq," in *Proceedings of the 2016 International Computer Music Conference*, Utrecht NE, 2016.
- [5] C. Jette and K. Kirchoff, "#chronicled: Two realtime concert reviews sonified," in *Proceedings of the 2014 Ammerman Center Arts and Technology Symposium*, 2014.
- [6] C. Jette, "In vitro oink: A composition for piano and wiiremote," *e Contact!*, no. 13.2, April 2011.
- [7] D. . P. Hay G.J., Marceau D.J. and B. A., "A multiscale framework for landscape analysis: object-specific analysis and up-scaling," *Landscape Ecology*, vol. 16, p. 471490, 2001.
- [8] C. Roads, "From grains to forms," in *Proceedings of the international Symposium Xenakis. La musique electroacoustique / Xenakis. The electroacoustic music*, M. Solomos, Ed., universite Paris 8, 2012.
- [9] [<http://jmpelletier.com/cvjit/>].
- [10] F. Grond and T. Hermann, "Aesthetic strategies in sonification," *AI and Society*, vol. 27, p. 214, 2012.
- [11] B. N. Walker and M. A. Nees, "Theory of sonification," in *The Sonification Handbook*, J. G. N. Thomas Herman, Andy Hunt, Ed. Logos Verlag Berlin GmbH, 2011.

HYUNDAI MOTORS SPECIAL SESSION

RECOGNIZABILITY AND PERCEIVED URGENCY OF BICYCLE BELLS

Lisa Frohmann, Marian Weger, Robert Höldrich

Institute for Electronic Music and Acoustics (IEM)
University of Music and Performing Arts
Graz, Austria
li.frohmann@gmail.com

ABSTRACT

Raising awareness about how alarm sounds are perceived and evaluated by an individual in traffic scenery is important for developing new alarm designs, as well as for improving existing ones. Bearing a positive contribution to road safety, cyclists and pedestrians especially can benefit from appropriate alarming bell and horn sounds. Primarily, the alarm signal should evoke a precise idea of what is the source of the warning and the desired reaction to it. Furthermore, it should not be masked by other noises thus going undetected by the ear. Finally, an appropriate warning signal should transmit the urgency of a given situation, while at the same time, it should not cause other road users and pedestrians to startle.

In two listening experiments, we examined the perception of commonly available bicycle bells and horns. Average typicality or recognizability as a bicycle bell among other everyday sounds has been investigated through a free identification task. In a second experiment, we tested perceived urgency of the warning sounds in relation to traffic noise. This article further provides a survey on non-verbal alarm design, as well as an analysis of acoustic properties of common bicycle bells and horns. Consequently, a linear regression model presents the relationship between named properties and perceived urgency.

It is our intention to give an insight into the often unattended but important issue of the perception of auditory warning sounds in our everyday acoustic environment.

1. INTRODUCTION

Acoustic warning signals are part of our everyday lives, designed to make hazardous environments and situations safer by giving accurate information about an expected risk at the appropriate time. We are confronted with such alarm signals in various situations - from road traffic, aviation monitoring and countless industries, to medical equipment in the hospital. Whether we hear the honk of a car horn or the constant beeping of an electrocardiogram monitor (ECG), all these sounds have a commonality - they bear an important message for their target and they manage to draw attention to an incident in a preferably short period of time.

In contrast to warnings using speech, non-verbal auditory alarms lack the feature of precise expression of what they are informing about and what the desired reaction to this alarm should

look like. Therefore, it is important to take into account temporal as well as spectral signal properties in order to design the best possible warning sound for a specific scenario. Originally established for alarm systems on military aircraft [1], several acoustic properties form criteria for adequate acoustic warning that suit all application areas. As proposed, warning sounds...

- must be unique in the noise environment,
- must be discriminable from other sounds of the surrounding,
- should convey the correct relative urgency for the associated priority level,
- should be presented at correct audio level for reliable detection.

Edworthy et al. [2] emphasize on the term of *Urgency* which makes a ranking of warning sounds possible. Its understanding is crucial for the success of alarm design:

"[...] there is often a serious mismatch between the perceived (psychoacoustic) urgency of a warning -its implicit urgency as a function of its sound parameters- and its situational urgency -the degree of urgency that the operator (e.g. pilot, nurse) has learned to associate with the warning as a function of the situation itself." [2]

The authors studied the effects of different signal parameters on perceived urgency of warning sounds. These parameters were based on an approach by Patterson [3] who constructed sound-bursts from impulse-trains which were modified in pitch, tempo, intensity, melody, and rhythm. Participants of a listening experiment then rated and ranked the perceived urgency of these synthetic alarms under laboratory conditions, leading to the following conclusions: The higher the pitch, the more irregular the harmonics, and the faster the pulse rate, the greater the perceived urgency [2].

These findings, however, could not be approved clearly for real, already existing alarms and for subjects being under high workload when alarmed. On the one hand, there seems to be a connection of acoustic characteristics and urgency on a low level. On the other hand, there has to be a higher level influence of auditory learning and acculturation. The perception of urgency (and therefore potential danger) is linked to a mental representation of the cause [4]. Without existing knowledge or education around the meaning of any specific alarm, it is difficult to find an appropriate reaction to it and the alarm may have a startling effect and lead to confusion.

A lot of research concerning intelligent alarming has been done in the fields of medical care, as well as aviation and automotive industry, e.g., [5, 6, 7, 8]. New car horn designs and electric

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

vehicle noises that match the road users' needs and the designers' intentions concerning branding in equal shares, have been explored [9]. As modern cars are built to be as soundproof as possible, new problems arise, and solutions have to be found to make ambulance alarms and police sirens audible again. Inventions such as additional low-frequency sirens [10], that penetrate the hard surfaces of the car body, are dealing with the situation. Bringing the warning directly inside the car via Radio Data System (RDS) yields another approach [11, 12].

Other road users such as cyclists and pedestrians are unjustly excluded in these developments. Alarm instruments that warn cyclists, as well as sounds that cyclists use for signaling others, are an important safety feature compulsory by law in most countries. However, these mandatory safety tools are poorly developed, unused and misinterpreted on a daily basis. The sound of a bicycle bell is often associated with a low level of threat or even with pleasure, which certainly interferes with the effectiveness of the warning [13].

In this article, we present two listening experiments in which commonly available bicycle bells and horns have been examined on the basis of the acquired knowledge about properties and design principles of auditory warning sounds. As the mental representation of an alarm's cause is of great importance to the reaction to it, a classification test was performed in order to reveal if the tested bicycle warnings could be identified as such among a set of everyday sounds (Sec. 2). Furthermore, the level of urgency that conventional bicycle bells carry to their target is of interest, as well as if an urgency ranking of different bell types can be obtained. Therefore, a second listening experiment on perceived urgency of bicycle warning sounds has been carried out (Sec. 3). Spectral and temporal properties of the bell sounds (perceived pitch, spectral centroid, roughness, onset time and relative intensity) are examined in Sec. 4 and set in relation to the measured urgency values, yielding an objective model formed by linear regression (Sec. 5). A general discussion and conclusion finally follows in Sec. 6.

2. EXPERIMENT 1: RECOGNIZABILITY TESTING

The aim of this listening experiment was to give information about how well the traditional bike bells as well as state-of-the-art bicycle warning tools were identified and accepted as such.

2.1. Procedure and Apparatus

Subjects. 17 subjects participated in Experiment 1: 5 females and 12 males with an average age of 24.5 years (ranging from 21 to 29). All of them were students in Audio Engineering at Graz University of Technology and Graz University of Music and Performing Arts. Nobody reported a hearing difficulty, and 15 of 17 had experience with listening experiments.

Apparatus. The experiment took place in an acoustically treated room of 4.3 m × 6.2 m × 3.4 m size (w × l × h), with reverberation times between 0.15 and 0.22 s in the relevant frequency range above 200 Hz. The technical setup included a Genelec 8020 CPM studio monitor connected to a Samsung 700Z laptop through an RME Fireface UCX audio interface. The stimuli were played back under Windows with Cubase 8 software. Subjects were seated in the center of the room with the loudspeaker located in a distance of 1.76 m in line of sight.

Procedure. The subjects were asked to freely identify 20 everyday sounds which they could encounter on the streets of a city.

Before starting the experiment, subjects were informed vaguely about this imagined scenario without revealing the exact stimuli. Each stimulus was presented twice, with a short pause in between. Then, before continuing with the next stimulus, there was a pause of about 5 seconds for writing down the answer. Subjects were not allowed to pause or stop the experiment, as their first impression and classification was desired. The order of the stimuli was randomized between subjects in order to avoid a possible influence of order.

Stimuli. The 20 stimuli were selected to include only 10 actual bicycle warning sounds. These common bell stimuli included two large (L.) and two medium sized bells that have to be flicked horizontally, i.e., twisted (TB), three bells with a hammer stroke mechanism (H), a bike horn, and the electric bicycle bell *Hornit* by Bullet Ventures Limited [14], which offers two sounds: a loud piercing siren for street use and a horn-like sound for more quiet settings such as cycling through a park. The other half of the stimuli consisted of 8 acoustic warning signals (2 car horns, 2 acoustic traffic lights for the blind, the ringing of a mobile phone, a tramway bell, a doorbell, and a whistle), and 2 miscellaneous sounds (the rattling sound of a keyring and a wind chimes sound). The latter sounds were included in the hope of making the experiment more interesting and less obvious to the participants. The stimuli had a duration of 3 to 6 seconds.

2.2. Results

Table 1 shows the answer matrix for the classification test in a reduced version. Sounds that were correctly identified by most of the subjects (e.g., tramway signal, rattling of a key ring, etc.) were excluded from the further evaluation as they were not the target of our examination but were included only for experiment design reasons.

The individual answers were grouped under umbrella terms. From these results, the recognizability as a bike bell was calculated for each stimulus as the percentage of answers which included both 'bicycle' and 'bell/horn'. The resulting values of bicycle typicality are shown in Tab. 1. 95 % confidence intervals are added, assuming a binomial distribution.

A hierarchical clustering analysis of the classification results revealed three main clusters: Cluster A consists of both large 'TB' type bells. Cluster B contains the small 'TB' bells as well as the 'H' type bells. Cluster C combines the bicycle horn and the two *Hornit* sounds. For further analysis we additionally split Clusters B and C in order to achieve five clusters which can even be formed instinctively by the physical mechanism of sound generation and thus typical auditory gestalt: (1) 'Ding-dong', (2) 'Ding', (3) 'Rrring', (4) 'Honk' and (5) 'Electronic'.

2.3. Discussion

The results of Experiment 1, investigating recognition as bicycle and clusterization, matched our assumptions made beforehand. Well known, most common and purchasable bell types such as hammer bells and small to medium sized twist bells with a rattling sound reached high bicycle typicality values. Larger twist bell models gained average to low scores, and the new *Hornit* bell tool did not create a mental connection of a bicycle for neither of the subjects. The low values of bells of group (1) (cluster B) could be explained by their resemblance to a doorbell and therefore by assignment to a different setting and cause. Tab. 1 displays this

group	Stimulus	Classification category										Bicycle typicality in % CI (95 %)		
		bicycle bell/horn	car horn	mobile phone	general horn	doorbell	bird chirp	general bell (small)	general bell (large)	toy sound	alarm/siren			other
1	TB l. yellow	5				10			2				29	10-55
	TB l. pink	8.5				6.5		1	1				50	27-77
2	H lilac	13						3				1	76	50-93
	H black 1	12						1	1			3	71	44-89
	H black 2	7		1				3	2			4	41	18-67
3	TB Dinotti	16						1					94	71-99
	TB Puch	16						1					94	71-99
4	Bicycle horn	7			9					1			41	18-67
5	Hornit Park		2		7					3	3	2	0	0-19
	Hornit Street				1		5			7	2	2	0	0-19
-	Doorbell					10		1	5		1		0	0-19

Table 1: Classification of stimuli in reduced form and calculated bicycle typicality with confidence interval (CI).

mismatch. Yet interestingly the original doorbell sound was never inversely mistaken for a bike bell. The bike horn with a recognizability of 41 % was often classified either as a typical children's bike horn or a common horn with no relation to a bicycle at all. This supports our personal impression that such bike horns are rare on adult bicycles, at least in the cultural environment of the tested participants.

The grouping of bell types will be used in Experiment 2 for the purpose of comparing types rather than individual bells. Considering the fact that the small groups (1) to (5) can be obtained by further division of clusters A and C, we will stick to the five small groups, even if they are not completely arguable perception-wise through cluster analysis. Most importantly, however, they give an adequate representation of bell types available on the market.

3. EXPERIMENT 2: URGENCY

3.1. Procedure and Apparatus

Subjects. The 17 subjects to participate in Experiment 2 were the same as in Experiment 1.

Apparatus. The experiment was conducted in the same room as the first experiment. The technical setup comprised four Genelec 8020 CPM studio monitors in the corners of a 3.53 m × 3.53 m square, connected to a Samsung 700Z laptop through an RME Fireface UCX audio interface. The experimentation software was implemented in Matlab, with audio playback through Pure Data (Pd). Subjects were seated in the center of the square, with the speakers in a distance of 2.45 m to the center of the head, in 1.2 m height, in azimuth angles of 45°, 135°, 225°, and 315°. Calibration was done with an NTI Audio M2210 measurement microphone and Brüel&Kjaer acoustical calibrator Type 4231.

Stimuli. Subjects were presented 12 bicycle bell and horn sounds.

10 of these were the same as in Experiment 1. The two additional ones were slightly modified versions of Stimuli 'TB large pink' and 'H black 2'. In particular, the spectrum of 'TB large pink' was altered by boosting the gain at 2.5 kHz, 7 kHz and 11.8 kHz by 12 dB in order to create a more piercing sound. Stimulus 'H black 2' was reduced in tempo, equivalent to a stimulus duration stretch from 3 to 6 seconds.

For the condition with background soundscape, different field-recordings of environmental noise at typical street crossings in the city of Graz were layered to form an indefinite and blurry version of a typical urban acoustic environment. The traffic noise was played back quadrophonic with all four loudspeakers; the sound pressure level measured at the listening position was set to match the one at an average street crossing.

In order to ensure an ecologically valid presentation of the stimuli, their sound pressure levels were adjusted to match the original ones as close as possible. In a first step, the sound pressure levels of two prototype bell stimuli – a traditional one and the horn – were calibrated to match the real ones at the loudspeaker position by comparing the levels measured with the measurement microphone at approximate head position. All other stimuli were then subjectively adjusted to match the amplitudes of these prototype bells.

Procedure. Experiment 2 was divided into two multi-stimuli tests. For part one, subjects had to blindly rate the urgency of all the stimuli over a Matlab-based GUI from 'not urgent' to 'very urgent' using horizontal sliders without scaling. The direction of sound playback differed pseudo-randomly over the four loudspeakers. Subjects were able to listen to all stimuli in their preferred order, enabling them to compare stimuli pairwise and also sort their answers. In part two, the urgency rating experiment of part one was repeated with the additional traffic soundscape, simulating real life conditions. As experiments excluding the acoustic environment of the warning signal have already been carried out [2], we were interested in testing both the noisy and the silent scenario for the bike bells to detect differences in urgency ratings caused by the context of traffic noise.

3.2. Results

Figures 1 and 2 show boxplot diagrams for the urgency ratings of the 12 individual bell stimuli in the two different conditions without and with background noise, with the mean values (bold black line) displayed additionally to the median (narrow line).

We proceed with a statistical analysis of these results. In both conditions, some outliers were identified by Grubbs' test and thus replaced by the mean value of the affected stimulus (3 stimuli of 1 participant in part 1, 2 stimuli of 2 participants in part 2). Liliefors test could not reject the null hypothesis of normally distributed data at a significance level of $p < 0.05$ for the majority of the data (25 % non-normal for stimuli without background noise, 8 % non-normal for stimuli with background noise). We therefore performed a two-way (Stimulus × Noise) analysis of variance (ANOVA).

ANOVA showed a highly significant main effect of stimulus type on urgency rating at the $p < 0.05$ level ($F(11,176)=15.69$, $p < 0.001$). This indicates that different bell types incorporate significantly different perceived urgency. There was also a significant main effect of traffic noise on urgency rating ($F(1,16)=4.50$, $p=0.034$). Furthermore, the interaction between stimulus and traffic noise was highly significant ($F(11,176)=3.13$, $p < 0.001$).

Figure 1: Urgency rating without background traffic noise. Alternating background filling indicates membership to the bell groups of Fig. 3.

In order to compare the individual stimuli's urgency values to each other, pairwise t-tests were performed. The differences between stimuli scores were examined averaged over background noise conditions. The Hornit type bells were not significantly different to each other, while they were significantly more urgent than all other bells ($p \leq 0.007$), except for the bicycle horn which did not differ significantly from Hornit Park ($t(16) = -1.808$, $p = 0.089$), but showed significantly lower urgency compared to the siren-like Hornit Street ($t(16) = -2.454$, $p = 0.026$).

The bicycle horn was rated significantly more urgent than the large TB bells ($p \leq 0.002$), as well as significantly more urgent than the hammer type bells ($p \leq 0.035$). However, the difference in urgency to the TB type bells of rattling sound character was not significant. Furthermore, the three large TB type stimuli were not significantly different to each other. All of these, however, were significantly less urgent than the clattering 'TB Puch' and horn stimuli ($p \leq 0.005$). Pairwise comparison between the two medium sized rattling twist bells revealed a significant disparity in urgency, meaning 'TB Puch' was significantly more urgent than 'TB Dinotti' ($t(16) = 2.312$, $p = 0.034$).

Inside the group of 'H' type bells, 'H black1' and 'H black2 long' were both significantly more urgent than the other two stimuli ($p \leq 0.037$), while other pairwise comparisons were not significant. Both 'H lilac' and 'H black2' were also significantly less urgent than the two rattling 'TB' bells ($p \leq 0.009$). In addition, 'TB large pink' was significantly more urgent than 'H lilac' ($t(16) = 2.317$, $p = 0.034$), and 'TB Puch' was significantly more urgent than 'H black1' and 'H black2 long' (≤ 0.027). All combinations that have not been named so far, were not significant concerning perceived urgency.

With concern to the significant effect of background noise on perceived urgency, a pairwise t-test showed that average urgency scores were significantly lower in a noisy scenario (mean=0.47) than without noise (mean=0.51) ($t(16) = 2.343$, $p = 0.032$).

In a next step, we also performed a statistical analysis of the pooled results in the five groups as derived in Experiment 1. Box-plots for both background noise environments are shown in Fig. 3.

Figure 2: Urgency rating with background traffic noise. Alternating background filling indicates membership to the bell groups of Fig. 3.

No outliers were detected by Grubbs' test for neither of the traffic noise conditions. Lilliefors test, however, yields 60 % non-normal data in quiet surrounding, but 0 % non-normal data in the noisy setup. We will therefore rely on the two-sided Wilcoxon rank-sum test in the following statistical analysis.

Figure 3: Urgency ratings of the bell type groups, without (a) and with (b) background traffic noise.

The noise condition had a significant effect only for stimulus group (5): The Hornit type stimuli were significantly more urgent without (mean=0.789) than with background noise (mean=0.654) ($p = 0.036$). Statistical analysis of pairwise rank-sum tests, averaged over noise condition was performed between the respective stimuli groups and can be summarized as follows.

Groups (4) and (5) did not show a significant difference in ur-

gency rating ($p=0.103$), but both were perceived significantly more urgent than all other groups ((3) vs. (4): $p=0.026$, all other combinations: $p\leq 0.001$). Similarly, groups (1) and (2) were not significantly different concerning perceived urgency ($p=9.912$); however, both were perceived significantly less urgent than all other groups ((1) vs. (3): $p=0.007$, all other combinations: $p\leq 0.001$).

The interaction between stimulus group and noise condition could be explained through the following observation: According to Wilcoxon rank-sum test, in the noisy condition, group (1) was significantly more urgent than group (2) ($p=0.039$, with means 0.435 and 0.351, respectively). In contrast to this, it was yet the other way around without noise (means 0.374 and 0.419). However, the latter observation was not significant ($p=0.191$). All other pairwise comparisons maintained the sign of their urgency difference throughout both noise conditions.

3.3. Discussion

Interpreting the obtained results, we can state that traffic noise is responsible for a drop of the average rating of urgency. One explanation could be that the additional traffic noise may generate a level of stress for the subjects and takes away the focus from the bell stimuli to other potentially dangerous sound sources such as cars – the bicycle bell sounds are perhaps put in relation to the possibly more urgent traffic noise. Somehow connected to this assumption, possibly more important information gained from traffic noise, might induce informational masking effects on the examined bell sounds. Another reason could be masking of spectral components of the bell sounds by the traffic noise, thus reducing their perceived urgency.

However, the noise condition affected urgency significantly only for the Hornit sounds which were the overall winners of the urgency rating experiment. Interestingly, without being asked, the majority of the subjects reported confusion, when having to rate the Hornit sounds (especially the Park mode, designed by the manufacturer for calm environments, hence aiming at lower urgency). Despite giving high values in the no-noise test, participants did not sense an urge to react for the same stimulus in noisy environment, stating that it sounded too far away to be of concern.

The ‘Rrring’ type bells of group (3) seem to form a stable class of average urgency, as the difference between noise conditions is rather small. We can see a similar picture for the bike horn, but with a greater spreading of subject ratings when having to judge urgency in background noise. That might be because the horn is often associated with a toy sound or a children’s bike bell, as well as because in the noisy setting, direction and source of the sound are not easily assigned.

The hammer bells of group (2) got lowest ratings in noisy environment, which seems logical as these consist of one pulse only and appear weak and insufficient for the environment. The traffic noise seems to absorb important features of the sound, which are available to the ear in silent surroundings, with the effect of lowering perceived urgency when these acoustic cues are less prominent due to spectral masking.

4. ACOUSTICAL DESCRIPTORS OF URGENCY

A spectral and temporal analysis of the given bell sounds in relation to the urgency estimation aimed to explore possible correlations. Given that the bell types incorporated a high grade of variation in sonic characteristics, adequate signal analysis methods are

desired in order to predict the urgency of the recorded bicycle bells from acoustical descriptors. We decided for five common signal attributes: perceived pitch, spectral centroid, mean relative roughness, onset time, and relative signal power. These attributes can be easily calculated or at least roughly estimated for any recorded bell sound with existing models and methods, illustrating both temporal (e.g., onset) and spectral (e.g., pitch, centroid) components which relate to perceived urgency.

4.1. Perceived Pitch

Estimating a precise value for perceived pitch of a bell sound or a burst of sound containing numerous frequencies is problematic. For many musical instruments such as cordophones or aerophones, it is easy to identify a fundamental frequency, but for said bells, chimes or percussive sounds, i.e., idiophones, the situation of finding pitch is remarkably unclear [15]. Evaluation of the spectrum does not always lead to the right solution, as the perceived pitch can be influenced by difference tones and virtual pitch suggested by the human brain. Despite these problems, there is no doubt that one instantly gets some impression of pitch, when hearing a bell sound, even if this pitch differs between individuals. A higher fundamental frequency does affect the sensing of urgency, although studies like [2] (and others cited therein) describing it as a minor influence.

To explore the case for bicycle bells, perceived pitch was measured by means of a frequency-matching experiment performed by the investigator, i.e. through adjustment of a pure sine wave to match the perceived pitch of the individual bell stimulus. For sounds that consist of two or more consecutive or simultaneous auditory events, the value refers either to the first event or to the overall impression on the sound respectively. As this is a very subjective approach, readers are warned that perceived pitch might differ for a larger pool of subjects conducting such a frequency-matching experiment. Perceived pitch is displayed in Tab. 2.

4.2. Spectral centroid

The spectral centroid is an indicator of the spectrum’s “center of mass”, its location in frequency influencing the perception of a sound’s brightness. Incorporating a significant amount of signal energy at high frequencies, we assume that a bright sound with a centroid in the human ear’s most sensible frequency range from about 2 to 5 kHz could favor alertness and induce higher perceived urgency.

The spectral centroid is generally calculated – following Eq. 1 – as the weighted mean of the frequency bins of a signal obtained by Fourier Transform (with $\frac{k f_s}{N}$ being the frequency of bin k), weighted with the respective magnitudes $|X(k)|$ for each bin. For this experiment’s computation, the bell sounds were additionally weighted with an A-weighting-filter before spectral density estimation was performed using Welch’s method (`pwelch` in Matlab). Spectral centroid values are displayed in Tab. 2.

$$Centroid = \frac{\sum_{k=0}^{\frac{N}{2}-1} \frac{k f_s}{N} |X(k)|}{\sum_{k=0}^{\frac{N}{2}-1} |X(k)|} \quad (1)$$

Stimulus	urgency rating		signal parameter				
	with noise	without noise	perceived pitch kHz	spectral centroid kHz	mean roughness	onset time ms	signal power relative to loudest Δ dB(A)
Hornit Street	0.72	0.78	1.28	2.95	6.53	82	-1.4
TB Puch	0.61	0.51	0.91	6.41	0.89	21	-7.5
Bicycle horn green	0.60	0.65	0.49	3.14	0.36	15	-3.3
Hornit Park	0.59	0.90	0.56	2.70	1.12	71	0.0
H black 1	0.52	0.42	1.67	4.23	0.01	2	-10.3
TB large pink	0.45	0.39	0.60	4.75	0.07	5	-6.7
TB Dinotti	0.43	0.50	1.04	8.06	0.27	24	-13.1
TB large yellow	0.43	0.36	0.59	3.61	0.12	3	-5.6
TB large pink EQ	0.43	0.37	0.59	3.96	0.20	6	-6.2
H black 2 long	0.36	0.48	1.34	8.87	0.00	17	-13.4
H lilac	0.26	0.32	1.58	3.38	0.01	2	-13.8
H black 2	0.20	0.45	1.34	10.19	0.01	3	-13.6

Table 2: Signal parameters and urgency rating. Stimuli are sorted by urgency rating with background noise in descending order.

4.3. Roughness

The property of roughness, as an effect of interaction of sinusoids close in frequency, is selected due to its potential of causing the “unpleasantness” to the ear. It might correlate to findings in [2], that inharmonic partials result in higher urgency ratings. In addition, most people urge to stop an annoying sound when it becomes too irritating, which could lead to a shorter reaction time for a high-roughness warning.

The roughness or sensory dissonance of A-weighted bell sounds was computed in Matlab using the `mirroughness` function of the *MIRtoolbox* [16]. This particular algorithm was implemented based on proposals of Plomp and Levelt for an estimation of roughness related to the beating phenomenon of sinusoids close in frequency to each other and Sethares’ model of sensory dissonance based on their work. Latter is obtained by computing the spectral peaks of a sound and taking into account all the dissonances between all possible pairs of peaks, as cited in [16] and the *MIRtoolbox* manuals. For further information about the model, please consult Appendix G of [15]. As the computed roughness of the stimuli is a function of time, the mean roughness of the bell sounds is calculated in a last step (see Tab. 2).

4.4. Onset time

Onset time could also be a relevant factor affecting urgency perception. Experiments carried out by [2] and [4] led to the conclusion that a slow onset is considered less urgent than the stated standard onset of 20 ms for synthesized alarm sequences. One explanation could be that a fast onset implies a quick approach of a moving sound source towards the test subject with a certain intention, contrary to a slow onset probably not bearing the same level of exigence. Furthermore, the output over the total duration is obviously higher for such short-onset stimuli [2].

In the case of our complex stimuli, onset time (or attack time) was defined as the time it takes for a sound’s envelope to rise from zero to the first maximum on the time axis. The temporal pro-

file of each bell sound was plotted and onset time was measured by hand for each stimulus individually by the investigator. For sounds consisting of two or more consecutive pulses, onset time was measured only for their first pulse accordingly. The measured onset times (in ms) are displayed in Tab. 2.

4.5. Signal power

As mentioned before, stimulus intensity was not the focus of our investigation and we therefore tried to neglect this factor, while it is undoubtedly one of the strongest known factors having an impact on urgency, as stated by [6] and others cited therein. Relative balancing of the sounds was done by ear by two investigators for said reason. Despite that, the actual signal power reaching the subjects ears at the center of the experimental chamber was examined. Regardless of all previous arrangements, signal power could have had an influence on the subjects’ rating preferences. Calculations of signal power were performed on the A-weighted bell sounds. Signal power Δ dB(A), relative to the loudest stimulus ‘Hornit Park’, is displayed in Tab. 2.

4.6. Discussion

Evaluating the calculations of the sound properties introduced in the previous sections aimed at finding simple parameters for any type of bicycle bell stimulus, that could render prediction of urgency ranking possible or at least give a hint on what could be a valid factor affecting urgency besides sound intensity.

For the bell sounds’ perceived pitch, no reasonable connection to the urgency values could be found at first sight. Bells within one group – which incorporate similar mechanisms of sound production and bell sizes – possess similar perceived “fundamental frequencies”. In between the bells of a bell group, there seems to be a small positive effect, but there is no overall relationship for the individual stimuli. Experiments in [2] did find some minor relation of fundamental frequency and urgency, though it does not seem to play a big role in the perception and rating process, especially if

other salient spectral and temporal attributes are determined. The presence of other strong signal characteristics might be the reason or at least an influencing factor to our results.

As stated in Sec. 4.2, a higher spectral centroid implies higher brightness of a sound and might support the perception of a stimulus' message in terms of urgency. However, for bells with a spectral centroid above 6 kHz, urgency seems to decrease in both noise conditions. Below this line though, there is a great spread in urgency ratings. The Hornit bell sounds and the bike horn, with spectral centroids around 3 kHz, achieved far higher urgency scores than the large twist bells and two of the hammer bells with centroids in the range of about 3.3 to 5 kHz. Our assumption lacks an explanation for said low values, hence it is not possible to draw a conclusion for urgency behavior from the spectral centroid analysis only.

A high value of roughness of a bell sound corresponds clearly to a high urgency rating for – in descending roughness order for traffic noise condition – the Hornit sounds, TB Puch, the bike horn and TB Dinotti (both twist bells are of group (3) with a rattling character to the sound). In this observation, hammer bells and large twist bells indicate very small roughness values compared to the other groups. Interestingly, without traffic noise, hammer bells appear slightly more urgent than the large twist bells, despite holding the lowest average roughness scores. Subjects confusing the large bike bell with a doorbell sound might be a simple explanation for this order of the stimulus groups. In a noisy environmental context, however, the two named groups switch places, leaving most hammer bells on the bottom of the roughness scale.

The results for onset time obtained from our experiments could not support the findings mentioned in Sec. 4.4. Slow onsets for both Hornit bell sounds did not correlate with low, but with extremely high urgency scores (1st and 4th place). The rattling TB Puch with an onset close to the standard value of 20 ms even reached the second highest rating on the scale. The bike horn with a faster onset gained 0.6 in traffic soundscape, while all other bells with the fastest onsets around 5 ms are piled on the bottom of the scale. It seems that other properties than the temporal approach of onset time are responsible for the rank obtained by the individual stimuli's ratings.

Finally, signal power estimation was related to the urgency ratings. There is a tendency for bell sounds of high signal power to be considered more urgent than ones of low power values. The large bell sounds again are the exception to this finding in the setting without background noise. Despite having higher power estimates than the group of hammer bells, urgency ratings appear too low as to fit the assumptions. Again, the mismatch of mental representation of sound source could be a reason for this, similar to the observations concerning roughness. For urgency ratings with background noise, there is a better fit of the large TBs in the power-urgency relation. Despite some changes in stimuli ranking due to other properties interfering, signal power certainly has great influence on urgency rating.

5. OBJECTIVE URGENCY MODEL

Following the computation of the signal parameters, we are interested if it is possible to model the relationship between urgency and the explanatory variables using multiple linear regression. All computations were performed in Matlab using the Statistics toolbox. As it is reasonable for psychoacoustic application, logarithmic scaling is used for the regressors pitch, centroid and rough-

Figure 4: Boxplot of subjective urgency, plotted against objective model urgency. Circles correspond to mean (dark) and median (light). The model is shown as linear regression line.

ness. Urgency rating with background soundscape is selected as the dependent and ecologically valid variable, representing a real-life traffic situation. The corrected Akaike Information Criterion for small sample sizes is applied for deciding which regressors are used in the model to avoid over-fitting. As we suspected, onset time and spectral centroid are not among the chosen parameters. The case of regressors roughness, pitch and signal power (loudness) is displayed in Fig. 4. All variables show positive coefficient values, thus meaning that increasing the roughness, pitch and loudness of a bell sound leads to greater urgency. The mean urgency value pairs are presented using dark circles, while the smaller light dots depict the median values on the boxplot, accompanied by a best-fit line.

R^2 , the coefficient of multiple determination, provides a measure of how well the observed outcomes are simulated by the model. As there is a greater variance for urgency ratings between the 17 subjects for some stimuli than others, calculation of R^2 from raw data only yields a rather small value of 0.27. Therefore the corrected R^2_c is used, which is referring to the mean of the subjective ratings, compensating the mentioned effect and leading to a value of 0.84.

6. GENERAL DISCUSSION AND CONCLUSION

We presented an evaluation of the perceived urgency and recognizability of commonly available bicycle bells in a silent environment as well as with background traffic noise.

A free identification task revealed that not all these warning signals are similarly identified as corresponding to a bicycle. While the twist type bell with the classic 'Rrring' sound achieved high recognizability values, hammer bells ('Ding') and large rotation bells ('Ding-dong') achieved lower but still high ratings. The toy horn, often found on children's bikes, yielded low recognition as bicycle, while an electric bike alarm of type Hornit was never identified as corresponding to a bicycle by the participants of the listening experiment.

An urgency rating experiment showed that the tested bell types differed significantly in urgency ratings. The highly irritating,

siren-like electronic sounds of the Hornit bike bell got highest ratings for both no-noise and noise settings. However, in background noise not all test subjects were drawn to these sounds with the same strength as they were in silence.

The honk of the bike horn also achieved high urgency values for both traffic noise conditions, while average values were obtained by the group of rattling 'Rrring' sound bells. The most common bell types of group (2) 'Ding' and (1) 'Ding-dong' character find their place at the bottom of the urgency scale. Especially, in a noisy environment these sounds seem to lack the ability to stand out and attract attention.

No matter how high urgency scores were for each bell sound, the connection of recognizability and urgency must be considered. Bells which provide a good compromise between respectable urgency and high bicycle typicality, would be the classic 'Rrring' bells with the rattling character of group (3), demonstrating an even better performance under traffic noise conditions. As stated before, most people are not familiar with the Hornit sounds used as a bike bell, therefore these sounds do not appear suitable if one wants to be recognized as a cyclist. Hammer bells are definitely recognizable as bike bells, but their urgency is below average, in contrast to the bike horn with better urgency ratings but poor bicycle typicality scores.

Drawing conclusions from acoustical analysis of the bells' signal characteristics did not turn out to be done easily. Evaluating the chosen signal parameters of Sec. 4 did not always lead to explicit practical results, but it certainly does provide some insight.

Onset time and spectral centroid do not seem to play a major role in urgency prediction at all and perceived pitch seems to have only a little influence. Roughness appears to have an effect on urgency, but the sensory dissonance on its own can not explain the values of the test data entirely. A higher signal power surely influences the perception of urgency, even for our relatively balanced stimuli setup.

The simple objective urgency model showed a relationship for three of the parameters chosen: pitch, signal power, and roughness. In the case of the latter especially, the results support our assumption made in the experimental planning process, that the roughness of a sound has a positive effect on its perceived urgency.

For further investigation, other signal parameters need to be found and examined. Gaining satisfying results from observing only one parameter at a time is not possible for these already existing bell sounds. Many factors from spectral composition, temporal behavior, environmental noise level and, in particular, the acculturation and the listener's knowledge of meaning form a sound's perceptual urgency.

The sense of hearing will always remain important for the human being as a tool for localization and evaluation in possibly dangerous environments such as urban traffic. This is why research and new developments in the field of auditory warning design are of great value and importance. Improving the status quo should be encouraged and supported, thus guaranteeing a safer environment for everybody. Determining and grading warning signal urgency by spectral or temporal manipulation could be an effective approach leading to the desired objective of adequate warning sound design.

7. REFERENCES

- [1] S. James, "Audio warnings for military aircraft," *Audio Effectiveness in Aviation*, 1997.
- [2] J. Edworthy, S. Loxley, and I. Dennis, "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency," *Human factors*, vol. 33, no. 2, pp. 205–231, 1991.
- [3] R. D. Patterson, *Guidelines for auditory warning systems on civil aircraft*. Civil Aviation Authority, 1982.
- [4] A. Guillaume, C. Drake, M. Rivenez, L. Pellieux, and V. Chastres, "Perception of urgency and alarm design." Georgia Institute of Technology, 2002.
- [5] A. Pirhonen and K. Tuuri, "Communicative functions of sounds which we call alarms," in *Proceedings of the 16th International Conference on Auditory Display ICAD 2010*, 2010, pp. 279–286.
- [6] A. Guillaume, "Intelligent auditory alarms," *The Sonification Handbook, Logos (Berlin, 2011)*, pp. 493–508, 2011.
- [7] S. Langlois, C. Suied, T. Lagear, and A. Charbonneau, "Cross cultural study of auditory warnings." International Community for Auditory Display, 2008.
- [8] J. Edworthy and N. Stanton, "A user-centred approach to the design and evaluation of auditory warning signals: 1. methodology," *Ergonomics*, vol. 38, no. 11, pp. 2262–2280, 1995.
- [9] G. Lemaitre, P. Susini, S. Winsberg, S. McAdams, and B. Letinturier, "The sound quality of car horns: Designing new representative sounds," *Acta acustica united with Acustica*, vol. 95, no. 2, pp. 356–372, 2009.
- [10] Federal Signal: Rumbler Siren, last access on 2018-05-03. [Online]. Available: <https://www.fedsig.com/product/rumbler>
- [11] O. Pflüger, "Platz da! (engl. 'Make way!')," *brand eins*, vol. 19, pp. 14–15, 09 2017.
- [12] "Ambulances to jam car radios in Sweden," BBC News, last access on 2018-05-03. [Online]. Available: <http://www.bbc.com/news/technology-38650441>
- [13] A. Petocz, P. E. Keller, and C. J. Stevens, "Auditory warnings, signal-referent relations, and natural indicators: Rethinking theory and application." *Journal of experimental psychology: Applied*, vol. 14, no. 2, p. 165, 2008.
- [14] The Hornit dB140, last access on 2018-05-03. [Online]. Available: <https://www.thehornit.com>
- [15] W. A. Sethares, *Tuning, timbre, spectrum, scale*. Springer Science & Business Media, 2005, p. 32.
- [16] O. Lartillot and P. Toiviainen, "A matlab toolbox for musical feature extraction from audio," in *International Conference on Digital Audio Effects*. Bordeaux, FR, 2007, pp. 237–244.

A PSYCHOACOUSTIC AUDITORY DISPLAY FOR NAVIGATION

Tim Ziemer

University of Bremen
Bremen Spatial Cognition Center
Enrique-Schmidt-Str. 5, 28359 Bremen, Germany
ziemer@uni-bremen.de

Holger Schultheis

University of Bremen
Bremen Spatial Cognition Center
Enrique-Schmidt-Str. 5, 28359 Bremen, Germany
schulth@uni-bremen.de

ABSTRACT

A psychoacoustic auditory display for navigation in two-dimensional space is presented. The auditory display is examined in an experiment with novice users. Trajectory analysis indicates that users were able to a) accurately find sonified targets b) analyze the sonification axis-by-axis c) integrate the sonified dimensions to approach the target on the shortest path. Techniques developed in this work appear to work equally well with three-dimensional coordinates.

1. INTRODUCTION

Many specialized operations have in common that human-machine interaction is supported by visual navigation assistance. Examples include car parking, piloting, remote vehicle control and minimally invasive surgery. Auditory displays have been proposed as complement or even as alternative for visual assistance in such operations [1, 2, 3, 4]. However, these auditory displays communicate only sparse information about the spatial context.

In the present study, a novel, psychoacoustic auditory display is derived, implemented, and examined by means of an experiment with 18 novice users. It acts as a standalone-solution for navigation in 2-dimensional space, without the need for additional visualization.

2. BACKGROUND AND RELATED WORK

A major problem associated with multi-dimensional auditory displays is ambiguity due to perceptual interactions between orthogonal physical audio parameters [5]. For example, physical frequency affects both perceived loudness and pitch and even physical amplitude may affect both loudness and pitch perception [6]. Consequently, these physically independent parameters cannot serve to communicate orthogonal dimensions to a human operator, because they are not orthogonal in perception. It has been observed that a persisting problem in auditory display research is that auditory display design is often arbitrary or based on engineering convenience, and the need to thoroughly consider auditory perception, e.g. in terms of psychoacoustics, has been expressed [7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. An overview about psychoacoustics is given in [6].

Some theoretic approaches towards psychoacoustic sonification can be found in [8, 17, 12]. The first leverages pitch as height, timbre as angle, and brightness as radius in a cylindrical coordinate system, following the example of color space. [8] points out the distinction between objective physical and subjective perceptual aspects of sound and identifies several obstacles for the conduct of his approach. [17] define a framework that translates physical gesture input over gesture-perception and auditory perception to physical audio output. They identify potential psychoacoustical parameters, such as pitch, loudness, timbre aspects, like brightness, roughness, vibrato and formants, as well as their temporal evolution in terms of derivatives. [12] identify loudness, sharpness, roughness, beating and pitch as potential psychoacoustic parameters for perceptual sonification and suggest to map two input data streams to two sensations for two-dimensional sonification. However, the authors realized that finding the right audio parameters that create the desired perceptual outcome is an inverse problem. They suggest lookup tables to find one possible constellation of audio parameter settings to create the desired perceptual sound impression. The risk here is that continuous changes in the data input may cause sudden jumps of audio parameters.

From these studies we derive three critical demands on auditory displays for navigation:

- (α) The auditory display is interactively interpretable
- (β) The auditory parameters of each axis are orthogonal in perception
- (γ) The auditory axes can be integrated, i.e., interpreted together

These enable users to

- (a) Accurately find a target
- (b) Interpret each axis individually
- (c) Reach a target on the shortest path,

which are critical demands on a navigation system.

Even though these studies treat the topic well, formulate problems, necessities and suggest solutions, they lack experimental evaluation. Pioneering work has been done by [18]. Here, different strategies were applied to sonify the distance of a target using psychoacoustic quantities, such as loudness, pitch, brightness, beating and inharmonicity. After the participants carried out a training, the approaches were examined in an experiment with several tasks. A target point lay between 20 and 27.5 cm to the right of a starting point on a tablet computer. One task was to find the target as quickly as possible with a stylus. In another scenario, participants had to find the target point as precisely as possible. All participants

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

were able to find the sonified target locations. Even though only the distance along one dimension is sonified, i.e., the approach is not multi-dimensional but 0.5-dimensional, the results serve as a benchmark for our own study in which we examine navigation in a 2-dimensional space, i.e. including distance and direction along two Cartesian dimensions. Details on the results of their study are resented in Sec. 5 and compared to the results of our own study.

The remainder of this paper is structured as follows: first, the psychoacoustic auditory display is briefly introduced. Then, we present important results of our pilot study, followed by the description of our current experiment. The results are presented and discussed against the results of [18]. The conclusion is followed by a brief outlook on further developmental steps.

3. PSYCHOACOUSTIC AUDITORY DISPLAY

In contrast to many auditory displays, our psychoacoustic auditory display does not map orthogonal spatial dimensions to independent audio parameters, like amplitude and fundamental frequency. Instead, it maps the dimensions to independent perceptual auditory qualities through signal processing. In principle, the sound informs the user where the target lies, relative to the current location by means of one auditory stream with different perceptual auditory qualities. This means the mapping is user-centric. We describe the mapping principle in colloquial terms to make it understandable for non-experts in psychoacoustics. This terminology is the same that we have used to describe the auditory display to the participants in our experiment.

The mapping principle is illustrated in Fig. 1 showing the same three exemplary cursor-target-constellations from two perspectives. The graphics indicate the sound features for cursor locations when the target lies in the center (a) and for target locations when the cursor lies in the center (b). We presented both versions to the participants of our experiment while explaining the mapping principle to them. Details on both the psychoacoustic background and the technical implementation are out of scope of this paper but can be found in [19, 20, 21]. Demonstration videos of the auditory display for some simple trajectories can be found on the first author's YouTube channel¹.

The target is a circular region represented by pink noise. In Fig. 1 targets are represented by red circles. The horizontal direction of the target relative to the current location is mapped to the direction of pitch: when pitch rises, the target lies to the right, when it falls, to the left. The distance is mapped to the speed: the faster the rise or fall, the further away the target lies within the horizontal x -dimension. At the center of the target, pitch is constant. In psychoacoustic terms not pitch per se is increasing or decreasing; in fact, chroma [22] is altered either clockwise or anticlockwise, while height is kept constant. This is achieved by a so-called *Shepard-tone* [22], which creates the auditory illusion of an infinitely rising or falling pitch even though it is a cyclic repetition of a sweep sequence. The speed of pitch rise or fall is actually the cycle speed of the sequence repetition. This Shepard tone consists of octaves only, so the sound does not exhibit roughness or beating and both brightness and loudness are equal for each pitch. The highest cycle speed is lower than 10 Hz, so that physical cycle duration equals perceived cycle duration [6, ch. 12].

The vertical y -dimension is divided in two. When the target lies below the current location the original Shepard tone is manip-

ulated to sound rough: the further away, the rougher the sound. At the target height, the sound is smooth. When the target lies above the current location, the sound will remain smooth, but beating, i.e., regular loudness fluctuation is audible. The further away, the faster the beating. However, even the fastest beating is so slow that it will not be perceived as roughness. At the target height, the loudness is steady. Since a very slow beating as well as a very subtle degree of roughness are barely audible, the target height is indicated by an additional audible click, which represents the x -axis in a target-centered coordinate system. In technical terms, all frequency components of the original Shepard tone act as carrier frequencies in a frequency modulation synthesis [23] to create the roughness impression. Here, the modulation frequency is high enough to create no audible vibrator effect but sidebands near the carrier frequency. The further the target lies below, the higher the modulation depth, i.e. the higher the number and the amplitudes of the sidebands. Perceptually, this does not only increase the degree of roughness but also makes the sound increasingly inharmonic and noisy [24]. The modulation depth is restricted to the region in which the sound preserves high pitch salience [25], so that the pitch metaphor for the x -dimension can even be interpreted clearly at the highest degree of roughness. The beating impression is achieved by amplitude modulation with a low depth. The higher the target, the higher the amplitude modulation frequency. However, it is kept not only well below 15 Hz, where the impression of beating starts fading into the perception of roughness [6, ch. 11], but even below 10 Hz, where perceived duration equals physical duration [6, ch. 12]. When the central target location is reached, a smooth sound with steady pitch and loudness is heard, together with background noise.

The sonification and the additional elements, i.e., the click and the noise, are segregated auditory streams [26]. According to [26] it is easy to notice the presence of one auditory stream while listening to another auditory stream but almost impossible to recognize and follow two auditory streams at once. This is why the triggered click and noise only carry binary information: "central target height reached" and "target reached", respectively.

If the psychoacoustic principles are implemented correctly, the auditory display fulfills the three necessities mentioned above:

a) The sonification is interactively interpretable, which α) enables users to accurately find targets. The interpretability is mainly a matter of b) and c).

b) The sonification is perceived as one auditory stream, i.e., one sound with several characteristics, which are interpreted to derive the target location. At most locations and motions in the two-dimensional space, pitch alterations co-occur with alterations of beating frequency or roughness intensity. But these perceptual auditory qualities are largely independent from each other. Consequently, β) users should be able to either hear out information about the axes separately, and navigate axis-by-axis. For example, they could interpret the speed and direction of pitch change first, to derive the distance and direction along the x -axis. After navigating to the sonified x -coordinate, they interpret the degree of roughness or the speed of beating to reach the corresponding y -coordinate.

c) Alternatively, users could mentally combine the sound characteristics; e.g., a subtly rough sound with a quickly rising pitch means that the target is slightly below and far to the right. This way, they could γ) interpret the location of the target and approach it on the shortest path. The ability to interpret both dimensions together (shortest path) is what makes a multi-dimensional navigation superior to a succession of one-dimensional navigations

¹<https://tinyurl.com/yowmdh8r>.

Figure 1: Mapping principle describing the sound for different locations of the user (cursors) with the target in the center (a) and of multiple targets (circles) with the user in the center (b).

(axis-by-axis).

4. METHOD

We carried out a pilot study to evaluate the learnability and interpretability of the pure sonification. Then, we examined the suitability of the whole auditory display for navigation in an interactive experiment.

4.1. Pilot Study

In a pilot study [19, 20, 27], we examined whether users are able to interpret the psychoacoustic sonification after a short explanation.

7 participants were introduced to the psychoacoustic sonification in a five-minute explanation with demo sounds. In the main experiment 19 sounds were played to them successively. These were 7 s-long sonifications of a static target, followed by 7 s of silence. Within these 14 s the participants had to assign the sound to one of 16 fields on a map, shown in Fig. 2. Even though one participant performed at chance level, an overall average of 41% of the sounds have been assigned to the correct field, 83% to the correct quartile. The figure indicates how often the individual field and each quartile have been assigned correctly. The outcome of this study provided evidence that people are able to quickly learn to interpret the sonification in a passive listening test. Based on this pilot study, some improvements of the sonification could be realized.

The results of this study motivated us to carry out an interactive experiment to evaluate the suitability of the whole auditory display for a navigation task. In the pilot study the participants were passive and had 7 s to listen and additional 7 s to think about the meaning of the sound. However, in an interactive navigation

Figure 2: Map with 16 fields and percentages of correctly assigned field (gray level) and quartile (numbers in the corners).

situation, the user is active and the sonification dynamic. This scenario is the basis of the navigation experiment.

4.2. Navigation Experiment

The psychoacoustic auditory display for navigation in two-dimensional space is evaluated in an experiment with novice users. Based on theoretic considerations and the results from the pilot study we hypothesize that a) the auditory display is interactively interpretable, b) the two sonified dimensions are orthogonal in perception c) the two dimensions are perceptually integratable, i.e., they are perceived as one auditory stream with several characteristics. If so, users were able to α) accurately find the targets, β)

analyze the sonification axis-by-axis, and γ) approach the target on the shortest path. So, we analyze the cursor trajectories of the participants to explicitly examine α), β), and γ) to implicitly provide evidence for a), b), and c), which we consider as a necessity for auditory navigation and as the main achievement of our psychoacoustic auditory display. A brief explanation of the experiment and initial findings can be found in [28]. A comprehensive paper about the experiment will be available soon in [21].

4.3. Setup

The experiment setup is illustrated in Fig. 3. Participants sat in front of a monitor and use a computer mouse to move a visible cursor on a screen towards an invisible target. The described auditory display guided them. The screen is located in the x - y -plane.

Figure 3: Experiment setup. A mouse is moved over a table to control a visible cursor on a otherwise blank screen.

4.4. Training

To get familiar with the auditory display, the experiment started with a training, always containing a verbal explanation, followed by an interactive demonstrator. The 7 demonstrators are depicted in Fig. 4. Since the participants were not familiar with psychoacoustic terms, we used more colloquial language. First, we played the target noise to the candidates. In demonstrator 1 they could move the cursor to trigger the target sound that appeared when a visible icon was reached. Next, we explained the pitch metaphor to them. In demonstrator 2 the candidates moved the cursor along the visible x -axis with the target in the center. They were advised to stop the mouse motion from time to time to listen closely to the sonification of a constant cursor-target-constellation and not only to the sonification of variable cursor-target-constellations, i.e., motion. Next, we explained the vertical dimension, i.e., the beating for targets above and the roughness for targets below the cursor, and the audible click at the target height. In demonstrator 3 candidates could move the cursor along the y -axis and interrupt their motion from time to time to get a feeling for a variable and a static cursor-target constellation. In demonstrator 4 candidates could move the mouse up and down along three additional vertical lines to the left and three to the right of the y -axis to experience the interactive beating/roughness dimension at different constant rates

of pitch change. In demonstrator 5 they could move along 7 horizontal lines to experience the interactive pitch mapping at different constant beating rates or degrees of roughness. What followed was demonstrator 6 where candidates could activate single rows or columns out of the 7 horizontal and 7 vertical lines. In demonstrator 7 candidates could freely move through two-dimensional space with a fixed target in the center. Here, they were free to carry out diagonal or circular motions, zigzag lines or alike.

Figure 4: 7 demonstrator screens for the training. The circle represents the target. The sonification is active on/in the gray lines/area.

The training was concluded with an intermediate test, which resembles the test in the previous experiment. The candidates had to assign 10 sounds to the corresponding field on a map with 8 fields. The sonified targets lay in the center of the corresponding field. These fields were located slightly to the left/right and far up/down or far to the left/right and slightly up/down, illustrated in Fig. 5. Then, they described the sound characteristics: *Is pitch going up or down? Quickly or slowly? Is the sound rough? Slightly or heavily? Or is it beating? Quickly or slowly?* After their description, candidates were allowed to revise their decision on the field assignment. Candidates passed the test when they allocated 5 out of 10 sounds correctly and at least 8 in the correct quartile. 18 out of 25 candidates passed the test and participated in the main experiment.

4.5. Task

For the main experiment, 16 targets with a radius of $b = 5$ mm were distributed equally on a square field with a side length of $a = 20$ cm on a monitor. The participants' task is to find 20 invisible targets as quickly as possible and click on them. First, the 16 targets appeared in random order, then, 4 random targets appeared again. The cursor started in the center of the screen and was moved by the participant to the anticipated target location. Here, the participant performed a click, which saved the mouse trajectory to a file, reset the cursor to the center of the screen, and loaded the next target. Participants were not informed whether they actually hit the target.

5. RESULTS AND DISCUSSION

Two quantitative measures serve to test the hypothesis that participants were able to find the target accurately due to the psychoacoustic auditory display. First, hits were counted, i.e., how frequently users found the targets. Second, the trajectories were analyzed in terms of the time to reach the target. Shorter duration for nearer targets indicates that the targets were found by interpreting

Figure 5: Map showing 8 fields in 4 quadrants.

the auditory display and not by systematically or randomly scanning the whole two-dimensional space. Qualitative, visual inspection of the trajectories provides additional evidence for or against a systematic or random scan of the whole space. The visual inspection also serves to test hypotheses b) and c). If participants navigate towards the target axis-by-axis, this is evidence that they are able to interpret each auditory dimension separately. If trajectories approximate the shortest path, this is evidence that participants are able to combine the two auditory dimensions.

Statistics of individual targets are summarized in Figs. 6 and 7. The large disks are centered at their corresponding target. The cross represents the spatial axes. In Fig. 6 the numbers indicate the mean hit rates $\pm$ the estimated standard errors. For a better overview disk shading redundantly indicates how frequently a target was hit. The darker the disk, the higher the hit-rate. The numbers in the corners give the mean hit-rate for the respective quartile. Most targets were hit in 90 to 100% of all trials with an average hit rate of 91.8%. One-way ANOVA revealed no significant effect of quartile on the hit-rate ($F(3, 12) = 0.85, p = 0.49$). This implies that the combinations of chroma change with beating and roughness were interpretable similarly well.

In Fig. 7 the number and the disk shading indicate the mean time to reach the target. The darker the disk, the longer the needed time. The numbers in the corners give the mean duration for the respective quartile. It took participants about 20.9 s to reach a target, ranging from 14.7 to 31.2 s. We found no significant relationship between the hit rate and the time to reach the target. A general trend can be observed that the further away the target, the longer it took participants to reach it. However, even though the outermost targets are about 3.5 times further away than the nearest, it took participants only twice as long to reach them. So in relation to the distance, near targets were reached relatively slower. On average, targets in all quartiles were reached within a similar amount of time. One-way ANOVA revealed no significant effect of quartile on the time to reach the target ($F(3, 12) = 1.23, p = 0.34$). This

Figure 6: Hit rate $\pm$ estimated standard error for each target and the four quartiles.

supports the finding from the hit-rate, i.e., that the combinations of chroma change with beating and roughness were interpretable similarly well.

Figure 7: Mean time to reach each target $\pm$ standard deviation for each target and the four quartiles.

Statistics of the 18 individual performances are summarized in Figs. 9 and 8. The participants hit between 75% and 100% of the targets. On average, it took participants between 8 and 34 s to reach the targets, except for one individual, who needed no less than 54 s on average. We found no significant linear relationship between the hit-rate and the needed time. We observe that the longer the mean time to reach the target, the larger the standard de-

viation. The two correlate significantly (Pearson's correlation coefficient $r = 0.934347$, $F = 109.99$, $p = 1.4108 \times 10^{-8}$). This rather common relationship states that the slowest participants are not consistently slow. Their performance varies stronger, which may indicate that they are more insecure. After the experiment, some participants confessed that they confused left with right and up with down from time to time. One reason for that may be our misleading graphical representations of the auditory display during the training. We presented Fig. 1 (a) to them, which describes the sound when the target is in the center and the cursor is somewhere else in 2-dimensional space. When the cursor is far to the right and above the target, the sound exhibits fast pitch decrease and roughness. But we also presented Fig. 1 (b) to them, which describes the sonification when the cursor is in the center and the target is somewhere else in 2-dimensional space. When the target is far to the right and above the cursor, the sound will exhibit fast pitch increase and beating. These seemingly contradictory representations may have been the source of confusion. We hypothesize that an unambiguous explanation of the mapping principle and longer familiarization with the sonification metaphors and more experience with the auditory display may improve and stabilize the users' performance and their resoluteness.

The observed training effect supports this hypothesis. We compared the mean value of the first 6 with the last 6 valid runs for all users. Paired samples t-test revealed that the time to reach the target significantly improved by 9.8 s from 26.3 ± 8.3 s to 16.5 ± 3.0 s ($t(17) = 2.964947$, $p = 0.0043395$).

Figure 8: Hit-rate (mean value and estimated standard error) for each participant (user).

Some exemplary trajectories are plotted in Fig. 10. Here, the circles indicate the exact extent of the target in relation to the two-dimensional space. At least one trajectory is plotted for each of the 16 targets. It can be seen that some paths approach the target axis-by-axis, either finishing one axis after the other, or interchanging between them. Other trajectories aim relatively straight towards the target. It can be observed frequently that participants oscillate around the target height to repeatedly trigger the click to confirm that they are still at the correct height. Even comparably long trajectories are still targeted and do not resemble a systematic or random scan of the whole two-dimensional field. The trajectories of the worst performances, i.e., the six longest times to reach a target are plotted in Fig. 11. Here, some axis-by-axis motions and oscillations around the target height can be observed. These trajectories do not seem to scan the whole field. They only cover a comparably small part of the field and rather tend to concentrate on rough target x and/or target y -coordinate.

Figure 9: Performance (mean value and standard deviation) of each participant (user).

Figure 10: Typical trajectories. Some participants solve the x -axis first, others the y -axis. Some switch between motions along the two. Some trajectories approach the target relatively straight. Many trajectories oscillate around the target height to repeatedly trigger the click as confirmation that their height is still correct.

The data confirms our hypotheses that the auditory display enables users to reliably find invisible targets, analyze the sound axis by axis or integrate the information about both axes to aim for the shortest path. The participants find the targets faster and more well-aimed than with a random scan of the whole space. This is true for targets in all quartiles, i.e., the combination of pitch with beating and the combination of pitch with roughness work equally well.

Many auditory displays lack experimental evaluation [15]. Therefore, it is difficult to quantify the benefit of our psychoacoustic auditory display, because no benchmark exist. The only benchmark is the study by [18] for a half-dimensional guidance task. Note that one should be wary of the comparison results, due to different experiment setups, trainings, scenarios, tasks and populations: The participants of our study took 20.9 s to find a target that covered $2 \times \pi b/a^2 = 0.031\%$ of a two-dimensional space. When their task was to find the target as precisely as possible, users

Figure 11: Trajectories of the 6 longest times to reach a target. They occupy only a small region of the field.

in [18] took a similar amount of time to find a target with a precision of $0.97\% \pm 0.43\%$ along a half-dimensional space, i.e., along one line to the right. Their best methods achieved a precision of 0.1% and 0.2%. When their task was to find the target as quickly as possible the participants took 4.5 ± 0.1 s to approach the target with an error of 10.2 ± 1.9 mm, i.e., $3.68\% \pm 0.68\%$. Obviously, our psychoacoustic auditory display enabled users to find a smaller target in less time even though a two-dimensional space is sonified instead of a half-dimensional space. However, participants in our study took roughly 30 minutes of training, including explanation, demonstrators, and intermediate test. Participants in [18] only took 7 ± 2 trials, on average, to train the task.

6. CONCLUSION

In this paper a psychoacoustic auditory display for navigation in two-dimensional space has been presented and experimentally evaluated. The sonification principle is a mapping of relative direction and distance to a combination of direction and speed of pitch change with the speed of beating or the degree of roughness. The approach considers psychoacoustics in terms of chroma [22], pitch and pitch salience [25] [6, ch. 5], perceived duration [6, ch. 12], beating and roughness [6, ch. 11] [24], inharmonicity and noisiness [24].

After only 30 minutes of explanation and training, targets were a) accurately found way faster and more targeted than a systematic scan of the whole two-dimensional space. Users were able to b) approach the target axis-by-axis or c) on the shortest path. These observations as evidence that our psychoacoustic auditory display is α) interactively interpretable, that β) the axes are orthogonal in perception, and γ) the sonification is perceptually integrated as one auditory stream and thus both axes can be interpreted together to derive the target direction and distance. The performance of our participants is comparable or even better than reported for a half-dimensional navigation task in [18]. Furthermore, a highly significant training effect could be observed, indicating that the users

improve their performance with further practice, so we expect even better results for more experienced users.

7. OUTLOOK

Naturally, many navigation tasks take place in 3-dimensional spaces. We included constants in our signal processing, which could act as variables to map the third dimension to perceived *brightness* [29] and *fullness*, *volume* or *sonority* [29], [30], [31, p. 31]. Interactive experiments could serve to validate that this dimension is well interpretable and orthogonal to the other two.

The exact mapping from the physical input to the audio parameters which create the desired perceptual output is based on general knowledge from the field of psychoacoustics. The mapping from the physical distance to the speed of pitch change and of beating is linear, as it is kept within the range in which the perception of duration is reported to be linear [6, ch. 12]. The roughness-mapping is a combination of a linear and an exponential term, which seemed necessary in consideration of roughness perception as reported in the literature [24, 32, 6] to enable users to distinguish several magnitudes in all regions from a low to a large distance. Here, implementing psychoacoustic roughness models as in [24, 32] to derive the optimal mapping function has the potential to perfect the mapping towards perceived linearity and continuity.

To date, psychoacoustic models tend to be valid for static sounds, like [24, 6, 25]. Others have been validated with dynamic sounds, like notes of musical instruments [32, 29]. However, there is an urgent need for psychoacoustic models in interactive scenarios with dynamic sounds and active users [11, 12], not only in the field of auditory display research but also in musicology, audiology and psychoacoustics. The psychoacoustic auditory display could serve as a tool to develop such psychoacoustic models in interactive scenarios.

Our experiment served as a proof-of concept and highlights its strengths and weaknesses. A direct comparison of auditory displays for navigation may be an interesting topic for the future. However, a drawback of a direct comparison of methods is that a number of sonification principles have to be learned by users. This is very time consuming and may cause fatigue and confusion of mapping principles. Therefore, we analyzed the users' trajectories by several means, which could serve as benchmarks. These benchmarks enable to compare navigation performance between different studies. Details can be found in [21].

8. ACKNOWLEDGMENT

The authors thank David Black, Frederik Nagel, Thomas Sporer, Robert Mores and Stefan Weigand for fruitful discussions about the psychoacoustic principles, auditory display evaluations, and future research. We also thank Wasutorn Sanamchai, who programmed the software used in our experiment. We are grateful to the two anonymous reviewers whose ideas and formulations improved our paper.

9. REFERENCES

- [1] A. Lundkvist, R. Johnsson, A. Nyknen, and J. Stridfelt, "3d auditory displays for parking assistance systems," *SAE Int. J. Passeng. Cars Electron. Electr. Syst.*, vol. 10, pp. 17–23, 04 2017. [Online]. Available: <https://doi.org/10.4271/2017-01-9627>

- [2] D. S. Brungart and B. D. Simpson, "Design, validation, and in-flight evaluation of an auditory attitude indicator based on pilot-selected music," in *Proc. 14th International Conference on Auditory Display (ICAD2008)*, Paris, Jun 2008, p. 8 pages. [Online]. Available: <http://hdl.handle.net/1853/49897>
- [3] A. Vasilijevic, Z. Vukic, and N. Miskovic, "Teleoperated trajectory tracking of remotely operated vehicles using spatial auditory interface," in *IFAC PapersOnLine*, vol. 49, no. 23, 2016, pp. 079–102.
- [4] D. Black, J. Hettig, M. Luz, C. Hansen, R. Kikinis, and H. Hahn, "Auditory feedback to support image-guided medical needle placement," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–9, 2017. [Online]. Available: <http://dx.doi.org/10.1007/s11548-017-1537-1>
- [5] J. E. Anderson and P. Sanderson, "Sonification design for complex work domains: Dimensions and distractors," *Journal of Experimental Psychology: Applied*, vol. 15, no. 3, pp. 183–198, Mar 2009. [Online]. Available: <http://dx.doi.org/10.1037/a0016329>
- [6] H. Fastl and E. Zwicker, *Psychoacoustics. Facts and Models*, 3rd ed. Berlin, Heidelberg: Springer, 1999. [Online]. Available: <https://doi.org/10.1007/978-3-540-68888-4>
- [7] G. Kramer, "Some organizing principles for representing data with sound," in *Auditory Display: Sonification, Audification, and Auditory Interfaces*, ser. Santa Fe Studies in the Science of Complexity, Proc. Vol. XVIII, G. Kramer, Ed. Reading, MA: Addison-Wesley, 1994, pp. 223–252.
- [8] S. Barrass, "A perceptual framework for the auditory display of scientific data," in *Proc. 2nd International Conference on Auditory Display (ICAD1994)*, Santa Fe, Nov 1994, pp. 131–145. [Online]. Available: <http://hdl.handle.net/1853/50821>
- [9] S. Smith, "Representing data with sound," in *Proc. IEEE Visualization*, Piscataway (NJ), 1990.
- [10] S. M. Williams, "Perceptual principles in sound grouping," in *Auditory Display: Sonification, Audification and Auditory Interfaces*, G. Kramer, Ed. Reading (MA): Addison Wesley, 1994, pp. 95–126.
- [11] B. N. Walker and G. Kramer, "Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making," in *Ecological Psychoacoustics*, J. G. Neuhoff, Ed. Amsterdam: Elsevier, 2004, ch. 6, pp. 149–174.
- [12] S. Ferguson, D. Cabrera, K. Beilharz, and H.-J. Song, "Using psychoacoustical models for information sonification," in *Proc. 12th International Conference on Auditory Display (ICAD 2006)*, London, Jun 2006. [Online]. Available: <http://hdl.handle.net/1853/50694>
- [13] J. P. Bliss and R. D. Spain, "Sonification and reliability — implications for signal design," in *Proc. 13th International Conference on Auditory Display (ICAD 2007)*, Montreal, Jun 2007, pp. 154–159. [Online]. Available: <http://hdl.handle.net/1853/50028>
- [14] A. Hunt and T. Hermann, "Interactive sonification," in *The Sonification Handbook*, J. G. N. Thomas Hermann, Andy Hunt, Ed. Berlin: COST and Logos, 2011, ch. 11, pp. 273–298. [Online]. Available: <http://sonification.de/handbook/>
- [15] N. Degara, F. Nagel, and T. Hermann, "Sonex: An evaluation exchange framework for reproducible sonification," in *Proc. 19th International Conference on Auditory Display (ICAD 2013)*, Lodz, Jul 2013. [Online]. Available: <http://hdl.handle.net/1853/51662>
- [16] T. Bovermann, J. Rohrerhuber, and A. de Campo, "Laboratory methods for experimental sonification," in *The Sonification Handbook*, T. Hermann, A. Hunter, and J. G. Neuhoff, Eds. Berlin: COST and Logos, 2011, ch. 10, pp. 237–272. [Online]. Available: <http://sonification.de/handbook/>
- [17] D. Arfib, J. Couturier, Kessous, L., and V. Verfaillie, "Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces," *Journal of Organised Sound*, vol. 7, no. 2, pp. 127–144, 2002.
- [18] G. Parsehian, C. Gondre, M. Aramaki, S. Ystad, and R. Kronland-Martinet, "Comparison and evaluation of sonification strategies for guidance tasks," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 674–686, April 2016.
- [19] T. Ziemer, D. Black, and H. Schultheis, "Psychoacoustic sonification design for navigation in surgical interventions," *Proceedings of Meetings on Acoustics*, vol. 30, 2017. [Online]. Available: <https://doi.org/10.1121/2.0000557>
- [20] T. Ziemer and D. Black, "Psychoacoustically motivated sonification for surgeons," in *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. (Suppl 1):1, Barcelona, Jun 2017, pp. 265–266. [Online]. Available: <https://link.springer.com/article/10.1007/s11548-017-1588-3>
- [21] T. Ziemer, H. Schultheis, D. Black, and R. Kikinis, "Psychoacoustical interactive sonification for short range navigation," (*accepted for*) *Acta Acust. united Ac.*, vol. 104, 2018. [Online]. Available: <http://www.ingentaconnect.com/content/dav/aaua>
- [22] R. N. Shepard, "Circularity in judgments of relative pitch," *J. Acoust. Soc. Am.*, vol. 36, no. 12, pp. 2346–2353, 1964. [Online]. Available: <https://doi.org/10.1121/1.1919362>
- [23] J. Chowning and D. Bristow, *FM Theory & Applications. By Musicians for Musicians*. Tokyo: Yamaha Music Foundation, 1986.
- [24] W. Aures, "Berechnungsverfahren für den sensorischen wohlklang beliebiger schallsignale (a model for calculating the sensory euphony of various sounds)," *Acustica*, vol. 59, no. 2, pp. 130–141, 1985.
- [25] E. Terhardt, G. Stoll, and M. Seewann, "Algorithm for extraction of pitch and pitch salience from complex tonal signals," *J. Acoust. Soc. Am.*, vol. 71, no. 3, 1982. [Online]. Available: <http://doi.org/10.1121/1.387544>
- [26] A. S. Bregman, *Auditory Scene Analysis*. Massachusetts: MIT Press, 1990.
- [27] T. Ziemer and D. Black, "Psychoacoustic sonification for tracked medical instrument guidance," *J. Acoust. Soc. Am.*, vol. 141, no. 5, p. 3694, 2017. [Online]. Available: <http://dx.doi.org/10.1121/1.4988051>
- [28] T. Ziemer and H. Schultheis, "Perceptual auditory display for two-dimensional short-range navigation," in *accepted for: Fortschritte der Akustik (DAGA2018)*: 44.

Deutsche Jahrestagung für Akustik, Munich, Mar 2018. [Online]. Available: <https://www.dega-akustik.de/publikationen/online-proceedings/>

- [29] J. W. Beauchamp, "Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones," *J. Audio Eng. Soc.*, vol. 30, no. 6, pp. 396–406, 1982.
- [30] A. Schneider, "Perception of timbre and sound color," in *Springer Handbook of Systematic Musicology*, R. Bader, Ed. Berlin, Heidelberg: Springer, 2018, ch. 32, pp. 687–726. [Online]. Available: http://doi.org/10.1007/978-3-662-55004-5_32
- [31] J. Meyer, *Acoustics and the Performance of Music*. New York, NY: Springer, 2009. [Online]. Available: <http://doi.org/10.1007/978-0-387-09517-2>
- [32] M. Leman, "Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (sim)," in *Proc. COST G-6 Conference on Digital Audio Effects (DAFX-00)*, Verona, Dec 2000.

SOUNDTRAD, A METHOD AND TOOL FOR PROTOTYPING AUDITORY DISPLAYS: CAN WE APPLY IT TO AN AUTONOMOUS DRIVING SCENARIO?

Doon MacDonald

CHERISH-DE
The Cherish Digital Economy Centre
Swansea University, UK
d.g.macdonald@swansea.ac.uk

Tony Stockman

Queen Mary University of London
Centre for Digital Music
London, UK
t.stockman@qmul.ac.uk

ABSTRACT

This paper presents SoundTrAD, a method and tool for designing auditory displays for the user interface. SoundTrAD brings together ideas from user interface design and soundtrack composition and supports novice auditory display designers in building an auditory user interface. The paper argues for the need for such a method before going on to describe the fundamental structure of the method and construction of the supporting tools. The second half of the paper applies SoundTrAD to an autonomous driving scenario and demonstrates its use in prototyping ADs for a wide range of scenarios.

1. INTRODUCTION

In this paper we describe a method and toolkit for prototyping auditory displays (ADs). We argue that there is a lack of methods for supporting novice designers who want to design ADs, specifically a lack of methods that enable early-stage prototyping of an AD design. The method we describe, called SoundTrAD, takes, as its basis, a conceptual blending [1] of techniques from soundtrack composition and from user interface design, to form a novel and playful approach to AD design for the human-computer interface. In this paper we describe its current application and future potential rather than going into specific detail of how SoundTrAD was developed. Interested readers can reference [2] and [3] to learn more around how SoundTrAD was conceptualised, developed and evaluated.

In this paper (section 4) we propose one hypothetical design problem and apply SoundTrAD to exemplify its use. This involves the creation of an auditory display for an autonomous vehicle scenario. We demonstrate that SoundTrAD can help novice designers think about certain design issues for ADs such as the employment of use-case scenarios, sound masking and aesthetics. We also argue that its flexible in the level of support it can offer the designer depending on their experience. We conclude by discussing ideas for further design and development of SoundTrAD.

This work is licensed under Creative Commons Attribution Non-Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

2. BACKGROUND

There are many approaches to creating auditory interfaces in existence. These range from approaches that focus on the conceptual stages of the design problem [4], to those that focus more on specific mapping techniques of information to sound [5]. However, after Frauenberger, Stockman and Bourguet [6] carried out a survey and collated approaches to auditory display design, the first and second author still concluded in a later publication that, despite the number of approaches, ‘unsupported design decisions were found on all levels of the design process’ [7]. It has been observed that there is the lack of methods that enable the designer to move successfully from a conceptual design to a physical prototype of their interface [8] and it is arguable that this could be down to the lack of centralised design knowledge as proposed by [7]. As Brazil and Fernstrom [4] observed, there is no single method that can deal with the complex socio-cultural context of auditory display design. It is thought that the methods for creating auditory displays are ad-hoc, scattered across a number of sources and not accessible to novice interface designers [7, 8] and that sound design for user interfaces is often based on intuition and available technology, rather than a framework or theoretical analysis [9]. It has also been argued that methods that enable a novice designer to build an auditory display creatively are not fully developed and that design knowledge is not disseminated outside of auditory display design experts. This is exemplified by the observation from Stephen Barras and Christopher Frauenberger that ‘the knowledge of how to build a ‘good’ auditory display is hidden in the experience of the experts and creativity of the artists’ [10]. This highlights the importance of de-mystifying the process of creating an auditory display for novice designers. There are, after all, potential benefits to considering the role of the novice designer, as it could lead to the importance of sound in user interfaces becoming more widely appreciated.

It has been argued that sounds need to be consistent across the interface and need to be tested for effectiveness and increased interface usability [11].

3. WHAT IS SOUNDTRAD?

SoundTrAD is motivated by the need for a method that is accessible to novice audio interface designers and furthermore, introduces the field of audio interface design to a wide group of people. SoundTrAD also caters for the need for aesthetics when designing ADs, something that Vickers [12] argues is missing. SoundTrAD is based on a blending of ideas from user interface design

and soundtrack composition. Parallels between the two disciplines that underpin the SoundTrAD method are illustrated in Figure 1 and a summarising explanation of how this ‘blend’ was formed to inform the method stages and steps of SoundTrAD is documented in section 3.3. Section 3.4 then describes the tools that were designed to support the method stages and steps.

3.1. Stages of Interface Design

Existing guidelines for interface design are often made up of several methodological stages. The following stages are derived from a range of sources on interface development, including [13, 14, 15, 16, 17] and represent a summary of interface stages accounted in those references. The different sources vary to some degree in terminology and emphasis they place on each of the stages, however, these stages provide a broad structure from which to approach the design of an interface.

Requirements Gathering Events

The first of the stages is the requirements investigation or requirements gathering stage. This stage involves identifying who the users are, the tasks to be supported by the interface and the context of use. Often at this stage a designer will identify actions and objects within the interface that relate to the human interaction [15].

Conceptual Design

The second stage, the conceptual design stage, involves considering the overall form of the interface including the modes of interaction to be supported and how communications between the users and the system are to be organised.

Detailed Design Stage

The third stage is the physical or detailed design stage whereby the interface is mapped, refined, implemented and prototyped. When discussing these stages of developing interfaces specifically for auditory displays, Peres et al. [16] point out that in ‘whatever form it takes, the specification should detail how the interface will be organised, how it will sound and appear and how it will behave in response to user-input well enough to prototype or implement the auditory task to a point that is sufficient for subsequent development and evaluations’ (p.173).

Evaluation

Finally, there is the evaluation stage, whereby the interface is tested with end-users/ stake holders [17, 16]. It is usual that evaluation is done iteratively throughout the development of the interface, so the designer effectively iterates between the last 3 stages in order to better inform the design process.

3.2. Soundtrack Composition

The following section provides a review of the functions and principles of creating soundtracks. This review is based on multiple sources including [18, 19, 20, 21, 22, 23]. The different sources vary to some extent in the way they describe the principles and functions of soundtrack composition, but the review given here draws on what is considered to be the most frequently utilised and widely held principles. Once again, given the overall goal of the SoundTrAD method, we particularly focus on the concepts in soundtrack composition practice that could be seen to parallel

practices in interface development.

Spotting

Composers will analyse a scene by ‘spotting’ it for places that could be enhanced by audio. Hollywood sound designer, David Sonnenschein wrote that within every on-screen character, object and action there is potential to generate a sound that can enhance the narrative and story [19]. Sonnenschein identifies the following ‘voices’ to listen out for in order to identify key storytelling elements that can be ‘amplified by sound’; namely by identifying and circling explicit words and phrases: People, objects, actions, environments, emotions and transitions.

Arranging

Once the composer has annotated their ideas for sounds, the suggested ideas are classified into categories of dialogue, music, sound effects (D-M-Es). Additionally, the Foley track is created (whereby actions sounds are created to sync with on-screen actions). What is important in a soundtrack is the consideration of how these sounds work together and how they form the bigger soundtrack by relating to one another and to the story being told. Within a film, the soundtrack contains not only the musical score, but ambient sound, dialogue, sound effects, and silence. To underline the importance of considering the soundtrack and all the sounds that make it up as a whole, Lipscomb and Tolchinsky argued for the analysis of the entire soundtrack because ‘musical sound, dialogue, sound effects, silence, and some sounds that fall in the cracks between traditional categories all exist for the purpose of enhancing the intended message of the motion picture’ (p.5) [24]

Composing/Sampling/Syncing

This stage involves sourcing the sound samples and/or composing original music. The stage also involves arranging the Foley track and creating sound effects.

Evaluation/Auditioning

This is inherent in soundtrack composition as it forms a natural part of auditioning the sounds and going back to previous stages to revisit and revise arrangements and ideas.

3.3. Similarities: the Stages of SoundTrAD

Figure 1 presents four major stages of interface design and four major stages of soundtrack composition and steps within these stages. Figure 1 portrays existing similarities between how a film scene and an HCI scenario are analysed. Specifically, the importance that both soundtrack composition and auditory interface design place on identifying and working with actors (users) or characters as part of the spotting stage and requirements gathering stage respectively. Identifying actions, objects, locations and context is also important to both disciplines. Both disciplines, as part of the detailed design stage and composing stage, respectively, also place importance on considering suitable audio onto which events can be mapped. Finally, both disciplines depend on the ability to evaluate and iterate their respective design processes.

A Blended Space

In their recent studies on designing natural and intuitive interfaces [25] described conceptual blending as ‘the ability to combine multiple conceptual spaces into one emergent one, called a ‘blend’.

Auditory Interface Stages and Steps	Soundtrack Stages and Steps
Stage: Requirements Gathering - scenarios and task analysis with <i>users/actors/characters, events, objects, actions, context</i>	Stage: Spotting the scene - stories with <i>characters, actions, objects, transitions, locations</i>
Stage: Conceptual Design - thinking about interface arrangement, and what parts need sonifying and how it is laid out	Stage: Arranging ideas and cues, sketching, establishing and iterating ideas
Stage: Detailed Design - mapping events to audio	Stage: Composing/Designing original music and sourcing sound samples to map to the cues/events
Stage: Evaluation	stage: Evaluation

Parallels Between The Two Disciplines

Figure 1: Parallels Between Soundtrack Composition and Interface Design

Fauconnier and Turner also argued that ‘a conceptual blend operates in two input mental spaces to yield a third space, the blend’ and that ‘partial structure from the input spaces is projected into the blended space, which has an emergent structure of its own.’ [1].

Figure 2 presents an initial bringing together of these stages and steps based on the parallels identified. This represents a candidate blended space between the two disciplines and formed the basis of SoundTrAD. The principle method stages are consistent with the stages used for interface design: requirements gathering, conceptual design, detailed design and evaluation. This supports the goal of using SoundTrAD to develop auditory interfaces. The finer steps within these stages are synthesised from both disciplines of interface design and soundtrack composition.

Stages for SoundTrAD	Steps	
	1. Requirements Gathering	Work with scenarios. Understand users (actors or characters) involved. Think about their context or location. Within the scenario identify events as made up of objects and actions.
	2. Conceptual Design	Layout of interface, arranging events within the scenario
	3. Detailed Design	Mapping events to audio. Choosing, designing or composing suitable sounds
	4. Evaluation	Evaluating and iterating the design and auditioning sounds as part of an iterative design process.

A Blended Space

Figure 2: Framework for SoundTrAD-method stages and steps

3.4. The Tools

There are, to-date, a set of prototype tools that were developed along side the method concept. They are based on techniques employed in soundtrack composition and were designed to support important aspects of the method and act as a proof of concept.

3.4.1. Cue Sheet

The cue sheet (Figure 3) helps the designer/user document the information that needs to be sonified within the scenario. Specifically, the events, their timings and their types. This makes the cue sheet a good document to refer back to at a later date. It is particularly useful if the events are being moved around (see timeline) as part of the arranging stages of the method and the designer wants to see the original layout, justification for design, thoughts etcetera or wants to share ideas with other sound designers and capture design rationale. The columns on the cue sheet, from left to right are labelled as follows:

1. Time.
2. Description /Value.
- 3.Cause and Associated Events.
4. Number of events to be sonified.

The cue sheet is filled out from top to bottom whereby each requirement is mapped out on its own horizontal line. The user can enter whatever they like into the columns and rows, however, some guidelines are offered when it comes to filling out the column labelled ‘cause and associated events’.

Events: The current version of SoundTrAD identifies the following causes and associated events that the user can select from. The idea is that they will map to suggestions in the database of sound ideas. The causes that are listed below initially are based on event descriptions in research by [26].

The causes are as follows:

- Start
- Stop
- State Changes
- Threshold

The events are as follows:

- User Actions (UA): a user triggers an event
- System Actions/Events (SA): a system triggers an event
- Environmental Events (EE): natural occurrences are classed as events
- Continuous Events (CE): an event that lasts all or most of the duration of the use case scenario

The aim of the cue sheet is to capture design rationale and to gather both qualitative and quantitative data about each requirement before they are mapped to suggested sounds.

3.4.2. The Database

After the cue sheet is filled out with the requirements, the designer can access a database of sound suggestions.

When relating auditory display to soundtrack composition, then its important to think about how the sounds blend and work

Time	Description / Value	Cause and Associated Events	Number of Events to be sonified

Empty Cuesheet

Figure 3: An Example Cue sheet

together to help communicate what ever is they are meant to communicate within the scenario. The stages and steps of SoundTrAD (the blended space, outlined above) should support this attention to detail with the mapping whilst a larger method structure is kept in place. In turn this helps the designer in the design process (not lose track of design thinking, ideas, iterations etcetera whilst focussing on sound and mapping).

All the different ‘types’ of sound that make up a soundtrack are catered for and represented in a database of sound samples: music, background, Foley, sound effects and ambient soundscapes.

Additionally, familiar soundtrack composition techniques are utilised such as the use of concrete sounds to represent actions, motifs to help thematic development and and drones to enhance a sense of space and a sense of time passing. Importantly all the different types of sounds can be constructed and layered to build a soundtrack that users can customise when mapping.

A set of guidelines for adding sounds to the interface were devised by Hildebrandt and Rinderle-Ma [26]. These are used as guidelines within SoundTrAD they were included to cater for different types of sounds and are particularly aligned to guidelines for soundtrack composition. They are cited below:

1. ‘Users should be able to customise the mapping from data to sound’.
2. ‘If concrete auditory representations for the occurring events are available, the use of Auditory Icons can produce positive results’.
3. ‘When occurring events are mapped to Earcons, complex timbres (possibly based on real-world instruments) should be preferred over simple timbres (like sine waves)’.
4. ‘The Earcons should take concepts from the areas of motif design and melodic contours into consideration and adhere to “musical” concepts (such as the western tonal system)’.
5. ‘If motifs are being applied, they should differ not only in pitch but also in rhythm and intensity’.
6. ‘Different concepts can be conveyed by using different motifs (possibly hierarchically structured) and/or different timbres’.
7. ‘In general, rhythm and percussion should be included in sonifications’.
8. ‘Continuous sounds (such as drones) should be used to convey the duration of ongoing activities’.

For the first investigation of the method and tools described here, we used a database of 80 sounds. The sounds were based on the guidelines quotes above and were categorised as follows:

- Sound Effects/Foley
- Ambience, environmental and background drones
- One off chimes, ‘dings’ and beeps within set scales (minor, major)
- A set of motifs based on these scales.

For the investigation of a specific case research should be undertaken to investigate whether there is empirically evidence on which to base the choice of sounds, this should be particularly the case where safety is involved, or the system is critical or unstable state. It should, be remembered at this stage the system is providing a prototyping platform and again, in the case of safety critical system will be subjected to exhaustive simulation.

3.4.3. The Timeline

The timeline, as shown in Figure 4, caters for the ability to audition sounds, re-document ideas and move events around within a given use-case to cater for non-linear interactions in time. The events are represented from left to right and from top to bottom. Time is represented from left to right which means that events that occur at the same time will appear vertically in line and will sound together. Specifically, the timeline supports the modelling of different event sequences and combinations and repeated interactions.

This gives the designer the chance to check for issues with aesthetics and masking. The events can be moved using the mouse and the start and stop buttons can control the playback and moving grid line.

Blank Timeline

Figure 4: The Timeline. Events have not been filled in yet. The number of event (E) lines, represented here on the vertical axis) can be added to and taken away. The time can be shortened and lengthened and changed to a suitable scale to cater for the designer’s needs

3.5. Latest Findings and Potential Applications

An evaluation of SoundTrAD with the steps and tools described above is described in [3]. 11 participants took part in the study and were mostly new to AD design.

At the end of the study we asked the participants to write down any scenarios to which they felt they could apply SoundTrAD. The responses demonstrate the wide-range of scenarios and applications that participants felt the system could be used for. Listed below are accounts of these participant (P) testimonials as potential application areas for SoundTrAD at this stage of its development.

P1: "Sport events, data sonification, live data performance tool".

P2: "I would like if it can respond to real time events. Then it can be used in public areas such as cafeterias. The system can be related to the play list the cafe is using. When there are more customers in the shop, play some pop music or music with a quick tempo, customer might eat faster and leave quicker. When there are only few people in the cafe, play some classic music to ease the pressure of customers."

P3: "Analysis of any complex system e.g. scientific research"

P4: "For tasks or cases that need constant monitoring and need to fire alerts if changes happen, while users are occupied doing something else. Almost like a background monitor."

P5: "Petrol consumption. When client payments made to bank account (might be too complex but would be really helpful) Metered water consumption. When I need coffee!"

P6: "I think it can be applied to use with any storyline/events that needs auditory display"

P7: "Music education, aiding system for sensory-impaired people, mobile application, animal monitoring, and so on."

P8: "Transport hubs like railways stations and airports, either for people (queues building up, need more staff on check in desks etc) or for the logistics of moving bags (airports) or rolling stock (sending in more carriages) or even having an alert for a problem with trains and getting rail replacement busses."

P9: "Digital archives. Revision aids. Teaching aids especially for students with special educational needs'. I would really like to use with historical data being archived online. I think this kind of audio display would be great for a variety of students needs and could make education more accessible for all. There are endless possibilities here I think'

P10: "To help workers 'keep an ear' on their work. To help supervisors monitor staff activity. To help managers review data and plan accordingly. To add auditory and notification to already existing IT systems (in vehicles/computers/classroom environments). Hugely beneficial to people with impairments."

P11: "in vehicles for optimum performance and monitoring. For use in medical and sporting monitoring equipment. for monitoring energy use in the home/business. for use in extreme locations for monitoring yourself/the environment. to aid a person with disability/medical condition. For anyone that is multitasking any situation and needs to monitor and respond to situations while engaged in another activity"

4. APPLYING SOUNDTRAD TO A DRIVING SCENARIO

In this paper we apply SoundTrAD to the design of an in-vehicle AD. We feel that a driving scenario is very suitable to exemplify the potential of SoundTrAD, particularly in regards to demonstrating how soundtrack composition can influence the design of the auditory display.

For example, a driver will often maintain two or more tasks whilst driving, such as navigating the road whilst listening to music and paying attention to collision warning sounds or sounds that indicate a seat-belt is not done up, or the fuel is low. Additionally, it is arguable that the natural sounds from the road and car engine also play an important part when it comes to successful and safe driving.

This scenario therefore already has a potential construct of Fo-

ley sounds, environmental/location/ contextual sounds and music and possibly speech all vying for bandwidth. Consequently, there will be issues with masking, aesthetics and acceptability of the interface [11].

A soundtrack is normally made up of these elements and designers, if using SoundTrAD, will 'spot' the scene to map out sounds and then use a cue sheet. Therefore, it is felt that SoundTrAD (as a method and tool that takes ideas from soundtrack composition) gives the designer a better chance of addressing these issues when designing the auditory display for this scenario.

4.1. The steps and the Tool (applied):

4.1.1. Stage 1: A Scenario

1. **The Scenario:** Autonomous Motorway Driving
2. **The User:** A solo driver
3. **Things you Want to Find Out:** When as a driver you need to pay attention to the auditory display, giving you the option to be aware of choices the car is making.
4. **Parameters of the Situation:** 1. Travelling at a default 70mph, 2. Other Vehicles and hazards on the road.
5. **Alerts/Sounds for:**
 - (a) Indication of speed change and why (weather, accident/hazard, traffic density/flow).
 - (b) Lane changing and why (accident/hazard, maintain speed),
 - (c) Continuous state of the car- why - normal environment/status.
 - (d) Manoeuvring off the motorway environment - why - handing the driving over to the user.

This scenario is now broken down and entered into the cue sheet. Figure 5 illustrates how this has been done.

Time	Description / Value	Cause and Associated Events	Number of Events to be sonified
1pm.	Travelling at 70mph	Start / CE	1
1.04pm	Change of speed due to rain	State Change / SA and EE	2
1.10pm	Lane change-hazard	State Change / SA and OUA	2
1.12pm	Speed Up-rain stopping	State Change / SA and EE	2
1.15pm	Off motorway hand over	State Change / SA and UA	2

Filled out cuesheet

Figure 5: A filled out cue sheet for the autonomous driving scenario. Each requirement is marked out from top to bottom. The number of events that can be used to represent these requirements of the interface are indicated in column 4. Column 2 represents a place where qualitative and quantitative information can be documented.

4.1.2. Stage 2: Conceptual Design/Arranging

At this stage in the design a series of sound suggestions are presented based on the criteria listed above. The designer has the option to audition different sounds and 'upload' them to the events

mapped out on the timeline. The sounds are based around the following basic mapping principles, however, it should be noted that it is clear that they are suggestions and the designer can of course use any sound they feel suitable. In addition to the list below, it is a feature of the database of sounds that the sounds will harmonically relate to one another. For example, the drone will have a pitch and the motifs will relate harmonically to that pitch. As mentioned, it is up to the designer to make this decision, but one important feature of SoundTrAD is that the sounds are categorised with that available option.

- User Actions /other user actions: Foley sounds, everyday sounds.
- System actions: Foley sounds, everyday sounds, motifs (melodic and rhythmic variations.). Short sounds (beeps, dings, chimes, and percussive hits) all with the ability to be pitched.
- Environmental events: natural, everyday sounds ranging from organic sounds to industrial.
- Continuous events: drones, soundscapes and ambient-based sounds, ranging from organic sounds to industrial. All with a pitch or ability to be pitched.

4.1.3. Stage 3: Mapping and Auditioning

Once the user/designer has uploaded the sounds to the timeline (see Figure 6) they can play the events on the timeline and listen to their selections. Specifically, they can listen to how different events may sound together and subsequently pay attention to aesthetics and potential masking of information. For example, in the driving scenario we know that event 2 and 3 have to represent the fact that the car changes state by slowing down (a system action) due to the rain (an environmental action/event). We are essentially needing to sonify two pieces of information - the slowing down of the car and the start of the rain. The events are placed on the timeline in such a way that the designer can hear these two bits of information that are needed for this requirement simultaneously and subsequently make a decision as to whether the sounds communicate whatever it is they need to be communicating. It may be, for example, the designer wants the sounds to play shortly one after another or overlap. In this case they can re-arrange the events to audition this idea. Another important feature of the timeline is the ability to repeat events. For example, the rain event is represented on line 2 of the timeline and occurs twice during the scenario. However, it occurs alongside different events (the first being the slowing down of the car and the second being the speeding up of the car). As a result, the 'rain sound' can be used twice but may in fact sound very different when played with the 'slowing down sound' and 'speeding up sound', respectively. Figure 6 demonstrates this.

4.1.4. Stage 4: Usability Evaluation /Auditioning

The events can be added, auditioned, moved, lengthened, shortened and removed. At any time during the design process the designer and or the end user, or ideally both together, can revisit the scenario and the cue sheet to re-examine their initial design ideas. Importantly, SoundTrAD caters for different use-cases within a given scenario. For example, the events can change order given a different order of user interaction and subsequently different sound

Figure 6: The Timeline for the autonomous driving scenario. Events can be moved about, lengthened and shortened. More lines can be added.

combinations can be tested given different possible combinations of events.

5. DISCUSSION AND FUTURE WORK

This paper has presented SoundTrAD, a method and tool for prototyping ADs, and applied it to a hypothetical driving scenario. We have argued, based on evidence from a previous study, that SoundTrAD can cater for a wide range of scenarios. In this paper we have applied it to a specific driving scenario to demonstrate its application.

SoundTrAD supports the designer in breaking down a given scenario in to events and systematically mapping them out in order to reconfigure the scenario. As a result, the auditory display can a) cater for different use-cases and b) cater for the application and auditioning of audio and potential issues with masking and aesthetics.

In this example we have taken a limited scenario to demonstrate how a novice AD designer can build an auditory interface that supports an end-user when responding to an autonomous device.

This paper concludes by proposing that this application of SoundTrAD, to a hypothetical driving scenario, can be expanded in order that the driver (end-user) features in the decision making process. Specifically, that the end-user can make decisions around how the AD is incorporated into their interaction with the vehicle. We feel that there are benefits to giving the driver choices over sounds and bespoke personalised alerts. The sound designer could set parameters for the end-user, for example, to introduce change so that the end-user does not become desensitized to the alerts. It could be suggested that giving the end-user a perception of control over the audio in the car will add to their overall sense of confidence, familiarity and trust with the autonomous device. An exciting area of research that we feel SoundTrAD could be applied to.

6. ACKNOWLEDGMENT

The majority of research in this paper was funded by the EPSRC, digital economy Media Arts Technology Doctoral Training Cen-

tre at Queen Mary University of London. The later stages of the work are supported by CHERISH-DE, an EPSRC next stage digital economy research Hub at Swansea University.

7. REFERENCES

- [1] G. Fauconner and M. Turner, "Conceptual blending, form and meaning," *Recherches en communication*, vol. 19, no. 19, pp. 57–86, 2003.
- [2] D. MacDonald and T. Stockman, "The development of a method for designing auditory display based on soundtrack composition," in *Proc. of ICAD 2013, Meeting of the International Conference on Auditory Display (ICAD)*, 2013.
- [3] D. MacDonald and Stockman, "Lowering the barriers to the creation of interactive auditory displays: an experimental investigation," in *Proc. of ISON (Interactive Sonification Workshop)*, 2016.
- [4] E. Brazil and M. Fernström, "Subjective Experience Methods for Early Conceptual Design of Auditory Displays," in *Proceedings of the 15th International Conference on Auditory Display (ICAD)*, 2009, pp. 1–8.
- [5] T. Hermann, A. Hunt, and J. G. Neuhoff, "Chapter 1 Introduction," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin: Logos Publishing House, Berlin, Germany., 2011, ch. Chapter 1, pp. 1–6.
- [6] C. Frauenberger, T. Stockman, and M.-L. Bourguet, "A survey on common practice in designing audio in the user interface," in *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it*, vol. Volume 1. British Computer Society., 2007, pp. 187–194.
- [7] C. Frauenberger and T. Stockman, "Auditory display design: an investigation of a design pattern approach," *International Journal of Human-Computer Studies*, vol. 67, no. 11, pp. 907–922, November 2009.
- [8] E. Brazil, "A review of methods and frameworks for sonic interaction design: Exploring existing approaches." Brazil, E. (2010). A review of methods and frameworks for sonic interaction design: Exploring existing approaches. In *Auditory Display* (pp. 41–67). Springer Berlin Heidelberg., 2009, pp. 41–67.
- [9] A. Pirhonen, E. Murphy, G. McAllister, and W. Yu, "Non-speech sounds as elements of a use scenario : A semiotic perspective," in *Proceedings of the 12th International Conference on Auditory Display (ICAD)*, 2006, pp. 134–140.
- [10] S. Barrass and C. Frauenberger, "A communal map of design in auditory display," in *Proceedings of the 15th International Conference on Auditory Display*, 2009, pp. 1–10.
- [11] S. Brewster, P. Wright, and A. Edwards, "The application of a method for integrating non-speech audio into human-computer interfaces," University of York, Department of Computer Science, 1995., Tech. Rep., 1995.
- [12] P. Vickers and B. Hogg, "Sonification Abstraite / Sonification Concrete : An 'Æsthetic Perspective Space' for classifying Auditory Displays in the Ars Musica Domain," *Proceedings of the 12th International Conference on Auditory Display 2006*, pp. 210–216, 2006.
- [13] A. Dix and S. Brewster, "Causing trouble with buttons," in *Ancillary Proceedings of HCI'94*, D. England, Ed., 1994.
- [14] J. Carroll, *Making use: scenario-based design of human-computer interactions*. Cambridge, Massachusetts. London, England: The MIT Press, 2000.
- [15] D. Benyon and C. Macaulay, "Scenarios and the hci-se design problem," *Interacting with Computers*, vol. 14, pp. 397–405, 2002.
- [16] C. Peres, V. Best, D. Brock, B. Shinn-Cunningham, C. Frauenberger, T. Hermann, G. Neuhoff, L. Nickerson, and T. Stockman, "Audio Interfaces," in *HCI Beyond the GUI. Design for Haptic, Speech, Olfactory and Other Nontraditional Interfaces*, Kortum, P, Ed. Morgan Kaufman Publishers, 2008, ch. 5 Auditory Interfaces.
- [17] D. Benyon, *Designing Interactive Systems: A Comprehensive Guide to HCI and Interaction Design*, 2nd ed. Pearson Education Ltd., 2010.
- [18] M. Chion, *Audio Vision. Sound on Screen*, C. Gorbman, Ed. Columbia University Press, 1994.
- [19] D. Sonnenschein, *Sound Design: The Expressive Power of Music, Voice and Sound Effects in Cinema*. Michael Wiese Productions, 2001.
- [20] R. Beauchamp, *Designing Sound for Animation*. Focal Press, 2005.
- [21] T. Holman, *Sound for Film and Television*, 3rd ed. Elsevier/Focal Press, 2010.
- [22] D. Ventura, *Film Music in Focus*, 2nd ed. Rhinegold Publishing, 2010.
- [23] K. Kalinak, *Film Music: A Very Short Introduction*. Oxford University Press, Inc, 2010.
- [24] S. Lipscomb and D. Tolchinsky, "The role of music communication in cinema," *Musical communication*, pp. 383–404, 2005.
- [25] S. Bødker and C. Nylandsted Klokmoose, "Dynamics, multiplicity and conceptual blends in hci," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 2538–2548.
- [26] T. Hildebrandt and S. Rinderle-Ma, "Toward a sonification concept for business process monitoring," in *International Conference on Auditory Display*, 2013, pp. 323–330.

MAPS, LOCATION AND NAVIGATION

REIMAGINING HUMAN CAPACITY FOR LOCATION-AWARE AURAL PATTERN RECOGNITION: A CASE FOR IMMERSIVE EXOCENTRIC SONIFICATION

Ivica Ico Bukvic

Virginia Tech,
SOPA, DISIS, ICAT,
24061 Blacksburg, VA, USA
ico@vt.edu

Gregory D. Earle

Virginia Tech,
Electrical and Computer Engineering,
24061 Blacksburg, VA, USA
earle@vt.edu

ABSTRACT

The following paper presents a cross-disciplinary snapshot of 21st century research in sonification and leverages the review to identify a new immersive exocentric approach to studying human capacity to perceive spatial aural cues. The paper further defines immersive exocentric sonification, highlights its unique affordances, and presents an argument for its potential to fundamentally change the way we understand and study the human capacity for location-aware audio pattern recognition. Finally, the paper describes an example of an externally funded research project that aims to tackle this newfound research whitespace.

1. INTRODUCTION

Human interfaces to the natural world are inherently multisensory [1]. In simulated environments we often mimic our interaction with the natural world by combining sensory mechanisms to broaden cognitive bandwidth [1], [2], as well as to reinforce comprehension [3] and learning [4], [5]. A 1997 report to the National Science Foundation [6] defines sonification as “the use of nonspeech audio to convey information”. In this paper the authors refine this definition only slightly, using the term sonification to describe *the process of using sound to convey information that is not inherently auditory*.

Despite decades of concerted research, sonification remains vastly underutilized. In 1999 Hermann and Ritter suggested that sonification is an “underused perceptual channel for man-machine-interaction” [7], and in 2007 Nasir and Roberts stated “that researchers have not fully utilized the maximum potential of spatial sound” [8]. More recently, a 2014 paper by Thomas Hermann suggests that sonification is still in its infancy [9]. This continued challenge presents us with a seemingly obvious question: after decades of rigorous research, why is sonification still in its infancy? There are many facets to this complex issue. By leveraging the aforesaid immersive exocentric approach this paper proposes a path forward.

2. THE EXOCENTRIC ENVIRONMENT

Central to this paper is the idea of creating and exploring an exocentric environment through sonification. This naturally leads to the question: what is the working definition of the

term “exocentric environment”?

The term “exocentric environment” has its origins within the context of immersive environments where it suggests an out-of-body experience [11], or an immersive environment that completely encompasses the user [17], [18]. What is particular to this definition is that the observer is not considered the center of the experience (as would be the case if we were to observe the immersive environment through their point of view), but rather as one subject with a unique and adjustable vantage point. In the context of sonification an exocentric environment does not treat the listener as the center of the listening environment. Instead the listener is a mobile actor who can navigate the space and adaptively interact with the real world to enhance their capacity to localize and discriminate various concurrent sounds. This specific meaning of “exocentric environment” is used throughout this paper. We will return to it at the end of the paper where we propose a new set of guidelines for studying the human capacity for location-aware audio pattern recognition.

3. LITERATURE REVIEW

To better understand the current state of sonification research and its limits we present a cross-disciplinary review of 21st century literature in the field of sonification in a number of scholarly communities, including the Association for Computing Machinery (ACM), the Audio Engineering Society (AES), the Institute of Electrical and Electronics Engineers (IEEE), and the International Conference on Auditory Displays (ICAD). The overall goal is to identify similarities, patterns, obstacles, and promising areas for future studies.

The Association for Computing Machinery (ACM) science cohort’s early research has yielded promising results hinting at sonification’s potential both as an independent form of representing data and, when coupled with a visual feedback, as a means of broadening cognitive bandwidth [2], [10]. Since 1999 the entire community has produced 28 publications on the topics of sonification and spatialization [11], a number of which mention spatialization in contexts outside of audio environments. Most of these reduce spatial data to a single-dimensional output [12], or rely on approaches (e.g., headphones) that decouple aural perception from other sensory inputs [13]. This is unlike any other domain of human sensory perception. For example, in studying visual perception, even if the experimenters immobilize the user’s head, the user still retains the ability to create a spatial image through eye movement. Some research has gone as far as to suggest that sound is an inherently non-spatial medium [14], although this argument is easily refuted.

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

Recent areas of research include assistive technologies [15]–[20], reflective practice [21] and biofeedback that all but ignores our human auditory spatial capacity [22]–[28], or implements a spatial component in an egocentric way [29]–[31]. Studies of this type often yield discouraging data [32]. A few researchers have focused on studying infrastructure for the delivery of spatial information [33]. There are notable exceptions [34]–[36] that focus on perception and cognition of spatially delivered stimuli, and even go as far as simulating elements of an exocentric environment [34]. However, none of these studies focus on spatial aural perception in a true exocentric environment.

Similarly, a review of the Audio Engineering Society (AES) community’s output over the past 6 years offers three notable publications on the topic of sonification [37]–[39], none of which utilize the exocentric environment. A number of publications [37]–[45] either do not concern themselves with audio spatialization, or study it in isolated scenarios not directly related to sonification. Recent Institute of Electrical and Electronics Engineers (IEEE) literature echoes the ACM and AES trend. Topics can be categorized as fundamental studies [46]–[49], aggregators [50], [51], contextual spatialization algorithms [52], spatially sonifying data using virtual solutions [3], [26], [53]–[55], or sonification without any spatial component [56]. The best resource for sonification and spatialization research is found in the International Conference on Auditory Displays (ICAD) community, which annually publishes 50-70 papers that focus on sonification, spatialization, or both. Below we cite those most relevant to our theme.

The reasons for the apparent disparity between research and broad adoption are undoubtedly numerous and not yet fully annotated. Gregory Kramer talks about “clear parallels between the composer’s role in [auditory displays] and the graphic artist’s role in data visualization.” He goes on to state that “Improved aesthetics will likely reduce display fatigue” [57], suggesting that part of the challenge in sonification processes is the ability to deliver content that is not tiring. Solving it will in great part depend on context-aware case-specific solutions.

There are other more pervasive challenges in the current sonification literature that are inherently tied to human ecology [58], or the ways in which we perceive our environment. Studying how we perceive spatial sound in a natural environment makes it difficult to decouple different sensory mechanisms, so the results of such studies may differ significantly from outcomes in which such mechanisms are studied in isolation. Despite this apparent bias, a vast majority of sonification projects and studies focus on simulating virtual environments, often by relying on an implementation that decouples the user from their immediate environment and eliminates input from other senses [59]. The most common examples of this are studies that rely on a headphone-based delivery system that utilizes discrete [60], stereo [26], [28], [61]–[63], or some form of head-related transfer function (HRTF)-based implementation [5], [54], [64]–[67], [68].

The aforesaid citations represent a snapshot of a larger volume of examples, so they should not be considered exhaustive. Instead, they are meant to offer examples of each of the aforesaid categorizations. A considerably smaller number of research projects explore spatial sound [55], [69], with a majority continuing to be linked primarily to kinematic data [70]. Interestingly, one paper [55] considers spatialization approaches like Vector Based Amplitude

Panning (VBAP) [71] to be inadequate because of implicit assumptions that data perception should be decoupled from the perceiver’s vantage point. In contrast, spatialization studies can be approached as an opportunity for users to traverse a simulated exocentric environment akin to the way they interact with the real world [34], [72], using their vantage point in a natural way to attain greater clarity. For instance, a user approaching one part of an audio display can naturally generate and process amplitude disparities in the overall image, allowing them to more easily separate sources of particular interest. The result is akin to the cocktail party effect [73], in which we tend to stand in proximity to key points of interest, facing them to ensure we have a clear understanding of their locus and that they stand out above the environmental noise floor.

4. DISCUSSION

The aforesaid literature review offers two noteworthy observations. First, there appears to be little overlap in institutional knowledge between different scientific communities. This limits the impact of the ensuing scholarship and leads to inefficiency and redundancy. This issue is certainly not limited to the sonification research, and it undoubtedly stifles progress.

Second, of all the extant research only a small subset is concerned with studying human spatial aural perception. This is surprising considering how integral the immersive spatial component is to human auditory perception. Further, of those projects that do explore the spatial context, virtually all focus on some form of simulation of spatial content, most commonly utilizing HRTF or similar approaches to spatializing sound. In doing so, such implementations are hampered by idiosyncrasies. For example, headphones restrict the ability to convincingly place sound in front of the listener, and in and of themselves are unable to address the front-back confusion. These issues are naturally mitigated by head movement in real-world applications, yet in many studies they are rendered ineffective by the delivery methodology. Because simulations often do not account for head movement, such ambiguities may be compounded. Finally, the addition of other simulated spatial sound perception (e.g. head tracking) may create its own idiosyncrasies, such as those stemming from latency.

The vast majority of research in sonification is focused on placing the listener at the center of a space with optimal loudspeaker distances and phasing. The goal is often to identify limits of human perception when the audio component is decoupled from the other senses. This kind of foundational knowledge is essential to understanding the core capacity of human auditory perception, but it is hardly representative of the way in which humans interact with auditory cues in the real world. Humans naturally process data concurrently from different senses to mitigate limitations that stem from one sensory input. For instance, head movement is a simple kinematic approach that can help minimize the impact of the cone of confusion and/or front-to-back ambiguity in spatial auditory perception. One’s location and orientation can help to identify critical details that may stimulate a more comprehensive understanding of the surrounding environment and its stimuli. Instead of placing the listener in an egocentric environment where they are ideally equidistant to every loudspeaker around them, allowing user to move and disrupt the loudspeaker distance equilibrium recontextualizes such a space into an immersive

exocentric environment where user's interaction with the surrounding stimuli is more similar to the way we naturally perceive the world around us. To date, only one study [74] has shown interest in exploring such an exocentric environment, and it has yet to provide any tangible or reproducible data.

5. A CASE FOR IMMERSIVE EXOCENTRIC SONIFICATION

We posit that studying the way we perceive the world through sound should be inherently holistic. It is worth noting that the spatial dimension of human hearing capacity is inextricably linked with stimuli perceived using our other senses, particularly sensations related to head rotation, head position, and visual cues related to location awareness. For instance, to minimize the so-called cone of confusion for sounds to the left and right of our head where there pinna is unable to create filtering variance, we commonly rotate our head to change our perspective relative to the sound source and better pinpoint a sound. Inability to do so vastly limits our capacity to locate sounds, and may result in errant conclusions. For example, studies that do not account for this capacity may conclude that our spatial hearing capacity seems lower than it actually is.

Even in a reductionist approach to studying human sight, we do not prevent users from moving their eyes or head. These are, after all, critical elements that enhance our ability to see because our eyes only have a very narrow area where we can clearly perceive details; the edges of our field of view are blurry, but still sensitive to motion and change. It is the brain that constructs a holistic image as our eyes scan our surroundings using eye muscles, and this process is further enhanced by our head and body motion. In studying aural perception it therefore makes sense that we should allow and encourage movement, particularly in immersive three-dimensional environments, thereby encouraging ecological validity of the ensuing research data. By moving within the space we refine and improve our ability to locate sound sources. For example, those of us who have placed calls to our cellphones in order to find them in cluttered environments can vouch for the advantage of walking around while listening for its ring. As we search, the ability to move throughout the space enables us to use the loudness of the ringtone, as well as any potential reflections to our advantage.

Coupling head position, body movement and position awareness with our hearing perception enhances the way we interact with our environment, and it comes naturally to us. We don't have to train for it. We simply interact.

Now, consider a walkable High Density Loudspeaker Array (HDLA) environment with loudspeakers all around you, including above and below. As you move around the space, you are able to get closer to a particular sound, thereby naturally attenuating the perception of other sounds while concurrently using your spatial memory to generate a comprehensive spatial map. In such an environment you can interact with sonified data in the same way you interact with the real world. The ensuing immersive exocentric environment offers unique opportunities for immersive exocentric sonification. Here the sonification emanates from the space perimeter and does not require the user to remain in one location to explore the volume and phase characteristics of the immersive sound field.

When considering exocentric sonification, it is worth noting that the varying distance between the loudspeakers and a listener navigating the space is not detrimental. Rather, it is an advantage that more closely resembles the way we interact with the world. There may be phasing artifacts and other unforeseen interactions. Yet, the same are commonly experienced in the real world and while they may adversely affect the overall aural image, the advantage of cross-pollinating sensory input, as well as the ensuing amplitude differences are likely to offset any such shortcomings by significantly improving our spatial resolution. Consider the way we perceive movies on traditional cinema screens that, depending on where we sit, may be too large to fit within our peripheral vision. Under such circumstances, some of the areas on the edge of the large screen are much farther from our eyes than others, leading to unusual distortions to the image's perspective in respect to our vantage point. Yet, our brain understands this is a flat image and we happily reconstruct it as such in our heads without any concern for its observed physical distortion. By extension, akin to Layer Based Amplitude Panning algorithm (LBAP) [75], in the exocentric environment speakers within the same horizontal layer share the same elevation, much like a row of pixels on a screen. Once we have the overall awareness of the visual (e.g. TV) or aural (e.g. HDLA front defined by its loudspeakers) canvas, our brains are capable of compensating for such physical inconsistencies in favor of maintaining consistent relationship between the individual loudspeakers, or, as is the case with its visual counterpart, between the pixels on a screen.

The immersive exocentric environment and, by extension, immersive exocentric sonification focuses solely on the space's perimeter where the loudspeakers are located. While there are ways of simulating sounds within the space, particularly when using the wave field synthesis (WFS) [76] and the ambisonics [77], they are prone to idiosyncrasies that limit the human ability to move and study such sources from different vantage points. If we were to consider the ensuing speaker front or perimeter as one canvas, we could project and move data across it. For initial studies the most obvious data choices may be inherently spatial data, such as geographical or geospatial data that limit the need for arbitrary assignment of variables to the spatial component. We posit that immersive exocentric sonification presents an opportunity to enhance and consolidate sonification research. This approach has not been adequately studied, and it has the potential to change the foundation of our understanding of human spatial aural pattern recognition.

6. REAL-WORLD APPLICATIONS

Studying human auditory perception in isolation is an important step towards understanding the limits of human sensory mechanisms and is a critical precursor to the immersive exocentric sonification. Yet, the application of such findings in real-world scenarios without considering other complementing sensory dimensions may lead to misleading conclusions, including lower perceived human spatial auditory perception capacity. As such, one could argue that studying human spatial hearing without incorporating head and more importantly body movement and location may be just as misleading as studying human vision while disallowing head, body, and eye movement. The immersive exocentric approach to sonification described

above aims to address this potential pitfall and is rooted in the way we interact with our world. As a result, it feels natural and intuitive.

Its applicability can be seen both on micro and macro scales, as well as in natural and human-made environments. Let us return to the aforesaid example of how we may want to search for a misplaced smartphone in our home. Commonly we resort to calling it to hear the source of its ring. As the smartphone rings, rather than standing in one place we move around and typically closer to the source to improve our spatial resolution that is often impeded by physical and acoustic obstacles. We may want to consider this a micro scale natural environment.

Now imagine a large space equipped with a cutting-edge technology capable of monitoring human presence, motion, location, and user input, while offering a multisensory immersion in a complex dataset. Such an immersive environment is commonly referred to as the “decision theater” [78]. This macro scale human-made environment presents another case scenario where both the spatial aural and multisensory data immersion can take place and where listener location, motion, and orientation can help amplify their understanding of the data and the ensuing patterns. For instance, as they move closer to the edge of the space populated by an aural (e.g. a loudspeaker array) display, listeners attain greater spatial resolution and can effectively amplify specific sources of interest and consequently the perceived signal (source of interest) to noise (other sounds) ratio due to change in their proximity and orientation.

As evidenced by the literature review, virtual immersive spatial sonification has greatly benefited from the advances in immersive technologies, allowing for the incorporation of head rotation. Its utilization is steadily increasing. Yet, even in virtual immersive auditory environments that focus solely on the head rotation, it is worth considering that rotation also includes translation, allowing for ears to be closer to the source. If such scenarios were to utilize virtual sources placed in close proximity to the listener’s head, such an aural spatial image stands to benefit from the increased spatial resolution due to head movement. This, arguably micro-scale human-made environment could have a far-reaching impact from spatial aural cues in operating equipment (e.g. cars and airplanes), to data rich environments and “decision theaters” that provide multisensory feedback with the goal of broadening cognitive bandwidth.

The immersive exocentric sonification may have applications in a number of scientific domains, both as a novel methodology for data analysis and as a teaching tool. To determine the extent of its utility the initial research needs to focus on using physical, real-world environments and inherently spatial datasets to the extent possible, thereby minimizing reliance on the virtual and its limitations and idiosyncrasies. Such an approach requires an immersive HDLA facility. One such facility is Virginia Tech Cube. It is a configurable immersive studio measuring 50x40x32 feet. It is equipped with 24 motion-capture cameras and 149 of individually addressable loudspeakers distributed over the ceiling and walls to create a flexible and powerful immersive audio environment. Since its introduction in 2013, the facility has been used in dozens of research projects, ranging from scientific to artistic.

To assess the impact of the proposed immersive exocentric sonification, the authors are currently leveraging the said Cube through an externally funded effort to study its utility for scientific data analysis, as well as to compare its

capacity to the existing approaches, including virtual counterparts. Of particular interest is the capacity of the immersive exocentric sonification and its unique affordances to reveal correlations and inter-relationships between measured variables that are not easily revealed by conventional analysis techniques, such as visual inspection of standard two- and three-dimensional graphs. Our project title is SADIE, an acronym for Spatial Audio Data Immersive Experience. The main objectives of the study are to experiment with various sonification techniques, to uncover differences in spatial aural perception between the traditional egocentric approaches and an immersive exocentric environment, and to determine whether test subjects within the exocentric environment can experience an enhanced ability to detect subtle relationships between variables that represent real-world data.

7. SADIE

The SADIE project utilizes inherently spatial data representing an esoteric realm in Earth’s geophysical environment, the ionosphere. We focus on the altitude range occupied by satellites in low-Earth orbit (LEO), where satellites, suborbital rockets, and ground-based radars have accumulated data since the dawn of the space age. This medium is inherently complex, consisting of three distinct particle populations: neutral gases, ions, and electrons. All three gases occupy the same volume of space simultaneously, and they interact with each other via collisions, pressure gradients, and electromagnetic forces. These forces affect the temperatures, velocities, densities, and global distributions of all three species. In addition, the geomagnetic field constrains the motions of the charged particles, but does not affect the neutrals. All of the physical parameters described above vary as functions of season, location, solar activity levels, and time. The result is an extremely complex domain that is only partially understood despite decades of research.

The dataset is inherently four-dimensional, and is commonly represented with latitude, longitude, altitude, and time as the independent variables. By initially restricting our study to a specific range of altitudes we collapse the 4D space into a 3D domain that can be represented as sounds in the Cube. A test subject immersed in the environment hears sounds from all directions that represent the ion, electron, and neutral gas densities and temperatures. These are generally enhanced in areas where sunlight is most intense, so in a simulated (and temporally accelerated) 24-hour day a user standing at the center of the Cube hears the sound field representing these variables rotate around his/her observing point. In addition, the medium has natural variations in the fundamental variables that change significantly with latitude. For example, auroras occur primarily at high latitudes, while other geophysical phenomena are restricted to mid and equatorial domains. Naturally occurring changes in solar radiation fluxes, coupled with seasonal and diurnal variations of the polar axis relative to the ecliptic plane, provide a continually varying environment that is inherently complex, and difficult to analyze with conventional techniques.

Since the data maps naturally onto the 3D space represented by the Cube, we limit potential idiosyncrasies that could arise due to arbitrary spatial mapping. Our approach mimics as closely as possible the human interaction with real world stimuli, which can emanate from any direction with minimal technology-induced idiosyncrasies. SADIE users are free to navigate the ensuing immersive

exocentric environment, and by doing so they fundamentally change the amplitude relationships between spatially distributed aural stimuli, as well as azimuth and elevation relationships with the patterns of interest. By moving within the space users modify and enhance their natural interaction with the environment—they build a location-aware image of the sonified data and use motion to improve their ability to pinpoint specific sources. We see this initial case study as a fundamental step in uncovering the extent of the human capacity to perceive and process exocentric sonification of data, and we look forward to reporting on the outcomes at the next conference.

8. CONCLUSION

In this paper the authors have presented a review of the 21st century sonification literature from the several key scholarly communities. By reviewing similarities and differences among the cited research, the authors have presented a case for a new vector in sonification research—immersive exocentric sonification that leverages the way humans interact with the real world while minimizing any potential idiosyncrasies that may arise from interacting with the real world through egocentric technology (e.g. head motion and location). Further, by clarifying its purpose and potential, this paper presents immersive exocentric sonification as a holistic and potentially unifying approach to studying human spatial aural perception. We anticipate that this approach may challenge and redefine the traditional understanding of its limits. Lastly, we briefly describe an example project designed to investigate advantages in data analysis and understanding that may accrue from using a unique facility conducive to the newly proposed approach.

9. ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. 1748667.

10. REFERENCES

- [1] K. Paul A., “Cognitive load theory: implications of cognitive load theory on the design of learning,” *Learning and Instruction*, vol. 12, no. 1, pp. 1–10, Feb. 2002.
- [2] S. A. Brewster, “Using Non-Speech Sound to Overcome Information Overload,” *Displays*, vol. 17, pp. 179–189, 1997.
- [3] K. Hussein, E. Tilevich, I. I. Bukvic, and S. Kim, “Sonification design guidelines to enhance program comprehension,” in *IEEE 17th International Conference on Program Comprehension*, 2009. ICPC ’09, 2009, pp. 120–129.
- [4] P. Lennox and T. Myatt, “Concepts of perceptual significance for composition and reproduction of explorable surround sound fields,” 2007.
- [5] N. Mariette, “Mitigation of Binaural Front-Back Confusions by Body Motion in Audio Augmented Reality,” Jun. 2007.
- [6] G. Kramer et al., “Sonification Report: Status of the Field and Research Agenda,” *Faculty Publications*, Department of Psychology, 1999.
- [7] T. Hermann and H. Ritter, “Listen to your data: Model-based sonification for data analysis,” *Advances in intelligent computing and multimedia systems*, vol. 8, pp. 189–194, 1999.
- [8] T. Nasir and J. C. Roberts, “Sonification of spatial data,” 2007.
- [9] T. Hemann, “Taxonomy and Definitions for Sonification and Auditory Display,” in *14th International Conference on Auditory Display*, Paris, France June 24 - 27, 2008, Paris, France, 2008, p. 8.
- [10] S. Bly, “Presenting information in sound,” in *Proceedings of the 1982 conference on Human factors in computing systems*, 1982, pp. 371–375.
- [11] ACM Digital Library, “Results ACM DL: sonification spatialization.” [Online]. Available: <http://dl.acm.org/results.cfm?query=%252Bsonification+%252Bspatialization>. [Accessed: 09-Nov-2016].
- [12] W. Heuten, N. Henze, and S. Boll, “Interactive Exploration of City Maps with Auditory Torches,” in *CHI ’07 Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, 2007, pp. 1959–1964.
- [13] S. Harada, H. Takagi, and C. Asakawa, “On the Audio Representation of Radial Direction,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2011, pp. 2779–2788.
- [14] K. M. Franklin and J. C. Roberts, “A path based model for sonification,” in *Eighth International Conference on Information Visualisation*, 2004. IV 2004. Proceedings, 2004, pp. 865–870.
- [15] R. L. Klatzky, J. R. Marston, N. A. Guidice, R. G. Golledge, and J. M. Loomis, “Cognitive Load of Navigating Without Vision When Guided by Virtual Sound Versus Spatial Language,” *Journal of Experimental Psychology*, Vol. 12, No. 4, p. 9, 2006.
- [16] R. Ramloll, W. Yu, S. Brewster, B. Riedel, M. Burton, and G. Dimigen, “Constructing sonified haptic line graphs for the blind student: first steps,” in *Proceedings of the fourth international ACM conference on Assistive technologies*, 2000, pp. 17–25.
- [17] M. E. Brittell, “Line Following: A Path to Spatial Thinking Skills,” in *CHI ’11 Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, 2011, pp. 1753–1758.
- [18] M. Jeon and B. N. Walker, “Spindex (Speech Index) Improves Auditory Menu Acceptance and Navigation Performance,” *ACM Trans. Access. Comput.*, vol. 3, no. 3, p. 10:1–10:26, Apr. 2011.
- [19] A. Pendse, M. Pate, and B. N. Walker, “The accessible aquarium: identifying and evaluating salient creature features for sonification,” in *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, New York, NY, USA, 2008, pp. 297–298.
- [20] N. Schaffert, K. Mattes, and A. O. Effenberg, “A Sound Design for Acoustic Feedback in Elite Sports,” in *Auditory Display*, S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, Eds. Springer Berlin Heidelberg, 2010, pp. 143–165.
- [21] G. E. J. Bolton, *Reflective Practice: Writing and Professional Development*, 3 edition. Los Angeles, Calif: SAGE Publications Ltd, 2010.
- [22] T. Smith et al., “Exploring Gesture Sonification to Support Reflective Craft Practice,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in*

- Computing Systems, New York, NY, USA, 2015, pp. 67–76.
- [23] J. Stienstra, K. Overbeeke, and S. Wensveen, “Embodying complexity through movement sonification: case study on empowering the speed-skater,” in Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction: Facing Complexity, New York, NY, USA, 2011, pp. 39–44.
 - [24] J. Harris, S. Vance, O. Fernandes, A. Parnandi, and R. Gutierrez-Osuna, “Sonic Respiration: Controlling Respiration Rate Through Auditory Biofeedback,” in CHI ’14 Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA, 2014, pp. 2383–2388.
 - [25] S. Hasegawa, S. Ishijima, F. Kato, H. Mitake, and M. Sato, “Realtime sonification of the center of gravity for skiing,” in Proceedings of the 3rd Augmented Human International Conference, New York, NY, USA, 2012, p. 11:1–11:4.
 - [26] M. Matsubara, H. Terasawa, H. Kadone, K. Suzuki, and S. Makino, “Sonification of muscular activity in human movements using the temporal patterns in EMG,” in Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific, 2012, pp. 1–5.
 - [27] Y. Tsubouchi and K. Suzuki, “BioTones: A wearable device for EMG auditory biofeedback,” in 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2010, pp. 6543–6546.
 - [28] L. Chiari, M. Dozza, A. Cappello, F. B. Horak, V. Macellari, and D. Giansanti, “Audio-biofeedback for balance improvement: an accelerometry-based system,” *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 12, pp. 2108–2111, 2005.
 - [29] A. Bikaki and A. Floros, “An RSS-feed Auditory Aggregator Using Earcons,” in Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound, New York, NY, USA, 2011, pp. 95–100.
 - [30] H. J. Song and K. Beilharz, “Aesthetic and auditory enhancements for multi-stream information sonification,” in Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts, New York, NY, USA, 2008, pp. 224–231.
 - [31] T. Nasir, “Geo-sonf: Spatial sonification of contour maps,” in IEEE International Workshop on Haptic Audio visual Environments and Games, 2009. HAVE 2009, 2009, pp. 141–146.
 - [32] R. McIlraith, P. Walton, and J. Brereton, “The Spatialised Sonification of Drug-Enzyme Interactions,” in The 21st International Conference on Auditory Display (ICAD 2015), 2015, pp. 323–324.
 - [33] S. D. Beck, “The Immersive Computer-controlled Audio Sound Theater: History and Current Trends in Multi-modal Sound Diffusion,” in SIGGRAPH 2009: Talks, New York, NY, USA, 2009, p. 41:1–41:1.
 - [34] Y. Vazquez-Alvarez, M. P. Aylett, S. A. Brewster, R. V. Jungenfeld, and A. Virolainen, “Designing Interactions with Multilevel Auditory Displays in Mobile Audio-Augmented Reality,” *ACM Trans. Computer-Human Interaction*, vol. 23, no. 1, p. 3:1–3:30, Dec. 2015.
 - [35] F. Ribeiro, D. Florencio, P. A. Chou, and Z. Zhang, “Auditory augmented reality: Object sonification for the visually impaired,” in 2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSp), 2012, pp. 319–324.
 - [36] A. Neidhardt and A. Rppel, “Multiplayer Audio-only Game: Pong on a Massive Multichannel Loudspeaker System,” in Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound, New York, NY, USA, 2012, pp. 130–134.
 - [37] L. Picinali, B. Menelas, B. F. Katz, and P. Bourdot, “Evaluation of a haptic/audio system for 3-d targeting tasks,” in Audio Engineering Society Convention 128, 2010.
 - [38] C. Bennett, C. N. Leider, and R. McNeer, “Application of audio engineering and psychoacoustic principles to audible medical alarms,” in Audio Engineering Society Convention 135, 2013.
 - [39] D. Avissar, C. N. Leider, C. Bennett, and R. Gailey, “An Audio Game App Using Interactive Movement Sonification for Targeted Posture Control,” in Audio Engineering Society Convention 135, 2013.
 - [40] G. Courtois, P. Marmaroli, H. Lissek, Y. Oesch, and W. Balande, “Binaural hearing aids with wireless microphone systems including speaker localization and spatialization,” in Audio Engineering Society Convention 138, 2015.
 - [41] S. Lepa, S. Weinzierl, H.-J. Maempel, and E. Ungeheuer, “Emotional Impact of Different Forms of Spatialization in Everyday Mediatized Music Listening: Placebo or Technology Effects?,” in Audio Engineering Society Convention 136, 2014.
 - [42] M. J. Morrell and J. D. Reiss, “Inherent doppler properties of spatial audio,” in Audio Engineering Society Convention 129, 2010.
 - [43] R. Johannes, J.-W. Beh, W.-S. Gan, and E.-L. Tan, “Investigation of 3-D Audio Rendering with Parametric Array Loudspeakers,” in Audio Engineering Society Convention 128, 2010.
 - [44] Z. Seldess, “MIAP: Manifold-Interface Amplitude Panning in Max/MSP and Pure Data,” in Audio Engineering Society Convention 137, 2014.
 - [45] J. Ratcliffe, “MotionMix: A Gestural Audio Mixing Controller,” in Audio Engineering Society Convention 137, 2014.
 - [46] J. J. Lpez, A. Gonzlez, and L. Fuster, “Room compensation in wave field synthesis by means of multichannel inversion,” in Applications of Signal Processing to Audio and Acoustics, 2005. IEEE Workshop on, 2005, pp. 146–149.
 - [47] W. Zhang and T. D. Abhayapala, “Three Dimensional Sound Field Reproduction Using Multiple Circular Loudspeaker Arrays: Functional Analysis Guided Approach,” *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 22, no. 7, pp. 1184–1194, Jul. 2014.
 - [48] M. Sikora, B. Ivancevic, and K. Jambrosic, “Use of acoustic simulation and visualization for revitalization of ancient buildings,” in Video/Image Processing and Multimedia Communications 4th EURASIP-IEEE Region 8 International Symposium on VIPromCom, 2002, pp. 121–125.
 - [49] W. Page, B. Schmidt, and P. Driessen, “Rendering sound and images together,” in 2013 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), 2013, pp. 395–399.
 - [50] A. A. A. Ibrahim, R. M. Andrias, G. S. Leng, D. Nizam, and L. Penn Jack, “Reliability of Usability Inspection

- for Sonification Applications,” in 2009 Fifth International Conference on MEMS, NANO, and Smart Systems (ICMENS), 2009, pp. 203–207.
- [51] A. A. A. Ibrahim, F. M. Yassin, S. Sura, and R. M. Andrias, “Overview of design issues and evaluation of sonification applications,” in 2011 International Conference on User Science and Engineering (i-USER), 2011, pp. 77–82.
- [52] R. Sarkar, S. Bakshi, and P. K. Sa, “Review on image sonification: A non-visual scene representation,” in 2012 1st International Conference on Recent Advances in Information Technology (RAIT), 2012, pp. 86–90.
- [53] K. M. Franklin and J. C. Roberts, “Pie chart sonification,” in Seventh International Conference on Information Visualization, 2003. IV 2003. Proceedings, 2003, pp. 4–9.
- [54] M. Pec, M. Bujacz, P. Strumillo, and A. Materka, “Individual HRTF measurements for accurate obstacle sonification in an electronic travel aid for the blind,” in International Conference on Signals and Electronic Systems, 2008. ICSES '08, 2008, pp. 235–238.
- [55] R. R. A. Faria, M. K. Zuffo, and J. A. Zuffo, “Improving spatial perception through sound field simulation in VR,” in Proceedings of the 2005 IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2005. VECIMS 2005, 2005, p. 6 pp.-pp.
- [56] A. Stefik, K. Fitz, and R. Alexander, “Layered Program Auralization: Using Music to Increase Runtime Program Comprehension and Debugging Effectiveness,” in 14th IEEE International Conference on Program Comprehension, 2006. ICPC 2006, 2006, pp. 89–93.
- [57] G. Kramer, Auditory display: Sonification, audification, and auditory interfaces. Perseus Publishing, 1993.
- [58] G. L. Young, “Human Ecology as an Interdisciplinary Concept: A Critical Inquiry,” in Advances in Ecological Research, vol. 8, A. MacFadyen, Ed. Academic Press, 1974, pp. 1–105.
- [59] P. D. Adamczyk, “Seeing Sounds: Exploring Musical Social Networks,” in Proceedings of the 12th Annual ACM International Conference on Multimedia, New York, NY, USA, 2004, pp. 512–515.
- [60] G. Baier, T. Hermann, and U. Stephani, “Event-based sonification of EEG rhythms in real time,” Clinical Neurophysiology, vol. 118, no. 6, pp. 1377–1386, Jun. 2007.
- [61] R. MacVeigh and D. Jacobson, “Increasing the Dimensionality of a Geographic Information System (GIS) Using Auditory Display,” Jun. 2007.
- [62] T. Stockman, N. Rajgor, O. Metatla, and L. Harrar, “The design of interactive audio soccer,” Jun. 2007.
- [63] D. R. Begault, E. M. Wenzel, R. Shrum, and J. Miller, “A virtual audio guidance and alert system for commercial aircraft operations,” in Proceedings of the Third International Conference on Auditory Display, 1996, vol. 117122.
- [64] B. Carty and V. Lazzarini, “Binaural HRTF based spatialisation: New approaches and implementation,” in DAFx 09 proceedings of the 12th International Conference on Digital Audio Effects, Politecnico di Milano, Como Campus, Sept. 1-4, Como, Italy, 2009, pp. 1–6.
- [65] B. E. Treeby, R. M. Paurobally, and J. Pan, “Decomposition of the HRTF from a Sphere with Neck and Hair,” Jun. 2007.
- [66] J. C. B. Torres, M. R. Petraglia, and R. A. Tenenbaum, “HRTF modeling for efficient auralization,” in 2003 IEEE International Symposium on Industrial Electronics, 2003. ISIE '03, 2003, vol. 2, pp. 919–923 vol. 2.
- [67] G. Wersenyi, “Simulation of Small Head-Movements on a Virtual Audio Display Using Headphone Playback and HRTF Synthesis,” Jun. 2007.
- [68] Y. Kawai, M. Kobayashi, H. Minagawa, M. Miyakawa, and F. Tomita, “A support system for visually impaired persons using three-dimensional virtual sound,” TRANSACTIONS-INSTITUTE OF ELECTRICAL ENGINEERS OF JAPAN C, vol. 120, no. 5, pp. 648–655, 2000.
- [69] M. Gröhn, T. Lokki, and T. Takala, “Localizing Sound Sources in a CAVE-Like Virtual Environment with Loudspeaker Array Reproduction,” Presence: Teleoperators and Virtual Environments, vol. 16, no. 2, pp. 157–171, Mar. 2007.
- [70] G. Dubus and R. Bresin, “A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities,” PLOS ONE, vol. 8, no. 12, p. e82491, Dec. 2013.
- [71] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,” Journal of the Audio Engineering Society, vol. 45, no. 6, pp. 456–466, 1997.
- [72] Y. Vazquez-Alvarez, M. P. Aylett, S. A. Brewster, R. von Jungfeld, and A. Virolainen, “Multilevel Auditory Displays for Mobile Eyes-free Location-based Interaction,” in Proceedings of the Extended Abstracts of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, New York, NY, USA, 2014, pp. 1567–1572.
- [73] B. Arons, “A review of the cocktail party effect,” Journal of the American Voice I/O Society, vol. 12, no. 7, pp. 35–50, 1992.
- [74] T. Schmele, J. A. Romero, T. A. Troge, N. Ruiter, and M. Zapf, Sonifying multichannel ultrasound data for periphonic loudspeaker array. Georgia Institute of Technology, 2015.
- [75] I. I. Bukvic, “3D TIME-BASED AURAL DATA REPRESENTATION USING D4 LIBRARY’S LAYER BASED AMPLITUDE PANNING ALGORITHM,” presented at the International Conference on Auditory Display, Canberra, Australia, 2016.
- [76] M. M. Boone, E. N. Verheijen, and P. F. Van Tol, “Spatial sound-field reproduction by wave-field synthesis,” Journal of the Audio Engineering Society, vol. 43, no. 12, pp. 1003–1012, 1995.
- [77] J. Ahrens and S. Spors, “Focusing of virtual sound sources in higher order ambisonics,” in Audio Engineering Society Convention 124, 2008.
- [78] R. Edsall and K. L. Larson, “Decision making in a virtual environment: Effectiveness of a semi-immersive ‘decision theater’ in understanding and assessing human-environment interactions,” in Proceedings of AutoCarto, 2006, vol. 6, pp. 25–28.

SEEKING A REFERENCE FRAME FOR CARTOGRAPHIC SONIFICATION

Megen Brittell

Spatial Computation, Cognition, and Complexity Laboratory
Department of Geography, University of Oregon
Eugene, Oregon, USA
megen@uoregon.edu

ABSTRACT

Sonification of geospatial data must situate data values in two (or three) dimensional space. The need to position data values in space distinguishes geospatial data from other multi-dimensional data sets. While cartographers have extensive experience preparing geospatial data for visual display, the use of sonification is less common. Beyond availability of tools or visual bias, an incomplete understanding of the implications of parameter mappings that cross conceptual data categories limits the application of sonification to geospatial data. To catalyze the use of audio in cartography, this paper explores existing examples of parameter mapping sonification through the framework of the geographic data cube. More widespread adoption of auditory displays would diversify map design techniques, enhance accessibility of geospatial data, and may also provide new perspective for application to non-geospatial data sets.

Index Terms— cartography, geography, geospatial data, map, parameter mapping, sonification, audio

1. INTRODUCTION

Geospatial data are characterized by their position in space and time. Consideration of the spatio-temporal properties of data reveals patterns that could be missed when treating locations or time values as generic numeric variables in a multi-variate data set. Typical assumptions of independence between observations do not hold due to spatial and temporal autocorrelation. Further, parameter mappings that overlook organization of data along spatial or temporal dimensions compel the listener to mentally reconstruct that organization. The strong temporal qualities of audio have provided a natural parameter mapping for time series data (e.g., [1]); but, the effective representation of (geo-)spatial data and the use of spatial audio remain open questions. Organization of data and sonification into binary categories of “spatial” and “non-spatial” [2] take an initial step toward addressing the peculiar needs of geospatial data, but further examination of the “spatial” category is warranted.

Specialized systems have emerged to handle geospatial data. Geographic information systems combine data structures and

Population Density in Oregon (2010)

Figure 1: Maps of population density in the state of Oregon are graphic examples of thematic maps. These maps use color to depict population density (*attribute*); darker colors indicate higher population density. Geographic coordinates (*location*) determine the positions of features within the map. Notably, both maps depict the same underlying data. Data recorded by Census tracts (left) have been aggregated by county (right) to demonstrate one type of cartographic generalization. *Data Source: U.S. Census Bureau*

computer algorithms to support the unique needs of geospatial data storage, processing, and efficient query. Statistical methods that accommodate spatial and temporal autocorrelation have been designed (e.g., [3]). Evidence from psychology and neuroscience suggests that the human brain has specialized mechanisms for encoding and processing spatial data (e.g., [4]). And, design and use of maps have been studied as communication channels for geospatial data (e.g., [5]).

Thematic maps and reference (or navigation) maps constitute two common types of geographic map. Thematic maps depict the location of attribute values over geographic space. Among their varied purposes, thematic maps encourage the map reader to notice spatial patterns. For example, a map of Oregon that depicts population density shows the majority of the population living on the west side of the state (Figure 1). The map does not explicitly declare this pattern; the map reader interprets the information they perceive from the map. In contrast, navigational maps facilitate route learning, guide movement through the physical environment. Experience navigating through physical space inspires a metaphor for exploration of a data space [6], creating a connection between the two map types. But this paper challenges that connection and emphasizes sonification of geospatial data in thematic maps.

Auditory display of geospatial data is not a new idea. Although the majority of modern cartographic display techniques fall in the realm of graphic design, widespread and inexpensive

This work supported in part by NSF DDRI Grant #1634086.

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

Figure 2: Conceptual categorization of data across location, time, and attribute (left) is an established framework for organizing geospatial data [7]. Each of these categories may be further subdivided across multiple dimensions (center). The number of dimensions depends on the specific data set under consideration. Displays of geospatial data create parameter mappings that stay within or translate across the axes of the geographic data cube (right). For example, depiction of data location to screen coordinates demonstrates a mapping within the location axis (A). And among many possible examples of translations: turn-by-turn directions translate spatial information into a temporal sequence (B), time series data represented as line graphs convert time into a spatial location on the x-axis (C), scatterplots translate attribute values into locations (D), and time series data may be represented with a color attribute in a static graphic (E).

sound synthesis in the early 1990s prompted cartographers to consider auditory displays. In parallel with expanded use of audio in more general human-computer interfaces, cartographers used sound to represent uncertainty in remote sensing data [8] and to highlight anomalies in data [9]. However, initial interest and optimism about the use of audio in geospatial data displays has waned. Recent availability of browser-based audio tools revived some interest in augmenting web-based interactive maps with auditory elements (see review in [10]). Still, proliferation of visual displays and challenges in the design of effective parameter mappings for sonification of spatial data, have meant that the use of sonification in cartographic design has been low.

To support more widespread adoption of sonification in cartography, this paper draws attention to parameter mappings that cross the conceptual category boundary between data location and the temporal dimension of an auditory display. The next section provides a brief overview of geospatial data, associated data transformation techniques, and example cartographic design guidelines. The third section considers existing sonification approaches to (geo-)spatial data, highlighting the role of space and time in those displays. A final section reflects on the current status of sonification in cartography, posing an open question about the implications of cross-category parameter mappings.

2. GEOSPATIAL DATA AND CARTOGRAPHY

Geographers and cartographers have a long history of organizing, transforming, and representing geospatial data. The discussion revolves around the location and time axes of the geographic data cube as a context in which to examine the application of parameter mapping sonification to geospatial data.

2.1. The Geographic Data Cube

An emphasis on location, or position within a two (or three) dimensional reference frame, distinguishes geospatial data from other multi-dimensional data sets. The geographic data cube [7], an established conceptual framework from geography, organizes data in three categories: attribute, time, and spatial location. The

three axes represent three categories of data inherent and necessary to any geospatial data set (Figure 2, left). Omitting any category is detrimental to interpretation of the data set. For example, a temperature attribute value carries little information without the context of when and where the observation was recorded. While all three categories are necessary, a category may be held constant for a given map product. As a case in point, the maps in Figure 1 depict a range of locations and attribute values, while holding time constant (the year 2010).

The three categories each consist of one or more dimensions (Figure 2, center). Of particular interest in the case of geospatial data is the location axis that has two or more dimensions. A location may be recorded as two-dimensional coordinates on the surface of the earth, and elevation may be added as a third location dimension. Notably these dimensions of spatial location are orthogonal to one another, and values along these dimensions tend to be autocorrelated: “everything is related to everything else, but near things are more related than distant things” (Tobler’s first law of geography, [11]).

2.2. Cartographic Processing

Cartographers have developed techniques to transform geospatial data in preparation for display in geographic maps. Two such techniques, projections and generalization, are described here. Projections, or systematic translations of spherical geographic coordinates (e.g., latitude and longitude) into two-dimensional planar (page or screen) coordinates, are standardized data transformations. Locations in the physical world have a one-to-one mapping with locations in the two-dimensional frame (although relationships between locations are inevitably, but predictably, distorted). Even though projection introduces error in the location data, the resulting two-dimensional model has been applied to good effect in helping people conceptually understand and reason about phenomena on the surface of the earth.

Print and screen technologies have traditionally necessitated the dimension reduction from three to two dimensions for map production. Even though the physical world is three dimensional and 3D rendering technologies are emerging, the most common map displays – maps printed on paper, displayed on computer

screens – are still constrained to a two dimensional spatial plane. Although this spatial dimension reduction affords a simplified special case of geospatial data, projection may not be required for auditory representations [2] and may be propagating a visual bias into auditory display design.

As a second example, generalization eliminates unnecessary clutter or emphasizes specific characteristics of the data [12], and may be applied to data from any of the three data cube categories. Cartographers have developed many generalization routines for geospatial data in visual displays, and analogous approaches have also been applied in sonification, e.g., by emphasizing “distinct” data values [6]. Generalization transforms data into a simpler or alternative form. The degree to which such tools are applied depends on several factors including display technology and modality.

Generalization helps focus attention on the message that the map was designed to convey, but introduces error and removes detail (compare, e.g., the two maps in Figure 1). In some sonification prototypes, heavy-handed generalization or simplification has made evaluation feasible, but researchers recognize that it is not realistic data (e.g., [13]) and is not functionally equivalent to its visual or tactile counterparts. While arbitrarily reducing data complexity is not a long term solution, generalization may still play a role in the design of an effective parameter mapping.

2.3. Symbolization of Geospatial Data

Map symbolization, and more generally graphic or sonification design processes, creates relationships between data values and display values. This section briefly describes the treatment of location and time data in typical cartographic design, and considers how these approaches do or do not apply to auditory displays. In some cases, a direct relationship exists between the data and an analogous display parameter, but a direct relationship is not strictly necessary, and mappings may cross the boundaries outlined by the categories of the geographic data cube (Figure 2, right).

Cartographic design commonly presents geospatial locations as corresponding locations within a graphic map display. The spatial arrangement of light receptors in the retina of the eye and the projection of that two dimensional organization into higher level processing areas of the brain [14] further support a direct mapping of the location of geospatial data to location within a visual display. Similarly, this “easy” choice for representing location is also observed in tactile map graphics: the position of symbols on the map correspond to location in the real world.

The relative ease with which humans visually perceive spatial relationships has also lead to the use of location to depict non-spatial data. For example, *attribute* values are represented by a *location* in space in a scatterplot (“spatialization”, [15]) or iconographic display [16]. While the translation across axes of the geographic data cube can be effective, there is still a correlation between the ease and usability of a display and the alignment between the dimensionality of the data and that of the display.

However, this approach does not directly generalize to auditory displays. Despite what may appear at a cursory glance to be an easy direct mapping of location attributes of the data to monaural or binaural spatial audio cues, on closer inspection, several limitations become apparent. On the plus side, spatialized audio targets the human ability to localize sounds [17] and can

“leverage the natural affordances of the space and the user’s location within the sound field” [18]. When applied to sonification of non-spatial data, spatialized audio has been reported to facilitate segregation between data streams [6, 19] and to provide orienting cues for the use of a haptic mouse in the absence of visual feedback [13]. But, spatialized audio is relative to the listener, and relies on either distance cues or elevation cues to determine a position in a two dimensional space. The egocentric perspective that relies on distance cues may be sub-optimal for communicating relationships between data points (e.g., [20]) and accuracy of perceived location varies across different axes of physical space (e.g., poor resolution in conveying elevation [21, 22]). These nuances highlight open questions about the use of spatialized audio to depict geospatial data.

As an alternative to spatialized audio, cartography has explored depiction of one of the spatial dimensions on the time axis. Time has been used to depict both temporal and spatial data. Interactive maps have provided functionality to produce animations of geospatial data that change over time (see review in [23]). Animated visual displays are consistent with the general recommendation to use time to represent temporal data [24]. As cartographers and sonification designers explored stand-alone auditory displays of geospatial data, the time dimension was also co-opted for the display of location. In 2000, Saue [6] proposed the idea of spatial data “temporalization,” which translates location data into the time dimension of an auditory display using a metaphor of walking through an environment. The depiction of location data over time has since been a common approach to sonification of geospatial data (either alone or with redundant location information from other sensory modalities). A drawback to this approach however, is that the reduction of two dimensional space to a linear sequence takes longer to perceive than a visual display of the same data and data that is spatially proximal may be separated by extended time intervals. The listener faces the challenging task of remembering a long sequence of data values and mentally reconstructing two dimensional space. And in the context of accessibility, the resulting display, which may take more than a minute to render depending on the size of the data set, lacks functional equivalence with its visual counterpart, which might take only a few seconds to perceive and mentally process.

The next section expounds on the auditory display of geospatial data, reviewing a number of sonification examples.

3. SONIFICATION OF GEOSPATIAL DATA

As sound production and real-time audio rendering became possible in computer hardware and later through software, sonification emerged as a feasible data display modality across many application domains [25, 26]. Although researchers have recognized the challenge of presenting multiple variables in a single audio stream (e.g., [24]), there are also many success stories. Sonification has been used to depict large, complex, and multi-dimensional scientific data sets including recordings of solar winds [27] and climate data [28]. Still under investigation is effective and efficient sonification of geospatial data.

Amid ongoing advances in technology and expanding adoption of sonification across multiple disciplines, geographers too considered ways to incorporate audio into map design. Audio was viewed as a “largely untapped medium for the communication of cartographic data” [8] and a “means of expanding the representational repertoire of cartography and visualization” [9].

Geographic applications began pairing auditory representations of non-spatial data with visual map displays. For example, Fisher [8] represented uncertainty information associated with classified remote sensing data, and Krygier [29] augmented animated graphics with redundant information through natural sounds. Over the following decades, however, interest among cartographers waned.

Although the introduction of audio capability in web browsers spurred interest in multimedia mapping (see review in [10]), initial difficulties adopting audio into cartographic design had already set the tone. Some geographers and cartographers came to doubt the potential of auditory displays to represent of spatial data over non-trivial spatial extents [30], or grew skeptical of any non-visual display of geospatial data [31]. Rather than justification for jumping ship and abandoning sonification, however, this skepticism could indicate a lack of an appropriate design. And, not all have abandoned auditory map display: “Rather, the use of sound forces us to rethink the very concept of the map as primarily a visual image of space that serves as a simple conveyer of information” [32].

The remainder of this section explores three groups of examples, organized by the role that audio plays in the display: audio-enhanced displays, multimodal displays, and stand-alone sonification.

In audio-enhanced displays, sonification of attribute data accompanies or enhances another display modality. Location information is conveyed through, e.g., vision [30, 10]. Interaction with a mouse, touchpad, or stylus triggers playback of an audio recording or renders a data value in audio that is specific to the selected location. For example, in an interactive web map the map user triggers rendering of a parameterized note or playback of a pre-recorded audio clip by selecting a location using a mouse-click or tapping a touchscreen. Within this group of audio-enhanced displays, the role of sonification is limited to the display of isolated non-spatial data values. The audio component of the display cannot stand alone; the display relies on an alternate display modality to communicate spatial location - the aspect of the data that makes it (geo-)spatial. And, the map reader must mentally integrate disparate sensory input streams to interpret the complete set of spatial and non-spatial data attributes.

In other multimodal displays, the auditory component of the display conveys partial or redundant information about spatial location. Location data is depicted in a two dimensional plane through, e.g., proprioceptive feedback [33, 34, 35, 36, 37, 38, 39], or a haptic device [40, 13]. While such audio-enhanced displays have found some success communicating spatial data, evaluations have found that users have difficulty interpreting spatial patterns, without a companion visual or tactile display [41, 38] or without providing contextual clues about the specific layout [42].

The number of examples of stand-alone sonification of (geo-)spatial data is more limited. Stand-alone auditory displays encode sufficient information in the auditory stream to convey location information independent of other display modalities. Flowers, Buhman, and Turnage [43] used frequency and time as two axes for an auditory display of scatterplots, depictions of (non-geographic) data points within a two-dimensional space. Alty and Rigas introduced *AudioGraph* [42], which used pairs of notes to represent coordinate locations within a display. Timbre indicated which axis is being represented and frequency encoded location along that axis. A virtual cursor traced shapes in the display, playing a pair of notes at each vertex. Specifically to

display geospatial data, Zhao et al. implemented a stand-alone auditory display in *iSonic* [21] that traverses the two dimensional geographic space following pre-established scan patterns. A virtual cursor moves through the display as auditory feedback indicates the data value at the cursor location. As the sounds play, the listener must mentally assemble the individual notes to recreate the two dimensional arrangement of objects in the display. Reports from evaluation of the *AudioGraph* display indicate some success communicating spatial information, but also difficulty interpreting overall patterns [42]. Over several years of development, the *iSonic* interface seems to have moved away from the audio-only display in favor of an interactive display with a spatial input device [35]. Such findings are consistent with other evaluations of non-visual display in which a heavy burden is placed on users’ working memory [44].

The parameter mappings employed in the examples mentioned in this section are summarized in Table 1. The table lists the location and attribute categories present in the data set (none of the example sonifications depicted temporal data), along with the respective auditory dimension in which that data was encoded. Listed parameter mappings are available in the respective systems, but may not be concurrently available. The set of examples is not intended to be exhaustive, but to illustrate the variety of approaches that have been explored. The next section reflects on trends among these example systems and poses an open question for future research.

4. REFLECTION AND OUTSTANDING CHALLENGES

The examples of geospatial data sonification listed above show diversity among parameter mappings, and shows that translations across categories of data and display dimensions are common. The categories themselves were not mutually exclusive. Amplitude, for example was used in a way that mimicked distance (*location*) in the physical world with closer features represented as louder sounds [34, 13]. But, amplitude was also used in a way that mimicked magnitude (*attribute*) [37, 10]. Both uses employ intuitive metaphors, and there is no single rule to assign amplitude to a category of display dimension. Within the location category, special attention is drawn to the distinction between egocentric and allocentric perspectives. Differences between the two perspectives have implications for interpretation of the display. Beyond categorization as spatial and non-spatial, parameter mappings may also need to address or accommodate the alignment of perspectives between the data and the display. Acknowledging the limitations of a simplified account of the examples, this summary offers one interpretation and provides a basis for discussion; for full details of the parameter mappings and their respective display systems, readers are referred to the original papers.

With an exponential number of possible combinations, the selection of auditory dimensions to serve as such a reference frame is neither obvious, nor trivial. Selecting a parameter mapping is complicated by perceptual limits on the number of auditory events that can be processed concurrently [45] and interactions between auditory dimensions (e.g., [46, 24]). Further, results from empirical evaluation of sonification parameters by experiment must be applied with caution when removed from the laboratory and applied to real-world data [46, 13] or generalized from pure tones to more complex musical sounds [47]. A two-dimensional auditory reference frame to support effective and efficient auditory displays of geospatial data is still illusive.

Table 1: Existing implementations of (geo-)spatial data sonification provide examples of parameter mapping sonification representing location and attribute data (none of the example systems depicted temporal data). Grey highlighting and **bold font** emphasize egocentric location information and translation across axes of the geographic data cube, respectively.

Reference, Content Domain (<i>System</i>)	Location	Attribute
<i>Audio-enhanced Displays (no auditory display of spatial location)</i>		
Bearman and Fisher [30], elevation, uncertainty (<i>ArcGIS extension</i>)		data value → frequency
Brauen [10], multiple airborne pollutants		data value → amplitude, audio clip
<i>Audio-enhanced Displays (with complementary visual, proprioceptive, tactile, or haptic display of location data)</i>		
Krueger and Gilden [33], named polygon features (<i>KnowWhereTM</i>)	relative location → audio icon	data value → speech
Daunys and Lauruska [34], shapes of polygon features	distance → amplitude relative location → audio icon	data value → speech data value → audio clip
Zhao et al. [35], choropleth map (<i>iSonic</i>)	x-location → azimuth	data value → frequency, speech, timbre
Nasir [36], surface contours (<i>Geo-Sonf</i>)	direction → azimuth, direction → time , distance → frequency, number of notes	
Adhitya and Kuuskankare [37], multiple building attributes (<i>Sonified Urban Masterplan</i>)	movement → onset time , distance → duration	
Brittall, Young, and Lobben [38], choropleth map (<i>mGIS</i>)	relative location → audio icon	data value → frequency
Kaklanis, Votis, and Tzovaras [40], reference maps (<i>Open Touch/Sound Maps</i>)	direction → azimuth, distance → frequency , relative location → audio icon	data value → speech
Geronazzo et al. [13], guided pointing, object location	relative location → azimuth relative location → elevation distance → amplitude	
Schito and Fabrikant [39], elevation	x-location → azimuth, frequency y-location → waveform , distance → frequency relative location → time	data value → duration , frequency
<i>Audio Displays (stand-alone auditory display of spatial location)</i>		
Flowers, Buhman, and Turnage [43], scatterplot	x-location → time , y-location → frequency	
Alty and Rigas [42], geometric shapes (<i>AudioGraph</i>)	coordinate value → frequency , coordinate axis → timbre , shape → time/order	
Zhao et al. [21], choropleth map (<i>iSonic</i>) [temporalization]	location → onset time relative location → audio icon x-coordinate → azimuth y-coordinate → frequency	data value → frequency, timbre

In the case of geospatial data, the need for data patterns to be simultaneously interpreted across multiple dimensions poses a unique problem. The goal is not to encourage separation or perceptual streaming, but to organize data within a two (or three) dimensional frame and communicate patterns that occur across those dimensions (c.f., [6, 19]). Recall at this point that the

objective of thematic maps is to communicate a general spatial pattern.

A trend across the (geo-)spatial sonification examples is reliance on time or temporal order. The sonification follows a cursor through the display controlled by either an interactive spatial input device or a pre-established scan path. The supporting

metaphor of movement through a space occupied by sound sources offer both affordances from real world experience and appealing simplicity in implementation. But the resulting sequential exploration is susceptible to variable or ambiguous interpretations prompted by spurious depiction of distance between locations that are proximally located (pre-planned and automated scan patterns) or emergent texture and patterns that are intertwined with the speed of movement (user-controlled spatial input devices). Time that elapses during movement directly corresponds with the timing of auditory events. This interplay between space and time leads to emergent properties of the sound, creating textures and patterns [16, 24, 37] that can be difficult to anticipate from the underlying data.

Harnessing the power of these emergent patterns means striking a balance between a parameter mapping that is “exact and rigorous” [48] and one that is more fluid: “It should be stressed that the sound tracks need to be constructed according to their own sound logic and cannot simply be reduced to the structure of data or other map variables” [49]. Again for an application in which general patterns are of interest, the ability to extract a single data value with high precision is not a goal of the display. For sonification designers who are steeped in a visual tradition, such as that of cartography, finding such a balance will require conscious effort to mitigate visual bias (e.g., [50]).

Even though geospatial data are not inherently visual, an occular-centric trend in cartography and GIS has emphasized visual displays. Mainstream cartographic designs often focus on graphic displays, and occasionally later append an auditory component of the display. The auditory display is an afterthought. It is subject to the decisions that were made to optimize the graphic display. A conscious effort is required to avoid temptations to translate visual displays into audio. Without diminishing the value of visual displays, the focus on map graphics allows implicit assumptions in the way that sighted researchers think about geospatial data to persist unnoticed and to creep into the tools for map production. For example, the implementation of the GeoTools library [51] tightly integrates the spatial data representations with the Swing graphics library [52], which is good for code optimization in implementing visual map displays but can complicate efforts to explore auditory displays that still rely on functionality to handle geospatial data. Map production that targets auditory displays earlier in the design process can help reverse some of the embedded visual bias.

In practice, the dominance of visual maps in print and on computer screens has lead to a scarcity of alternative cartographic display techniques and a reduction in accessibility of geospatial data, particularly for people with disabilities [53]. As noted in the context of multidimensional astronomy data, however, auditory display enhanced both accessibility for researchers who are blind, and conveyed patterns beyond those apparent in visual displays [25, 54]. Visual displays of geospatial data are good, but not sufficient to support a diverse population of map readers and growing volumes of scientific data. With efficient and usable designs, auditory displays could make a substantial contribution to the cartographer’s toolbox.

Using the geographic data cube to describe the structure of cartographic sonification reveals an open question: how does translation across the conceptual boundaries of the geographic data cube influence communication of geospatial data through sonification? By investigating this question, geographers and cartographers can join musicians and sound designers in pursuing

ways to think about auditory patterns that emerge from sonification of multi-dimensional data sets. A pursuit that seeks a reference frame for cartographic sonification at the intersection of exact science and expressive art.

5. ACKNOWLEDGEMENT

This work was developed in part through writing for my doctoral comprehensive exam; I would like to thank Dr. Amy Lobben and all members of my Dissertation Advisory Committee for their feedback and insights that helped shape this work.

6. REFERENCES

- [1] R. L. Alexander et al. “The bird’s ear view of space physics: Audification as a tool for the spectral analysis of time series data”. In: *Journal of Geophysical Research: Space Physics* 119.7 (2014), pp. 5259–5271. DOI: [10.1002/2014JA020025](https://doi.org/10.1002/2014JA020025).
- [2] T. Nasir and J. C. Roberts. “Sonification of Spatial Data”. In: *Proceedings of the 13th International Conference on Auditory Display*. Ed. by W. L. Martens, R. Quesnel, and G. Scavone. Montréal, Canada, June 2007. URL: <http://hdl.handle.net/1853/50035>.
- [3] D. G. Krige. “Essential basic concepts in mining geostatistics and their links with geology and classical statistics”. In: *South African Journal of Geology* 102.2 (1999), pp. 147–152.
- [4] A. Gomez et al. “Differential hippocampal and retrosplenial involvement in egocentric-updating, rotation and allocentric processing during online spatial encoding: an fMRI study”. In: *Frontiers in Human Neuroscience* 8.150 (2014). DOI: [10.3389/fnhum.2014.00150](https://doi.org/10.3389/fnhum.2014.00150).
- [5] A. Koláčny. “Cartographic Information—a Fundamental Concept and Term in Modern Cartography”. In: *The Cartographic Journal* 6.1 (1969), pp. 47–49. DOI: [10.1179/caj.1969.6.1.47](https://doi.org/10.1179/caj.1969.6.1.47).
- [6] S. Saue. “A model for interaction in exploratory sonification displays”. In: *Proceedings of the 6th International Conference on Auditory Display*. Ed. by P. R. Cook. Atlanta, Georgia, USA, Apr. 2000. URL: <http://hdl.handle.net/1853/50666>.
- [7] B. J. L. Berry. “Approaches to Regional Analysis: A Synthesis”. In: *Annals of the Association of American Geographers* 54.1 (1964), pp. 2–11. DOI: [10.1111/j.1467-8306.1964.tb00469.x](https://doi.org/10.1111/j.1467-8306.1964.tb00469.x).
- [8] P. F. Fisher. “Hearing the Reliability in Classified Remotely Sensed Images”. In: *Cartography and Geographic Information Systems* 21.1 (1994), pp. 31–36. DOI: [10.1559/152304094782563975](https://doi.org/10.1559/152304094782563975).
- [9] J. B. Krygier. “Sound and Geographic Visualization”. In: *Visualization in Modern Cartography*. Ed. by A. M. MacEachren and D. R. F. Taylor. Oxford, UK: Pergamon, 1994. Chap. 8, pp. 149–166. ISBN: 0-08-042416-3.

- [10] G. Brauen. "Interactive Audiovisual Design for Cartography: Survey, Prospects, and Example". In: *Developments in the Theory and Practice of Cybercartography Applications and Indigenous Mapping*. Ed. by D. R. F. Taylor. Vol. 5. Academic Press, 2014. Chap. 10, pp. 141–159. DOI: [10.1016/B978-0-444-62713-1.00010-6](https://doi.org/10.1016/B978-0-444-62713-1.00010-6).
- [11] W. R. Tobler. "A Computer Movie Simulating Urban Growth in the Detroit Region". In: *Economic Geography* 46 (1970), pp. 234–240. URL: <http://www.jstor.org/stable/143141>.
- [12] J. B. Krygier and D. Wood. *Making Maps: A Visual Guide to Map Design for GIS*. 2nd edition. New York, New York, USA: Guilford Press, 2011. ISBN: 978-1-60918-166-6.
- [13] M. Geronazzo et al. "Interactive spatial sonification for non-visual exploration of virtual maps". In: *International Journal of Human-Computer Studies* 85 (2016). Data Sonification and Sound Design in Interactive Systems, pp. 4–15. DOI: [10.1016/j.ijhcs.2015.08.004](https://doi.org/10.1016/j.ijhcs.2015.08.004).
- [14] C. Constantinidis and X.-J. Wang. "A Neural Circuit Basis for Spatial Working Memory". In: *The Neuroscientist* 10.6 (2004), pp. 553–565. DOI: [10.1177/1073858404268742](https://doi.org/10.1177/1073858404268742).
- [15] A. Skupin and S. I. Fabrikant. "Spatialization Methods: A Cartographic Research Agenda for Non-geographic Information Visualization". In: *Cartography and Geographic Information Science* 30.2 (Apr. 2003), pp. 99–119. DOI: [10.1559/152304003100011081](https://doi.org/10.1559/152304003100011081).
- [16] S. Smith, R. D. Bergeron, and G. G. Grinstein. "Stereophonic and Surface Sound Generation for Exploratory Data Analysis". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Seattle, Washington, USA, 1990, pp. 125–132. DOI: [10.1145/97243.97264](https://doi.org/10.1145/97243.97264).
- [17] J. C. Middlebrooks and D. M. Green. "Sound Localization by Human Listeners". In: *Annual Review of Psychology* 42 (1991), pp. 135–159. DOI: [10.1146/annurev.ps.42.020191.001031](https://doi.org/10.1146/annurev.ps.42.020191.001031).
- [18] S. Bland. "Researchers from music and engineering team up to turn big data into sound". In: *Virginia Tech News* (Nov. 2017). URL: <https://vtnews.vt.edu/articles/2017/11/icat-datasound.html>.
- [19] L. Brown et al. "Design Guidelines for Audio Presentation of Graphs and Tables". In: *Proceedings of the 9th International Conference on Auditory Display*. Ed. by E. Brazil and B. Shinn-Cunningham. Boston, MA, USA, July 2003. URL: <http://hdl.handle.net/1853/50454>.
- [20] S. Harada, H. Takagi, and C. Asakawa. "On the audio representation of radial direction". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Vancouver, British Columbia, Canada, 2011, pp. 2779–2788. ISBN: 978-1-4503-0228-9. DOI: [10.1145/1978942.1979354](https://doi.org/10.1145/1978942.1979354).
- [21] H. Zhao et al. "Sonification of Geo-referenced Data for Auditory Information Seeking: Design Principle and Pilot Study". In: *Proceedings of the 10th International Conference on Auditory Display*. Ed. by S. Barrass and P. Vickers. Sydney, Australia, July 2004. URL: <http://hdl.handle.net/1853/50918>.
- [22] B. F. G. Katz et al. "The Effect of Spatialization in a Data Sonification Exploration Task". In: *Proceedings of the 14th International Conference on Auditory Display*. Ed. by P. Susini and O. Warusfel. Paris, France, June 2008. URL: <http://hdl.handle.net/1853/49961>.
- [23] A. Lobben. "Classification and Application of Cartographic Animation". In: *The Professional Geographer* 55.3 (2003), pp. 318–328. DOI: [10.1111/0033-0124.5503016](https://doi.org/10.1111/0033-0124.5503016).
- [24] J. H. Flowers. "Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions". In: *Proceedings of the 11th International Conference on Auditory Display*. Ed. by E. Brazil. Limerick, Ireland, July 2005. URL: <http://hdl.handle.net/1853/50188>.
- [25] S. Barrass and G. Kramer. "Using sonification". In: *Multimedia Systems* 7.1 (Jan. 1999), pp. 23–31. DOI: [10.1007/s005300050108](https://doi.org/10.1007/s005300050108).
- [26] S. Barrass and C. Frauenberger. "A communal map of design in auditory display". In: *Proceedings of the 15th International Conference on Auditory Display*. Ed. by M. Aramaki et al. Copenhagen, Denmark, May 2009. URL: <http://hdl.handle.net/1853/51285>.
- [27] R. Alexander et al. "Sonification of Ace Level 2 Solar Wind Data". In: *Proceedings of the 16th International Conference on Auditory Display*. Ed. by E. Brazil. Washington, D.C., USA, June 2010. URL: <http://hdl.handle.net/1853/50049>.
- [28] S. St. George et al. "Making Climate Data Sing: Using Music-like Sonifications to Convey a Key Climate Record". In: *Bulletin of the American Meteorological Society* 98.1 (2017), pp. 23–27. DOI: [10.1175/BAMS-D-15-00223.1](https://doi.org/10.1175/BAMS-D-15-00223.1).
- [29] J. Krygier. "Mapping with Sound". In: *Proceedings of the Association of American Geographers*. 1993.
- [30] N. Bearman and P. F. Fisher. "Using sound to represent spatial data in ArcGIS". In: *Computers & Geosciences* 46 (2012), pp. 157–163. DOI: [10.1016/j.cageo.2011.12.001](https://doi.org/10.1016/j.cageo.2011.12.001).
- [31] C. Brown and J. McKinney. "Accessibility and Section 508". In: *Proceedings of the Federal GIS Conference*. Washington, D.C., USA, Feb. 2015. URL: http://proceedings.esri.com/library/userconf/fed15/papers/fed_123.pdf.
- [32] S. Caquard et al. "Designing sound in cybercartography: from structured cinematic narratives to unpredictable sound/image interactions". In: *International Journal of Geographical Information Science* 22.11-12 (Nov. 2008), pp. 1219–1245. DOI: [10.1080/13658810801909649](https://doi.org/10.1080/13658810801909649).

- [33] M. W. Krueger and D. Gilden. "KnowWhere: an audio/spatial interface for blind people." In: *Proceedings of the 4th International Conference on Auditory Display*. Ed. by J. Ballas. Palo Alto, California, USA, Nov. 1997. URL: <http://hdl.handle.net/1853/50747>.
- [34] G. Daunys and V. Lauruska. "Sonification System of Maps for Blind". In: *Universal Access in Human-Computer Interaction. Ambient Interaction*. Ed. by Constantine Stephanidis. LNCS 4555. Springer Berlin / Heidelberg, 2007. Chap. 16, pp. 349–352. DOI: [10.1007/978-3-540-73281-5_37](https://doi.org/10.1007/978-3-540-73281-5_37).
- [35] H. Zhao et al. "Data Sonification for Users with Visual Impairment: A Case Study with Georeferenced Data". In: *ACM Transactions on Computer-Human Interaction* 15.1 (May 2008), 4:1–4:28. DOI: [10.1145/1352782.1352786](https://doi.org/10.1145/1352782.1352786).
- [36] T. Nasir. "Geo-Sonf: Spatial sonification of contour maps". In: *IEEE International Workshop on Haptic Audio visual Environments and Games*. Nov. 2009, pp. 141–146. DOI: [10.1109/HAVE.2009.5356116](https://doi.org/10.1109/HAVE.2009.5356116).
- [37] S. Adhitya and M. Kuuskankare. "The Sonified Urban Masterplan (SUM) Tool: Sonification for Urban Planning and Design". In: *Proceedings of the 17th International Conference on Auditory Display*. Budapest, Hungary, June 2011. URL: <http://hdl.handle.net/1853/51918>.
- [38] M. Brittell, M. Young, and A. Lobben. "The mGIS: A Minimal Geographic Information System Accessible to Users Who Are Blind". In: *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Orlando, Florida: ACM, 2013, pp. 554–557. DOI: [10.1145/2525314.2525329](https://doi.org/10.1145/2525314.2525329).
- [39] J. Schito and S. I. Fabrikant. "Exploring maps by sounds: using parameter mapping sonification to make digital elevation models audible". In: *International Journal of Geographical Information Science* (2018), pp. 1–34. DOI: [10.1080/13658816.2017.1420192](https://doi.org/10.1080/13658816.2017.1420192).
- [40] N. Kaklanis, K. Votis, and D. Tzovaras. "Open Touch/Sound Maps: A system to convey street data through haptic and auditory feedback". In: *Computers & Geosciences* 57 (2013), pp. 59–67. DOI: [10.1016/j.cageo.2013.03.005](https://doi.org/10.1016/j.cageo.2013.03.005).
- [41] F. Delogu et al. "Non-visual exploration of geographic maps: Does sonification help?" In: *Disability and Rehabilitation: Assistive Technology* 5.3 (May 2010), pp. 164–174. DOI: [10.3109/17483100903100277](https://doi.org/10.3109/17483100903100277).
- [42] J. L. Alty and D. Rigas. "Exploring the use of structured musical stimuli to communicate simple diagrams: the role of context". In: *International Journal of Human-Computer Studies* 62.1 (2005), pp. 21–40. DOI: [10.1016/j.ijhcs.2004.08.003](https://doi.org/10.1016/j.ijhcs.2004.08.003).
- [43] J. H. Flowers, D. C. Buhman, and K. D. Turnage. "Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples". In: *Human Factors* 39.3 (1997), pp. 341–351. DOI: [10.1518/001872097778827151](https://doi.org/10.1518/001872097778827151).
- [44] M. Rice et al. "Design Considerations for Haptic and Auditory Map Interfaces". In: *Cartography and Geographic Information Science* 32.4 (2005), pp. 381–391. DOI: [10.1559/152304005775194656](https://doi.org/10.1559/152304005775194656).
- [45] D. K. McGookin and S. A. Brewster. "Understanding Concurrent Earcons: Applying Auditory Scene Analysis Principles to Concurrent Earcon Recognition". In: *ACM Transactions on Applied Perception* 1.2 (Oct. 2004), pp. 130–155. DOI: [10.1145/1024083.1024087](https://doi.org/10.1145/1024083.1024087).
- [46] J. G. Neuhoff, G. Kramer, and J. Wayand. "Sonification and the interaction of perceptual dimensions: Can the data get lost in the map?" In: *Proceedings of the 6th International Conference on Auditory Display*. Ed. by P. R. Cook. Atlanta, Georgia, USA, Apr. 2000. URL: <http://hdl.handle.net/1853/50681>.
- [47] P. Newell. "Sound, Decibels and Hearing". In: *Recording Studio Design*. 3rd edition. Focal Press, 2012. Chap. 2, pp. 13–33. ISBN: 978-0-240-52240-1. DOI: [10.1016/B978-0-240-52240-1.00002-8](https://doi.org/10.1016/B978-0-240-52240-1.00002-8).
- [48] L. Asquith. "Listening to data from the Large Hadron Collider". In: *TED[®] Zurich* (Dec. 2013). URL: <https://www.youtube.com/watch?v=iQiPytKHEwY>.
- [49] P. Théberge. "Sound Maps: Music and Sound in Cybercartography". In: *Cybercartography: Theory and Practice*. Ed. by D. R. F. Taylor. Amsterdam, Netherlands: Elsevier, 2005. Chap. 17, pp. 389–410. ISBN: 9780444516299.
- [50] C. Frauenberger et al. "Interaction patterns for auditory user interfaces". In: *Proceedings of the 11th International Conference on Auditory Display (ICAD2005)*. Ed. by Eoin Brazil. Limerick, Ireland, 2005, pp. 154–161. URL: <http://hdl.handle.net/1853/50161>.
- [51] The Open Source Geospatial Foundation (OSGeo). *GeoTools*. URL: <http://www.geotools.org/>.
- [52] Oracle. *Java Swing*. URL: <https://docs.oracle.com/javase/8/docs/technotes/guides/swing/index.html>.
- [53] A. Lobben, M. Brittell, and N. Perdue. "Inclusive Cartographic Design: Overcoming Ocular-Centric Cartographies". In: *Cartography - Maps Connecting the World: Proceedings of the 27th International Cartographic Conference*. Ed. by C. Robbi Sluter, B. C. Madureira Cruz, and M. P. Leal de Menezes. Springer International Publishing, 2015. Chap. 7, pp. 89–98. DOI: [10.1007/978-3-319-17738-0_7](https://doi.org/10.1007/978-3-319-17738-0_7).
- [54] W. L. Diaz-Merced et al. "Sonification of Astronomical Data". In: *Proceedings of the International Astronomical Union* 7.S285 (2011), pp. 133–136. DOI: [10.1017/S1743921312000440](https://doi.org/10.1017/S1743921312000440).

USING A MULTIMODAL IMMERSIVE ENVIRONMENT TO INVESTIGATE PERCEPTIONS IN AUGMENTED VIRTUAL REALITY SYSTEMS

Samuel Chabot, Wendy Lee, Rebecca Elder, and Jonas Braasch

Graduate Program in Architectural Acoustics
Rensselaer Polytechnic Institute
Greene Building, 110 8th St,
Troy, NY 12180, USA
chabos2@rpi.edu

ABSTRACT

The Collaborative-Research Augmented Immersive Virtual Environment Laboratory at Rensselaer is a state-of-the-art space that offers users the capabilities of multimodality and immersion. Realistic and abstract sets of data can be explored in a variety of ways, even in large group settings. This paper discusses the motivations of the immersive experience and the advantages over smaller scale and single-modality expressions of data. One experiment focuses on the influence of immersion on perceptions of architectural renderings. Its findings suggest disparities between participants' judgment when viewing either two-dimensional printouts or the immersive CRAIVE-Lab screen. The advantages of multimodality are discussed in an experiment concerning abstract data exploration. Various auditory cues for aiding in visual data extraction were tested for their affects on participants' speed and accuracy of information extraction. Finally, artificially generated auralizations are paired with recreations of realistic spaces to analyze the influences of immersive visuals on the perceptions of sound fields. One utilized method for creating these sound fields is a geometric ray-tracing model, which calculates the auditory streams of each individual loudspeaker in the lab to create a cohesive sound field representation of the visual space.

1. INTRODUCTION

As sets of data become ever larger and more complex, tools and practices for extracting information must also adapt and develop. One such emerging tool is that of the immersive virtual environment. These state-of-the-art spaces present users with enveloping visuals often combined with three-dimensional audio reproduction systems. Size and degree of immersion in the visual or auditory field can vary, ranging from the incredibly large and nearly all-encompassing to the more individual and limited scale. The AlloSphere [1] at The University of California, Santa Barbara, would be an example of the former, supporting upwards of 30 simultaneous on a bridge spanning the center of a spherical projection screen. Currently, 54 loudspeakers form three rings around the sphere for spatial audio using vector-based amplitude panning (VBAP) and Ambisonics techniques [2]. Many of the CAVE (Cave

Automatic Virtual Environment) systems reside in the latter category and are generally more appropriate for a single user or two [3]. Audio playback is often done simply through a 5.1 surround sound system.

An immersive workspace at Rensselaer Polytechnic Institute has been developed to facilitate large groups of users while not being rendered intractably complex: The Collaborative-Research Augmented Immersive Virtual Environment Laboratory, or CRAIVE-Lab. A function of this lab is its multimodality which puts equivalent emphasis on both the auditory display and visual display. This workspace measures 10 m by 12 m to accommodate groups of users both small and large. A nearly 360° projector screen surrounds the perimeter of the space and reaches 4.3 m tall. The screen is rectangular in shape, with rounded corners between the four sides [$r=1.52\text{m}$]. A sole PC equipped with the software Pixelwarp Evo¹ warps and blends eight overlapping high-performance projectors to produce a single continuous desktop display with a final resolution of 15360 by 1200 pixels. Therefore, content-creation is a rather straight-forward process. Visuals need only be created to encompass a screen width of 15k pixels. For many programs and resources, such as Photoshop or HTML, this only requires defining the workspace with the correct pixel dimensions. Rapid prototyping of immersive recreations of real environments and presentations of abstract and symbolic data is therefore possible. Figure 1 and 3 show respective examples of this. The projection screen itself is made of a microperforated PVC material, allowing it to remain acoustically transparent for the horizontal array of 128 loudspeakers behind it.

The array is positioned at approximately ear-height and follows the perimeter of the lab. Six additional loudspeakers are hung from the ceiling and directed downward into the workspace to add a third dimension to the auditory display. Matching the height and length of the projector screen, a heavy acoustic curtain hangs behind the horizontal array to provide strong damping of the physical room's response. Carpeting has been installed over the previously concrete floor to provide additional dampening of sound and reduce reflections of light. The entire loudspeaker array is controlled using a single Mac Pro equipped with an RME HDSPE MADI FX sound card². This allows channels to be accessed and addressed individually. Production of sound is therefore extremely flexible, and multiple approaches can be taken. For example, individual

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

¹<http://pixelwix.com/warp-and-blend-software/10-pixelwarp-evo-standard-wide-screen-warper.html>

²http://www.rme-audio.de/en/products/hdspe_madi_fx.php

Figure 1: Panoramic rendering of realistic environment featuring two users

channels can be explicitly targeted with auditory streams. Due to the high-density nature of the array, the system also allows for complex spatialization techniques such as higher-order ambisonics and wave field synthesis, which require channels to work in conjunction to form a cohesive sound field. For a much more detailed overview of the CRAIVE-Lab and its hardware and software considerations, refer to this previous ICAD paper [4].

A variety of studies, including use-cases, have been performed in the lab. One study demonstrated the advantages of the space's multimodality in an experiment involving abstract data exploration. Participants were presented with a large volume of visual data. Auditory cues and presentation techniques were tested for aiding visual data extraction. A variety of cues were chosen, including sine tones, white noise and camera shutter sounds, and three playback groupings were available: local, regional, or global sounds. Participants were scored on their performance, which was rated for speed and accuracy of identification.

Focusing on the advantages of human-scale immersive presentations of realistic scenes, another study investigated the perceptions of architectural renderings when viewed both on two-dimensional printouts and within the immersive visual display. Participants were presented with three architectural renderings that had been "inserted" into real scenes and asked to rank their preferences (first with the printouts, second using the immersive display). Its findings suggest disparities between participants' perceptions of the renders when using either two-dimensional printouts or the immersive CRAIVE-Lab screen.

Another project seeks to study the connection and influences of immersive imagery on users' perceptions of immersive sound fields. Renderings of realistic scenes are generated and paired with acoustic simulations of spaces, produced using panoramic imagery and a ray-tracing model. To do so, a geometric model is imple-

mented using MATLAB to calculate individual impulse responses at each loudspeaker location. These impulse responses are then used to create the auditory streams of each loudspeaker for a combined sound field recreation of the modeled space. Users' perceptions of the artificial sound fields will first be surveyed without any immersive imagery. Later, users will be presented with paired visual scenes (sometimes with correctly matched and sometimes with disparate scenes) to again be assessed on their perceptions of the sound fields amidst the immersive visuals.

2. AUDITORY CUES FOR EXTRACTING VISUAL DATA

The question this psychophysical study sought to address was which auditory cue and cue type would most effectively assist visual data extraction in a complex, immersive visual display system [5]. Previous works in this area both discuss human visual and auditory attention [6] and find auditory information altered the perceptions of visual information [7]. In this study, it was hypothesized that auditory cues that provide the strongest localization information (while also inflicting the least attentional distractions) would provide the most effective assistance for speed and accuracy of data extraction. An objective of the experiment was to assess various auditory cues and presentation techniques to determine those combinations that most helped participants' performances.

2.1. Experimental Design

Seven participants (various graduate students; 4 male, 3 female; ages between 24-33) were recruited for testing. The task presented to each user was to extract visual data from the immersive CRAIVE-Lab screen using auditory cues and answer ques-

tions based on the visual data as quickly and accurately as possible, modeled on the test method described in *The Handbook for Multisensory Processes* as “spatial discrimination tasks” [8]. To do so, the CRAIVE-Lab panoramic screen was populated with icons which depicted a variety of symbols: letters, numbers, animals, foods, etc. Figure 2 depicts a sample. The icons were arranged in a 92-by-9 grid (for a total of 828) on a plain white background across the entirety of the screen. After assessing various layouts, this configuration was chosen for its combination of legibility and density of icons.

For the auditory cues to be tested, ten different sounds were chosen for assessment of their localization and distraction levels [9]. Each cue was repeated over a question's entire 20-second duration (with the exception of one, a camera shutter cue, which only played a single instance). The variety of the ten cues included the following: sine tones at 500 Hz and 1000 Hz, presented as bursts or continuous streams, white noise, also presented as bursts or a continuous stream, a click train, bell ding, tech alert, and camera shutter (former seven self-produced, latter three sourced from Zapsplat³). The cues were presented in one of three ways: locally, where cues were presented from a single loudspeaker at the target location; regionally, where cues were presented using an area of 21 loudspeakers surrounding the target location; or globally, where cues were routed directly to all 128 loudspeakers in the lab. Regional presentation utilized the loudspeaker at the target location and the ten loudspeakers to the left and to the right. For reliability, all ten auditory cues were tested using all three presentation methods three times each, for a total of 90 test questions throughout the experiment. Users interacted with the test questions on a separate laptop using an interface created in MATLAB, pictured in figure 4. This interface would present users with the test questions and automatically record the users' speed and accuracy upon answering. The MATLAB interface took advantage of the Open Sound Control (OSC) protocol to appropriately engage each auditory cue for each question. The cues were handled by Max/MSP on the lab's Mac Pro linked to the loudspeaker array. To avoid potential biases, the order of the cues and presentation methods was randomized (and documented) for each user by MATLAB before outputting to Max.

³<https://www.zapslat.com>

A participant's session went as follows: The participant was briefed with instructions on the laptop before entering the lab. The end of the instructions had the participant bring the laptop into the lab, where the visuals were on display, and place it on a (movable) stand in the center of the room. An 8-ft by 8-ft box marked by white tape on the floor indicated the space a participant may move about but was required to remain within. This was to prevent the user from drifting too close to both the screen and loudspeaker array, which would add bias in both the auditory localization and the visual search. Upon clicking "Begin" on the laptop interface, a 3 second "rest" page appeared. These appeared between each question. Then, accompanying the question "What color is the following icon on the big display?", one of the icons was shown with all color removed and a multiple-choice option of four colors. An example of one of these pages is shown in figure 4.

One of the ten cues was presented using one of the three methods, determined by the randomized order. Participants answered the question as quickly and accurately as possible and were instructed not to guess. (It was determined that arbitrary guesses do not inform results of accuracy and speed based on the auditory cues.) Each question was given a maximum of 20 seconds. This was done to prevent users from eventually abandoning the auditory cues and relying solely on visual search. If the question was not answered in time, the cue would stop and the laptop would move to the next “rest” page. No feedback was provided after each question. After each set of ten questions, the participant was allowed an untimed break to prevent fatigue. Upon completing all 90 questions, the participant was asked psychological questions concerning the ten auditory cues. The interface posed three prompts for each cue as it was replayed from a loudspeaker: “Describe the sound in your own words,” “How stressed do you feel after hearing the sound?” and “Do you find the sound pleasant?” Participants did not have a time limit for this section.

Most of the participants correctly identified the sine tones and white noise as such, while the other cues were perceived more ambiguously. For example, participants responded with perceptions of a temple wood block, metronome, and rain drops for the click train. The bell ding and tech alert were most often associated with phone rings, but others described sounds such as bells and water droplets. Participants appeared to draw on familiar conventional sounds to describe the cues. The camera shutter had the greatest consensus, with most responses including the word “camera.” The variety in responses reveals potential biases in interpretations, and warrants further research with a larger sample size. Some participants commented on the perceived ability to localize cues. Sine tones were said to offer poor localization information, while more complex cues, such as the bell and shutter, were perceived as localized more easily. The Franssen Effect points out that sine tones lacking a strong onset are particularly difficult to localize by the human ear [10]. White noise bursts and clicks saw the highest accuracy, and both constant sine tones the lowest. The constant white noise and clicks provided the quickest responses, and the bell ding and 500 Hz tones the slowest. Figure 5 shows the mean accuracy of each auditory cue sorted by presentation type. Each bar is an average of the total 21 responses (3 responses each of the 7 participants). Figure 6 shows the mean speed of the same pairings. This second graph considers specifically those questions answered

Figure 3: Participant in the CRAIVE-Lab with the final panoramic image with all 828 icons aligned in a 92-by-9 grid layout

Figure 4: Example of MATLAB interface presenting questions to participant

correctly, as questions not answered were counted as incorrect and are included in the accuracy tabulation.

Cues presented locally most often led to both the quickest and most accurate responses. Regional cues provided nearly the same accuracy for some and dramatic drops for others. For the tones, all cues not presented locally saw spikes in identification times, but the more complex tones did not see as large a shift until presented globally. The presentation of cues globally saw a drastic reduction in accuracy and spikes in response times. Cues presented locally provided the best performance, however regional cues also did relatively well. This may indicate that the auditory cues were most useful in informing participants of the general area of the visual target, but that vision took over for the searching technique once the target was regionalized. Unexpected was the high performance elicited from the camera shutter cue. Its accuracy exceeded nearly all other cues for each respective presentation type, and speed of identification did not see a dramatic difference from the other complex tones. It is possible that the single cue was less distracting while still providing adequate localizing information. When asked how stressed a user felt after hearing each sound, only the four more complex sounds (click, bell, tech alert, shutter) averaged below a 3 on a 5-point scale. Five of the six tones and noise scored an average of 3.5 or above, with both constant sine tones reaching a score of approximately 4.8. In response to a cue’s pleasantness, the four complex cues received the most favorable answers.

Figure 5: Accuracy of responses of each cue type, shown for each of the three presentation methods (local, regional, global). Each bar is the mean of the total 21 responses (3 responses per participant).

Figure 6: Mean time for correctly answered questions of each cue type, shown for each of the of three presentation methods (local, regional, global).

3. IMMERSIVE ARCHITECTURAL RENDERINGS

Another investigation into the advantages of an immersive system encompassing a congruent audio and visual experience explored questions of architectural design and how it can be better analyzed when presented in a physically immersive space. Through the spaces immersion and multimodality, it is thought that designs can be interpreted at greater depth and more true to reality [11]. Currently, many renderings suffer one or more of the following characteristics: glorified aesthetic; unattainable vantage points; unrealistic environments; no reference to senses other than the visual. Buildings are most often rendered from a birds-eye view, a perspective most clients will never see in reality, along with other ideal conditions, such as perfect lighting, coloring, and atmospheres. Another consequence of most architectural renderings is the two-dimensional presentation of displays or printouts. Clients are then asked to imagine the finished product at full human scale. An immersive augmented environment is the closest form for presentation to human scale.

3.1. Site Data

For this experiment, sites of urban landscapes were reviewed for their diversity of design techniques and implementations. These sites were also reviewed for the activity of their soundscapes. Characteristics desired of each site included public access, close proximity to transportation hubs, and surrounding foliage. The three sites chosen were all located in Manhattan: the plaza at the southeast corner of Central Park at the William Tecumseh Monument, Foley Square, and Washington Square Park. Collection of the site data consisted of two parts. One was the retrieval of the visual footprint. This entailed taking a number of photographs from the same location and ensuring enough photographs were taken to cover the entire 360 degree of the horizontal plane. The camera was positioned at eye-level in order to recreate the most realistic perspective of the scene (versus a bird's eye view, etc.). Post-processing on these images in Photoshop stitched and edited them together to form a cohesive 360 degree image for presentation on the CRAIVE-Lab screen. The other part of site data collection focused on the soundscape. This was done using two Zoom H2s each offset by 45 degrees to capture eight surrounding channels to make use of the CRAIVE-Lab speaker array. To present the most cohesive representation possible, visuals and audio were captured during the same site visit. Architectural renderings were sourced from available student designs within the School of Architecture. Three designs were chosen, ranging from simple to the more abstract and unfamiliar. These three architectural renderings were then "placed" in the visual scenes taken at each site in Manhattan. A location within the image was chosen for the new replacement building. Using Rhino, the architectural software, and Photoshop, the renderings were "inserted" into the panoramic scenes so as to realistically. Figure 7 shows the three renderings within the Site 3 location.

3.2. Testing

The experiment consisted of three tests: the first involved only auditory stimuli; the second involved only a printout of the visual location; and the third combined the audio with the panoramic imagery for the complete immersive experience. A total of 15 participants were tested (graduate students with and without architectural training; ages between 24-33). In Test 1, participants were

Figure 7: Three renderings placed into Site 3 to be evaluated for their perceived appropriateness in the location

asked to stand in the center of the CRAIVE-Lab. The soundscapes recorded at each of the three sites, lasting two minutes each, were played. This was done prior to exposing the participant to the visual data to prevent individuals from becoming biased toward the visual data while answering perceptual questions about the audio data, such as identifying 3-5 specific sources in the soundscape (e.g. car horn, mechanical noise, etc.). Participants also ranked the soundscapes on a scale of 1 to 5 for 17 pairs of attributes (e.g. wide/narrow, far/nearby, etc.). Test 2 surveyed participants' visual perceptions of each design at each site using a two-dimensional printout. It asked participants to rank the designs by how well they fit in with the surrounding environment. The third test combined the soundscapes previously heard with the panoramic imagery of each site. This provided the participants with much more context about each location and how each rendering fit into the scene. Participants were again asked to identify sound sources and describe

the soundscape, as well as choose the renderings they thought most appropriate for the scene and provide feedback.

3.3. Results and Analysis

Participants' subjective answers for sound identification and ranking may have been altered with the introduction of visuals. Some identifications were more precise when a visual aid was available, such as multiple incorrect identifications of keys rattling in Test 1, but the subsequent correct identifications of dog tags in Test 3 (where a dog was visible in the scene). For Site 1, the identification of people talking as one of the predominant sources during Test 1 rose from 60% of participants to 86% during Test 3 (where multiple groups of people are clearly visible in the foreground). Sites 1 and 3 also saw more consensus in the subjective ranking: the number of pairs where there was over 50% agreement on the score rose from 2 to 8 (of the total 17) and 5 to 11 for Sites 1 and 3, respectively.

The results of this experiment suggest that the presentation style of architectural renderings can play a role in the perceptions of that rendering. The clearest example of this occurred for the scene of Washington Square Park. During Test 2 where participants only had the two-dimensional printout, replacement building input B was chosen by 33% of participants. However, during Test 3 in the immersive setting, this share shot up to 73%. Participants cited the greater level of context when surveying the scene as having swayed their decision. Two additional buildings around Washington Park with warmer color profiles (like that of replacement building input B) are introduced in the panoramic image versus the two-dimensional printout. Overall, 70% of participants declared the introduction of bi-modal sensory stimuli as influential on their judgments. Figure 8 shows the overall results for each site during both Test 2 and Test 3. It can be seen that responses during the third test do not always match up with those in the second test.

4. INFLUENCES ON USERS' PERCEPTIONS OF SOUND FIELDS BY IMMERSIVE VISUAL IMAGERY

This project is the expansion of the soundscape analysis in the immersive rendering work. It seeks to study the connections and influences of immersive imagery on users' perceptions of immersive sound fields. Namely, how are one's perceptions of a spatialized acoustic environment altered in the presence of human-scale visual environments? A previous study by Yuko Wani [12] presented users with imagery on a single projection screen and played a stereo presentation of an auralization of the space pictured in the imagery. This experiment would build on the CRAIVE-Lab's immersive, multimodal features by presenting users with various immersive sound fields, both with and without paired immersive visuals. In order to investigate this, the sound fields of a variety of spaces must be created and paired with (matching and disparate) panoramic imagery.

4.1. Procedure

To do so, a ray-tracing model was implemented to generate impulse responses for the individual loudspeakers of the Lab. A geometrical model of a space is defined in MATLAB. Figure 9, an example of Cologne Cathedral in Germany, shows sound-reflecting walls (black), a sound source (red), and a receiver (blue). A set of rays is sent out from the sound source in every direction within the

Figure 8: Overall preferences of participants broken down by site and rendering during Test 2 and Test 3

horizontal plane at equiangular distances of 5° . Each ray is then traced, and every time a ray meets a wall it is reflected back using Snell's law considering that the outgoing angle equals the incoming angle. The ray is traced until the 20th reflection occurs, unless the ray exits the geometrical model. At every reflection, the sound level is attenuated by 2 dB across frequency to simulate acoustic wall absorption. The sound intensity is also attenuated over distance based on the inverse square law, assuming the sound source to be of omnidirectional character. The collection of rays is shown in figure 9 as gray lines such that the rays become lighter in color with distance and decreasing sound pressure. All rays are then collected at the receiver position assuming a spatial window matching the dimensions of the CRAIVE-Lab. Each calculated ray is tested if it intersects the spatial window at the receiver position. For each intersecting ray, it is then calculated how far it traveled from the source position to the receiver position, at which azimuth angle it arrives at the receiver position and how many times it had been reflected (reflection order).

Based on these data a room impulse response is calculated in which each loudspeaker contributes individually to the entire

Figure 9: Schematic of a recreated space (Cologne Cathedral) used in the geometric ray-tracing model. Red dot is source location, blue dot is receiver location

sound field those reflections that crossed the threshold of the spatial window at the loudspeaker location. Each impulse contains the rays calibrated to the amplitude each should have based on distance traveled and number of wall reflections. In addition, a late reverberation tail is generated at a constant level (assuming a statistically evenly distributed diffuse reverberation field) using an exponentially decaying Gaussian noise burst.

Using these generated data, one can convolve the impulse responses of each loudspeaker with anechoic tracks to create a sound field that users are able to explore and walk around while maintaining the perceived appropriate source image location.

4.2. Tentative Experimental Setup

The experiment will follow a similar flow as that in the immersive architectural renderings experiment. Participants are to be surveyed on the sound fields and panoramic visuals separately before they are paired together. First, participants will be presented with the generated sound fields to determine their unbiased (by the visuals) perceptions of what (or, rather, where) they are hearing. Various questions will be asked in a thorough questionnaire to determine each sound field's perceived acoustic scene. Second,

participants will be shown various panoramic scenes and asked to describe the spaces and their expectations about each acoustic scene. Later, participants will be brought back to experience combinations of sound fields and panoramic scenes. Some of the sound fields will be paired with their corresponding scenes while others will be intentionally mismatched. The questioning for this is two-fold: to what extent must an acoustic scene match its visual environment, and can participants be convinced of mismatched sound fields that do not belong to the immersive visuals they are presented with?

5. FUTURE WORK

In the near future, the full experimental setup for analyzing the influence on users' perceptions of immersive sound fields with or without the presence of immersive visuals will be complete and ready for participants. More data across all three studies will be collected as more users participate to better identify patterns in trends in user perceptions. Continuing work seeks to explore the consequences of the congruent immersive visual and immersive audio systems and the perceptual effects that occur when users are presented with both systems simultaneously. Another discussed project would pair panoramic imagery with a real-time user-adjustable acoustic room model. Users would then be instructed to adjust the acoustic profile of the room to match what they perceive to be the most correct profile. Responses would be compared with other users to determine an overall consensus or lack thereof.

6. CONCLUSION

The CRAIVE-Lab fits into the field as a resource for presenting both real and symbolic data, capable of multimodality. The advantages of having simultaneous immersive visual and auditory displays working in conjunction are being studied as qualities unique to environments such as this. These preliminary studies hint at strong relationships between the modalities. Ongoing research will continue to analyze the effects and consequences of such a system on human-computer interactions.

7. ACKNOWLEDGMENT

This work is supported by the National Science Foundation (NSF #1229391) and the Cognitive and Immersive Systems Laboratory (CISL) at Rensselaer directed by Hui Su.

8. REFERENCES

- [1] X. Amatriain, J. A. Kuchera-Morin, T. Höllerer, and S. Travis Pope, "The AlloSphere: Immersive multimedia for scientific discovery and artistic exploration," *IEEE Multimedia*, vol. 16, no. 2, pp. 64–75, 2009.
- [2] A. Cabrera, J. Kuchera-Morin, and C. Roads, "The Evolution of Spatial Audio in the AlloSphere," *Computer Music Journal*, vol. 40, no. 4, pp. 47–61, 2016.
- [3] S. Manjrekar, S. Sandilya, D. Bhosale, S. Kanchi, A. Pitkar, and M. Gondhalekar, "CAVE: An emerging immersive technology - a review," *Proceedings - UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014*, pp. 131–136, 2014.

- [4] S. Chabot and J. Braasch, "An Immersive Virtual Environment for Congruent Audio-Visual Spatialized Data Sonifications," in *Proceedings of the 23rd International Conference on Auditory Display*, Happy Valley, PA, 2017.
- [5] W. Lee, "Using Auditory Cues to Perceptually Extract Visual Data in Collaborative, Immersive Big-Data Display Systems," M.S. thesis, Sch. of Arch., Rensselaer Polytech. Inst., Troy, NY, 2017.
- [6] B. G. Shinn-Cunningham, "Object-based auditory and visual attention," *Trends in Cognitive Sciences*, vol. 12, no. 5, pp. 182–186, 2008.
- [7] B. Cowan, D. Rojas, B. Kapralos, K. Collins, and A. Dubrowski, "Spatial sound and its effect on visual quality perception and task performance within a virtual environment," *Proceedings of Meetings on Acoustics*, vol. 19, no. 1, p. 050126, 2013.
- [8] G. A. Calvert, C. Spence, and B. E. Stein, Eds., *The Handbook on Multisensory Processes*. Cambridge, Mass.: MIT Press, 2004.
- [9] J. M. Terhune, "Localization of pure tones and click trains by untrained humans," *Scandinavian Audiology*, vol. 14, no. 3, pp. 125–131, 1985.
- [10] W. M. Hartmann and B. Rakerd, "Localization of sound in rooms IV: The Franssen effect," *The Journal of the Acoustical Society of America*, vol. 86, no. 4, pp. 1366–1373, 1989.
- [11] R. Elder, "Influence of Immersive Human Scale Architectural Representation," M.S. thesis, Sch. of Arch., Rensselaer Polytech. Inst., Troy, NY, 2017.
- [12] Y. Wani, T. Terashima, and Y. Tokunaga, "Effect of visual information on subjective impression for sound field in architectural space," vol. 19, no. 1, p. 040102, 2013.

AD THEORY

SONIFICATION: A PREHISTORY

David Worrall

Audio Arts and Acoustics Department,
Columbia College Chicago,
600 South Michigan Avenue Chicago, IL 60605,
USA.
dworrall@colum.edu

ABSTRACT

The idea that sound can convey information predates the modern era, and certainly the computational present. Data sonification can be broadly described as the creation, study and use of the non-speech aural representation of information to convey information. As a field of contemporary enquiry and practice, data sonification is young, interdisciplinary and evolving; existing in parallel to the field of data visualization. Drawing on older practices such as auditing, and the use of information messaging in music, this paper provides an historical understanding of how sound and its representational deployment in communicating information has changed. In doing so, it aims to encourage a critical awareness of some of the socio-cultural as well as technical assumptions often adopted in sonifying data, especially those that have been developed in the context of Western music of the last half-century or so.

1. INTRODUCTION

We are at a time in the evolution of humans and their tools when the power of digital information processing and algorithmic decision-making is demonstrating an ability to radically change our lives: From genetic finger-printing, gene-splicing and pharmacology, to driverless vehicles, patterns in our consumption and how we amuse ourselves. Even now, so early in this new Dataist era, organizations with networked computational intelligence already have access to more data about ourselves that we ourselves have access to, and are beginning to demonstrate a power to make better decisions for us than we make for ourselves. What, then, one might reasonably ask, is the use of exploring such ancient, intensely-human-centered approaches to information-gathering and decision-making as listening? How old-fashioned; how quaint! This is an increasingly pertinent question and the answers are not necessarily obvious as they lie at the heart of the difference between a conception of life merely in terms of information flow and data storage as might be imagined by the Cognitivists [20], and one in which mind, body and (now technologically enhanced) consciousness play a fundamental role in active perception, knowledge acquisition, meaning-creation and decision-making.

In a contemporary media-saturated environment, sound plays a wider variety of different social and communicative roles today than it ever did in the past. Although computer-

generated data sonification is a relatively recent formalized practice, cultural antecedents can be identified in all periods of history. In the spirit of the adage “know from whence you came,” anecdotes and general trends are instructive because they provide insights into what people thought sound was and how it could be used to communicate.

The idea that sound can convey non-speech information predates the modern era, and certainly the computational present. By considering some ways in which sound has been used in the past to convey specific information, this paper attempts to provide a focused historical understanding of ideas about sound and how its deployment in communicating information has changed in the modern era. In doing so, it is hoped that this will assist us to critically examine the socio-cultural assumptions we have adopted, particularly with regard to the current practice of data sonification. Whilst acknowledging their Eurocentricity, there is no suggestion that the ideas discussed here do not have wider applicability.

In the process of uncovering the accuracy of our assumptions about human perception, it is easy to be seduced into a type of sense war, in which hearing and listening are pitted against vision and seeing. Each have their place and so to be encumbered with such a burden is not useful, particularly when a focus on one sense assumes or leads to the decline of another. Such casting can take many forms. What is to be gained by Marshal McLuhan’s myopia, for example?

The ear is hypersensitive. The eye is cool and detached. The ear turns man over to universal panic while the eye, extended by literacy and mechanical time, leaves some gaps and some islands free from the unremitting acoustic pressure and reverberation [16, p168].

Thankfully, the discussion has moved on, at least in some circles. This is not to deny that there are sensory differences; in fact it can be instructive to identify them. For example, it is hearing, not vision, that affords omnidirectional coherent perceptual experiences. For all moving creatures, including our ancestors, both ancient and modern, in situations where sight is obscured, spatial auditory clarity plays a vital survival role in determining both from where the predator is approaching and to where the prey has escaped. On the other hand, we can suppose that, to a creature sleeping in a cave, being alerted early by the amplifying echoic resonance of the space that something of a certain mass was entering, was more important than details of its exact position.

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License.

The full terms of the License are available at
<http://creativecommons.org/licenses/by-nc/4.0/>

2. HISTORICAL USES OF SOUND TO CONVEY INFORMATION

There are many reasons why, in certain circumstances, sound might be the preferred representation and communication medium for information, including the known superiority of the hearing sense to discriminate particular kinds of structures. For example, it is easy for most of us to personally verify that a purely visual side-by-side comparison of two sets of written records requires high levels of concentration and that doing so is very prone to error, especially over extended periods of time. On the other hand, listening to vocalizations of such representations is much easier. The presence of such auditing can be inferred from records of Mesopotamian civilizations going back as early as 3500 BCE. To ensure that the Pharaoh was not being cheated, auditors compared the “soundness” of strictly independently-scribed accounts of commodities such as grains moving in and out, or remaining, in warehouses [3]. When granary masters, otherwise strictly isolated from each other, assembled before the Pharaoh and alternated in their intoning of such records, differences in their accounting records could be easily identified aurally. A faster and more secure method that eliminates any “copy-cat” syndrome in such alternation, is to have the scribes read the records simultaneously—a type of modulation differencing technique. Although we have no hard evidence that these techniques were practiced, such a suggestion does not seem unreasonable, and would represent possibly the earliest form of data sonification.

Sound has also played an important role in both theoretical and empirical inquiry for millennia. At least as far back as Pythagoras (born ~569 BCE), the concept of The Harmony of the Spheres played a unifying role in the development of the arts and sciences, and incorporates the metaphysical principle that mathematical relationships express qualities or “tones” of energy ratios. The application of number relations (i.e. ratios) and sound have been integral to the conceptualization and realization of Western music over all periods in radically different ways: From the development of richly fecund investigations in tuning and temperament to multiple voice polyphony and to highly chromatic polytonalities; serialized additive rhythms to stochastic mappings controlled by Poisson’s distribution, for example.

3. THE EAR AS THE ORGAN OF THE CHURCH

In pre-medieval Europe, the Roman Catholic Church eventually filled the political and spiritual vacuum caused by the collapse of the Roman Empire around 500 CE and went to extraordinary lengths to establish supreme papal power, culminating in the construction of massive reverberant cathedrals, which replaced the echoic caves of the past to become temples in which the resonant voice of an omnipresent God, through the pope and his priests and their choirs, was delivered to a largely passive, observing audience; silenced by the unintelligible Latin and the presence of the holy sacrament, for the safety of their immortal souls. Even though the church almost universally forbade the use of all instruments of “profane” music in worship, sometime during the tenth century large *Blockwerk* organs began to be permanently installed in churches and cathedrals. By producing the lowest-pitched and most

powerful musical sounds of the period, these organs resonated the buildings in which they were installed, and impressed on congregations a power found only in the church. In addition, these modern caves were engineered to produce an awe-inspiring, comforting, inclusive “community” feeling through spatial aural reverberant incoherence, and by which potentially questioning or dissenting individuals were subsumed into the anonymity of the Mass.

When Johannes Gutenberg (1398–1468) invented his mechanical movable-type printing press in 1440, the closely connected (visual) activities of writing and reading were engaged in only by the educated, ruling classes. The overwhelmingly illiterate common people’s only connection to writing was when it was read aloud to them. They maintained their profane cultures by telling each other stories from within their rich oral/aural traditions and playing instrumental music. Gutenberg’s invention flooded Europe with printed material. It made the Bible more available and encouraged an increase in literacy rates which eventually resulted in dissent from the Roman church’s authority in the form of the sixteenth century’s Protestant Reformation. The spread of literacy was accelerated by the emphasis on formal education and made way in the seventeenth century for an intellectual rebirth in Europe in the form of the Renaissance which eventually led, over an extended eighteenth century, to the Age of Enlightenment in which reason was advocated as a means of establishing an authoritative system of government, religion, ethics and aesthetics. This supported an intellectual and scientific impetus to obtain objective truths about the whole of reality, the spread of learning to the masses, and laid the material basis for modern knowledge-based economies.

Breaking the mental clutches that the church imposed by leveraging illiteracy and the use of sound to surround and subsume independent voices, required (men) of vision, using empirical techniques, to “look” outside the religious dogma of the church for objective theories of the natural world “out there”. It also required the invention of observational instruments and techniques “favorable to the progress of the arts and sciences” [20, p227]. For example, Galileo Galilei (1564-1642) built his own telescope in 1609, a year after it was patented in the Netherlands [15, p15]. Another such invention was the (re)discovery of the laws of linear perspective by the Filippo Brunelleschi of Florence (1387-1446) which enabled the depiction of visual depth on a planar surface. Lacking a theory of mathematical perspective, artists of the Middle Ages were more concerned with the static depiction all-encompassing religious or spiritual metaphors rather than depicting the real, physical world oriented towards an individual viewer.

In parallel to these advances in visual representation, the dominance of angelic voice-only sacred monophonic plainchants of the church were challenged by, and eventually subsumed into, the polyphonic complexities of the compositions of increasingly individually-recognized composers who, in composing order on the cathedral-induced echoic aural incoherence of multiple overlapping lines of the chants, explored, notated and extemporized new musical structures. These structures were not arbitrarily abstract however, but developed from careful attention to both the sounds and accents of the words being sung as well as their meaning in the texts.

4. MUSIC AS LANGUAGE AND RHETORIC

The belief, that since music is a language and can be consciously treated as such, is not confined to post-medieval Europe. Many cultures have developed melodic and rhythmic modes and dance gestures through which associated concrete meanings and/or affective states were communicated to understanding audiences. For example, the ragas and talas and mudras of classical India are used to clearly communicate specific ideas, events, actions, or creatures. The ancient Greeks had invented a whole systematic theory of how different musical elements affect the soul in different ways. The Renaissance of their ideals in Europe included the application of rhetorical devices, not only in speech, but in music and dance. A system of rhetorical devices, i.e. a representational vocabulary for making communication explicable and persuasive, was considered essential for organizing the syntax of (initially voice-only) compositions and, by addressing an audience's logical and emotional dimensions simultaneously; made the music semantically effective and thus able to communicate successfully. By the middle of the sixteenth century, such rhetorical devices had developed into often extravagant musical word paintings or madrigalisms as they were known. Nevertheless, the study and use of rhetoric to make musical discourse meaningful continued through the seventeenth and well into the eighteenth century. Haydn, often described by his contemporaries as "a clever orator" and "the Shakespeare of music," is probably the last major European composer whose music was regularly discussed by his contemporaries in terms derived from the classical tradition of rhetoric [2], [19].

5. THE RISE OF ABSTRACT INSTRUMENTAL MUSIC

During the Middle Ages, secular music consisted of heroic and lyrical minstrel songs and instrumental music to accompany dancing. During the Renaissance, the Church's attempted to suppress most forms of instrumental music as profane failed as regal courts became significant centers of social, economic and cultural economic activity outside of Church [4]. As monarchs and their courtiers required intellectual stimulation and entertainment, the composers serving in the courts became more experimental, using a widened variety of musical instruments, increasingly considered musical forms as modes of personal expression. These forms became increasingly non-rhetorically-referential to written or spoken texts. While initially following in the spirit of courtly dances, the decline on the reliance of textual subjects in favor of formal reflexivity [10, p94] became a means of embodying aesthetic affects.

As a consequence, there arose a belief that music was not a language in the sense that it no longer had properties of defining and referring to specific meanings about which interpretations and responses could be made. The need for some understanding of what we might call the semiological codes of music was apparent however, for without them it would have been difficult to claim for music to be no more than beautiful arrangements rather than important expressions of human experience and expression in the humanist tradition. In discussing musical language, Alan Durant [8, p10] says:

On the one hand, there is 'language' as an assumed range of properties and associated effects, determined ... by psychological or acoustical resources. These supply continuity for the 'language' and an overall framework for activity. Music made outside the framework is simply part of another 'language'. What would be most interesting about this 'language' would be the precise nature of the acoustic and psychological resources, even if these-failing some way of determining correspondences between forms and effects-will not explain very much about actual, particular pieces of music. In the other emphasis, there is language as a set of properties and associated effects (still conditioned by psychological and acoustic resources), whose most significant features are those of social, regional and historical variation around those resources.

And later (op. cit., pp 13-14), recalling Claude Debussy's "Works of art make rules but rules do not make works of art" [17, p590], he observes:

...In this latter emphasis, musical works do not simply exist within, or use, the musical 'language': they make and remake it, in particular realizations and directions of potential.

Following the rise of public concerts and music to be listened to for its own sake, the status of instrumental music grew in the eighteenth century, as did the Cartesian idea that abstract reflexive forms were the pre-eminent *raison d'être* of musical expression, so that "critics increasingly came to think of music as having emancipated itself from mimetic representation" [10:189]

6. "ORGANIZING THE DELIRIUM"

During the nineteenth and early twentieth century, there was a shift by composers from musically representing various gestural and aesthetic affects (as portrayed in the different formal dance forms, for example) to the representation of individuals' changing emotional states. The principal structural means of organizing both moment-to-moment flow within a work as well as the form of the entire composition was functional tonal harmony, which relies on a set of codes or functions to create expectations in the listener of, given what has just occurred, what might occur next. The German Romantic interest in the personal expression of intense emotional states led some composers to search for new expressive means such as rapidly shifting modulations and eventually the abandonment of functional harmony as means of organizing both moment-to-moment flow within a work as well as the form of the entire composition—and eventually of tonality itself.

In addition to its social and cultural imperatives, there arose a distinct trend for Western composers to conceive of their music as complex-patterned time-ordered series of acoustic events that vary in pitch, loudness and timbre; that are absorbed and elicit emotions when listened to. This paradigm is embedded in scored compositions that are abstractly composed and transmitted to listeners by expert musicians. Following the examples of the second Viennese school in the first part of the twentieth century, many composers used serial procedures based on permutation group orderings of the chromatic pitch gamut. This led to a short but intense period when each of the (now "parameterized") dimensions of a composition (pitch,

duration, dynamics and timbre) were all serially organized. This idea of music as a variable-parameterized space has been adopted by sonifiers engaged in parameter-mapping sonification (PMSon).

The composer/architect Iannis Xenakis, whose oblique contribution to sonification is discussed later, was clear that the serialists had made an important contribution to the search for new methods of organizing music by introducing abstraction. However, he did not agree with their declared necessity for using just serial techniques to organize musical material and was critical of linearity of serial thinking: the way pitch dominated musical structuring, even when its influence on the sound was only secondary; that under serialism as they defined it, duration was, in general, even less structurally organized than in traditional forms, and that the resulting disjunct linear polyphony destroys itself by its very complexity [24, p193].

7. FROM PROGRAM MUSIC TO PROGRAMMED MUSIC

Not all composers were caught up in the psychological angst afflicting the German romantics and expressionists, or the following existential modernist's position that ordinary musical listening was archaic and needed to be replaced with "structural listening" [9, p20]. So, the need to concretely "ground" musical expression in the earthly domain persisted in compositional references to the real, often rustic, world. In fact, compositions depicting nature in one way or another were so common that the term *Pastoral* is used as a description of a type of composition, and there are a number of recognized national Pastoralist schools.

Whilst over the last several centuries most aestheticians regarded depictions of birdcalls and battles in music as a debasement of the art, they were popular with the public, as were—and are—nature *atmospheres* such as Antonio Vivaldi's *Four Seasons* (1725), Ottorino Respighi's *Pines of Rome* (1917), and Charles Ives, *Three Places in New England* (1912-16). In fact, a compiled list of such works would include many of the best known works in the Western classical canon.

7.1.1. Although less common, the literal transcription of real-world sounds into music has been even more challenging to the abstract expressionists than Pastoral music. The cuckoo calls in Beethoven's Sixth Symphony (1808) and Olivier Messiaen's use of a wind machine in *Des Canyons aux étoiles* (1974) are obvious examples but the trend extended throughout the twentieth century with the inclusion of non-traditional sounds such as the fire siren in Edgard Varese's *Ameriques* (1921), the futurist's noise art and then, especially following the invention of the tape recorder in the 1940's, by *musique concrète*—the term itself being coined by Pierre Schaeffer to contrast the music he was making at *Radiodiffusion Française* in Paris with the "pure" (i.e. abstract) music of the period, especially that being produced by the German expressionist composers including those at the Westdeutscher Rundfunk (WDR) in Cologne. Schaeffer and others established the Groupe de Recherches de Musique Concrète, Club d'Essai de la Radiodiffusion-Télévision Française (GRMC- at RTF) in Paris in 1951 which attracted many notable composers of the period. He went on to establish the Groupe de Recherches Musicales (GRM) one of several theoretical and

experimental groups unified by the study of audiovisual communication and mass media.

8. ALGORITHMIC COMPOSITION AND DATA SONIFICATION

Xenakis' use of mathematics in composition included the application of group theory, game theory, symbolic logic and stochastics [24] was seminal in establishing a compositional "style" known as algorithmic composition which sowed the seed for the idea of representing abstract data relations in sound for investigative purposes. Xenakis was a participant at GRM but, following Schaeffer's refusal in 1963 to use mathematics and the computer in the studio there, established Centre d'Etudes de Mathématique et Automatique Musicales (Center for Mathematical and Automatic Musical Studies CEMAMu specifically in order to undertake research into the application of mathematical ideas to music composition.

The other notable early algorithmic music research of the period was by Lejaren Hiller and Leonard Isaacson at the University of Illinois at Urbana-Champaign. They used university's ILLIAC I computer to develop a "rules-based" system to create *cantus firmi*, four-voice harmonies, various rhythmic and dynamic schemes, and Markovian stochastic grammars which they realized in the *Illiac Suite* for string quartet in 1957 [13].

Such computational proceduralism developed rapidly during the second half of the twentieth century in league with the rise of cybernetics and cognitive science. The technical feasibility of being able to accurately repeat the synthesis of sounds by digital means enabled the birth of computer music, which was also heavily influenced by the "acoustic event" paradigm that became embedded in many of the compositional software tools used to create it. These tools have been widely adopted by researchers who use them in an attempt to obtain a better understanding or appreciation of relations in datasets of various sizes, dimensions and complexities—what is now called scientific sonification.

It is useful to distinguish data sonifications made for the purpose of facilitating the communication or interpretation of relational information in data, and data-driven music composition, ambient soundscapes and the like—the primary purpose of which is personal expression and other broader cultural considerations, whatever they may be. While scientific or pragmatic data sonifications and music compositions share a common reliance on the need to render structures and relations into sound, their purpose is often different, and so too the rationale for the evaluation of the sonic results. The current use of the term *sonification* to include such cultural concerns is somewhat unfortunate because it blurs purposeful distinctions. A desire to maintain these distinctions is not to suggest that there are not commonalities—the two activities can provide insights that are mutually beneficial. However, because the purposes of the activities are different, so too will be their epistemological imperatives and consequences, such as, for example, in the representational methodologies employed, in tool design, in user-interface and usability requirements and evaluation.

9. PURPOSEFUL LISTENING: MUSIC AND SONIFICATION

There is no one-way, or reason, to listen to music. Even different musics have different contexts and thus require different ways of listening that may involve whole complexes of social dimensions that are simply not relevant to the perceptualization of data relations for pragmatic purposes. Furthermore, though music may be composed of syntactic structures, there is no universal requirement that these structures be made explicit, even aurally coherent. In fact, stylistic or even dramatic considerations may require the exact opposite—in the orchestration of spectral mixtures by melding of instrumental timbres, for example.

In contrast, clarity in information sonification is essential and rendering techniques that produce a kind of “aural gumbo” can be more successful. Consequently, sonifications in which the user-driven real-time exploration of datasets using dynamic scaling in multiple dimensions, perhaps with auditory beacons [14, pp 185-221], may not result in what is currently understood to be musically coherent sound streams. Even if listened to as music, information sonifications may provoke critical commentary about musical issues such as the appropriateness or formal incompleteness of the resulting sonic experience. Perhaps, as Paul Polansky suggested, the closest point of contact between pragmatic data sonification and musical sonification is in compositions in which a composer intends to “manifest” mathematical or other formal processes [18]. This “classical” algorithmic motivation is explicitly enunciated by Xenakis in his seminal book, *Formalized Music* [24] and many of the cultural concerns in his thesis defense [23].

While numerous composers use mapping and other procedural techniques of one kind or another in their musical compositions, they are rarely interested in “featuring” the mapping explicitly. Nor do they use mapping in order to simplify the working process or to improve production efficiency, but in order to craft the emergence of musical forms. In order to gain a deeper insight into the way composers map conceptual gestures into musical gestures, Paul Doornbusch surveyed a select few composers who employ the practice in algorithmic composition [7, pp. 145-156].

I am not interested in projecting the properties of some mathematical model on to some audible phenomena in such a way that the model be recognized as the generator of some musical shape.

So, those interested in producing music of a certain complexity may shy away from simple mappings as they can be hard to integrate with other musical material of a substantial nature. On the other hand, as Larry Polansky explains,

...the cognitive weight of complex mappings degenerates rapidly and nonlinearly such that beyond a certain point, everything is just ‘complex’.

Even a suitably complex, structurally coherent mapping may not be musically sufficient if the composition relies on a (human) performer, as composer Richard Barrett (in *ibid.*) emphasizes:

In a score one is always dealing with the relatively small number of parameters which can be recorded in notation, and which interact with an interpreter to produce a complex, ‘living’ result.

The importance of this embodied “living” aspect of music has often been forgotten, ignored, or even dismissed in many discussions of Western art music, including by some composers. While there are historical reasons for and consequences of doing so, such an approach to data sonification could have a major impact on the intelligibility of computer-rendered mapping-encoded artifacts.

10. MUSICAL NOTATION AS REPRESENTATION

In the evolution of Western art music, notation evolved, along with the notion of the *work*, from a performer’s *aides-mémoire* to a tool of thought for defining works of increasingly abstract complexity [11]. Notated scores came to be thought of as the encoded representation of sounds, even as a somewhat definitive objectification of a composer’s thoughts. That we (at least in English) so frequently substitute the word ‘note’ for ‘tone’, ‘music’ for ‘score’, exemplifies the strength of this conceptual elision. Indeed, in a number of intricately notated works of the twentieth century, it seems the performer is sometimes considered an unfortunate necessity. In others, notation functions as encapsulated stimuli by which the performers, as they attempt the impossible task of playing it “note-perfectly”, enact a drama of physical and mental exertions. Theodore Adorno noted a tendency to consider the bodily presence of the performer as a kind of contamination of musical experience, as a manifestation of a commodity fetishism where the “...immaculate performance ... presents the work as already complete from the very first note. The performance sounds like its own phonograph record” [1].

Occidental art music today encompasses a wide range of motivations and listening practices, and reducing the intelligibility of such music to the conceptual level of scores and instruments has enabled an unprecedented level of complexity. However, there is a growing recognition among music researchers, supported by a significant body of research in neuroscience, that the conveyance of this complexity is reliant, at least to some extent, on embodied interpretation for effective communication [12]. It was not until it was technically possible to construct musical compositions without the assistance of embodied interpreters that it was possible to meaningfully speculate on the extent to which a listener’s perception of the structural characteristics of a piece of music are dependent on the sound-encoded gestures of performers, and not just the notated score. This has the unfortunate consequence that if sonifiers think of data as being the sonification equivalent of a parameterized musical score and follow the path of least resistance, as most have been apt to do, use composition software designed to produce abstract musical objects, their sonifications will lack the intelligence carrying capacity that is recognized as embodied in the (often micro-) gestures of musical performers. [22, pp. 52-59].

This should act as a caution that, while adopting tools of one domain into another can be a very empowering, such adoption does not come value-free. If the intelligibility of much music is bound, not only to text, rhetoric, metaphor and formal devices such as phrase structure and the semantics of harmonic tension and resolution, but to the transmission through sound of the embodied foreknowledge of performers, in establishing the foundations for the practice of information sonification, it is sensible to

embrace approaches to forming sounds and their relationships that are supported, yet as unencumbered as possible, by the conceptual boundaries placed by tools for computer music-making as they currently exist. This in turn has the potential to enrich the practices for which the tools were originally meant.

Sounds and sonic structures demonstrate their weak proclivity to bind to casual inferences [6, p108-109] and strongly to metaphorical representations: From the rustle of dried leaves caused by escaping prey to the echoic resonances of a cathedral and a crowded street; from the phonetic structures of speech to word-painting in Renaissance madrigals; from musical rhetoric and Pastorals to instrumental expressions of affective states; from the turbulence of gas molecules to gravity waves and the folding and cracking of the earth's crust; from the movement of objects relative to each other to the flow of data through a network. With careful attention to sonic details and their relationships to each other in context; to the isomorphic-heteromorphic bindings between sounds and their causes, sonification can more powerfully render vibrant voices for unseeable things.

11. REFERENCES

- [1] Adorno, T. 1991. "On the fetish character in music and the regression of listening" in *The Culture Industry*, ed. J. M Bernstein, London: Routledge.
- [2] Beghin, T and S Goldberg (ed), 2007. *Haydn and the performance of rhetoric*. University of Chicago Press: Chicago, IL, USA.
- [3] Boyd, E. 1905. *History of auditing* in R. Brown (ed.). *A History of Accounting and Accountants*. Edinburgh: T.L & E.C. Jack.
- [4] Brown, H. 1976. *Music in the Renaissance*. Prentice Hall History of Music Series. Englewood Cliffs, New Jersey; Prentice-Hall, Inc.
- [5] Cage, J. and Knowles, A. 1969. *Notations*. New York: Something Else Press.
- [6] Chion, M. 1961. *Sound: An Acoulogical Treatise*. J.A. Steintrager (trans). Duke University Press. First published in French in 1998.
- [7] Doornbusch, P. 2002. "Composers' views on mapping in algorithmic composition," *Organised Sound* Vol. 7 No. 02, August 2002.
- [8] Durant, A. 1984. *Conditions of music Language, discourse, society*, Macmillan.
- [9] Eisler, H. and Adorno, T. 1994. *Composing for the films*. London: Athalone.
- [10] Erlmann, V. 2010. *Reason and resonance: A history of modern aurality*. Zone Books: Brooklyn, NY.
- [11] Godøy, R.I. and Leman, M. (eds.) 2010. *Musical Gestures: Sound, Movement, and Meaning*. New York: Routledge.
- [12] Goehr, L. (1994). *The Imaginary Museum of Musical Works*. Oxford, UK: Oxford University Press.
- [13] Hiller, L.A and Isaacson, L.M. 1959. *Experimental Music: Composition With an Electronic Computer*. Second edition. New York: McGraw-Hill.
- [14] Kramer, G. 1994. "Some organizing principles for representing data with sound," in G. Kramer (ed.), *Auditory display: Sonification, Audification, and Auditory Interfaces*, Santa Fe Institute Studies in the Sciences of Complexity, Proceedings, Volume XVIII, Reading, MA: Addison Wesley Publishing Company, 1994.
- [15] Loker, A. 2008. *Profiles in Colonial History*. Solitude Press, Williamsburg, VA, USA.
- [16] McLuhan, M. 1968. *Understanding Media: The Extensions of Man*. New York: McGraw-Hill Book Company.
- [17] Paynter, J. 1992. *Companion to Contemporary Musical Thought*, Psychology Press Ltd, London.
- [18] Polansky, L. and E. Childs. *Manifestation and Sonification. The Science and Art of Sonification, Tufte's Visualization, and the 'slippery slope' to Algorithmic Composition. An Informal Response to Ed Childs' Short Paper on Tufte and Sonification; with additional commentary by Childs*, 2002. <http://eamusic.dartmouth.edu/~larry/sonification.html>. Accessed 20180201.
- [19] Saint-Dizier, P. 2014. *Musical Rhetoric: Foundations and Annotation Schemes*, John Wiley & Sons Inc., London.
- [20] Wilson, A. 1957. *Diderot: The testing years 1713-1759*. Available online at https://archive.org/stream/diderotthetestin001232mbp/diderotthetestin001232mbp_djvu.txt. Accessed 20180226.
- [21] Worrall, D. 2010. *Parameter mapping sonic articulation and the perceiving body*. In *proceeding of The 16th International Conference on Auditory Display*, Washington, D.C, USA, June 9-15, 2010.
- [22] Worrall, D. 2014. "Can Micro-Gestural Inflections Be Used to Improve the Sonificatory Effectiveness of Parameter Mapping Sonifications?" *Organised Sound* Vol. 19 No. 1 Cambridge University Press 2014..
- [23] Xenakis, I. 1985. *Arts/Sciences: Alloys*. New York: Pendragon Press.
- [24] Xenakis, I. 1991. *Formalized Music: Thought and Mathematics in Music*. New York: Pendragon Press. The First edition was published in English in 1971 by Indiana University Press.

INVESTIGATING EFFECTIVE METHODS OF DESIGNING SONIFICATIONS

Michael Quinton, Iain McGregor and David Benyon

Edinburgh Napier University, School of Computing,
Merchiston Campus, 10 Colinton Road, Edinburgh,
EH10 5DT, United Kingdom
{m.quinton, i.mcgregor, d.benyon}@napier.ac.uk

ABSTRACT

This study aims to provide an insight into effective sonification design. There are currently no standardized design methods, allowing a creative development approach. Sonification has been implemented in many different applications from scientific data representation to novel styles of musical expression. This means that methods of practice can vary a greatly. The indistinct line between art and science might be the reason why sonification is still sometimes deemed by scientists with a degree of scepticism. Some well-established practitioners argue that it is poor design that renders sonifications meaningless, in-turn having an adverse effect on acceptance. To gain a deeper understanding about sonification research and development 11 practitioners were interviewed. They were asked about methods of sonification design and their insights. The findings present information about sonification research and development, and a variety of views regarding sonification design practice.

1. INTRODUCTION

The purpose of this study was to investigate effective methods of designing and evaluating sonifications. Guidelines have been developed to facilitate practitioners in their work, but sonification design also enjoys an open ended and creative approach [1]. There are still difficulties for newcomers to the field to understand the basis for good design. Brazil [2] argued that it was hard for new practitioners' to understand how to design auditory displays due to guidance being limited and scattered across publications. It is often unclear for what purpose the sonification was designed and for whom. Sonifications are often restricted to contained studies without further development. A lot of publications give little information about the data gathering process and it is often hard to follow why certain design decisions were made. In order to reap a deeper understanding of the sonification design process it was thought necessary to interview practitioners about projects that they had worked on and to investigate their design processes. This publication addresses the results obtained from these interviews and discusses reflections obtained about sonification development and research.

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

1.1. Sonification and Methods of Sonification

Sonification can be defined as non-verbal sound used to represent data [1]. It is a subset of Auditory Displays, broadly defined by Walker & Nees [3] as any display using sound to relay information. Sound is a temporal phenomenon and sonification can be effective at representing time based data [4]. Scientific data is usually sonified using either: Audification or Parameter Mapping Sonification.

Kramer's [5] definition of audification cited by Dombois & Eckel [6] describes it as being the direct translation of data waveforms into sound. The waveform itself might not even belong to the sound domain. The use of amplitude, pitch or equalisation to modify a wave and reveal previously inaudible data, can be classified as an audification.

Parameter Mapping is when data values are mapped using various acoustic attributes (Pitch, Amplitude, Timbre, Rhythm and Spatial Dimensions). This is determined by the data source and available synthesis parameters. Appropriate data preparation determines the success, especially with a multivariate dataset. Dimension reduction techniques must be implemented or complementary derivatives can be added. Data Dimension Reduction is influenced by two factors. The dimensions of synthesis parameters must be utilised as fully and efficiently as possible. Noise or distortion resulting from the parameter mapping must be eliminated to ensure accurate perception. The mapping procedure poses two challenges. The first to ensure that a proper formalization connecting factual data to the elusive nature of the perceptual domain has been achieved. The second that there is good mapping ensuring that the data and its synthesis achieve perceptually valid results [7].

1.2. Sonification Design

Dubus & Bresin [8] citing Scaletti [9], state sonification only becomes relevant if it is communicating original information comprehensibly. Hermann [1] developed four steps that could aid effective sonification. Sound must reflect objective properties or relations to the input data, echoing Scaletti's [9] statement about the importance of the semantic accuracy of sonification portraying data. The second step calls for systematic transformation of data. In that it is precisely defined, and users can comprehend how changes in the data are directly related to changes in the sound. The third step states that the sonification should be reproducible. The same data and similar sonic interactions should sound identical to the original. For a sonification design to be precisely replicated it is important for the original designer to convey comprehensible and effective guidelines that others can easily follow. The fourth step states

that a sonification system should be able to be utilised repeatedly with different or identical data.

Sonification design has been compared to visualisation and argued that well established visual practices are partly to blame for sonification's restricted appeal amongst scientists. Listening to scientific data has been criticised as being broadly subjective and deemed unscientific [10]. The immersive and emotional qualities of sound seem to reinforce this claim negating sonification being recognised as an accepted scientific method, and seen merely as an art [11].

The comparison between sonification and visualisation does not only reflect differences but also similarities. Any form of design practice methods in sonification, visualisation or HCI design are a constant learning process that evolve over time. Advances in technology and social trends determine the way that design evolves. Accumulative, collected, knowledge acquired is usually compiled to reflect the design practice within a field and to establish codes of practice or methods. In the sonification community a 'Sonification Handbook' was collated by Hermann and Hunt in 2011 [12]. It contained the ideas of leading practitioners in the field and established definitions, terms and guidelines to assist practitioners and help bring about an established code of design practice.

There are however various publications about sonification research that tend to lack relevant information about the design approach used in the study. It is often hard to understand for what purpose the sonification was designed, for whom and how design decisions were taken. Frauenberger [13] analysed 82 submissions published in ICAD 2007 and found that only 23 of the publications described the design of the auditory display. Verona and Peres [14] identified that many sonifications had not been empirically evaluated adding further lack of understanding of how to design sonifications for specific tasks.

1.3. Requirements Gathering

The importance of a thorough investigation of a data-set is also echoed in Barrass's [15] sonification design method "TaDa" (Task-oriented, Data-sensitive method for auditory information design). This method is based on four design principles. The first is "Scenario Description" or a short narrative, in the form of a research question, describing the information process activity that the sonification is designed to support. The second step, "Requirements Analysis", means that requirements must be in line with the research question. It must provide an answer that can be analysed and from which data is characterised from the requirements process. The third principle is about "Representation Design" acquired from the requirements process that are useful to the task and true to the data. Once the investigation and understanding of the data has been established the "Realisation" of the sonification can be fulfilled. The importance of appropriate data gathering and implementation of task analysis methods has been emphasised in multiple sonification papers [15], [16], [3], [17], [18], [19].

1.4. Designing for End-users

Many sonifications have been created without involving endusers. This means that the user is unaware of what elements of the data the sonification parameter mappings are representing. Black *et al.* [20] describes how the use of auditory displays in image guided intervention has been largely neglected despite its benefits. Future work could include working more closely with auditory display designers to create

more meaningful displays [20]. The problem of not including end users in the design process was identified by Kramer *et al.* [21] in 1997 and has yet to be fully resolved. He stated that sonification research needs to be user and task centred. Although it had been successful in a broad range of application areas, it was still not clear as to how to design an effective, working sonification for a specific task. This could be a reflection of open ended design approach and not a failure in the design process. Kramer suggested that progress in sonification would require specific research directed at developing predictive design principles or design guidelines. Sonification is naturally task dependant, requires adequate representation for data portrayal and user interface interaction. The report describes a number of considerations that should be taken into account when designing a sonification (see table 1).

Table 1 Guidelines for designing a sonification user interface

Considerations	Descriptions
Control	Parameter controls for sound parameters that are efficient, effective & accessible
Mapping	Provides flexibility & the ability to design new sonification mappings allowing the user to have intuitive control over data dimensions in relation to sound parameters
Integrability	To allow the different formats of data from different disciplines to be imported into the system and then sonified
Synchrony	To allow easy integration with other display systems like VR systems and other visual or assistive technologies
Experimentation	To integrate a perceptual framework for testing overall mapping functions and sound synthesis

Source:- Kramer *et al.* [21]

1.5. Evaluation Methods used to test Sonifications

There is often an open-ended approach to sonification design but patterns are noticeable in the way practitioners map parameters [22]. A total of 22 publications were chosen and analysed to explore sonification design. These publications were selected since they reflected a variety of applications of sonification and followed different design methods. This indicated the diversity of sonification application. There were publications about sonifications that were created for the same field of study, showing the different design approaches taken by various practitioners to represent similar aspects of data. Eleven had not been tested and the papers only discussed their sound design methods [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33]. The other eleven publications [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [14] were evaluated by investigating their methods of data gathering, whether they chose audification or parameter mapping, if there was training prior to testing, the testing procedure itself, the results and any mentions of further development of the sonification.

1.6. Requirement Gathering Methods

Effective methods of practice typically follow a rigorous investigation to learn about the data. Most sonification designs are constructed by first obtaining information from relevant literature [34], [35], [38], [39], [40]. Other sonification designs were based on conducting interviews with the proposed endusers [36], [37], [41], [42], [43], [14]. Two publications did not explain clearly how the requirements or data gathering processes were conducted. These publications were "Not Informative" [36], [37]. Two studies gave detailed

explanations about the type of information that was gathered but were vague about how it was gathered. These publications were “Partly Informative” [41], [43]. There were two publications that gave a clear explanation of the data gathering investigations, which were “Informative” [42], [14]. Requirements and data gathering were often under-reported in publications. Out of the 6 reports that conducted interviews or involved an enduser in the design process, only two gave descriptive information about these procedures. Some publications gave ample information about the data but were scarce on reporting how it was obtained.

1.7. Parameter Mapping Techniques

Parameter Mapping descriptions are occasionally not accessible or easy to grasp. Table 2 gives an indication of mappings that were found in each of the publications that had some form of testing. Boschi *et al.* [34] has been included in the table to indicate the testing methods but used audification as a sonification technique to categorise seismic signals. Vertical-Component recordings of the November 2011 magnitude 5.6 Oklahoma Earthquake made at 17 stations (< 500 km) epicentral distances were speeded up from 40 Hz to 6 kHz in pitch and from 300 to 2 seconds of playback time.

The most common parameter mapping is the use of pitch. Most of the sonifications analysed for this study involved multiple mappings where more complex representations of the data were required. The exceptions Brewster & Murray [35] who only mapped pitch to movements in stock prices and Schaffert *et al.* [41] who mapped pitch using Middle C as a zero point and notes above and below being higher or lower to that point.

Table 2 Parameter Mappings, Test Methods & Results

Author	Data	Mappings	Participants	Training	Positive Results	Negative Results
34	Earthquake	* Audification	24	Yes	✓	✗
35	Financial	Pitch - up (higher) down (lower)	12	Yes	✓	✓
36	Financial	Pitch, Tremolo, Note length	12	Yes	✓	✗
37	Oil Well	Pitch, Tone, Amplitude	20	Yes	✓	✓
38	Alzheimers	Pitch, Rhythm, Tone	2	No	✓	✓
39	ECG	Pitch, Amplitude, Harmonics	22	Yes	✓	✗
40	Motor Skill Learning	Noise, Pitch, Rhythm	45	Yes	✓	✓
41	Rowing	Pitch	23	Yes	✓	✗
42	Piloting Aircraft	Pitch, Rhythm, Tone, Syllables	17	Yes	✓	✓
43	Internet Log Data	Tone, Note Length, Rhythm, Amplitude	Analysed by research team	No	✓	✓
14	Surface Electromyography	Pitch, Amplitude, Panning, Filter	43	No	✓	✗

1.8. Test Methods and Results

In some instances participant training prior to testing was needed (see table 2). Eight sonifications had some form of training. Negative results were only reported by 5 of the designers. By giving a more positive overview of sonification's effectiveness it could be portraying a false indication of the general effectiveness of sonification.

1.9. Further Development of the Sonification Design

Out of the 11 studies that were evaluated only 4 publications provide information about future work [36], [38], [34], [37]. In the case of Jamieson & Boase [43], a general overlook on the future of work on sonification is presented but there is no

specific information about further development of their sonification model.

1.10. Obtaining more information about Sonification design

To obtain more information about sonification design it was decided that it would be worth interviewing sonification designers. Grounded Theory was used as a method of investigation for this study. It investigates what occurs in the research setting and explores how they explain statements or actions. The data collected is then studied and compiled through an analytical process. The researcher has to be open to what is happening in the studied scenarios and the interview statements. Grounded theory methods consist of systematic, yet flexible guidelines for qualitative data collection and analysis which are used to construct theories ‘grounded’ in the data themselves. The data is simultaneously collected and analysed it is separated, sorted and synthesised through a process of qualitative coding. Analytical codes and categories are created from the data and not from pre-existing conceptualisations. The researcher seeks to discover any social processes reflected in the data. Once the data has been sorted into categories these are integrated into a theoretical framework [44].

2. METHOD

Grounded Theory was used as a method of investigation for this study. Semi Structured interviews were conducted to ask practitioners about the purpose of their sonification, the design, testing, results, and the outcome of the study. The 11 participants had designed sonifications for a variety of applications. Some of them have developed guidelines about the sonification design process. Participants were recruited by email. Eighteen people were initially contacted and 14 replied. Contact details were obtained from ICAD publications. A variety of sonification researchers were chosen with different degrees of experience in the field. One side of the scale consisted of well established members who have heavily contributed to the field. The other side were people just starting in sonification research and development. This was done with the intention of obtaining a rich spectrum of different perspectives, knowledge and ideas concerning sonification design. The University ethical procedures were followed. This included the provision of a participant information sheet and an informed consent form. The questions were based around the design processes of specific projects and sonification design in general.

A different set of questions were asked to one participant who was interested in talking about sonification design guidelines. These questions were aimed to learn more about the effectiveness of these guidelines, how these assist practitioners and whether or not they are being applied.

2.1. Procedure

Participants were contacted via Skype and the interviews were recorded on a portable audio recorder. Due to qualitative approach of the investigation no pilot study was conducted, allowing the interviews to evolve and be adaptable to each of the interviewees. Participants were asked questions about sonification design according to the outcome of the interview. Some people spoke about a specific project. Others discussed various projects giving details on how each sonification was designed. The interviews were approximately one hour long.

Each interview was transcribed and coded. The coding procedure passed through three different phases of extraction and refinement. The first phase was to obtain an initial set of codes and descriptors from the participants' responses. The second phase was to refine the codes and definitions and to see what was meaningful or redundant. The third and final phase was to check for consistency in the results. No form of inter-rater reliability was conducted due to the use of Grounded Theory as a method of investigation.

3. RESULTS

A total of 11 people were interviewed about sonification design. Participant's experience in sonification design varied. The more seasoned sonification designer's seemed to have a more global perspective about the workings of sonification design. Those who were newer to the field often spoke more about their sonification projects and the experiences encountered during the design processes.

3.1. Different Opinions and Common Ground

Paul Vickers (PV) made reference to the first ICAD conference: *"really you can trace it back to, in terms of where things really took off, was 1992 with the first ICAD."* Certain practitioners believe that sonification has been successful within this limited time frame. Bruce Walker (BW) clearly states that: *"I know absolutely, it is being used more and more."* Whereas Steven Barrass (SB) argues that the use of sonification is *"Pretty low or Zero really"*. When asked about effective sonification design Keith Nesbitt (KN) replied that: *"Some of the questions that you asked have been and are still being asked also about visualisation."* The common ground shared by the designers was that the sonification design process is challenging. Daniel Verona (DV) mentions how: *"It's not easy to do a direct correlation between the design criteria and the mapping"*. KN describes that even a well prepared design could sometime falter: *"Trying to come up with understanding how to design these displays, your best laid plans can often go astray"*. These views are echoed by Philart (P): *"This is what we are doing including users and do participatory design, but still it's not really perfect"*. In other disciplines of design such as HCI and Visualisation there are two key factors that are common practice. It is clear who the product is designed for, and users are involved in the design process. The second is that numerous iterations of the design are made. These practices have become more common in sonification design over the years, and all the participants spoke about these processes when describing data gathering and design exercises.

3.2. Data Gathering Process and Comprehension

All the practitioners discussed how they had consulted with people that they were designing for, or those that were related to the design process. In other instances the designers had run mock tests with preliminary designs to improve their models. Jamie Ferguson (JF) described how there was a continuous back and forth interaction with the European Space Agency, ESA, in the design of his sonification of star maps: *"So they were pretty much on the ball with you. They would give you feedback on it and you would improve the model according to the feedback."* Robert Alexander (RA) also described this process by stating: *"I needed to first learn their approach and what's unique about their dataset. And then it became a back*

and forth dialogue." RA's thorough investigation of the dataset he was sonifying led him to change his approach from using parameter mapping to audification: *"I had decided to go to the other route and actually, really use direct audification to try to find, really, again, you know, subtle and nuanced features."* It was this switch to audification that led to a ground-breaking discovery in the study of solar astronomy (RA): *"It was the audification work that led to the new scientific discoveries, which then led to the published papers"*. The discovery itself helped scientists to change their methods of research and it was through RA's sonifications that this became possible: *"I was listening to carbon 6 to 4 and what ended up happening was that no one had looked to deeply into the 6 to 4 ratio and then when we went back and plotted it, it was pretty easy to see that carbon is much kind of smoother curve."* As P had pointed out: *"Sometimes sonification is very successful and sometimes just...nothing"*. BW explains that: *"Getting it right requires careful planning, information architecture, a lot of knowledge in the field, mastery of the tools"*.

KN refers to the importance of iterative prototyping: *"The best approach these days is to try something and provide iterative prototyping"*. This means that numerous discussions are held between the users and the designers. Tests are conducted with different versions of the sonification and improvements are made according to the results.

P described the sonification design of emotional conveyance of dancer's movements. The team consists of dancers, musician and sound designers. The movements of the dancers are recorded whilst they danced to music that evoked certain emotions in them. These movements were then given to the musicians and the sound designers and designed sound and music that convey the emotions reflected in the dancers movements. A survey was then carried out with 20 participants to see which of the three final designs worked best. (P) *"We just finished a very basic survey type study with around 20 participants, non-dancers, they are just college students."* BW describes the complexity of designing any kind of interface and that inter-disciplinary knowledge is required for effective sonification design: *"It's hard work and requires inter-disciplinary teams and broad background to get it right."*

Gionfrida (LG) described how: *"We proposed different tune structures, thinking they would sound better at the time but they were more difficult for physicians to understand."* LG's final sonification reaped positive results: *"There was essentially an increase (improvement) between how they performed without sonification and how they performed with sonification"*. That is why the investigation and design process is essential for an effective sonification design, as clearly stated by SB: *"Design and design research is a real critical enabler in this field."*

Clearly identifying the scope of a sonification project is an important factor and DV emphasised this in his sonification design. DV was convinced that by using a Task Analysis approach a sonification could be much more effective. To explore this he designed two sonifications using a Task analysis technique known as GOMS and two were data-based designed. DV describes how: *"My task based sonifications performed quite a bit better than my non-task or data based sonifications"*. It proved that: *"Designing for a specific task can help listeners perform the task with the sonification"*.

One other essential part of the investigation process is reference to the literature. To obtain knowledge a designer has to learn about the subject by following currently existing models or prototypes that have been designed and tested. The interviews presented a number of arguments that reflected a weakness in this aspect of the sonification design process.

3.3. Limitations in Sonification Literature

There was shared concern about the lack of descriptions of sonification design processes in the literature of the field. DV described this by stating that: *"What I found so frustrating in sonification literature was, there was hardly ever any explanation of why a specific mapping or why a specific frequency range was chosen"*. SB also expresses certain limitations with the literature on the subject: *"A lot of their assumptions (the designers) are built into their decision making and they are the only ones that understand the rationale. And even when you read their papers with very little explanation about stuff, it's still assumptions that have never been tested"*. PV describes: *"If you look at the literature, a lot of it is still one off, bespoke, sonification systems with studies that are narrow in scope so the results are not that generalizable"*. This point was further emphasized by BW *"It's either part of a class or a Master's thesis and they pick it up over the course of a semester. They [students] are still obviously novices and even if they do very well, the chances are they haven't learned enough"*.

There was also the mention of a lack of vocabulary in sonification design. Mark Ballora (MB) clearly states that sonification: *"Just doesn't have the established vocabulary that visualisation does"*. He also refers to Carla Scaletti's comments in her ICAD 2017 keynote that also emphasizes this point: *"So she was talking about, that it doesn't have an established vocabulary (Sonification)"*. Other fields of study such as Human Factors and Ergonomics design and Visualisation have an established vocabulary. DV argues: *"If there is a language of design that is familiar to people outside our community, then suddenly there is a less of a gap to cross between our work and their work (Human Factors and Ergonomics design)"*.

3.4. Sound Design and Parameter Mapping

3.4.1. Mimicking Nature & Using Associative Anchors

MB states that: *"Natural sounds, sounds that mimic nature, sounds that are completely abstract, it's all project specific"*. Sometimes the obvious association to a sound is required in the sound design. SB describes it as: *"If people want to hear rainfall then they should hear rainfall. If the task is to hear the amount of rain then you should hear that as the amount of rain, not the pitch of a Cello"*. MB also used mappings that mimicked the sound of the wind in a sonification that represented Meteorological storm data: *"Sometimes the storms get more intense but they don't get more symmetrical and she can't see that so readily from the videos but she could certainly hear it from the sonifications"*. The use of familiar sounds was also voiced by Dyer (JD): *"Real world objects make sounds in certain ways and those can be a good starting point for design"*. The familiarity of sound often acts as an anchor and this is taken into consideration when designing a sound and mapping it. Ferguson (JF) describes this anchoring as: *"try to tap into that, when people hear these things, they can begin to*

make analogies about it". PV also describes how easily people make the connection with more natural sounding designs: *"When you do the sound design you have to be aware that any sound you create can have an association with the real world, in fact, people are possibly going to reach for it."*

3.4.2. Simple Mappings

Simple sonification design is at times the necessary way to effectively map a dataset. The neural mappings described by LG were systematically organised to allow simple parameter mappings to effectively represent the data and achieve highly positive results: *"You can get lost with a lot of information. That information has to be very simple, straightforward"*. The simple sonic representation employed by LG still revolved around a multi-layered sound design, not complex, but consisting of more elements to enhance the semantic representation of the data. DV describes how his sonification consisted of simple sound design of more complex representations than simple one on one mapping: *"We can ascribe meaning to a sound, but we're not going to ascribe meaning to a sound based on one dimension of sound. Not just based on pitch, it's a big combination of all these different things"*. JD had also expressed the effectiveness of simple sonification mappings: *"It was informative just because of its ease of use"*.

3.4.3. Psychoacoustic Elements

The psychoacoustic elements of sound design play an integral part in parameter mapping. There are elements that could be more effective than others (BW): *"if you are using something like frequency then you have the built in possibility of these logarithmic relationships"* and (BW): *"But amplitude is somewhat unreliable. People are quite poor at discerning levels of amplitude"*. Elements of Rhythmic representation are an easy association, and also have a temporal effect that are difficult to represent visually. JD describes how Rhythmical elements helped in coordinating movement of users in motion gesture sonification: *"Trajectories of movement when you're trying to time yourself quite precisely, tend to be more accurate if you go for this strategy, ballistic (Sound)"*.

3.4.4. Aesthetics

Aesthetics in Sound Design are a key element in the effectiveness of a sonification and MB clearly states that: *"I think the aesthetics are critical to it."* BW echoes this: *"Aesthetics absolutely plays a crucial role in any kind of technology. Sound is absolutely included in that. Something has to sound good, and also has to sound right"*. Sonifications can easily become unwanted sounds (KN): *"Things can sound like noise"*. It is this element of noise that has often led people to switch sonifications off (JF): *"What I've discussed with most people is that the problem with sonification is that most people turn them off"*. This doesn't necessarily mean that the sonifications sound unpleasant. Many of the researchers insisted that aesthetics does not mean that the sound designs have to necessarily sound pleasing. PV sees this as a problematic simplification *"a lot of sonification researchers view aesthetics simply through the lens of 'Does it sound nice?'"*. SB agrees with this: *"I don't think of aesthetics in terms of pleasantness or decoration. I'm much more vested in the idea of like I think that if it is efficient it will be aesthetic"*.

3.5. About Positive Results of Sonification Testing

The sonification designs discussed in these interviews were not all tested scientifically. Those that were reaped positive results. Those that weren't tested had some form of positive feedback. It can be agreed that sonification is most effective in representing temporal information. There are cases when sonification can out perform visualisation and this is usually when the user is unable to refer to a screen and needs some form of data feed in order to carry out a task. This is apparent with forms of motion guided systems. The literature reflects positive testing results but still seems to remain within the realm of isolated tests and this presents a false positive about the success rate of sonification and its use beyond testing.

3.5.1. Sonification Testing – An Isolated Success Story

The discussion concerning Sonification testing and results concluded that many were done in isolation. Positive outcomes can give a false positive as to the success of sonification. MB criticises this approach and its limited perception of the bigger picture: *“A lot of the papers that you read it's like kind of bleep bloopy and they did a pilot test with 20 people and they got this out of it so by golly this is worthy of further research.”* SB describes how the testing should be about evaluating current methods, as there is no proof that these mappings actually work, rather than constantly trying to prove that sonification works: *“the most common technique in sonification, it's still pitch mapping and it's hardly ever been evaluated”*. SB goes on to criticise current testing techniques: *“There's a big difference between hypothesis testing and knowledge building”*. PV describes a more daunting scenario: *“I don't think we've learnt very much about sonification design.”* This could be due to a lack of understanding from the designers part and PV goes on to say that: *“I think that we have a lot of sonifications that don't really do what they're doing with an understanding of the wider perceptual issues.”*

With all the testing that has taken place sonification still has not become a more widely accepted data analysis tool. This is probably the yard stick with which many designers in the Community are measuring the success of sonification.

4. DISCUSSION

The interviews give a clear picture of what sonification is, its potential when designed correctly and that despite these efficiencies, sonification has still not become a more accepted tool for scientific data representation and analysis procedures. There are those who have criticised sonification literature and have pointed to lack of certain details that could help convey vital knowledge to any that would be interested in using the tool. Brazil [2] had identified how daunting it could be for a person who is trying to learn about sonification design from the literature. Verona had expressed this frustration too in his interview comments and had also expressed a further frustration towards lack of clarity in most publications. His comments clearly stated that: *“Maybe because of that people just weren't designing for the task or maybe if they were, they weren't explaining how they did that in their papers”*. Frauenberger [10] had also found these ‘holes’ in sonification literature where the descriptions of auditory display designs were lacking.

One of the main issues that have been highlighted is that a lot of sonification designs are contained projects, often

dissertations or isolated studies. Many of these studies involve people who are new to the field, who have to learn as much as they can within a limited time frame and to create an effective sonification. This probably means that a lot of sonification designs lack the necessary knowledge, skillset or the lifespan to exist beyond the confines of that project. Walker had described how this was a common feature in all types of design scenarios: *“The vast majority of software, just in general, are built and go nowhere. I think that we also see the same kind of thing with sonifications.”*

This lack of knowledge from the designer's angle could be one of the reasons why many sonifications are redundant. It is like the metaphor of ‘the blind leading the blind’. And it could also be the reason why the field has not yet established a common vocabulary. This is where sonification differs from Human Interactive fields and Visualisation, which both have established vocabularies. The majority of sonification designs analysed for this study lacked Task Analysis design methods. Kramer *et al.* [18] had outlined this problem in 1997 and Barrass [12] also proposed Task Analysis design techniques in 1996. The emphasis for appropriate data gathering has been echoed in many publications [12], [13], [3], [14], [15], [16]. The only study which used Task Analysis in the papers that were analysed was the Verona & Peres [11] study. Task Analysis design methods could help to form a vocabulary for sonification design. This means that a lot still has to be learned. As Vickers commented: - *“I think our knowledge of sonification design and theory is still fairly primitive.”*

4.1. It Works

There are those that have a more positive outlook towards where sonification is at the moment and where it is heading. BW strongly believes that sonification is being utilised: *“Scientific sonification is being used more. It's being used in oil and gas exploration and circumstances where there's a lot of data. It's being used in financial services, investment context, engineering and space exploration. Most of these situations where even the best visual displays are still not doing it. Multimodal, time varying complex patterns, all the kinds of situations where we point to sonification and say that it's likely to succeed.”* One of the reasons why there is a lack of reporting of the more widespread use of sonification is that (BW): *“We don't hear about all of those contexts. Some of those things, especially the financial investing and oil discovery, the big companies are not going to advertise the fact that they are using these specialised tools.”* Philart was reluctant to measure the success of sonification because it depends on the context. The use of sonification in industry is often overlooked and as an example P describes the incident where he purchased a washing machine and found that the sound design was the same one he had created 10 years earlier: *“When the washing machine cycled down the pitch generated a melody that I composed....ok seems successful. It's like 10 years later they still use that sound.”* P did go on to say that in academia, the success of sonification is different: *“But it is really hard to see that type of consistency in academic research.”* LG sees a strong potential in the use of sonification, especially in instances where the user cannot look at a visual display due to the task that they are carrying out. This is extremely prevalent in the medical realm: *“What I have always been thinking is that in the medical domain, they actually always rely on the visual information and they should actually rely more on the auditory feedback as well.”*

4.2. Not Yet Maybe Later

There are those that are not seeing any current progress but strongly believe it is a potential tool and that it will come into acceptance in the future. This more positive view towards the direction of sonification is based on the fact that most sonification designers believe that people in general are not really taught how to listen. KN, in describing how the visual component in humans is more dominant states the following: *"This is part of the problem in terms of making sonification more broadly used, is that there is just a stronger visual preference for many people and their hearing senses are under-utilised, under-trained."* PV also identifies this issue: - *"What we should be aspiring to is much greater skill in our listening."* MB strongly believes that the sophistication of the human hearing faculty automatically leans towards sonification becoming an accepted data analysis tool: *"That's the thing about sound. That's why I think it is inevitable that this becomes a part of our environment because the human animal is so responsive to sound."*

5. SUMMARY

The results give an indication of how each step of the design process is currently being practiced by sonification designs. It also portrays the complexity of sonification design and the contrasting views regarding design, testing, the outcome and effect of sonification. The main elements that can be highlighted from these results is that effective sonification design requires in-depth investigation of the data, the inclusion of the user in the design process, a valid number of iterations in choice of sonification technique, parameter mapping and sound design. The testing of the sonification has to be well studied and should reflect the overall scope of the sonification design. One time testing is ineffective and gives false positives. The testing process should be considered as further iteration of the design procedure. To rate how successful sonification is in general, there is a divided outlook with regards to this, but there is an overall consensus that much more work has to be done to improve sonification design and implementation. Even though opinions are divided about the current status of sonification's present implementation in scientific research, there is a positive outlook and common belief that sonification will become an accepted and widely used means of data representation.

6. ACKNOWLEDGMENTS

Acknowledgement and thanks go to all those that participated in this study: James Ferguson, Robert Alexander, Myoungsoon 'Philart' Jeon, Bruce Walker, Mark Ballora, Steven Barrass, Keith Nesbitt, John Dyer, Letizia Gionfrida, Paul Vickers & Daniel Verona

7. REFERENCES

- [1] Hermann, T. (2008). Taxonomy & definitions for sonification & auditory display. In P.; Susini & O. Warusfel (Eds.), . Paris, France: International Conference on Auditory Display 2008 (ICAD 2008). Retrieved from <http://www.icad.org/Proceedings/2008/Hermann2008.pdf>
- [2] Brazil, E. (2010), A review of methods and frameworks for sonic interaction design: Exploring existing approaches. Auditory Display – 6th International Symposium, CMMR/ICAD 2009, Copenhagen, Denmark, May 18-22, 2009, Revised papers, Volume 5954 of lecture notes in Computer Science, pages 41-67. Springer, 2010
- [3] Walker, B., & Nees, M. (2011). Theory of Sonification. *Principles of Sonification: An Introduction to Auditory Display and Sonification*, The Sonification Handbook Chapter 2, (November), 1–32. Retrieved from http://sonify.psych.gatech.edu/publications/pdfs/2006preprint-WalkerNees-SonificationChapter_v09.pdf
- [4] Vickers, P., Holdrich, R (2017), "Direct Segmented Sonification of the data domain", arXiv pre print, arXiv:1711.11368, 2017, arxiv.org
- [5] Kramer, G. (1994). *Auditory Display: Sonification, Audification and Auditory Interfaces*. Addison-Wesley, Reading (MA). Proceedings Volume 18, Santa Fe Institute Studies in the Sci (Book 18), 672 pages, CRC Press (June 7, 1994), ISBN-10: 0201626047
- [6] Dombois, F., & Eckel, G. (2011). Audification. In T. Hermann, A. Hunt, & J. G. Neuhooff (Eds.), *The Sonification Handbook*. Logos Publishing House, Berlin. Retrieved from <http://sonification.de/handbook/index.php/chapters/chapter12/>
- [7] Grond, F. and Berger, J. (2011). Parameter Mapping Sonification. In T. Hermann, A. Hunt and J.G. Neuhooff (eds.) *The Sonification Handbook*. Berlin: Logos, 363–97.
- [8] Dubus, G., and R. Bresin. 2013. "A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities." *PLOS ONE* 8(12):e82491.
- [9] Scaletti C (1994) Auditory display: sonification, audification and auditory interfaces, Addison Wesley Publishing Company, chapter 8: Sound synthesis algorithms for auditory data representations. pp. 223–251
- [10] Supper, A. (2016) Lobbying for the ear, listening with the whole body: the (anti-)visual culture of sonification, *Sound Studies*, 2:1, 69-80, DOI: [10.1080/20551940.2016.1214446](https://doi.org/10.1080/20551940.2016.1214446)
- [11] Supper, A. (2014). Sublime frequencies: The construction of sublime listening experiences in the sonification of scientific data. *Social Studies of Science*, 44(1), 34–58. [http://doi.org/10.1177/0306312713496875](https://doi.org/10.1177/0306312713496875)
- [12] Hermann, T., & Hunt, A. (2011). *The Sonification Handbook* (1st ed.). Berlin: Logos Publishing House, Berlin. Retrieved from <http://sonification.de/handbook/>
- [13] Frauenberger, C. (2009), Auditory Display Design: An investigation of a design pattern approach. PhD dissertation., Interaction, Media, and Communication, Queen Mary University of London, 2009, p 59
- [14] Verona, D., & Peres, S. C. (2017). A Comparison Between the Efficacy of Task-Based vs. Data-Based sEMG Sonification Designs. In Proceedings of the 23rd International Conference on Auditory Display (ICAD 2017)(pp. 49 – 56)
- [15] Barrass, S. (1997). Auditory Information Design. The Australian National University Australia, 288. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.127.6228&rep=rep1&type=pdf>
- [16] Anderson, J., (2005), "Creating an Empirical Framework for Sonification Design," in *Proceedings of ICAD 05 - Eleventh Meeting of the International Conference on Auditory Display*, 2005, pp. 393–397.
- [17] Flowers, J., (2005), "Thirteen Years of Reflection on Auditory Graphing: Promises, Pitfalls, and Potential New

- Directions,” in *Proceedings of ICAD05 - Eleventh Meeting of the International Conference on Auditory Display*, 2005, pp. 406–409.
- [18] Sanderson, P., Anderson, J., Watson, M., (2000), “Extending Ecological Interface Design to Auditory Displays,” in *Proceedings of the 2000 Annual Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of the Ergonomics Society of Australia*, 2000, pp. 259–266.
- [19] Peres, S.C., & Verona, D (2016), “A Task-Analysis-Based evaluation of sonification designs for two sEMG tasks”, The 22nd International Conference on Auditory Display (ICAD 2016), July 2nd – 8th, 2016 Canberra, Australia
- [20] Black, D., Hansen, C., Nabavi, A., Kikinis, R., and Hahn, H., (2017) “A survey of auditory display in image-guided interventions,” *Int. J. Comput. Assist. Radiol. Surg.*, pp. 1–12, 2017. <http://doi.org/10.1007/s11548-017-1547-z>
- [21] Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., MIner, N., & Neuhoﬀ, J. (1997). *Sonification report: Status of the field and research agenda*. Retrieved from <http://www.icad.org/websiteV2.0/References/nsf.html>
- [22] Ibrahim, A. A. A., Yassin, F. M., Sura, S., & Andrias, R. M. (2011). Overview of design issues and evaluation of sonification applications. *2011 International Conference on User Science and Engineering (I-USER)*. <http://doi.org/10.1109/IUSER.2011.6150541>
- [23] Ness, S., R., Reimer, P., Krell, N., Odowichuck, G., Schloss, W., A., & Tzanetakis, G. (2010). Sonophenology: A tangible interface for sonification of geo-spatial phonological data at multiple time scales. The 16th International Conference on Auditory Display (ICAD-2010), 335–341
- [24] Ben-Tal, O., Berger, J., Cook, B., Daniels, M., Scavone, G., Cook, P., (2002), “SONART: The Sonification Application Research Toolbox”, *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan, July 2002
- [25] de Campo, A., Dayé, C. (2006), “Navegar e preciso, viver nao e preciso”, The 12th International Conference on Auditory Display, June 20-23 2006, London
- [26] Kadkhodaie, A. Rezaee, R. (2017), “Have you ever head the sound of well logs or reservoir data?”, *Journal of Petroleum Science and Engineering*, 156 (2017) 340 – 347
- [27] McGee, R., Rogers, D., 2016. Musification of seismic data. In: The 22nd International Conference on Auditory Display (ICAD-2016). Canberra, Australia.
- [28] Polli, A. (2005), *Atmospherics/weather works: A spatialized meteorological data sonification project*. The 10th International Conference on Auditory Displays (ICAD), pages 31–36, Sydney, Australia, July 2004.
- [29] B. Sturm (2000), "Sonification of particle systems via de Broglie's hypothesis," *Proceedings of the 7th International Conference on Auditory Display*, Atlanta, GA, April 2-5, 2000
- [30] Vogt, K. (n.d.), “Sonification and Particle Physics”, Department of theoretical physics, Institute for physics, Karl – Franzens – University Graz Austria, Institute for Electronic Music and Acoustics, University for Music and Dramatic Arts Graz, Austria, katharina.vogt@uni-graz.at, Retrieved 16th April 2018, from, https://www.researchgate.net/profile/Katharina_Vogt/publication/228592750_Sonification_and_particle_physics/links/02e7e51ca83485c5f5000000/Sonification-and-particle-physics.pdf
- [31] Winton, J., Gable, T. M., Schuett, J. & Walker, B. N. (2012), “A sonification of kepler space telescope star data,” in *Proceedings of the 18th International Conference on Auditory Display*, Atlanta, GA, 2012.
- [32] Barrass, S., & Zehner, B. (2000). Responsive sonification of well logs. In P. Cook (Ed.) *Proceedings of the 7th International Conference on Auditory Display* (pp. 72 - 80). Atlanta, GA: International Community for Auditory Display, April 2-5, 2000
- [33] Baier G, Hermann T, Stephani U (2007) Event-based sonification of EEG rhythms in real time. *Clinical Neurophysiology* 118: 1377–1386.
- [34] Boschi, L., Pate, A., Holtzman, B., Le Carrou, J.L. (2015), “Can auditory display help us categorize seismic signals?”, The 21st International Conference on Auditory Display (ICAD-2015), July 8-10, 2015, Graz, Austria
- [35] Brewster, S., Murray, R., (2000): Presenting Dynamic Information on Mobile Computers. *Personal Technologies* 4(2), 209–212
- [36] Janata, P., and Childs, E. (2004). Marketbuzz: Sonification of real – time financial data. Paper presented at the 10th International Conference on Auditory Displays, Sydney, Australia. July 2004
- [37] Quintero, G., 2013. Sonification of oil and gas wireline well logs. In: The 19th International Conference on Auditory Display. Pp. 301–306, 6-10 July, Poland.
- [38] Gionfrida, L., Agnieszka, R., Keary, J., Mohanraj, H., Friedman, Kent P., (2016), The Triple Tone Sonification Method to enhance diagnosis of Alzheimer’s Dementia, the 22nd International Conference on Auditory Display. July 2016, Canberra, Australia
- [39] Kather, J-K. Hermann, T. Bukschat, Y. Kramer T., Schad, L., & Zäüllner, F-G (2017), “Polyphonic sonification of electrocardiography signals for diagnosis of cardiac pathologies.” *Scientific Reports* 7 (2017), 44549.
- [40] Dyer, J. Stapleton, P. Rodger, M. (2016), “Sonification movement for motor skill learning in a novel bimanual task: Aesthetics & Retention Strategies”, School of Psychology & Sonic Arts Research Centre, Queens University Belfast, The 22nd International Conference on Auditory Display (ICAD-2016), Canberra, Australia
- [41] Schaffert N, Gehret R, Mattes K (2012) Modeling the rowing stroke cycle acoustically. *Journal of the Audio Engineering Society* 60: 551–560.
- [42] Valery, B., Scannella, S., Peysakhovich, V., Barone, P., Causse, M., (2017), Can an aircraft be piloted via sonification with an acceptable attentional cost? A comparison of blind and sighted pilots, *Applied Ergonomics*, Volume 62, July 2017, PP 227-236, <https://doi.org/10.1016/j.apergo.2017.03.001>
- [43] Jamieson, J. Boase, J. (2017), “Listening to Social Rhythms: Exploring logged interactional data through sonification”, *The Sage handbook of social media research methods*, Thousand Oaks, CA: Sage Publications
- [44] Charmaz, K. (2006), “Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis”, SAGE Publications Ltd 1 Oliver's Yard 55 City Road London EC1 Y1SP, ISBN-10 0-7619-7352-4

Towards Ecological and Embodied Design of Auditory Display

Stuart Duncan Haffenden Cornejo

HighWire Doctoral Training Centre
Lancaster University
Bailrigg, Lancaster, Lancashire, LA1 4YW, UK
s.haffenden1@lancaster.ac.uk

ABSTRACT

Auditory display research has been criticised over a perceived lack of progress in tackling key issues relating to usability and user-experience. However, emerging trends in design-thinking present new tools for addressing the usability concerns that have long beleaguered this field of inquiry. In this paper, we provide an in-depth analysis on the emergence of design-based approaches in auditory display research by mapping out the progression of current research in the field. Through an ecological and embodied approach to perception and cognition, we then evaluate user-centric design strategies as tools for better understanding complex design spaces and improving usability. We then present a discussion to elucidate the benefits to auditory display research of employing user-centric design strategies for future projects.

1. INTRODUCTION

Data has become a valuable asset and driving force in many aspects of our daily lives. The reliance of our global society on connected devices, data and information has spearheaded rapid development in many sectors of industry and society, bringing both radical innovation, as well as disruption from new and complex problems. Technologies like auditory interfaces have presented innovative opportunities for assisting the analysis and practical use of the vast range of data types that are now possible to collect. However, despite all their promise in a number of applications, auditory displays still face several key hurdles that have stifled their path towards ubiquity in our everyday environments [1].

The field of Auditory Display (AD) has long struggled to overcome key issues relating to user-experience [2]. This has been well documented by an ever growing list of publications concerned with the discussion of aesthetics and usability in AD research [3]. In response, it has been suggested that the prevalence of cognitivist and information-processing approaches in AD has led to applications that are limited in scope and ignore critical issues in design such as the aesthetic, environmental and emotional sensibilities of end-users. In turn, this has led to poor usability from designs that may work in controlled settings, but do not translate well to real world environments [4], [5].

Despite the importance of aesthetics and usability in AD [6], many challenges remain unaddressed before auditory interfaces become viable everyday technologies. Critics have noted that overall, AD research appears perpetually mired, attempting to evaluate the pros and cons

between cognitivist/scientific frameworks over more human-oriented/artistic approaches to sonification [2]. This has arguably produced stagnation in the development of innovative ideas and arguments in the field [4], as researchers struggle to reconcile the different perspectives that arise from different schools of thought. Furthermore, due to the interdisciplinary nature of auditory display the prospect of developing an all-encompassing set of guidelines that are applicable to all possible contexts of auditory display applications is unrealistic [7].

To address these aesthetic and usability concerns in auditory display design, recent research has turned its focus towards user-centric design approaches [8] more commonly associated with Human-Computer Interaction and Design Research [9]. These recent developments have the potential of providing researchers with better tools for informing design, by enabling better understanding of users' environments and experiences [10].

2. BACKGROUND

2.1. Aesthetics and Usability

Auditory displays have their origin in scientific and engineering disciplines such as, computing, medicine and aviation [4]. Early AD research consisted mostly of experiments exploring sound as a medium for codifying and communicating information for utilitarian purposes (i.e. medical applications, sonar, etc.). The results of these early experiments and applications in AD are characterized by very functional sonifications of little or poor aesthetic quality (for example, the Geiger counter). These sound signals, albeit informative and appropriate within their given context, are monotonous and induce display fatigue over long periods of exposure [11].

However, as the concepts of sonification and AD have become more widely known in other fields, experimentation with these data analysis techniques outside traditional scientific disciplines has flourished. This particularly the case with the sonic arts. Sonic artists and sound designers began experimenting very early on with data driven audio for producing innovative artistic works [4]. Data driven sonic art presented an opportunity to affect and captivate listeners by presenting novel, tangible and engaging experiences of data [12]. This experimentation, alongside a desire to improve the poor user-experience in early sonification applications, brought about the consideration of aesthetics and interdisciplinarity to the forefront of the discussion on auditory interfaces, as seen at the 2002 edition of the ICAD conference [6].

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

The interest in AD by disciplines grounded in the arts and humanities, gave rise to tensions within this historically scientific field, where practitioners have struggled to value or incorporate the contributions of aesthetic approaches to AD research [13]. The dichotomous discussion of science vs. aesthetics in AD has been the dominating approach, mainly centred around the ambiguity introduced to data streams when codified using a heavily aesthetic approach. In contrast, the more functional and unambiguous signals produced for scientific-oriented sonifications, have been valued due to a preference in the field for quantifiable unambiguous data [1]. With the emergence of this discussion, some insights into the role aesthetics plays in the user-experience of auditory displays has been gained. However, Barrass and Vickers [4] argued that this also presented a barrier slowing progression the field by preventing the true interdisciplinary collaboration which is needed for addressing the critical issues of usability and user-experience in AD design. Despite this stagnation, it is clear the important role aesthetics play in the design of a successful auditory interface, from how it impacts the communication of the codified data, to the overall experience of the user. But what are ‘good’ AD designs? Or appropriate aesthetics? And how does a designer achieve usability?

Over time, considerable effort has gone into classifying different strategies in AD into categories and techniques, such as: audification, parameter-based, and model-based sonification, with the aim of converging and standardizing knowledge in the field [14]. However, a similar classification of design strategies and techniques is not possible to standardize due to the ambiguous and complex nature of this paradigm, and therefore AD designers continue to face a great challenge when choosing a sound palette suitable for the particular data set they wish to sonify. It thus becomes clear that sound selection in AD is a complicated process dependent on the contextual and cultural aspects of listening that influence the perception and engagement of a user with sound [8].

To address the challenge of producing auditory displays that tackle the issues of usability highlighted above, it has become necessary to adopt approaches that place greater emphasis on the user and their experience. Consequently, there has been a rise in ‘design-thinking’ approaches, which have been suggested as possible means for improving usability and user experience in technology design [15]. These approaches break away from the polarising and hierarchical dichotomy of scientific versus artistic representation and instead frame AD design as a dynamic and adaptive interdisciplinary exercise, requiring active engagement with end-users on a more human and ecological level [16]. In the following sections we describe the development of design thinking as a practice and elucidate why it is of interest to recent AD research trends.

2.2. Design Thinking, User-centred Design and Human-Computer Interaction

The idea of ‘design thinking’ can be understood as the careful consideration and application of cognitive processes that culminate in design action [17]. These cognitive processes arise and are inspired from the direct observation and collaboration with end-users using ethnographic tools, with the goal of understanding the user’s context and by

consequence the intended design space in its natural state [18].

Historically, the conceptualization of design thinking developed through four distinct phases, where the focus of the cognitive processes involved in design varied radically [19]. The first phase, covering the periods before 1960, conceived of the designer as an artist involved in an intuitive process. As such, design was traditionally viewed as a step in the latter stages of the product development process, whereby a designer’s role would be limited to the enhancement of the aesthetics of an already developed product, and did not include providing input during the innovation/ideation stages of development [18].

The second phase of development (1960s-1980s) characterized the designer in a more cognitivist light, as a rational and logical agent involved in planning and analysis of scientific rigor [20]. Here, designers aimed at controlling or minimizing the irrational, fuzzy and abstract aspects of design and viewed design as a ‘problem solving’ process dissectible through the rationality of an information-processing mind.

During the third phase of development (1970s-1980s) it was realized that for pure rationality to exist, a designer must have the impossible quality of being able to attain all existing information regarding a particular design problem. As such, certain human actions that endure any level of uncertainty cannot be explained by a rationalist approach and thus the term bounded-rationality was conceived as a midpoint between pure determinism and pure rationality [21]. In this state, similarly to the second phase, the designer applies logic and rationality, albeit in a more dynamic way, to different problems as they arise and without full knowledge of all information beforehand.

The final paradigm shift towards design thinking occurred during the 1980s when the designer’s role was recast to that of the reflective practitioner. This post-rationalist characterization aimed at placing the designer in a more natural state of ‘reflection-in-action’, whereby a designer is no longer solely dependent on preconceived knowledge or theory and technique, but instead embraces uncertainty and reflects on the situation at hand using experience and dialogue to construct a new theory for that unique case [22]. During this last stage of development, design thinking rejected the information-processing view of the ‘problem-solution’ framework to design, and instead embraced the concept of ‘design situation’ or ‘design space’, welcoming the consideration of human and environmental factors more commonly examined by the social sciences [19]. In this paradigm, the designer must excel at: empathy, integrative thinking, optimism, experimentalism and collaboration, as well as self-reflection [23], [24], [24] to better understand the dynamic interactions between designers, users and their environments [18].

Design thinking is inherently a user-oriented activity underpinned by the user-centred design (UCD) methodology. UCD aims to realize the goals of usability through the careful and empathic consideration of user-needs, experiences, environments and other contextual elements at every stage of the design process. Recently UCD has become a widely employed framework in many fields of research and has been adopted as a central tenet of disciplines concerned with technology design, such as, human-computer interaction (HCI), that have recognized

the need to address design-spaces from the point of view of user-experience [25].

As the field of HCI has matured and incorporated design practices, there have been calls to address key issues regarding the consolidation and integration of design knowledge into this traditionally scientific field [26]. The purpose of this integration, as with most intellectual disciplines, is to form conventions and standards that unify a field under an identity and provide legitimacy to practices within it [26], [27], as well as to produce coherent language to describe and scrutinize research under a common banner. However, due to the interdisciplinary nature of design practice, as well as its tendency to produce results that embrace ambiguity and the ill-defined nature of many design problems such formalizations are not possible and indeed counterproductive to the goals of UCD and design in general. Despite the relative resistance HCI presented at first, practitioners have been quite successful in addressing the incorporation of design-based practices into the field by pushing for an abandonment of the ‘scientism’ in HCI culture and instead explore mechanisms to transform the ambiguous data that design-based research in HCI produces, into actionable paths towards design [25], [26], [28].

As such, UCD has proliferated within HCI and resulted in an interdisciplinary and dynamic field of research embracing innovative frameworks and methods. Two recent examples of such frameworks can be seen in the emergence of ecological and embodied approaches to design within HCI. These two notions, grounded in psychology and cognitive science, offer fresh perspectives on how to better understand user experience, perception and behaviour in their natural environments. In the following section we present an account of these two concepts to frame current UCD approaches within HCI and design.

2.3. Ecological and Embodied approaches to Design

The ecological perceptual theory proposed by James Gibson in 1966, came as a reaction to the dominant cognitivist information-processing mentality at the time. Gibson rejected the idea of the world as a chaotic source of meaningless sensory stimuli, onto which perceiving organisms impose structure and meaning through acquired mental representations of the world within their cognitive apparatus [29]. He instead approached perception from a unique perspective for perceptual and cognitive science, by emphasising the situatedness of the perceiver in a naturally structured environment, full of ambient information and controlled by the laws of physics [30]. He characterized perception as an active and exploratory state, constantly orienting the organism towards the environment through action, in order for the organism to respond accordingly to the highly complex, but organized features of said environment. As such, perception here is considered a self-tuning/adapting process in which the discernment of environmental information by sensory organs reinforces and optimizes the perceptual system, tuning it more and more to the subtle invariant features of the evolving yet stable environment it inhabits [5].

Arguably the most fundamental concept introduced by Gibson’s ecological approach is the ideation of *Affordances*. Affordances provide a mechanism for understanding the interrelation between the environment and perceiver, by explaining how the inherent/invariant

properties of objects in the environment and the context of what is occurring *affords* the perceiver with cues on how to regulate its behaviour appropriately in regards to the context without the need of any prior knowledge or cognitive process [31]. The ecological approach makes no distinction between the natural and the cultural environment. As such, perception is not limited to physical objects, and it is possible to extract actionable meaning from abstract phenomena such as language, sounds and symbols [5]. By emphasizing the situatedness of a perceiver in a larger sensory-rich environment through the perception-action cycle, Gibson’s account of perception, not only highlighted the importance of a perceiver’s ecology in cognition, but also nurtured the notion of embodiment.

Up until developments in psychology and philosophy of the mind during the 1960s, the cognitive sciences had traditionally viewed the body of a perceiving organism as peripheral to cognition. However, through the proposition of embodiment, cognition began to be viewed as intrinsically dependant on aspects of the perceiver’s body other than the brain, resulting from its existence in an environment requiring both sensing and acting [30].

Today, in the current technological landscape of wearables, virtual/augmented reality and the Internet of Things, the body has become of central concern for design as we seek to expand its boundaries and capabilities. As such, embodied design aims to leverage the user’s body and its experience as a whole to inform design ideation [32]. The assets afforded to designers by the ecological and embodied approaches discussed here are of great importance in developing comprehensive user-centric design frameworks, particularly in perceptual research like AD, whereby the interrelationship between users, environment, body and product can be carefully scrutinised in order to make appropriate design decisions [33].

3. IMPLICATIONS FOR AUDITORY DISPLAY

From the descriptions of usability and aesthetics in AD, Design thinking and user-centric design in HCI, as well as ecological and embodied approaches to design, it can be seen that the field of auditory display shares many common threads with these disciplines. A common theme that emerges from the discussion in this paper, has been the progressive rejection of positivist approaches in different fields as mentality shifts towards user-centric ideas. This common evolution within disciplines, concerned with the production of new technologies and products, suggests a certain degree of desirability in integrating design-based approaches, despite obvious tensions that arise from traditionally opposing philosophies.

Interestingly the adoption of user-centric approaches to design seen in HCI has been much slower for AD research. Only until recent editions of the International Conference for Auditory Display have we begun to see explicit examples of user-centric approaches such as participatory design [34], [35]. In these examples we begin to observe the recognition of the potential of user-centric approaches in AD, as well as the use of the associated vocabulary and techniques already established within fields such as HCI and Design. This suggests an ongoing convergence amongst these fields of research, that may ultimately help homogenize the language and approaches employed in AD research, with the aim of producing a more coherent

discussion and exploration of the issues this field is faced with.

An initial example of this clear shift in focus towards user-centric approaches in AD can be observed in the work of Stephen Barras [8]. In this work Barras uses the example of how, what he terms the ‘aesthetic turn in visualization’, acted as a catalyst for innovative, participatory and user-centric approaches in the field. However, more importantly Barras describes how this aesthetic turn empowered individuals outside scientific and engineering disciplines to engage socially in producing novel and more personally relevant data visualizations. He proposes that for AD to experience the effects of such an ‘aesthetic turn’ in the field, design-based approaches are the most suitable for “*dissolving division between scientific and artistic methods*” [8] and producing AD designs that are usable and of mass-appeal.

Further evidence of the use of user-centric approaches that push the AD field in the direction envisaged by Barras, is exemplified in the work presented by Wolf et al. [15]. Wolf et al. present a sonification approach aimed at decentralizing AD design by providing end-users with a sonification system that can be tweaked and refined after the initial implementation by the researchers. This is achieved by providing users with an interactive semi-autonomous sonification system, based on previous well documented sonification techniques, that produces an initial base ‘soundscape’ for which users can then modify the sonification mapping parameters to suit their preferences. In this paper the authors make explicit their adaptation of techniques from HCI to develop a UCD approach to sonification. However, despite the potential of UCD highlighted in this paper, the authors acknowledge the need for further in-depth investigation of methodologies akin to the one they present.

In the two examples discussed above, we see clear evidence of UCD approaches developing in AD. Further examples can be found that trace the convergence of HCI and design methodologies with AD in the work of Goudarzi [7]. In this paper the author makes use of the participatory design workshop method for developing UCD oriented sonifications. This paper is of interest to this study, as it makes direct use of an HCI/design-based tool, for achieving UCD in sonification design. In the workshop, the authors aimed at creating a dynamic environment between AD designers and end-users based around dialogue and interaction. This dialogue provided a means of elucidating the common threads and frictions between designers and users. This common ground served as a basis for the cultivation of a symbiotic design relationship at the hands of the end-user. The authors describe the challenge of placing users and designers in direct contact with each other to produce participatory design. However, their results suggest that the workshop model provided an adequate mediation for producing a greater appreciation for each other’s perspectives and engendering the desired user-designer relationship [7].

Finally, while we have observed a clear causality between the emergence of aesthetic considerations in AD and the development of UCD approaches, we can further observe the convergence of AD with the more current trends in HCI research in the work of Landry and Jeon [35]. In this paper the authors employ a participatory design methodology that is heavily focused on the situatedness and embodiment of the user. By placing direct emphasis, not

only on the user’s needs, but on the actual physical experience of the user’s body within its environment as a source of design knowledge, this work presents a unique example of embodied and ecological design in AD. The authors highlight the complexity and unfamiliarity of employing this particular approach for the design of an auditory interface but encourage further exploration thanks to the rich source of data derived directly from end-users. As such, despite the difficulties faced by the researchers in implementing this approach, this paper serves to further highlight the recent convergence of AD and HCI theory, vocabulary and technique in the context of our discussion of ecological and embodied design of AD.

While relatively new to the discussion in AD research, notions of ecological and embodied approaches are not entirely unfamiliar to the field of auditory research. One clear historical example is the notion of listening modes envisaged by Pierre Schaeffer and refined by Michel Chion [36], [37]. In his *Traité Des Objets Musicaux: Essai Interdisciplines*, Schaeffer formalizes his ideas on the listening state of mind by introducing a taxonomy of 4 distinct modes of listening: *écouter, entendre, comprendre and ouïr*. This categorization laid the groundwork for modern musicological research into the listening and perception of organized sound. Since the *Traité Des Objets Musicaux*, the taxonomy of listening modes has been variously expanded and reduced in different attempts at reconciling some of the more obscure aspects of Schaeffer’s original depiction and adapting it to more modern perspectives of music perception [37], [38].

As the overall philosophy and approach to AD design has shifted towards trying to better understand users’ listening experiences, the discussion around the listening mode taxonomy initially proposed by Pierre Schaeffer has become of great relevance to the field [10], particularly from an ecological point of view, as this conception of listening aimed at creating an understanding of how listening is affected by different contexts. Other examples of ecological considerations in AD can be found in attempts to apply Gestalt Theory to understand how design auditory displays that discretely fit in within their environments [39].

From these instances and throughout the discussion in this paper, it can be observed that the AD community has been shifting towards more user-centric approaches in recent times. However, while much attention has been given to understanding the listener in the domain of music and organized sound, the same cannot be said of AD research [40]. This is of particular importance as greater emphasis on user-experience has become paramount to current AD design. As such, we believe it is important to encourage AD practitioners to explore UCD as a tool for producing vibrant and innovative research. Furthermore, we encourage the implementation of ecological and embodied frameworks as tools for engaging users fully and directly in their environments, with the aim of creating more meaningful dialogues between designers and end-users. We hope that through this discussion we have emphasised the need to address key usability concerns in AD and presented a theoretical background to do so.

4. CONCLUSION

In this paper we framed the current state of affairs in auditory display research by contextualizing and elucidating the user experience and usability concerns within the field. We provided a background into the emergence of key design-based approaches, namely design thinking and user-centric design, and their critical role into human-computer interaction studies. We also introduced the ecological and embodied approaches to design, as innovative elements for informing the design of AD applications in order to support better usability and user experience. Through the discussion in this paper we hope to encourage AD designers to experiment with the approaches discussed herein.

5. ACKNOWLEDGMENT

I would like to thank my partner, my parents and my supervisors for supporting my endeavours to talk and write about sound and music ad nauseam! But most importantly, I thank HighWire, Lancaster University and the EPSRC without whom this research would not have been possible.

6. REFERENCES

- [1] M. Ballora, "Sonification, Science and Popular Music: In search of the 'wow,'" *Organised Sound*, vol. 19, no. 1, pp. 30–40, Apr. 2014.
- [2] T. Hermann, A. Hunt, and J. G. Neuhoff, *Chapter 1: Introduction | The Sonification Handbook*. Isd, 2011.
- [3] B. F. G. Katz and G. Marentakis, "Advances in auditory display research," *J. Multimodal User Interfaces*, vol. 10, no. 3, pp. 191–193, Sep. 2016.
- [4] S. Barras and P. Vickers, *Chapter 7: Sonification Design and Aesthetics | The Sonification Handbook*. Isd, 2011.
- [5] E. Clarke, *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford University Press, 2005.
- [6] R. Berry and N. Osaka, "Art GAllery," in *ICAD 2002*, Kyoto, Japan, 2002.
- [7] V. Goudarzi, "Exploration of Sonification Design Process Through an Interdisciplinary Workshop," in *Proceedings of the Audio Mostly 2016*, New York, NY, USA, 2016, pp. 147–153.
- [8] S. Barras, "The aesthetic turn in sonification towards a social and cultural medium," *AI Soc.*, vol. 27, no. 2, pp. 177–181, May 2012.
- [9] S. Barras, "An Annotated Portfolio of Research through Design in Acoustic Sonification," *Leonardo*, vol. 49, no. 1, pp. 72–73, Jul. 2015.
- [10] F. Grond and T. Hermann, "Interactive Sonification for Data Exploration: How listening modes and display purposes define design guidelines," *Organised Sound*, vol. 19, no. 1, pp. 41–51, Apr. 2014.
- [11] T. Carlile, A. Hunt, and J. G. Neuhoff, *Chapter 3: Psychoacoustics | The Sonification Handbook*. Isd, 2011.
- [12] M. Akiyama, "Dataffect: Numerical Epistemology and the Art of Data Sonification," *Leonardo Music J.*, vol. 24, no. 1, pp. 29–32, Nov. 2014.
- [13] M. Filimowicz, "Peircing Fritz and Snow: An aesthetic field for sonified data," *Organised Sound*, vol. 19, no. 1, pp. 90–99, Apr. 2014.
- [14] F. Grond and T. Hermann, "Aesthetic strategies in sonification," *AI Soc.*, vol. 27, no. 2, pp. 213–222, May 2012.
- [15] K. Wolf, G. Gliner, and R. Fiebrink, "End-user Development of Sonifications using Soundscapes," presented at the International Conference on Auditory Display (ICAD) 2015, Graz, Austria, 2015, pp. 281–288.
- [16] G. Dubus, "Evaluation of four models for the sonification of elite rowing," *J. Multimodal User Interfaces*, vol. 5, no. 3–4, pp. 143–156, May 2012.
- [17] N. Cross, K. Dorst, and N. Roozenburg, *Research in design Thinking*. Delft University Press, 1992.
- [18] Tim Brown, "Design Thinking," *Harv. Bus. Rev.*, vol. 86, no. 6, p. 84, Jun. 2008.
- [19] Rabah Bousbaci, "Models of Man in Design Thinking: The Bounded Rationality," *Des. Issues*, vol. 24, no. 4, pp. 38–52, 2008.
- [20] H. Lindinger, Ed., *Ulm Design: The Morality of Objects*, 1st MIT Press Ed edition. Cambridge, Mass: MIT Press, 1990.
- [21] A. Newell and H. A. Simon, *Human Problem Solving*, Reprint edition. Englewood Cliffs, N.J: Longman Higher Education, 1971.
- [22] D. A. Schön, *The Reflective Practitioner: How Professionals Think in Action*, New Ed edition. Routledge, 2017.
- [23] C. Sas and A. Dix, "Designing for Reflection on Experience," in *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, 2009, pp. 4741–4744.
- [24] C. Sas and C. Zhang, "Do Emotions Matter in Creative Design?," in *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, New York, NY, USA, 2010, pp. 372–375.
- [25] C. Sas, S. Whittaker, S. Dow, J. Forlizzi, and J. Zimmerman, "Generating implications for design through design research," 2014, pp. 1971–1980.
- [26] W. Gaver, "What Should We Expect from Research Through Design?," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2012, pp. 937–946.
- [27] P. Dourish, "Implications for Design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2006, pp. 541–550.
- [28] I. Koskinen, J. Zimmerman, T. Binder, J. Redstrom, and S. Wensveen, *Design Research Through Practice: From the Lab, Field, and Showroom*. Elsevier Science, 2011.

- [29] J. J. Gibson, *The senses considered as perceptual systems*. Boston: Houghton Mifflin, 1966.
- [30] R. A. Wilson and L. Foglia, "Embodied Cognition," in *The Stanford Encyclopedia of Philosophy*, Spring 2017., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2017.
- [31] N. Hirose, "An ecological approach to embodiment and cognition," *Cogn. Syst. Res.*, vol. 3, no. 3, pp. 289–299, Sep. 2002.
- [32] D. Wilde, A. Vallg'a arda, and O. Tomico, "Embodied Design Ideation Methods: Analysing the Power of Estrangement," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2017, pp. 5158–5170.
- [33] C. Sas and R. Chopra, "MeditAid: A Wearable Adaptive Neurofeedback-based System for Training Mindfulness State," *Pers. Ubiquitous Comput.*, vol. 19, no. 7, pp. 1169–1182, Oct. 2015.
- [34] O. Metatla, N. Bryan-Kinns, T. Stockman, and F. Martin, *Sonifications for digital audio workstations: Reflections on a participatory design approach*. Georgia Institute of Technology, 2015.
- [35] S. Landry and M. Jeon, "Participatory Design Research Methodologies: A Case Study in Dancer Sonification," Jun. 2017.
- [36] "Traité des objets musicaux, Pierre Schaeffer, Sciences humaines - Seuil." [Online]. Available: <http://www.seuil.com/ouvrage/traite-des-objets-musicaux-pierre-schaeffer/9782020026086>. [Accessed: 19-Sep-2017].
- [37] M. Chion and W. Murch, *Audio-vision: Sound on Screen*. Columbia University Press, 1994.
- [38] K. Tuuri and T. Eerola, "Formulating a Revised Taxonomy for Modes of Listening," *J. New Music Res.*, vol. 41, no. 2, pp. 137–152, Jun. 2012.
- [39] N. Rönnerberg, J. Lundberg, and J. Löwgren, "SONIFYING THE PERIPHERY: SUPPORTING THE FORMATION OF GESTALT IN AIR TRAFFIC CONTROL," 2016.
- [40] P. Vickers, "Ways of listening and modes of being: electroacoustic auditory display," *J. Sonic Stud.*, vol. 2, no. 1, pp. 1–23, May 2012.

EXAMINING THE LEARNABILITY OF AUDITORY DISPLAYS: MUSIC, EARCONS, SPEARCONS, AND LYRCONS

Kay Tislar, Zackery Duford, Brittany Nelson, Madeline Peabody, and Myounghoon Jeon

Department of Cognitive and Learning Sciences
Michigan Technological University
1400 Townsend Dr., Houghton, MI 49931 USA
{ktislar, ztduford, bnelson1, mjpeabod, mjeon}@mtu.edu

ABSTRACT

Auditory displays are a useful platform to convey information to users for a variety of reasons. The present study sought to examine the use of different types of sounds that can be used in auditory displays—music, earcons, spearcons, and lyricons—to determine which sounds have the highest learnability when presented in sequences. Participants were self-trained on sound meanings and then asked to recall meanings after listening to sequences of varying lengths. The relatedness of sounds and their attributed meanings, or the intuitiveness of the sounds, was also examined. The results show that participants were able to learn and recall lyricons and spearcons the best, and related meaning is an important contributing variable to learnability and memorability of all sound types. This should open the door for future research and experimentation of lyricons and spearcons presented in auditory streams.

1. INTRODUCTION

Auditory notifications and representations of tasks, objects, and warnings are becoming ever more prevalent, particularly with screen real estate being at a premium as devices become smaller. There are a host of benefits to presenting information in auditory form, including the fact that it can provide additional information to a user by not overloading the visual channel and by instead capitalizing on the often under-used auditory channel. In addition, when a user is occupied with a task, sound can be a very effective way of capturing a user's attention [1]. For example, one study indicated that people who perform a task while monitoring data perform the task more accurately when the data are presented in auditory rather than in visual form [2].

In order to be effective, auditory displays must be highly learnable. If users cannot learn the sound meanings easily, they may choose not to use the display even if plenty of learning time is available [3].

An important concept in using auditory displays is that of streaming. As defined by Bregman and Campbell, an auditory stream is “a sequence of auditory events whose elements are related perceptually to one another, the stream being segregated perceptually from other co-occurring auditory events” [4]. The auditory events can be variables (sounds) that are mapped to meanings, with auditory dimensions also changing depending on the meaning. The

related auditory dimensions include qualities such as pitch, duration, intensity, and timbre. For example, researchers have experimented with using streams to sonify weather-related data. Schuett, Winton, Batterman, and Walker [5] sonified weather variables including temperature, humidity, wind speed, wind direction, and cloud cover. In some of the experimental conditions, these made up five streams, but in other conditions they only made up three. In the latter case, two of the variables were combined in a stream with an auditory dimension changed to indicate the second variable and its status. For example, temperature and humidity were combined. Temperature was represented by a sound that varied in pitch. Humidity was designated by applying a tremolo effect to the temperature stream, which indicated low or high humidity. Therefore, just one stream was used to represent two variables, each changing in a different auditory dimension.

Prior research has been conducted to compare various sound types. When comparing earcons, auditory icons, and spearcons, users preferred earcons but remembered auditory icons better. When comparing all of those three sound types plus speech, spearcons and speech were both verified to be extremely learnable, and earcons were the most difficult to learn [3]. In comparisons among auditory icons, earcons, and speech, speech was the easiest to learn, auditory icons took the longest to learn, and earcons were the most error prone [6]. According to Ballas, sound identifiability is related to factors such as how easy a listener can form a mental picture of the sound, how familiar the sound is, and sound clarity [7].

Research has also been done to determine how many auditory streams users can be exposed to while still responding correctly to the information they receive. When one dimension changed, users were more accurate when monitoring just one stream. When two dimensions changed, users were just as accurate monitoring one or two streams [8].

The current experiment was intended as a first step in extending this research by examining how many auditory streams containing different sound types users can recall. In addition to music, three other types of sounds were evaluated:

*Earcons (ear + icons) [9] use short musical motives as symbolic representations of objects or events, in which the sound has no relation to the object it represents. Earcons are very flexible in terms of design, but can be difficult to learn because they use arbitrary mappings.

*Spearcons (speech + earcons) [13] consist of compressed speech that may not be comprehensible. They can be nearly as easy to comprehend as speech with only brief training required.

*Lyricons (lyrics + earcons) [10] are a relatively new type of display, in which lyrics and associated earcons are played concurrently. The point is to train users to recall the

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

lyricon so well that the lyrics can be removed, but people will mentally fill in the lyrics themselves.

In the present study, the primary goal was to examine the learnability of different types of sounds, namely which specific sound mappings were remembered best and how many streams from a sequence could be recalled. This was done in order to generate a set of highly learnable sounds that could later be manipulated to form streams that incorporate different sound types. We focused on testing non-speech auditory displays, but we excluded auditory icons because many of the real-world items do not have natural sounds.

This study tested the following two hypotheses:

H1—People will remember more sound meanings of all types (music, earcon, spearcon, lyricon) when the meanings are related to the sounds (in other words, they are intuitive).

H2—People will remember more sound meanings when sequences of sounds contain a variety of sound types (music, earcon, spearcon, lyricon), as opposed to sounds of all the same type. This is based on our assumption that varying the sound type would aid in mentally separating the sounds during listening and recalling.

2. METHODS

2.1. Participants

Twenty undergraduate and graduate students volunteered to participate in the study, with some undergraduate students receiving credits in a psychology course and the other students receiving no compensation. Ten female and ten male students participated, all older than the age of 18. Fifteen of the participants were native English speakers.

2.2. Apparatus and Equipment

Windows® desktop workstations were used to deliver the experiment, and participants listened to the sounds through over-the-ear headphones.

As a way of allowing users to train themselves on the sound mappings, a simple HTML file containing a table was developed. Using the file, participants could click a sound button in a row to hear the sound corresponding to the word displayed in that row (See Figure 1.). Thirty-two sounds were mapped to different words.

The experiment itself was written as a script in the Psychology Experiment Building Language (PEBL) [11], [12]. Participants clicked a **Play Sound** button to play a sound sequence; all 32 sounds were used to build 21 audio sequences containing varying numbers and types of sounds (See Table 1.).

Click Arrow Button to Play Sound	What the Sound Means
▶ 0:00 / 0:04	Date
▶ 0:00 / 0:06	Power Off
▶ 0:00 / 0:02	URL
▶ 0:00 / 0:03	Word
▶ 0:00 / 0:00	Fast Forward
▶ 0:00 / 0:01	Delete
▶ 0:00 / 0:00	PDF
▶ 0:00 / 0:01	Play

Figure 1. Sample HTML training file.

After the sequence finished playing, a list of all of the individual words to which the sounds had been mapped was displayed on the screen as checkboxes. Participants clicked the box corresponding to each word whose sound they remembered hearing in the sequence, clicked the **Done** button, and repeated the process for all 21 auditory sequences (See Figure 2.).

For each participant, PEBL created a data file that contained information on the value of each checkbox (checked or not) for every sound sequence. Later, a researcher converted each data string into correct/incorrect response values, the dependent variables. The file also contained response time data for each sound sequence.

2.3. Stimuli Design

For the stimuli, we selected 32 words representing either tasks or objects and mapped them to sounds, as agreed on by the research team. Eight music sounds, earcons, spearcons, and lyricons were created.

To find music, we searched royalty-free websites to locate songs that could be used to represent abstract ideas, and then spliced phrases as necessary. Earcons were created using the FL Studio Digital Audio Workstation (DAW) from Image Line. To create spearcons, we pre-recorded sound files of spoken words using the online text-to-speech (TTS) engine, and then applied the SOLA algorithm to create the spearcon files [13]. For the lyricons, we searched lyric websites to find keywords that would represent tasks, then isolated the vocal track, and spliced audio around the sung phrase.

Order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Length	4	4	4	4	4	4	4	4	8	8	8	8	4	4	8	8	12	12	4	8	12
Type of Sound	E	L	M	S	E	L	M	S	E	L	M	S	C	C	C	C	C	C	C	C	C
Related/Unrelated	U	U	U	U	R	R	R	R	C	C	C	C	R	U	R	U	R	U	C	C	C

Table 1. Number and types of sounds included in each sequence. Row 1 contains the sequence number, and row 2 indicates the number of separate sounds the sequence contained. Rows 3 and 4 show the sound type and whether or not the sounds and their mappings were intuitive. E: Earcon, L: Lyricon, M: Music, S: Spearcon, U: Unrelated (Non-intuitive), R: Related (Intuitive), C: Combination.

Figure 2. Sample experiment screen in PEBL.

For this study, we did not vary the auditory dimensions and did not join sounds in a stream. A sequence of four sounds therefore represented four streams. The streams were played consecutively, rather than concurrently, to make it easier for users to distinguish which sound they were hearing.

Twenty-one auditory sequences were created using the Audacity® open-source audio software package; we attempted to use each individual sound and sound type approximately the same number of times overall and an equal distribution of related (intuitive) and unrelated (non-intuitive) sounds. A half second time gap was inserted between the sounds in each sequence. All sequences were saved in WAV format.

Table 1 shows the characteristics of each of the 21 sequences, which were delivered in the same order to all of the participants.

2.4. Procedure

Participants were tested one at a time by a single researcher. After a participant signed a consent form and provided demographic information (age, gender, education level, and native language), the researcher asked her or him to be seated at a computer. The researcher then explained the experimental procedure.

Next, the participant was provided with headphones and asked to familiarize herself or himself with the sound mappings using the training file. Users were allowed a maximum of five minutes of training time and could repeat all of the sounds as many times as needed.

Finally, the researcher started the experiment and explained the user interface. The participant then proceeded through the experiment, playing the sound sequences and selecting the sound mappings, until he or she had completed all 21 sequences.

3. RESULTS

As previously noted, some of the participants were not native English speakers. To ensure that native language was not a confounding variable, we tested for a relationship between native language and accuracy. A Spearman's rank-order correlation was run to determine the relationship between native language and overall score accuracy across conditions; it suggested that native language was not a statistically significant predictor of accuracy ($r_s = -.03, p = .89$).

Because spearcons are fast-paced speech in English, the relationship between native language and accuracy in the spearcon condition was also tested separately. Results once again suggested there was not a statistically significant relationship between native language and accuracy ($r_s = .24, p = .30$). Therefore, participants who did not speak English as their first language were retained in the following analyses.

A Pearson product-moment correlation coefficient was computed to assess the relationship between training time and accuracy. Results suggested there was not a statistically significant relation between training time and overall accuracy ($r = -.09, p = .7$). However, a visual inspection of the scatter plot suggested there may be a curvilinear relationship to the data. Therefore, a curvilinear regression analysis was performed using hierarchical multiple regression with one bend in the regression line. When a curvilinear relation is accounted for, the model neared statistical significance, $R^2 = .16, F(1, 19) = 3.12, p = .09$. Future testing with a greater number of participants could produce statistically significant results.

Figure 3. Means and standard errors of accuracy by relatedness and sequence type.

A two-way repeated measures Analysis of Variance (ANOVA) was conducted on the influence of two independent variables (relatedness and single versus mixed sound types in the sequences) on accuracy. Both the main and the interaction effects were statistically significant. Results suggested there was a main effect of relatedness on accuracy, $F(1,19) = 67.91$, $p = .001$. Results also suggested there was a main effect between single versus mixed sound type on accuracy, $F(1,19) = 16.25$, $p = .001$. Finally, there was a statistically significant interaction between both independent variables on accuracy $F(1,19) = 4.93$, $p = .039$ (See Figure 3.). Because the interaction was statistically significant, further analyses were conducted to test if the main effects were still statistically significant across conditions. This was done using paired samples t-tests.

Paired samples t-tests showed that there was a statistically significant difference in accuracy between related and unrelated sounds' scores when sequences consisted of single type sounds ($M = .15$, $SD = .18$), $t(19)=3.862$, $p = .001$, and when sequences contained a combination of different types of sounds ($M = .275$, $SD = .16$), $t(19)=7.678$, $p = .001$. Paired samples t-tests also indicated a statistically significant difference in mean accuracy between single and combined sound sequence types when the sounds were related ($M = .12$, $SD = .23$), $t(19)=2.4$, $p = .027$, and when sounds were unrelated ($M = .243$, $SD = .247$), $t(19)=4.420$, $p = .001$.

To test if there was a statistically significant difference in the accuracy by sound type, a number of paired samples t-tests were performed. A statistically significant difference between scores was found when only spearcons were used ($M = .48$, $SD = .20$) compared to when only earcons were used ($M = .26$, $SD = .15$), $t(19)=6.07$, $p = .001$. A statistically significant difference was also found between the spearcon ($M = .48$, $SD = .20$) and music ($M = .28$, $SD = .13$) conditions, $t(19)=5.44$, $p = .001$. In addition, statistically significant differences were found between the lyricon ($M = .54$, $SD = .14$) and earcon ($M = .26$, $SD = .15$) conditions; $t(19)=7.39$, $p = .001$. The mean accuracy between the lyricon ($M = .54$, $SD = .14$) and music ($M = .28$, $SD = .13$) conditions was also statistically significant, $t(19)=8.77$, $p = .001$.

Figure 4. Means and standard errors of accuracy by sound type.

Between mixed and single sound sequences, there was a statistically significant difference between mixed sequences ($M = .54$, $SD = .16$) and single-sound earcon conditions, $t(19)=5.69$, $p < .001$. Finally, there was a statistically significant difference between mixed-sequence accuracy ($M = .54$, $SD = .16$) and single-sound sequence music ($M = .28$, $SD = .13$) conditions, $t(19)=6.24$, $p < .001$. No other t-tests were statistically significant. Overall, accuracy in the mixed sequences was equivalent to accuracy in the spearcon-only and lyricon-only conditions, but was better than in the earcon-only and music-only conditions. (See Figure 4.).

To test if there was a statistically significant difference in the accuracy by number of sounds per sequence, additional paired samples t-tests were performed. There was a statistically significant difference between scores when sequences contained four sounds ($M = .45$, $SD = .13$) compared to sequences that contained eight sounds ($M = .53$, $SD = .16$); $t(19)= -2.02$, $p = .05$. A statistically significant difference was also found between sequences when they contained four sounds versus 12 sounds ($M = .30$, $SD = .10$); $t(19)=4.63$, $p = .001$. Finally, statistically significant differences were found between conditions using sequences of length eight and length 12; $t(19)=5.25$, $p = .001$. All other relationships were not statistically significant (See Figure 5.).

The differences in accuracy according to number of individual sounds could be at least partially due to the fact that the shorter sequences (four and eight sounds) were presented early in the experiment, with most of the four-sound sequences coming at the beginning; on the other hand, the 12-sound sequences came near the end. Participants may have become familiar with the sounds in the four-sound sequences and remembered them when encountering the 8-sound sequences. However, by the time they reached the 12-sound sequences, they may have exceeded their working memory capacity, leading to a decrease in performance.

To test if there was a statistically significant difference in accuracy based on whether the mapped words were related to the sounds (the relationship was intuitive), further paired samples t-tests were performed. There was a statistically significant difference between related ($M = .42$, $SD = .10$) and unrelated sound ($M = .32$, $SD = .13$) conditions; $t(19)=3.4$, $p = .001$. (See Figure 6.)

Figure 5. Means and standard errors of accuracy by number of sounds.

Figure 6. Means and standard errors of accuracy by sound relatedness.

4. DISCUSSION

While limited by design, the current study yielded several significant results that prompt further experimentation in the area of auditory streams and memory. The results concluded that eight sounds was the optimal length when compared to sequences of four or 12 sounds, regardless of sound type and relatedness. Furthermore, related sounds were recalled significantly more often than unrelated sounds or mixed related/unrelated sound sequences. Contrary to our second hypothesis, participants did not always perform better when sequences contained a variety of sound types; in certain cases, they performed as well with a single sound type.

The present results showed that lyricons and spearcons are recalled more frequently than music and earcons when people are given a short amount of time to remember many sounds. This is likely due to the fact that both are based on speech and, therefore, are more intuitive and processed more easily than are the arbitrary mappings of earcons and music. One limitation of the current study is that the design did not enable us to accurately measure training time for individual sounds so that we could compare training time for the various sound types.

Lyricons and spearcons also have the highest learnability; these findings suggest that future experimentation could separate them from music and earcons, and examine sequences and streams containing mixes of lyricons and spearcons. In addition, relatedness of sounds is important for user recall. Sounds should be related to the meaning they are conveying to the user; inter-rater reliability proved effective in the current study to judge relatedness of sounds.

Another limitation of this study is that the audio sequences used are not necessarily representative of what a person would encounter in an actual task. Future research is needed to create streams that convey an overall meaning to the participant, instead of sequences that contain related and unrelated individual word meanings. The present study identifies that certain types of sounds perform better in sequences and relatedness is a key component to memorability and learnability. Further experiments should create streams that can be used in different task environments, such as surgical operating rooms. Research is needed on the effectiveness of multiple streams being played at once, and the effectiveness of streams containing sounds that vary in dimensions such as pitch, rhythm, and register.

5. CONCLUSIONS

The present study identified that lyricons and spearcons have the highest recall when presented in varied sound sequences. Relatedness of meaning and sound is also a key component in memorability and learnability. Further research is needed to examine streams containing only lyricons and spearcons, and to integrate these sound sequences into streams.

6. REFERENCES

- [1] S. A. Brewster, P. C. Wright, and A. D. Edwards, "An evaluation of earcons for use in auditory human-computer interfaces," in *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, pp. 222-227, April 1993.
- [2] M. Watson and P. Sanderson, "Sonification supports eyes-free respiratory monitoring and task time-sharing," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 46, no. 3, pp. 497-517, 2004.
- [3] T. Dingler, J. Lindsay, and B. Walker, "Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech," in *Proceedings of the 14th International Conference on Auditory Display*, 2008.
- [4] A. Bregman and J. Campbell, "Primary auditory stream segregation and perception of order in rapid sequences of tones," *Journal of Experimental Psychology*, vol. 89, no. 2, pp. 244-249, 1971.
- [5] J. H. Schuett, R. J. Winton, J. M. Batterman, and B. N. Walker, "Auditory weather reports: demonstrating listener comprehension of five concurrent variables," in *Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound*, ACM, p. 17, 2014.
- [6] P. A. Lucas, "An Evaluation of the Communicative Ability of Auditory Icons and Earcons," in *Proceedings of the 2nd International Conference on Auditory Display*, Addison-Wesley, 1994.
- [7] J. A. Ballas, "Common factors in the identification of an assortment of brief everyday sounds," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 19, no. 2, pp. 250-267, 1993.
- [8] J. H. Schuett and B. N. Walker, "Measuring comprehension in sonification tasks that have multiple

- data streams,” in *Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound*, ACM, p. 11, 2013.
- [9] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, “Earcons and icons: Their structure and common design principles,” *Human-Computer Interaction*, vol. 4, pp.11–44, 1989.
- [10] M. Jeon, “Lyrics (Lyrics+ Earcons): Designing a new auditory cue combining speech and sounds,” in *Proceedings of the International Conference on Human-Computer Interaction*, Springer Berlin Heidelberg, pp. 342-346, 2013.
- [11] S. T. Mueller and B. J. Piper, “The Psychology Experiment Building Language (PEBL) and PEBL Test Battery,” *Journal of Neuroscience Methods*, vol. 222, pp. 250–259, 2014.
- [12] S. T. Mueller, The Psychology Experiment Building Language, Version 0.14. Retrieved from <http://pebl.sourceforge.net>, October 2016.
- [13] B. N. Walker, J. Lindsay, A. Nance, Y. Nakano, D. K. Palladino, T. Dingler, and M. Jeon, “Spearcons (Speech-based earcons) improve navigation performance in advanced auditory menus,” *Human Factors*, vol. 55, no. 1, pp. 157-182, 2013.

REHABILITATION

“MUSICAL EXERCISE” FOR PEOPLE WITH VISUAL IMPAIRMENTS: A PRELIMINARY STUDY WITH THE BLINDFOLDED

Ridwan Ahmed Khan, Myounghoon Jeon

Tejin Yoon

Mind Music Machine Lab,
Michigan Technological University,
1400 Townsend Dr., Houghton, MI, USA
{ridwank, mjeon}@mtu.edu

Neuromechanics Lab,
Michigan Technological University
1400 Townsend Dr., Houghton, MI, USA
tyoon@mtu.edu

ABSTRACT

Performing independent physical exercise is critical to maintain one's good health, but it is specifically hard for people with visual impairments. To address this problem, we have developed a Musical Exercise platform for people with visual impairments so that they can perform exercise in a good form consistently. We designed six different conditions, including blindfolded or visual without audio conditions, and blindfolded or visual with two different types of audio feedback (continuous vs. discrete) conditions. Eighteen sighted participants participated in the experiment, by doing two exercises - squat and wall sit with all six conditions. The results show that Musical Exercise is a usable exercise assistance system without any adverse effect on exercise completion time or perceived workload. Also, the results show that with a specific sound design (i.e., discrete), participants in the blindfolded condition can do exercise as consistently as participants in the non-blindfolded condition. This implies that not all sounds equally work and thus, care is required to refine auditory displays. Potentials and limitations of Musical Exercise and future works are discussed with the results.

1. INTRODUCTION

Physical activity or exercise is necessary for maintaining good health physically [8] and mentally [24]. However, accessibility to exercise is a critical issue for population with disabilities [19, 30]. For example, people with visual impairments suffer from health issues due to lack of exercise [25] [2] [12]. They need active guidance from another person to help them learn and practice the exercise [25]. Even if they have learned the exercise with the help of another person, they may have to do it by themselves sometimes. In such a case, guidance or at least a feedback system telling them if they are doing the exercise correctly can help improve their exercise. Technological interventions can assist people with visual impairments doing the exercise [23]. However, designing an intuitive feedback system to ensure that the user does the exercise correctly and consistently is a difficult goal to achieve. The feedback system should be intuitive and real-time so that people with visual impairments can easily understand how they are doing while exercising. Also, the feedback system needs to be non-visual but clear enough at the

same time. The feedback system also needs to be enjoyable so that people with visual impairments get motivated sufficiently. To address these issues, we have designed and evaluated the Musical Exercise system. It is a non-wearable exercise detection system using Microsoft Kinect which uses sound as real-time feedback. The Microsoft Kinect device is placed in front of the user and it tracks the movement of the user's joints. Musical Exercise then analyzes the movement data and decides whether it is a valid exercise. At the same time, Musical Exercise provides sound feedback which conveys the information of how the user is doing in the exercise. For the first implementation, we have selected two exercises - squat and wall sit for the experiment and designed two different audio profiles - continuous sound and discrete sound to test the Musical Exercise system.

In the following sections, after providing related work, we described the design of the Musical Exercise system. Then, we detailed an experiment of the system with 18 sighted participants. The results of the experiment show that Musical Exercise can have potential benefits to people with visual impairments with a proper sound design. The results also indicate that sighted people can also benefit from the Musical Exercise system with intuitive non-obtrusive sound feedback.

2. HEALTH & EXERCISE TECHNOLOGY

There are a huge number of fitness applications [18] that have immense benefits for behavior modification [11], older adults [9, 27], and chronic diseases [29]. Microsoft Kinect is one of the most frequently used devices. It tracks users' hand movements [4], sit posture [28], or running [5], and provides feedback or corrects the exercise form [6]. In a similar line, Kinect or Nintendo Wii has been used in a game for exercise of people with disabilities. Research has been conducted for motor rehabilitation exercise of patients having Parkinson's disease [1, 22], for full body motion of wheel chairs users [10], rehabilitation of patients' movement disorders [17], and upper extremities rehabilitation [13].

In this line of research, multimodal and auditory displays have been an important component. Newbold et al. [21] designed a mechanism by which users can be musically informed of their position in the stretching exercise. This musical information intuitively prevents the users from over stretching. Another study [14] showed that multimodal feedback can be helpful to older adults for performance gain while doing an activity. A system with auditory feedback was also developed by Singh et al. [26] to support exercise for people with chronic pain. The researchers found that when the audio was tailored to one's personality, exercise became psychologically motivating.

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

3. TECHNOLOGICAL ASSISTANCE FOR EXERCISE OF PEOPLE WITH VISUAL IMPAIRMENTS

Assistive technologies can also help people with visual impairments do exercise or physical activity easily. One such a project is Gobi [20], which helps swimmers with visual impairments to stay on the right way while swimming. It is a wearable device on the thigh which tracks the swimming lane with a camera and guides the swimmers to swim left or right if they are swimming away. Winoto and Tang [31] developed a helmet attaching Arduino and speakers for people with visual impairments to play games or move around. Once the users wear the helmet which has five directional speakers, the system generates the guidance using sound feedback via one of the speakers and haptic feedback as vibration via a mobile phone. Al Zayer et al. [32] explored the idea of using an aerial quadrotor to guide a blind runner in running. Morelli et al. [19] designed an exercise game for people with visual impairments which is similar to a popular tennis game for sighted people. Tactile display was used as their primary feedback, while auditory display was used as the secondary feedback. Rector et al. [23] used a Kinect device with auditory feedback to help people with visual impairments doing yoga exercise. Their system tracks users' body and limb positions and provides speech instructions for a specific yoga pose.

As shown above, researchers have accommodated different feedback systems for people with visual impairments. Tactile or haptic feedback systems are common, but using this type of feedback system comes with some caveats, e.g., the tactile or haptic device needs to be placed on users' body which seems obtrusive. Sometimes, it may distract users because they have physical responses from these devices. On the other hand, some researchers used auditory feedback to convey information about the activity. It is a non-obtrusive approach and users can hear the instructions and follow. However, speech instructions can be limited in number as these instructions are pre-recorded and not in real-time and may not meet the requirement of instructing every possible case the users can face themselves in. Also, speech provides abstract information; e.g., "move your right hand". The distance to move the right hand may vary for different users. Therefore, instructional speech may not cater to the specific need of a specific user. Musical Exercise aims to address these issues with non-speech sounds or musical tones. The benefit of using non-speech is that it can cater to the need of specific users. Users learn to hear the collection of musical notes of a proper form of the exercise. When they are deviating from the proper form, instantly the notes sound different. The sound of one user's proper form may differentiate from another user. In that way, Musical Exercise addresses the issue of giving individualized feedback elaborating non-speech sounds which are not confined by speech sounds.

4. MUSICAL EXERCISE FOR PEOPLE WITH VISUAL IMPAIRMENTS

4.1. Design Goals

To help people with visual impairments do the exercise in a more comfortable and effective way, we designed and developed Musical Exercise. Below is the list of main design goals:

a) Motivate to do exercise: Musical Exercise was designed to motivate people to keep people regularly do exercise. With

this system, people listen to the music of their body movement. Music can create positive emotions and affect in this situation [16]. Thus, by doing exercise which creates the music, people may become more interactive and engaged to the task. An engaging and interactive task creates motivation in people.

b) Learn exercise: Another design goal of Musical Exercise is to make it simple and easily accessible for people with visual impairments to learn different exercises. Once people are introduced to the system and the exercises, they can start learning about those exercises by themselves. They can simulate the sound of the proper form of each exercise.

c) Guide to do exercise: After learning the exercise using Musical Exercise, people can continue using the system as their guidance system while doing exercise. The sound intuitively guides them to maintain the proper form of the exercise and do the exercise in a correct and effective way.

d) Perform the exercise as good as a sighted person: One of the motivations behind Musical Exercise was to enable people with visual impairments to do the exercise as good as sighted people can do. Musical Exercise utilizes Microsoft Kinect to track motions of the user and then uses the tracking data to assist the user perform efficiently through sound feedback. Sound or music is intuitive feedback and instigates more physical responses from the user [7]. Also, sound is non-obtrusive feedback for people with visual impairments. Therefore, people with visual impairments can do exercise as well as sighted people using the Musical Exercise system.

e) Rely on assistive technology at home: People with visual impairments do not need to go to the gym or take much help from other people. They can use the Musical Exercise system and do exercise with less assistance. More importantly, they can use Musical Exercise at their home at their convenient time.

4.2. Research Hypotheses

Based on the design goals of Musical Exercise, we have posed the following hypotheses:

-Hypothesis 1: Musical Exercise will be rated as a usable exercise assistance system by the blindfolded participants. This question will be answered by the System Usability Scale (SUS) score.

-Hypotheses 2: Musical Exercise will help blindfolded participants do exercise as effectively as sighted participants do. This question will be answered based on the effects on consistency, completion time, and workload by comparing blindfolded group with audio conditions and sighted group without audio condition.

Musical Exercise has two distinct audio profiles. Given that the effects of different types of auditory displays vary depending on specific users, tasks, and contexts, we have more exploratory hypotheses about the differences of two audio profiles as follows.

-Hypotheses 3a: Continuous sound will assist the blindfolded participants to do the exercise equivalent to those in the visual without audio condition. This will be answered by comparing the effects of continuous sound on exercise.

-Hypotheses 3b: Discrete sound will assist the blindfolded participants to do the exercise equivalent to those in the visual without audio condition. This will also be answered by comparing the effects of discrete sound on exercise.

4.3. Technical Details

In Musical Exercise, participants performed two exercises - squat and wall sit in front of a Microsoft Kinect device. Microsoft Kinect, which is a non-wearable device, can detect the participant's body movements and Musical Exercise used

the Kinect data to recognize the specific exercise movement. Kinect was placed on a desk in front of the participant. The musical exercise application measured how many repetitions the participant completed, how consistent these movements were and how much time the participant took to complete the movement. Also, our application provided audio feedback so that the participant could understand the quality of the exercise such as how fast and when the participant completed the full range of motion for that specific exercise.

Musical Exercise was developed in C# using Microsoft Kinect for Windows Software Development Kit (SDK) version 2.0. The Displaying body data template (source: <http://kinect.github.io/tutorial/>) was used to develop the application. This code template shows how to get BodyFrame from Kinect, using CoOrdinateMapper for 3D (x, y, z) positions of different body joints and displaying the human body as a skeleton using the XAML figure drawing. On top of this code template, we added our own algorithm for matching a specific exercise - either the squat or the wall sit exercise.

Figure 1: From left to right, movement of the exercise

To decide the requirements for a valid exercise repetition, we interviewed two experts (i.e., a physical exercise researcher and a professional competitor in exercise competitions). They both suggested squat and wall sit as a good full body exercise and recommended the full range of motion. If one's thighs are in parallel to the ground in squat or wall sit position, it is indicated as a good form of those exercises [15]. They also emphasized that the consistency in exercise movement is a crucial indicator of successful exercise. Following their advice, we implemented the exercise detection algorithm. Two knee joints and two hip points are detected and their coordinate positions are calculated. During each repetition of exercise, the time to complete the exercise is calculated and the lowest squat or wall sit with respect to the Kinect position is calculated.

The two exercises incorporate the movement which is from standing up to going down to the position when the knees are in straight line with the hip, which makes a 90 degree angle in the knee and has the thigh parallel to the ground. The exercise ends when going back to the standing position. The difference between squat and wall sit exercises is that in wall sit exercise, the user has to hold the sitting position longer and has back support of the wall. If the participant does not do the full range motion of an exercise, it is not detected as a successful exercise repetition.

While the participant is in exercise motion, one of the two audio profiles will be played based on the participant's state in the exercise motion. The continuous audio profile has eight notes. In the full range motion of an exercise, these eight notes were set to be played based on the participant's position or state in the exercise motion. These are MIDI notes ranging from MIDI number 60 to 71. First note is played when the distance value crosses first cut-off point (28 unit, 1 unit=1cm). Sequentially, the notes are played with decreasing polarity when the distance value crosses 22, 16, 10, 7, 4, 2 and less than 2. When the participant stands up to complete the movement, the same eight notes are played in the reverse order. The range

of motion is not big enough to accommodate eight MIDI notes sound separately with time gaps in between. Thus, this audio profile sounds like a continuous audio in practice. Hence, this audio profile is named continuous audio. If the participants are familiar with at what position which note is played, they can remember the note progression or sequence and try to mimic each time they do the exercise. In short, the sound in the continuous audio profile specifically produces feedback on if the participant is doing the exercise in a correct form. If the participant does not do the exercise in an ideal form, then the participant will hear different sound patterns compared to the sound of the ideal form of exercise.

The discrete audio profile was designed in a simplistic way with three MIDI notes. It has a starting note (G#-44) at the first cut-off points when the exercise begins, an intermediate note (E-64) when the participant is at an acceptable depth (i.e., when the distance between knee points and hip points is almost zero), and lastly an ending note (C-84) to mark the completion. All three MIDI notes sound separately and distinguishably while the participant doing the exercise. Hence, the name of this audio profile is discrete audio. We conducted an evaluation experiment comparing these two audio profiles with participants doing both exercises.

5. EVALUATION

5.1. Participants

Eighteen sighted participants (*Age* = 20 years, 19-22; 13 M, 5 F) were recruited using the Psychology Subject Recruitment System at a university. As a proof of concept, we conducted a preliminary study with blindfolded sighted participants before conducting any evaluation study with people with visual impairments. According to Moll and Sallnäs [35], if any constraint is implied on the blindfolded sighted person, it is hoped to be applicable for people with visual impairments. Kanwal et al. [34] also showed in their experiment of a navigation assistance system, the walking speed of the blindfolded participants were less than blind people. Therefore, a system designed for people with visual impairments like Musical Exercise is likely to impose more restraints in terms of performance on a blindfolded sighted person than a person with visual impairments. Yet we recognize the fact this experiment can be best evaluated by conducting with people with visual impairments. In the pre-questionnaire, all participants expressed their interest to do exercise for better well-being. However, only ten wanted to use technological assistance to do exercise. Eight participants have used fitness applications before. Seventeen participants have done exercise in various forms (e.g., yoga, bodyweight exercises, playing games, barbell or dumbbell exercise or cardio) in their lives. All participants, at least, heard about squat and wall sit exercise before. Eleven participants do squat exercise sometimes and 13 participants do wall sit exercise sometimes in their daily lives.

5.2. Experimental Procedure and Conditions

Once the participants signed the consent form and answered the pre-questionnaire, they were briefed in short about the exercise application. The participants were then assigned to the following six conditions in random order: Visual without audio condition, blindfolded without audio condition, visual with continuous audio condition, blindfolded with continuous audio condition, visual with discrete audio condition, and blindfolded with discrete audio condition. The experiment used the within-subjects design. In the visual conditions,

participants saw the application running on a desktop monitor in front of them. The Kinect device was beside the monitor. The participants were asked to perform two exercises - squat and wall sit in random order. There were two sets of each exercise. In each set, the exercise needed to be repeated five times. There was 2-3 minute rest time between the sets. In the continuous audio conditions, the participants heard the continuous audio profile (eight MIDI notes) from the desktop monitor speaker while doing exercise. In the discrete audio conditions, the participants heard the discrete audio profile (three MIDI notes).

6. RESULTS

Firstly, we present how usable the Musical Exercise system is. Then, we report the effectiveness of Musical Exercise by discussing consistency of the exercise across different conditions. We also checked whether there is any effect of completion time among the conditions. After that, we show the workload measurement results across different experimental conditions. *Note that our goal is to show that blindfolded group using Musical Exercise can do exercise as effectively as non-blindfolded group without any audio aid. Thus, we aim to show statistically not significant results between those pairs.* Lastly, we present subjective questionnaire results.

6.1. System Usability

We used the System Usability Scale or SUS (source: <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>) to measure the usability of our system. SUS is a set of ten questions where each question has five answer options from “Strongly disagree” to “Strongly agree” scale. The score is calculated on a total of 100 points (for details about calculation, see [here-https://measuringu.com/sus/](https://measuringu.com/sus/)). SUS is a reliable and effective way to measure a system’s usability [3]. After completing all the conditions, each participant answered the ten SUS questions. The mean SUS score for Musical Exercise was 71.53 out of total score 100 with a standard deviation of 8.45 calculated from 18 participants’ data. The maximum SUS score was 85 and the minimum was 55 among 18 participants’ SUS scores. Based on the results from 500 studies, the average SUS score 68 or above is considered an “above average” system (sources: <https://measuringu.com/sus/> and <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>). Another study [3] reported that a SUS score of 71.4 with standard deviation of 11.6 can be described as a “good” system. Taken together, the Musical Exercise system stands as an “above average” or “good” usable system.

6.2. Effectiveness of the System

To measure the effectiveness of Musical Exercise, first we analyzed how consistently the participants did the exercise across the six conditions by calculating the variance data of distance values. Small variance means more consistency. However, consistency can also come from a bad form of exercise. For example, if the participants do the exercise consistently by maintaining a proper form almost all the time, then the variance of distance values will be low. In this case, this small variance would be good because the participants did the exercise constantly in the proper form. However, even though the participants consistently did the exercise, they might not sufficiently sit to the acceptable level as a successful position. Then, it would be consistently bad exercise. To distinguish these cases, we also analyzed whether consistency is good or bad by comparing mean distance values between

knee and hip when the participants sit in both exercises across conditions. Figure 2 shows the distance between knee and hip joints.

Deciding good or bad consistency needs a base condition with which we can compare the conditions in question. We selected the visual without audio condition as the baseline condition because if sighted people do not use any exercise assistance system, they would do the exercise just like in the visual without audio condition in this experiment. Therefore, we focused on comparing the mean distance values of the blindfolded with audio conditions (continuous and discrete) with the mean distance values of the visual without audio condition after checking consistency of the conditions. Because of this reason, we directly conducted *planned* pairwise comparisons using paired samples t-tests, instead of conducting an overall ANOVA first and then using the Bonferroni correction. Therefore, we maintained a traditional alpha level (0.05) [33]. Each mean distance value represented the mean of five repetitions’ distance values between the knee joints and the hip joints of a participant.

Figure 2: Distance between knee joints and hip joints

6.2.1. Effects on consistency

To measure the consistency of the exercise, we calculated the variance in distance between knee and hip positions when participants repeated the same exercise. This variance indicates how perfectly the participants did the exercise since equal or less than zero in the distance is considered a good form for both squat and wall sit exercises. Table 1 shows the variance data of squat exercise among six experimental conditions. From Table 1, we can see that the variance values of the visual with discrete audio condition and the blindfolded with discrete audio condition are the least among the six conditions if we consider total sum of variance, mean of variance, or even variance of variance values.

The visual with continuous audio and the blindfolded with continuous audio conditions have the biggest variance values among the six conditions. With discrete audio either blindfolded or not, the participants seemed to perform squat more consistently than with the visual without audio condition or the blindfolded without audio condition. However, with continuous audio either blindfolded or not, the participants seemed to perform squat more inconsistently than the visual without audio condition or the blindfolded without audio condition.

Conditions	Sum	Mean of variance	Variance of variance	SD
Visual w/o Audio	594.159	33.009	1322.933	36.372
Blindfolded w/o Audio	786.739	43.708	3017.291	54.929
Visual w/ continuous A	1228.879	68.271	15138.509	123.039
Visual w/ discrete A	445.287	24.738	837.089	28.932
Blindfolded w/ continuous A	1012.042	56.225	7581.276	87.071
Blindfolded w/ discrete A	590.074	32.782	718.998	26.814

Table 1: Variance in distance between knee and hip for squat in six experimental conditions

Conditions	Sum	Mean	Variance	SD
Visual w/o Audio	-466.583	-25.921	118.738	10.897
Blindfolded w/o Audio	-500.434	-27.802	92.724	9.629
Visual w/ continuous A	-461.375	-25.632	63.465	7.966
Visual w/ discrete A	-493.971	-27.443	55.033	7.418
Blindfolded w/ continuous A	-499.871	-27.771	109.025	10.441
Blindfolded w/ discrete A	-483.542	-26.863	102.882	10.143

Table 3: Mean data for squat exercise in six experimental conditions

For wall sit, results show different patterns (Table 2). The variance of the visual with discrete audio condition is the most similar to the visual without audio, but still higher. The variance of the blindfolded with discrete audio condition is much higher than the visual without audio. The blindfolded with both audio conditions show even higher variance compared to the blindfolded without audio condition. Again, the visual with continuous audio condition and the blindfolded with continuous audio condition show high variance.

In sum, in the visual with discrete audio condition, the participants seemed to perform relatively consistently for both exercises. When in the blindfolded with discrete audio condition the participants performed squat more consistently than either the visual without audio condition or the blindfolded without audio condition, but it is not the same case for wall sit exercise. Next, we show whether the consistent form is good or bad.

To determine if the participants performed in a consistently good form or bad form, we took the mean distance values of the exercise. Previously, while checking if the exercise is done consistently, we have found that in the visual with discrete audio condition, the participants performed more consistently. While the participants in the blindfolded with discrete audio condition performed, at least, as good as those in the visual without audio condition for squat exercise, this was not shown in the same way for wall sit exercise. Participants in the continuous audio conditions (either blindfolded or not) did not provide much consistent results in either of the exercises. Therefore, we focus on the paired samples t-tests between the blindfolded with discrete audio condition mean data and the visual without audio condition mean data.

Conditions	Sum	Mean of variance	Variance of variance	SD
Visual w/o Audio	457.724	25.429	801.7	28.314
Blindfolded w/o Audio	923.243	51.291	1996.086	44.678
Visual w/ continuous A	818.446	45.469	4795.157	69.247
Visual w/ discrete A	517.572	28.754	749.419	27.376
Blindfolded w/ continuous A	1001.747	55.653	7767.746	88.135
Blindfolded w/ discrete A	1026.737	57.041	5564.932	74.598

Table 1: Variance in distance between knee and hip for wall sit in six experimental conditions

Conditions	Sum	Mean	Variance	SD
Visual w/o Audio	-498.673	-27.704	52.048	7.214
Blindfolded w/o Audio	-454.719	-25.262	45.291	6.729
Visual w/ continuous A	-492.174	-27.343	98.564	9.928
Visual w/ discrete A	-491.446	-27.303	52.291	7.231
Blindfolded w/ continuous A	-466.774	-25.932	61.291	7.829
Blindfolded w/ discrete A	-437.104	-24.284	154.838	12.443

Table 4: Mean data for wall sit exercise in six experimental conditions

First for the squat exercise, Table 3 shows the mean values and standard deviation of mean distance data. Paired samples t-test between the blindfolded with discrete audio condition ($M = -26.863$, $SD = 10.14$) and the visual without audio condition ($M = -25.92$, $SD = 10.9$) shows no significant differences, $t(17) = 0.269$, $p = 0.79$. Similarly, for the wall sit exercise, Table 4 shows the mean values and standard deviation of mean distance data. The paired samples t-test between the blindfolded with discrete audio ($M = -24.28$, $SD = 12.44$) and the visual without audio condition ($M = -27.7$, $SD = 7.2$) shows no significant differences, $t(17) = -1.009$, $p = 0.32$. Both of the t-test results suggest that when in the blindfolded with discrete audio condition participants performed not significantly different from the proper form of the exercise when compared to the visual without audio condition as a baseline.

In short, in the blindfolded with discrete audio condition, the participants consistently performed the proper form of squat as same as in the visual without audio condition and performed inconsistently but still as good form of wall sit as in the visual without audio condition.

6.2.2. Effects on completion time

Musical Exercise calculated the time from the beginning of an exercise to the end. This is the completion time of the exercise. We wanted to observe if there is any significant difference in time when the same exercise is done for different conditions. Specifically, we wanted to find out if there was any significant difference of completion time during the exercises between the visual without audio condition versus the blindfolded with continuous and discrete audio conditions respectively.

Paired samples t-test between the blindfolded with continuous audio ($M = 1479.64$, $SD = 473.79$) and the visual without audio condition ($M = 1344.19$, $SD = 416.35$) for the squat exercise completion time shows no significant differences, $t(17) = -$

0.91, $p = 0.37$. Paired samples t-test between the blindfolded with discrete audio condition ($M = 1620.19$, $SD = 553.89$) and the visual without audio condition ($M = 1344.19$, $SD = 416.35$) for the squat exercise completion time also shows no significant differences, $t(17) = -1.69$, $p = 0.10$.

The above two t-tests show that there is no significant difference between the visual without audio condition and the blindfolded with both continuous and discrete audio conditions in terms of squat exercise completion time.

Likewise, paired samples t-test between the blindfolded with continuous audio condition ($M = 2070.22$, $SD = 790.65$) and the visual without audio condition ($M = 2369.94$, $SD = 1216.49$) for the wall sit exercise completion time shows no significant differences, $t(17) = 0.88$, $p = 0.39$. Paired samples t-test between the blindfolded with discrete audio condition ($M = 2522.58$, $SD = 1395.97$) and the visual without audio condition ($M = 2369.94$, $SD = 1216.49$) for the wall sit exercise completion time also shows no significant differences, $t(17) = -0.35$, $p = 0.73$. Therefore, whether it is squat or wall sit exercise, the exercise completion time is not significantly different in the blindfolded with one of two different audio (continuous and discrete) conditions than in the visual without audio condition. It means that while blindfolded with the aid of audios, the participant did not suffer from quickness or lag of completion time compared to a sighted person doing the exercise in the visual without audio condition.

6.3. Workload Measure

We used NASA Task Load Index or NASA-TLX for measuring the participants' perceived workload of Musical Exercise. After completing each of the six conditions, every participant answered the NASA-TLX questions. Based on their overall ratings of all the six experimental conditions, we conducted paired samples t-tests between the conditions. Table 5 shows the mean and standard deviation of all the six conditions.

Conditions	Sum	Mean	SD
Visual w/o Audio	639.96	35.55	18.38
Blindfolded w/o Audio	665.97	36.99	18.20
Visual w/ continuous A	645.67	35.87	19.24
Visual w/ discrete A	640.97	35.61	17.71
Blindfolded w/ continuous A	708.65	39.37	19.24
Blindfolded w/ discrete A	732.32	40.18	17.21

Table 5: Overall rating score of NASA-TLX for six experimental conditions

Paired samples t-test between the blindfolded with continuous audio ($M = 39.37$, $SD = 19.24$) and the visual without audio condition ($M = 35.55$, $SD = 18.38$) shows no significant differences, $t(17) = -0.61$, $p = 0.55$. Paired samples t-test between the blindfolded with discrete audio condition ($M = 40.18$, $SD = 17.21$) and the visual without audio condition ($M = 35.55$, $SD = 18.38$) shows no significant differences, $t(17) = -0.78$, $p = 0.44$. These two t-tests show that there is no significant difference of overall workload between visual without audio condition and blindfolded with both continuous and discrete conditions in terms of perceived workload measure.

6.4. Subjective Questionnaire

After all experimental conditions were completed, participants answered the post-questionnaire: For "Which sound condition do you like better doing exercise?" with options of with audio or no audio, 14 participants (77.78%) out of 18 participants chose the audio option (either continuous audio or discrete audio). For "Which exercise seemed difficult for you?" with options either wall sit or squat, 11 participants (61.11%) out of 18 participants answered wall sit. Table 6 shows the positive results of other post questionnaire questions based on the Likert scale where 1 being strongly disagree and 5 being strongly agree.

Questions	Mean	SD
"I think sound was source of information as feedback for exercise"	4.22	0.43
"The exercise application would improve your motivation to exercise."	3.00	1.03
"The exercise application's audio feedback was intuitive"	3.94	0.64
"The exercise application's audio feedback was useful."	4.22	0.81

Table 6: Subjective questionnaire scores

7. DISCUSSION

Musical Exercise is an exercise assistant application geared towards people with visual impairments. Our goal was to test if this system is usable and effective for people with visual impairments and if adding audio to the system works as an informative feedback medium. We conducted the experiment with sighted people being blindfolded. Though being blindfolded does not fulfill the premise of visual impairments, still, the experiment gives us insight on how people in general can respond to Musical Exercise. The SUS score of Musical Exercise showed that it is an "above average" or "good" system in terms of usability. For squat, the participants in the blindfolded with discrete audio condition tended to show more consistent (i.e., less variance) exercise than in the blindfolded without audio condition in terms of distance values. Also, the paired samples t-test showed that there was no significant difference between the blindfolded with discrete audio condition and the visual without audio condition. Therefore, with this discrete audio, Musical Exercise could improve consistency and assist the participants in maintaining a good form of squat exercise just as a sighted person would do in the visual without audio condition. However, for wall sit exercise, the results were different from the squat case. The performance of the participants was inconsistent both in the blindfolded with discrete audio condition and in the blindfolded with continuous audio condition. The reason seems to lie in participants being blindfolded and doing the wall sit exercise. Firstly, in wall sit exercise, the participants had back support which helped the participants do the exercise. In the blindfolded with audio conditions, the audio feedback and the physical contact to the wall may conflict with each other or even be combined into ambiguous feedback which might confuse the participants. Secondly, wall sit is a dependent exercise having the back-support component, whereas squat is an independent exercise. In squat, the participants developed their own rhythm with the audio feedback and followed it well, but in wall sit, the back support may have hampered the free movement of the participant which might contribute to the bad performance when in the blindfolded with audio conditions. Based on the results of the

experiment, using discrete audio feedback can be recommended only for the squat exercise. Further research and validation is needed to improve audio profiles for wall sit. Nonetheless, in both exercises, audio played an important role by maintaining consistency or at least to ensure the proper form of the exercise. Another important factor is that the blindfolded with discrete audio condition has no significant difference with the visual without audio condition in terms of completion time and workload measure. In effect, people could do exercise in the blindfolded with discrete audio condition as good as the visual without audio condition. On a side note, results also showed that adding the discrete audio profile can also enhance performance of the sighted group compared to the sighted group without audio. However, not all audios were helpful. For both squat and wall sit exercises, in the blindfolded with continuous audio condition, the participants performed worse than in the blindfolded without audio condition. This result demonstrates that good audio design is crucial because it can assist people with visual impairments to perform exercises in a better way. In conclusion, our first two hypotheses are supported in terms of usability and effective exercise assistance system. Between the third (H3a) and fourth (H3b) hypotheses, only the third hypothesis (discrete audio) was partly supported by the results. Even though we were unable to conduct an evaluation study with people with visual impairments, based on the blindfolded conditions, Musical Exercise seems to be promising to be helpful for people with visual impairments. Also, a large portion of the participants (77.78%) preferred to do the exercises with audio. They also commented that audio was intuitive and a source of information. Among the audio profiles, some participants liked the discrete audio profile, commenting that it helped “doing exercise in a correct form” or “ending repetition audio is good”, etc. Some participants found that “continuous audio was hard to follow or even distracting”. Only one participant who had music knowledge commented that the continuous audio profile was “more helpful in blindfolded conditions” and “could recognize a pattern in continuous audio”. The evaluation of Musical Exercise shows that the participants prefer to do exercise with audio rather than no audio. More importantly, the participants thought that the audio feedback was helpful and informational. What distinguishes Musical Exercise from previous research studies is that the feedback system is intuitive and non-obtrusive using audio. Moreover, the use of non-speech sounds in Musical Exercise makes it a unique system as the non-speech sounds customize the feedback for the participants. Another feature of Musical Exercise is a non-wearable system which can be used at home environment with less human assistance.

Though the discrete audio profile was better perceived by our participants, at the same time it indicates that further iterative user evaluation is needed. Audio feedback seems to be essential to guide users with visual impairments to do the exercise. It helps them consistently maintain the proper form of exercise. However, it can be done only when the audio is properly designed. In future, more exercises can be included in the Musical Exercise platform. We can also add more features to the current exercises (e.g., the degree of inclination of the back). It will require researchers to seek more experts’ opinions especially including iterative improvements of the system by evaluating and taking opinions of people with visual impairments on the design of the system. Also, Musical Exercise needs to be further evaluated by people with visual impairments. Overall, Musical Exercise has room for improvements.

8. CONCLUSION

Technology is continuously shaping our lives. Using technology in the area of solving accessibility and assistance issues for people seems to be promising. To use technology to its full potential, we need to create a usable and effective system. In the same line, the Musical Exercise platform aims to help people with visual impairments do exercise more independently. We designed the Musical Exercise system and tested it with two different types of audio feedback, compared to no audio condition. It seems that Musical Exercise is promising and heading to the right direction of being an effective and usable system for doing exercises not only for people with visual impairments, but also for sighted people. It not only worked as a feedback system but also was received positively so that it can also serve as a source of enjoyment during the exercises. Based on the results, we can make some design guidelines or recommendations for systems such as Musical Exercise. We hope that this system can play a major role in the field of using technology for people with disabilities and guide other researchers to design a more usable and effective technology.

9. REFERENCES

- [1] Albiol-Pérez, S., Lozano-Quilis, J., Gil-Gómez, H., Gil-Gómez, J., and Llorens, R. Virtual rehabilitation system for people with parkinsons disease. In 9th international conference on disability, virtual reality and associated technologies (ICDVRAT) (2012), pp. 423–427.
- [2] Augestad, L. B., and Jiang, L. Physical activity, physical fitness, and body composition among children and young adults with visual impairments: A systematic review. *British Journal of Visual Impairment* 33, 3 (2015), 167–182.
- [3] Bangor, A., Kortum, P., and Miller, J. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.
- [4] Bolton, J., Lambert, M., Lirette, D., and Unsworth, B. Paperdude: A virtual reality cycling exergame. In CHI ’14 Extended Abstracts on Human Factors in Computing Systems (New York, NY, USA, 2014), CHI EA ’14, ACM, pp. 475–478.
- [5] Buttussi, F., Chittaro, L., and Nadalutti, D. Bringing mobile guides and fitness activities together: A solution based on an embodied virtual trainer. In Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services (New York, NY, USA, 2006), MobileHCI ’06, ACM, pp. 29–36.
- [6] Conner, C., and Poor, G. M. Correcting exercise form using body tracking. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (New York, NY, USA, 2016), CHI EA ’16, ACM, pp. 3028–3034.
- [7] Elliott, D., Carr, S., and Orme, D. The effect of motivational music on sub-maximal exercise. *European Journal of Sport Science* 5, 2 (2005), 97–106.
- [8] Fletcher, G. F., Balady, G., Blair, S. N., Blumenthal, J., Caspersen, C., Chaitman, B., Epstein, S., Froelicher, E. S. S., Froelicher, V. F., Pina, I. L., et al. Statement on exercise: benefits and recommendations for physical activity programs for all Americans. *Circulation* 94, 4 (1996), 857–862.
- [9] Ganesan, S., and Anthony, L. Using the kinect to encourage older adults to exercise: A prototype. In CHI ’12 Extended Abstracts on Human Factors in Computing Systems (NY, USA, 2012), CHI EA ’12, ACM, pp. 2297–2302.

- [10] Gerling, K. M., Kalyn, M. R., and Mandryk, R. L. Kinectwheels: Wheelchair- accessible motion-based game interaction. In CHI '13 Extended Abstracts on Human Factors in Computing Systems (New York, NY, USA, 2013), CHI EA '13, ACM, pp. 3055–3058.
- [11] Gowin, M., Cheney, M., Gwin, S., and Wann, T. F. Health and fitness app use in college students: A qualitative study. *American Journal of Health Education* 46, 4 (2015), 223–230.
- [12] Houwen, S., Hartman, E., and Visscher, C. Physical activity and motor skills in children with and without visual impairments. *Medicine and science in sports and exercise* 41, 1 (2009), 103–109.
- [13] Huang, M.-C., Chen, E., Xu, W., and Sarrafzadeh, M. Gaming for up- per extremities rehabilitation. In Proceedings of the 2Nd Conference on Wireless Health (New York, NY, USA, 2011), WH '11, ACM, pp. 27:1–27:2.
- [14] Jacko, J. A., Scott, I. U., Sainfort, F., Barnard, L., Edwards, P. J., Emery, V. K., Kongnakorn, T., Moloney, K. P., and Zorich, B. S. Older adults and visual impairment: What do exposure times and accuracy tell us about performance gains associated with multimodal feedback? In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (NY, USA, 2003), CHI '03, ACM, pp. 33–40.
- [15] Kasim, P. Optimizing squat technique. *Strength and Conditioning Journal* 29, 6 (2007), 10.
- [16] Laukka, P., and Quick, L. Emotional and motivational uses of music in sports and exercise: A questionnaire study among athletes. *Psychology of Music* 41, 2 (2013), 198–215.
- [17] Lin, T.-Y., Hsieh, C.-H., and Lee, J.-D. A kinect-based system for phys- ical rehabilitation: Utilizing tai chi exercises to improve movement disorders in patients with balance ability. In Proceedings of the 2013 7th Asia Modelling Symposium (Washington, DC, USA, 2013), AMS '13, IEEE Computer Society, pp. 149–153.
- [18] Middelweerd, A., Mollee, J. S., van der Wal, C. N., Brug, J., and te Velde, S. J. Apps to promote physical activity among adults: a review and content analysis. *International Journal of Behavioral Nutrition and Physical Activity* 11, 1 (2014), 1–9.
- [19] Morelli, T., Foley, J., Columna, L., Lieberman, L., and Folmer, E. Vi-tennis: A vibrotactile/audio exergame for players who are visually impaired. In Proceedings of the Fifth International Conference on the Foundations of Digital Games (New York, NY, USA, 2010), FDG '10, ACM, pp. 147–154.
- [20] Muehlbradt, A., Koushik, V., and Kane, S. K. Goby: A wearable swimming aid for blind athletes. In Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (New York, NY, USA, 2017), ASSETS '17, ACM, pp. 377–378.
- [21] Newbold, J. W., Bianchi-Berthouze, N., Gold, N. E., Tajadura- Jimenez, A., and Williams, A. C. Musically informed sonification for chronic pain rehabilitation: Facilitating progress & avoiding over-doing. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2016), CHI '16, ACM, pp. 5698–5703.
- [22] Palacios-Navarro, G., García-Magarino, I., and Ramos-Lorente, P. A kinect- based system for lower limb rehabilitation in parkinson's disease patients: a pilot study. *Journal of Medical Systems* 39, 9 (2015), 103.
- [23] Rector, K., Bennett, C. L., and Kientz, J. A. Eyes-free yoga: An exergame using depth cameras for blind & low vision exercise. In Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (New York, NY, USA, 2013), ASSETS '13, ACM, pp. 12:1–12:8.
- [24] Salmon, P. Effects of physical exercise on anxiety, depression, and sensitivity to stress: a unifying theory. *Clinical psychology review* 21, 1 (2001), 33–61.
- [25] Shapiro, D. R., Moffett, A., Lieberman, L., and Dummer, G. M. Perceived competence of children with visual impairments. *Journal of Visual Impairment & Blindness* 99, 1 (2005), 15.
- [26] Singh, A., Klapper, A., Jia, J., Fidalgo, A., Tajadura- Jimenez, A., Kanakam, N., Bianchi-Berthouze, N., and Williams, A. Motivating people with chronic pain to do physical activity: Opportunities for technology design. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2014), CHI '14, ACM, pp. 2803–2812.
- [27] Taylor, D. Physical activity is medicine for older adults. *Postgraduate Medical Journal* 90, 1059 (2014), 26–32.
- [28] Wang, Q., Turaga, P., Coleman, G., and Ingalls, T. Somatech: An exploratory interface for altering movement habits. In Proceedings of the Extended Abstracts of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (NY, USA, 2014), CHI EA'14, pp. 1765–1770.
- [29] Warburton, D. E., Nicol, C. W., and Bredin, S. S. Health benefits of physical activity: the evidence. *Canadian Medical Association Journal* 174, 6 (2006), 801–809.
- [30] Weil, E., Wachterman, M., McCarthy, E. P., Davis, R. B., O'day, B., Iezzoni, L. I., and Wee, C. C. Obesity among adults with disabling conditions. *Jama* 288, 10 (2002), 1265–1268.
- [31] Winoto, P., and Tang, T. Y. Sensory substitution to enable the visually im- paired to play an affordable wearable mobile game. In Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Com- puting and Proceedings of the 2015 ACM International Symposium on Wearable Computers (New York, NY, USA, 2015), UbiComp/ISWC'15 Adjunct, ACM, pp. 193–196.
- [32] Zayer, M. A., Tregillus, S., Bhandari, J., Feil-Seifer, D., and Folmer, E. Exploring the use of a drone to guide blind runners. In International ACM SIGACCESS Conference on Computers and Accessibility (Reno, NV, October 2016), pp. 263–264.
- [33] Keppel, G., and Wicken, T. D. Design and analysis: A researcher's handbook. London, UK: Pearson Prentice Hall. (2004).
- [34] Nadia Kanwal, Erkan Bostanci, Keith Currie, and Adrian F. Clark, "A Navigation System for the Visually Impaired: A Fusion of Vision and Depth Sensor," *Applied Bionics and Biomechanics*, vol. 2015, Article ID 479857, 16 pages, 2015. doi:10.1155/2015/479857
- [35] Moll, Jonas, and Eva-Lotta Sallnäs. "Communicative functions of haptic feedback." In International Conference on Haptic and Audio Interaction Design, pp. 1-10. Springer, Berlin, Heidelberg, 2009.

SONIFYING SCIENCE AND TECHNOLOGY

SONIFICATION AND SCIENCE PEDAGOGY: PRELIMINARY EXPERIENCES AND ASSESSMENTS OF EARTH SCIENCE DATA PRESENTED IN AN UNDERGRADUATE GENERAL EDUCATION COURSE

Mark Ballora

Arts Design Research Incubator
The Pennsylvania State University
University Park, PA 16802, USA
ballora@psu.edu

Christopher Roman

Graduate School of Oceanography
University of Rhode Island
Narrangansett, RI 02882, USA
croman2@uri.edu

Robert Pockalny

Graduate School of Oceanography
University of Rhode Island
Narrangansett, RI 02882, USA
rpockalny@uri.edu

Karen Wishner

Graduate School of Oceanography
University of Rhode Island
Narrangansett, RI 02882, USA
kwishner@uri.edu

ABSTRACT

This paper describes preliminary investigations into how sonifications of scientific graphs are perceived by undergraduate students in an introductory course in oceanography at the University of Rhode Island. The goal is to gather data that can assist in gauging students' levels of engagement with sonification as a component of science education. The results, while preliminary, show promise that sonified graphs improve understanding, especially when they are presented in combination with visual graphs.

1. INTRODUCTION

Sonification is becoming an acknowledged informatics method, and its effectiveness for communicating complex datasets has been shown in many studies [1], [2]. However, its uptake has been slow for a number of reasons. We contend that is partly due to the fact that comparatively little auditory display work has focused on issues of education and learning. This collaboration was formed to explore the usefulness of sonification as a tool for science communication and education. Undergraduates, especially non-science majors, are often challenged by having to comprehend graphs that are unfamiliar to them. We created sonifications of a number of graph types that students typically encounter in science courses. All students were invited to submit online surveys before and after they were exposed to sonified versions of graphs. Section 2 of this paper describes related work done previously in the field. Section 3 describes the nature of the graphs used in this project and the designs of the sonifications. Section 4 describes the surveys and their results. Section 5 summarizes the results and describes how this collaboration will continue.

2. RELATED WORK

While the use of non-speech sound has been shown to be

informative in a number of contexts [2], in recent years we have noticed an increase of work done for purposes of outreach and communications. These projects are typically meant for student and non-specialist audiences, often in informal educational settings such as museums [3], [4]. The use of musical sound in informal STEM contexts reinforces recent findings that assert that different modalities of experiencing science are mutually reinforcing, and that integrated presentations of knowledge produce a more educated and robust workforce and research community than the traditional "specialist" model of education [5].

As for earlier examples of sonification being used in marine studies, this seems to be a fairly new domain. There is, of course, one of the most venerable use of auditory display, the sonar, which has for some time been essential to understanding underwater environments. And sonifications of ocean buoy data were presented in [6]. Beyond these two examples, we are not aware of other sonification work done in the area of oceanography.

3. DATASETS AND SONIFICATION DESIGNS

In our preliminary discussions of how to make use of sonifications in oceanography courses, we decided that a first step should be to create a variety of one-dimensional graphs through sound, thus making sonification a running theme of the team-taught general education science course OCG 111 Ocean Exploration. This course is typically populated by freshmen and sophomore non-science majors. It covers the basic tenets of oceanography and ends with a module on global change. Sonifications were created in consultation with Christopher Roman for modules that he taught.

The time series that were sonified for the course included:

- Monthly mean numbers of sunspots, spanning the period from March 1958 to November 2017 [7];
- The "Keeling Curve" [8] of monthly atmospheric carbon dioxide measurements, dating from March 1958 to September 2017;
- Tide levels at four recording stations in Louisiana, Newport, RI, San Diego, CA, and St. Petersburg,

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

FL, covering the period from July 2017 to September 2017 [9];

- Global mean sea levels, spanning the period from March 15, 1958 to December 15, 2013 [10];
- Monthly land and ocean temperature indices dating from March 1958 to October 2017 [11].

All sonifications were created by Mark Ballora with the SuperCollider audio programming language [12]. The approach is parameter-based, creating “sonic scatter plots” [2] in the same manner as work described in earlier publications [13], [14], [15]. The following sections describe the nature of the sonifications.

3.1. Sunspot Sonifications

The sunspot dataset consists of monthly mean counts from March 1958 to November 2017 (Figure 1). A throbbing sound made from filtered noise changes pitch according to the contour of the curve outlined by the mean sunspot counts, and is mapped to a range of half an octave. A light shimmer is added for aesthetic purposes, meant to suggest rays of sunlight. To suggest discrete spots, a clicking sound, somewhat like a Geiger counter, clicks at a varying rate within a range of zero to 30 clicks per second, varying in proportion to the number of mean spots per month.

Figure 1: Monthly sunspot counts plotted per month since 1958.

3.2. Atmospheric CO₂ Levels

The Keeling Curve is often introduced in contexts of climate change. It represents monthly atmospheric carbon dioxide levels. Levels rise and fall yearly with the seasons, while the steady ascent of the overall level is evidence of increasing global CO₂ emissions. (Figure 2).

Figure 2: Keeling curve, showing monthly CO₂ levels in parts per million.

The contour of the Keeling Curve is represented by a somewhat buzzy timbre, and a pitch that spans two octaves from the lowest to the highest data point, with four months playing per second. Since this alone would be a rather bland (although accurate) sound, a complementary layer is added that is derived from the spectrum of carbon dioxide [16] (Figure 3).

Figure 3: Spectrum of carbon dioxide (wavelengths in mm)

The wavelengths shown in Figure 3 are converted to frequencies by dividing them into the speed of light, then transposing and compressing these intervals so that they fall within the audible spectrum. The ascending curve is accompanied by a set of sine waves that are based on these relationships of the carbon dioxide partials. This secondary layer adds a bit of inharmonicity to the sound, although at a level quite a bit lower than the main “melody” instrument.

3.3. River Tide Levels

These datasets consist of tide level measurements taken every six minutes from July 2017 through September 2017. To emphasize the rhythmic but distinct tidal patterns that exist at different locations, four weeks of tidal activity were played (Figures 4a-d). The data points are mapped to pitches that fall within a span of just under two octaves, with the exception of the anomalous drop in levels in the St. Petersburg due to hurricane Irma (Figure 4d, bottom), which can be heard as a drop of an octave. The pitches are played by a sine-like tone that is meant to sound watery.

Figure 4a: Tide levels, Louisiana July 1-29, 2017

Figure 4b: Tide levels, Newport, RI, July 1-29, 2017

Figure 4c: Tide levels, San Diego CA, July 1-29, 2017

Figure 4d: Tide levels, St. Petersburg, FL
July 1-29, 2017 (top), September 2-30, 2017 (bottom)

3.4. Sea Levels

The sea level time series shown in Figure 5 is mapped to pitches that ascend three octaves, played by a windy/watery timbre. In addition to the main graph, which covers the period from 1880 to 2013, a shortened version starts at March 1958, which can be played simultaneously with the sunspot and Keeling Curve datasets.

Figure 5: Global mean sea levels, measured monthly, January 15, 1880 to December 15, 2013.

3.5. Earth Surface Temperatures

The global temperatures plotted in Figure 6 are mapped to a timbre consisting of a filtered combination of a sawtooth wave and Brown noise (sometimes called $1/f^2$ noise). The pitch and brightness reflect the sea levels, which span an octave over the course of the data. A second version of this sonification was created that renders the rightmost portion of the graph, from March 1958 to October 2017, so that it may be played simultaneously with the sonifications of the sunspots, Keeling Curve, and sea levels. The overall sound is meant to convey the short-term (yearly) variability in global temperature along with the steadier longer-term (decadal) increase occurring over the longer time period.

Figure 6: Monthly Land and Ocean temperature indices dating from 1880

4. METHODS

Students enrolled in an introductory oceanography course were invited to complete online questionnaires. No personal information was collected. The protocol was approved by the Institutional Review Boards (IRB) of both the Pennsylvania State University and the University of Rhode Island.

The tide levels were played for the students as an informal introduction to the idea of sonified graphs, during a course module that covered tides and currents. Four weeks of activity, corresponding to July 2017, were played, which covered two rising and falling tidal cycles. For St. Petersburg, a file was also rendered that consisted of the activity in September 2017, since this included the drop in water levels that resulted from hurricane Irma. Each tidal sonification was rendered to last 30 seconds. Students were presented with the sounds and then shown the graphs in Figure 4. They seemed generally able to match the sounds to

the correct graphs, and to have a good appreciation for the rates of tide change.

More formal evaluations, designed by Robert Pockalny, took place during the global change module that ends the course. This class session consisted of presentations of the Keeling graph, sunspots, surface temperatures, and sea levels. The playback time for the files that spanned 1958-2017 was 20 seconds. This duration allowed a number of files to be played back during a class period without jeopardizing students' attention spans. Files of the sea levels and surface temperatures dating from 1880 were rendered at a proportional timescale, which meant that they lasted 46 seconds. Two combinations were also played: 1) the Keeling curve/surface temperatures/sea levels and 2) surface temperatures plus the "clicking layer" of the sunspot sonification. Chris Roman created a combined display in MATLAB[®] software [17] that consisted of a moving dot progressing along the graphs as the sonifications played.

The audio files may be downloaded at <https://bit.ly/2wczYmB>.

4.1. Initial Survey

Before the sonifications were presented in class, students were asked to complete an online survey that was meant to characterize their perceived knowledge for various science and climate-related topics. Surveys were submitted by 53 out of 106 students. The results are shown in Tables 1-4.

Table 1: General Science Knowledge

Far below average	Below average	Average	Above average	Far above average
2%	13%	51%	26%	8%

Slightly above average (3.25 rating, 3 as expected average)
Standard deviation: 0.84

Table 2: Climate and Climate Change Knowledge

Far below average	Below average	Average	Above average	Far above average
4%	21%	47%	28%	—

Average (3 rating, 3 as expected average)
Standard deviation: 0.8

Table 3: Familiarity with Natural Cycles Related to Climate

Not at all familiar	Not so familiar	Somewhat familiar	Very familiar	Extremely familiar
14%	38%	40%	4%	4%

Below average (2.46 rating, 3 as expected average)
Standard deviation: 0.64

Table 4: Ability to Interpret Graphs

Far below average	Below average	Average	Above average	Far above average
2%	6%	64%	26%	2%

Slightly above average (3.2 rating, 3 as expected average)
Standard deviation: 0.92

4.2. Post-Sonification Survey

A second questionnaire was made available after students had been exposed to visual and auditory graphs. This was designed to get a sense of their understanding of material presented through sound. The second questionnaire was submitted by 23 students. The results are shown in Tables 5 through 11.

Table 5: Which answers describe the relationship between the sunspots and global temperature? (Select all that apply)

	Pre-Sonification (50 respondents)	Post-Sonification (23 respondents)
Sunspots occur on an 11-year cycle, and are likely causes of periodic changes in global temperature.	20%	30%
Sunspots occur on an 11-year cycle, but are not likely causes of periodic changes in global temperature.	22%	39%
An increase in sunspots frequency correlates with an increase in global temperature.	36%	43%
An increase in sunspots frequency does not correlate with an increase in global temperature.	14%	13%
What are sunspots?	8%	—
Don't know	32%	9%

Table 7: Which best describes the change in global temperature since 1880?

	Pre-Sonification (50 respondents)	Post-Sonification (23 respondents)
No significant change	2%	—
Initially not much change, but a more rapid increase recently.	66%	87%
Initially a rapid increase, but not much change recently.	10%	4%
Initially not much change, but a more rapid decrease recently.	6%	4%
Initially, a rapid increase, but not much change recently.	4%	—
Don't know.	12%	4%

Table 9: Effectiveness of “Climate Sounds” and graphics to understand patterns and trends of climate change

A great deal	15%
A lot.	23%
A moderate amount	50%
A little	4%—
None at all	8%

Table 10: Effectiveness of “Climate Sounds” and graphics to understand causes and effects of climate change

A great deal	19%
A lot.	23%
A moderate amount	38%
A little	15%
None at all	4%

Table 6: Which answers describe the relationship between the CO₂ record at Mauna Loa (i.e. Keeling Curve) and global temperature? (Select all that apply)

	Pre-Sonification (50 respondents)	Post-Sonification (23 respondents)
The overall increase in the CO ₂ trend follows the global temperature record.	44%	57%
There is no relationship between the CO ₂ trend and global temperature record.	6%	4%
The annual variability in CO ₂ is due to seasonal differences in wind patterns.	6%	22%
The annual variability in CO ₂ is due to seasonal differences in photosynthesis.	24%	17%
The annual variability in CO ₂ is due to seasonal differences in temperature.	6%	43%
There is no annual variability in CO ₂ .	—	—
What's the Keeling Curve?	12%	—
Don't know.	34%	17%

Table 8: How variable has the global temperature been since 1880?

	Pre-Sonification (50 respondents)	Post-Sonification (23 respondents)
Uniform	—	—
Overall, a general increase in temperature with 10- to 20-year variations/undulations in the overall trend.	73%	65%
Overall, a general decrease in temperature with 10- to 20-year variations/undulations in overall trend.	13%	13%
Uniform increase in temperature	7%	13%
Uniform decrease in temperature	7%	4%
Don't know.	—	4%

Table 11: Which approach did you find most effective for advancing your knowledge about patterns, trends and cause/effect relationships associated with climate change?

Graphs alone are most effective	23%
Sounds alone are most effective	8%
The combination of graphs and sounds are most effective	54%
All about the same	15%

5. DISCUSSION

5.1. Survey Results

The number of submissions of the second survey was lower than we would have preferred. There is no definitive way to identify why this was the case. The surveys were optional with no incentives offered for participation. While this simplified the IRB process, it left open the possibility of a low response rate. Without higher yields, results can only be considered preliminary. But the responses that were submitted indicate overall that: students found the addition of sounds to be effective, and the sound renderings were an aid in understanding the science content.

The results of the question about the relationship between the sunspots and global temperatures (Table 5) indicate potential misunderstandings for some respondents, shown in the first choice. There is an increased number of students that indicate that sunspots are likely causes of global temperature changes—20% indicating this as correct before hearing the sonifications, and 30% indicating this as correct after hearing the sonifications. In fact, the second and third answers are correct: an increase in the number of sunspots correlates with increases in global temperatures, but they do not cause them. There are approximately equal numbers of increased incorrect and correct answers on the second, post-sonification questionnaire.

The second question asked about the relationship between the Keeling Curve and global temperatures (Table 6). Here again, results were mixed. After the sonifications were played, there was an increase in the percentage of correct responses for the first choice, from 44% to 57%, which identified an increase in the CO₂ trend as following the global temperature record. There was also a decrease in the selection of the fourth (incorrect) choice, that annual CO₂ variability is due to seasonal photosynthesis, which dropped from 24% to 17%. After the sonifications there were increases in two incorrect answers: the third selection, that annual CO₂ variability is due to seasonal differences in wind patterns, rose from 6% to 22%; the fifth selection, that annual CO₂ variability is due to seasonal temperature differences, rose from 6% to 43%. Overall, there seemed to be an increase in knowledge of the Keeling curve and its relationship to temperature after the sonifications were played, although there were still 4 of 23 respondents (17%) on the post-sonification survey who indicated that they did not know whether there was a relationship between them.

The greatest improvement in correct selections in the post-sonification test had to do with identifying how global temperatures had changed since 1880 (Table 7), which rose from 66% to 87%. Those who completed the second survey were able to hear the rapid increase in temperatures, which reinforces an oft-quoted justification for the use of auditory displays: the strength of the ear in discerning patterns.

There was a slight shift from correct to incorrect answers on the question about variability in global temperatures since 1880 (Table 8). This suggests the possibility that some students misunderstood the term “variable.”

Tables 9 and 10 indicate that the majority of students found that the sonifications improved their understanding. Perhaps the most telling responses had to do with preferences of displays for understanding patterns, trends, and cause/effect relationships (Table 11). The great majority

found that the combination of visual graphs and auditory displays were the most effective, rather than one or the other.

5.2. Future Work

Now that a suitable audio design and format has been created, there is nothing to prevent these files from being used in future offerings of this course.

Preliminary work has also been carried out in sonifying populations of copepods in Oxygen Minimum Zones of the oceans. This is an area of active research in the field of oceanography, of which Karen Wishner is a leading figure [18], [19]. Oxygen Minimum Zones (OMZs) are mesopelagic ocean features with extremely low oxygen. They are predicted to expand geographically and in intensity with global climate change, which potentially impacts ecosystems, fisheries, and economies. Copepods are small, often microscopic-sized crustaceans. A variety of species are represented in terms of the average number per cubic meter of water, at increments of 100 meters, from the surface to a depth of 1000 meters. A topic of active investigation is the diel vertical migration (DVM) in which larger zooplankton and fish descend to about 300m depth during the day to avoid predators and ascend to the surface at night to feed. The extent of this migration compared to their body size is remarkable and not yet fully understood.

The next stage of work done under the auspices of this grant funding will be to sonify diel vertical migrations of various copepod populations within these zones. The educational intention will be to enhance learning of complex ocean processes in the undergraduate classroom. We anticipate that these auditory renderings can also help reveal correlations and unexpected interactions between zooplankton distributions and the environment that will be useful to researchers in the field.

5.3. Concluding Thoughts

The results, while preliminary, show promise. We regard this collaboration as aligning well with a long-term goal: to engender a generation of students who consider science as something that is understood through hearing as well as seeing. We expect that, for many students, musical sound can provide a more holistic and intuitive understanding of scientific phenomena than is possible through visual presentations alone.

We should stress, however, that by saying this we do not mean to imply that sonification is superior to or a substitute for time-honored methods of visual display. But there is good reason to feel that the addition of auditory displays to other graphing methods will add value to scientific engagement and understanding. In the keynote to ICAD 2017 [20], Carla Scaletti quoted Bruno Latour, who asserted that sciences produce objective knowledge through multiple arguments. Knowledge is never advanced by an isolated mapping. Rather there are always multiple mappings – instruments, theories, calculations, charts, graphs, etc. – that form a “cascade of inscriptions” that in a chain of reasoning that leads to a conclusion. New knowledge is the result of multiple interdependent inscriptions, which combine to legitimate new conclusions.

Our hope is that projects such as this one can be steps in a healthy “cascade of inscriptions,” steps that have a positive impact on science pedagogy by making scientific exploration appealing to wider audiences and by predisposing the next

generation of researchers to seek discoveries through hearing as well as seeing.

6. ACKNOWLEDGEMENT

Research reported here was supported by the National Academies Keck *Futures Initiative* of the National Academy of Sciences under the award number NAKFI DBS20. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Academies Keck *Futures Initiative* of the National Academy of Sciences.

7. REFERENCES

- [1] Kramer, Gergory, and Bruce Walker, Terri Bonebright, Perry Cook, John Flowers, Nadine Minder, Nadine, John Neuhoﬀ, Robin Bargar, Stephen Barrass, Jonathan Berger, Grigori Evreinov, W. Tecumseh Fitch, Matti Gröhn, Steve Handel, Hans Kaper, Haim Levkowitz, Suresh Kodha, Barbara Shinn-Cunningham, Mary Simoni, Sever Tipei. *Sonification Report: Status of the Field and Research Agenda*. Prepared for the National Science Foundation by members of the International Community for Auditory Display, 1999.
- [2] T. Hermann, A. Hunt, J. G. Neuhoﬀ. (Eds.). *The Sonification Handbook*. Berlin: Logos Publishing House, 2011.
- [3] J. L. Laughner, E. K. Canfield-Dafilou. "Illustrating Trends in Nitrogen Oxides Across the United States Using Sonification." Presented at the 23rd International Conference on Auditory Display (ICAD 2017), June 20-23, 2017, University Park, Pennsylvania, USA.
- [4] B. J. Tomlinson, R. M. Winters, C. Latina, S. Bhat, M. Rane, B. N. Walker. "Solar System Sonification: Exploring Earth and Its Neighbors Through Sound." Presented at the 23rd International Conference on Auditory Display (ICAD 2017), June 20-23, 2017, University Park, Pennsylvania, USA.
- [5] National Academies of Sciences, Engineering, and Medicine. *The Integration of the Humanities and Arts with Sciences, Engineering, and Medicine in Higher Education: Branches from the Same Tree*. Washington, DC: The National Academies Press, 2018. doi: <https://doi.org/10.17226/24988>.
- [6] B. L. Sturm. "Pulse of an Ocean: Sonification of Ocean Buoy Data." *Leonardo*, vol. 38, no. 2, pp. 143-14, April 2004.
- [7] SILSO data/image, Royal Observatory of Belgium, Brussels. <http://www.sidc.be/silso/home>
- [8] C. D. Keeling, C.D. and T.P. Whorf. "Atmospheric CO₂ concentrations derived from flask air samples at sites in the SIO network." In *Trends: A Compendium of Data on Global Change*. Oak Ridge, Tennessee, U.S.A.: Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, 2004. ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_mm_mlo.txt
- [9] Center for Operational Oceanographic Products and Services (CO-OPS) <https://tidesandcurrents.noaa.gov/>
- [10] J. A. Church, J. A. and N.J. White. "Sea-level rise from the late 19th to the early 21st Century." *Surveys in Geophysics*, vol. 32, nos 4-5, pp. 585-602, September 2011. <https://datahub.io/core/sea-level-rise> doi:10.1007/s10712-011-9119-1.
- [11] <https://data.giss.nasa.gov/gistemp/graphs/>
- [12] <http://supercollider.sourceforge.net>
- [13] M. Ballora, "Sonification, Science and Popular Music: In search of the 'wow.'" *Organised Sound*, vol. 19, pp. 30-40, April 2014. doi:10.1017/S1355771813000381
- [14] M. Ballora, "Two Examples of Sonification for Viewer Engagement: Hurricanes and Squirrel Hibernation Cycles." Presented at the 21st International Conference of Auditory Display (ICAD 2015), July 8-10, 2015, Graz, Austria.
- [15] M. Ballora, "Music of Migration and Phenology: Listening to counterpoints of Musk Ox and Caribou Migrations, and Cycles of Plant Growth." Presented at the 22nd International Conference of Auditory Display (ICAD 2016), July 2-8, 2016, Canberra, Australia.
- [16] <http://chemistry.illinoisstate.edu/standard/che362/handouts/362molvibs.pdf>
- [17] <https://www.mathworks.com>
- [18] L. Stramma, S. Schmidtko, L. A. Levin, G. C. Johnson. "Ocean oxygen minima expansions and their biological impacts." *Deep Sea Research Part I: Oceanographic Research Papers*, vol. 57, no. 4, pp. 587-595, April 2010.
- [19] K. F. Wishner, D. M. Outram, B. A. Seibel, K. L. Daly, R. L. Williams. "Zooplankton in the eastern tropical north Pacific: Boundary effects of oxygen minimum zone expansion." *Deep Sea Research Part I: Oceanographic Research Papers*, vol. 79, pp. 122-140, September 2013.
- [20] C. Scaletti, "Why Sonification is a Joke." Keynote address, 23rd International Conference of Auditory Display, June 20, 2017, The Pennsylvania State University. https://www.youtube.com/watch?time_continue=6&v=T0qdKXwRsyM

PLANETHESIZER: APPROACHING EXOPLANET SONIFICATION

Adrián García Riber

AGR Image and Sound Art,
Music Research and Sound Engineering,
Francisco Martí Mora 1-B, 22-3, IB 07011, Spain
adrian@imageandsoundart.com

ABSTRACT

The creation of simulations, sounds and images based on information related to an object of investigation is currently a real tool used in multiple areas to bring the non-specialized public closer to scientific achievements and discoveries. Under this context of multimodal representations and simulations developed for educational and informational purposes, this work intends to build a bridge between virtual musical instruments' development and physical models, using the gravitation laws of the seven planets orbiting around the Trappist-1 star. The following is a case study of an interdisciplinary conversion algorithm design that relates musical software synthesis to exoplanets' astronomical data - measured from the observed flux variations in the light curves of their star- and that tries to suggest a systematic and reproducible method, useful for any other planetary system or model-based virtual instrument design. As a result, the Virtual Interactive Synthesizer prototype *Planethesizer* is presented, whose default configurations display a multimodal Trappist-1, Kepler-444 and K2-72 planetary systems simulation.

1. INTRODUCTION

The exponential increase in virtual instruments' development and their popularity and establishment as part of the current workflow on audio productions, both in home studios and professional environments, creates a new opportunity in bringing science to the general public and designing audio tools based on physical or mathematical models of undoubted creative and artistic utility. There are numerous examples of composers who translate models based on nature or science into musical parameters to give their works a formal and structural solidity transcending their own compositional intuition. Their kind of approach, in addition to the publication of the discovery of the seven Earth-like planets orbiting around the Trappist-1 star in February 2017, inspired the idea of exploring the creation of a model-based virtual instrument. Developing an interactive astronomical data sonification tool useful in both creative and educational or informational contexts, this work also tries to suggest a possible line of interdisciplinary audio tools development.

2. ASTRONOMICAL DATA SONIFICATION REFERENCE WORKS

In an attempt to make a state-of-the-art review in terms of astronomical data sonification and to analyze the technology, processes, tools and techniques that are currently involved in

this field, it seems clear that Pythagoras (6th century B.C.) and his *Music of the Spheres* as well as Kepler's *Harmonices Mundi* (1619), represent the crucial starting point of the relation between occidental music and astrophysics [1]. Highly influenced by these fundamental works, the animated sonifications of the solar system from Twyman (2010) [2], and Mark Ballora's sonifications for the film *Rhythms of the Universe* (2013) [3] (a co-production by the ethnomusicologist and percussionist from the mythical band *Grateful Dead* and cosmologist George Smoot III from *Lawrence Berkeley Labs*), can be found. Quinton, M. et al. [4] went a step further with their perceptual research on the solar system's sonifications created for the planetarium Esplora (2016, island of Malta), proposing a planet properties test through sound recognition, realized on both specialized and non-specialized audiences. In this sense of probing the effectiveness of the astronomical data sonification processes, the work of Diaz Merced, W. L. (2013) [5], the *Supernova Sonification* [6] and the *xSonify* prototype project [7], built in collaboration with NASA, also stand out. Trying to equilibrate creative freedom and scientific accuracy to the original data sheet, it's worth mentioning the solar wind data sonifications of Alexander, R. et al. (2010) [8].

In a more creative context, Winton et al. [9] developed the sonifications of stellar data from the Kepler Telescope (2012), using *Sonification Sandbox* software [10] and *Matlab*'s interpolation functions to explore the aesthetical possibilities of direct unprocessed sonifications. Also interesting are the orchestral character coding sonifications of Quinn, M. for the exoplanets and stars information database of the *European Southern Observatory* [11], and the mapping sonification work of Jamie Ferguson in collaboration with ESA *From Hipparchus to Hipparcos* (2014) [12], that translate several historic star catalogues to sound in a similar way to John Cage's *Etudes Australes*. Finally, it's impossible not to mention Tim Pyle's *Trappist Transits* musical representation (2017) [13], illustrating data from NASA's *Spitzer* Telescope for informational purposes.

3. TRAPPIST-1 SYSTEM

The Trappist-1 system is located 39 light years (12 parsecs) away from the solar system within the constellation of Aquarius. It's formed by a red dwarf star, Trappist-1A, and seven planets named from Trappist-1b to Trappist-1h, as function of their radial distance to the star. In 2015 the *Transiting Planets and Planetesimals Small Telescope* (TRAPPIST) monitored the brightness of the star for 245 hours over 62 nights, from September 17th to December 28th, showing clear transit-like signatures later confirmed as planets b, c and d [14]. The system was completely identified using the *Trappist telescope*, the *Spitzer Space Telescope*, the *Very Large Telescope*, the *UKIRT*, the *Liverpool Telescope* and the

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

William Herschel Telescope [15]. The complete system discovery was finally published in February 2017 [16].

4. PHOTOMETRIC TRANSIT DETECTION

The photometric transit detection technique is based on the observation of the light flux variations of celestial objects. It basically consists of pointing a telescope at the star and obtaining its light curve, a graphic representation of brightness flux variations along time. When a planet passes in front of a star, it produces a partial eclipse and the consequent loss of flux perceived can be measured in terms of time and depth. As illustrated in *Figure 1*, the duration of total eclipse has a direct correspondence with the decrease of energy observed in the light curve and, according to (1), the transit depth is proportional to the surface ratios of the planet and the star, R_p/R_* .

Figure 1: Measure of flux variations over a generic light curve. Seager & Mallén-Ornelas, 2003. [17]

$$\frac{R_p}{R_*} = \sqrt{\Delta F}, \quad (1)$$

Equation (2) gives the orbital inclination of the planet, where a is the orbital radius (assuming circular orbit, Kepler's 3rd law), and b is the impact parameter which represents the increase of total eclipse duration for orbits close to the star's equator. It can also be calculated from the light curve using (3).

$$i = \cos^{-1} b \frac{R_*}{a}, \quad (2)$$

$$b = \sqrt{\frac{(1 - \sqrt{\Delta F})^2 - \left(\frac{t_F}{t_T}\right)^2 (1 + \sqrt{\Delta F})^2}{1 - \left(\frac{t_F}{t_T}\right)^2}}, \quad (3)$$

The orbital period of each planet can be calculated measuring the time interval between transits. At least three transits are needed to obtain conclusive results [18].

5. PLANETHESIZER VST PLUGIN DESIGN

The *Planthesizer* prototype was designed using Iain McCurdy's works [19], *LorentzSynth* (2015) and *Planet* (2012), as an inspiration. It can be defined as a vector synthesis virtual instrument implemented from Csound's *planet* opcode. This opcode (or computing operation code), generates the coordinates (x, y, z) of a planet orbiting a binary star system simulation whose outputs -calculated for each of the seven planets- are used to control seven software synthesizers. The instrument also provides an additional eighth synthesizer, controlled through a real-time MIDI keyboard, to improve

musical interaction. Csound [20] was used in combination with Cabbage front end [21] to implement a VST (Virtual Studio Technology) plugin and maximize compatibility with any DAW (Digital Audio Workstation), both on Mac and PC computers. Blender software [22] was used for the graphic design of the star.

The prototype can be downloaded for testing from: <https://archive.org/details/@agriber>

Figure 2: *Planthesizer* interface illustrating the axes used in the simulation's graphic display.

Figure 2 shows the prototype interface that allows end-users to control sound properties as a function of the radius, semi-axis, inclination, period and impact factor for the orbit of each planet or synthesizer in astronomical units. To reproduce the eclipse sequence cadence, each planet can be initially delayed introducing time differences expressed in days. Level control is allowed for master output, the keyboard's synthesizer as well as for every planet's module. The *Restart* button initializes the eclipse sequence and the *Init. Delay Off* button deactivates initial delay, allowing the synchronization of the planets for creative purposes. The prototype also includes a global tempo, global zoom control and one *on/off* button for each planet. Instrument default configuration represents a multimodal simulation of the Trappist-1 system. Saving and recalling presets is also allowed.

According to the research agenda proposed by Kramer et al. [23], the plugin's requirements relay on three main axes: adaptation to the language and needs of the research field, end-user control or interaction and prototype integrability. As described below, several targets were defined for the prototype:

- Provide a synchronized multimodal simulation for planetary systems
- Provide an effective and interactive control for every sound parameter expressed in astronomical units
- Provide end-users a flexible control of the simulation including global tempo and single planet, keyboard and master level controls
- Develop a micro-tonal and multi-timbral musical instrument
- Develop an extremely intuitive user interface
- Develop a Trappist-1 system sonification extendable to any other planetary system
- Allow real-time operation for all parameters
- Allow full integration in Digital Audio Workstations

The plugin design process was divided, as follows, into eight blocks attending to Iannis Xenakis's 'fundamental phases of a musical work' [24]:

1. Prototype concept definition
2. Sound elements definition
3. Physical-Musical-Mathematical algorithm design
4. Global control functions definition
5. Simplified prototype implementation, *b* planet
6. Full prototype implementation
7. Default configuration and presets implementation
8. Exporting to final format and testing under different DAW's

6. IMPLEMENTATION ALGORITHM

In order to maximize the intelligibility and simplicity of calculation, in *Planetesizer* every sound appears linked to each planet's graphic simulation, forming a multimodal representation with a high degree of intelligibility that minimizes its learning curve. *Figure 3* describes the interactive algorithm, parameters and variables used to control the planets' simulation. As shown in the block diagram, the *planet* opcode generates three-dimensional vectors whose coordinates (*ax*, *ay*, *az*) can be manipulated in real-time by end-users. Delayed variables (*axd*, *ayd*, *azd*), are scaled and converted to Csound's control variables (*kx*, *ky*, *kz*) to achieve the final visual representation and to allow synchronized sonification.

Figure 3: Block diagram of the synthesizer's control algorithm.

Global *Tempo* control changes the speed of the complete eclipses sequence. *Initial delay* creates the desired eclipses time sequence in concordance to observed transits, as described below in *Table 1*. *Zoom* modifies all planet size and amplitude to increase realism and interactivity. *Impact.f* is commonly expressed in units related to Jupiter's radius and in the prototype, positive factors represent orbits above the star's equator. Negative factors were used for *d,e,f* and *h* planets as a uniformity convention. Orbital *Period* is measured in days. *Semi-axis* is measured in astronomical units (distance Earth-Sun = 1AU = 14,959,7871 Km). *Inclination* has a multiplying factor that makes the simulation differ from observed data. For aesthetical reasons, 90 degrees correspond to horizontal axis. The planet's *Radius* control modifies both the timbre or pitch and size of the planet representation and its measure is related to Earth's radius.

As described in the Canonical Csound Reference Manual [25], *planet* opcode uses eleven arguments to generate the above-mentioned output coordinates (*ax*, *ay*, *az*). These arguments were initialized as follows:

- *kmass1*: Represents the mass of the first star in the simulated binary star system.
- *kmass2*: Represents the mass of the second star.
- *ksep*: Represents distance between the two stars.

After intense testing with a one planet simplified prototype, *kmass1* and *kmass2* values were respectively fixed to 11, 60.458 and 0.001, adapting the opcode capabilities to the prototype's simulation needs.

- *ix, iy, iz*: Introduce the initial coordinates of the planet. Initializing *ix* to -1, the planet's rotation is set counter clockwise. 0.1 and 0.3 values were fixed to *iy* and *iz* in prototype's behavior optimization process.
- *ivx, ivy, ivz*: Allow to introduce the initial vector's speed and were set to 0, 0 and 3.35 to assure stability.
- *idelta (ih)*: Controls the model's velocity as function of orbital period divided by 100,000. This factor was introduced to maintain planet's visualization.
- *ifric*: Introduces a gradual attenuation into planet's orbit. It was initialized to zero to achieve the perpetual intended simulation.

Figure 4: Block diagram showing the six parameters involved in each planet sonification. Note that *kx* and *ky* were normalized to assure sound stability.

As a descriptive reference point, the Trappist-1 sonification's left-placed planets represent lowest frequency sounds, analogously to the frequency axis of spectral analysis or the left hand in keyboard instruments. Sound duration was related to the visible hemisphere of the orbits and therefore to orbital period *and* *x* coordinate in the visual representation (*kx_norm* Csound's variable). *Global tempo* also modifies this variable. Pitch was related to *y* coordinate (*ky_norm* Csound's variable), in order to achieve a highly intuitive design and to ease the planets' auditory recognition. Highest sounds correspond to highest *y* values and by modifying the orbit's inclination control, tonal sweeps are generated. The *z* coordinate (*kz* Csound's variable) was related to sound intensity allowing a fade in and fade out synchronized audiovisual representation. Global zoom and planets' faders also affect this variable. The following magnitudes were dismissed in the implementation of the plugin but, if needed, they could be included in future prototypes:

- The Surface temperature of each planet could be color-coded for visual differentiation and their Irradiation could be related to the opacity of the representations.
- The Mass and Density could be mapped to the amount of an additional sound effect such as flanger or phaser.
- The Eccentricity was delegated on *planet* opcode, assuming circular orbits.
- The transits duration doesn't correspond to observed data since the star simulation is not affected by zoom control.

Intensive testing with sound generation *opcodes* described in Csound's Manual [25] was also done. Final choice required a compromise solution between processing CPU load and sound richness. Finally, the Csound's generators shown on *Table 1* were chosen to create a multi-timbral instrument and to increase clarity and precision in the visual to sound association of the multimodal display. Different timbres have been searched with the exception of planets *e*, *f* and *g*. These planets generate their sound using identical oscillators whose fundamental frequency can be controlled acting on the radius parameter. This makes possible the creation of micro-tonal chords expanding the synthesizer's creative possibilities.

Table 1. Csound opcodes, synthesis types and initial delay (in days), used for Trappist-1 System Sonification.

Planet	Opcode	Synthesis	Delay
b	fof	Granular	0.6
c	vco2 Moogladder	Wavetable LPF	0.5
d	buzz	Harmonic sinusoids	0.9
e	poscil3	Hi resolution sinusoid	1.2
f	poscil3	Hi resolution sinusoid	0
g	poscil3	Hi resolution sinusoid	0
h	hsboscil	Wavetable	9.5
Keys	foscil	FM	-

To create the desired simulated planetary sequence, a time conversion from seconds (inside the simulation) to days (in observed data) was needed. This conversion was made comparing the duration of 15 complete periods of different planets in the prototype and time intervals between observed Trappist-1 exoplanets shown in *Table 1*. As a result, an initial delay, expressed in days, can be entered for each planet of the prototype in concordance with orbital period units. This delay can be globally switched off for creative purposes.

The next tables show the mapping conversion used for the opcodes listed in *Table 1*. Opcode's arguments are fully described in the Canonical Csound's Reference Manual [25]. For a complete analysis, Planthesizer's source code is available through Cabbage or Csound software installation.

Table 2. Conversion used in *b planet* sonification.

fof opcode arguments	Mapping value
xamp	$2 * \text{kamp} * \text{kv} * \text{kMast}$
xfund	$\text{cpsmidinn}(\text{knum})$
xform	$\text{cpsoct}(\text{kcfoct})$
koct	kb
kband	$100 * (\text{kRp})$
kris/kdur/kdec	$0.002 * (\text{kRp}) / 0.2 / 100$

Several of Csound's control variables and functions were used during the prototype implementation process. As shown on *Tables 2 to 7*, to introduce amplitude, fader and master level control in the opcodes' amplitude argument (*xamp* and *kamp*) respectively *kamp*, *kv* and *kMast* variables were used. *Impact factor* control is introduced by *kb* variable to control octave (*koct*), fundamental frequency (*kcps* and *xcps*) or the number of harmonics (*knh* in *d planet's buzz* opcode) synthesized. The planet's *Radius* is represented by *kRp* and introduces sound character or pitch variations. Some factors were also applied in the algorithm to optimize the simulation. In order to maintain auditory display under a controlled audio frequency range, *knum* and *kcfoct* are respectively scaled versions of *kx_norm* and *ky_norm*. The *cpsmidinn* and *cpsoct* functions convert *knum* and *kcfoct* values to cycles per second, allowing

to control the fundamental frequency and octave range of the generated sound.

Table 3. Conversion used in *c planet* sonification.

vco2 opcode arguments	Mapping value
kamp	$2 * \text{kamp} * \text{kv} * \text{kMast}$
kcps	$\text{cpsmidinn}(\text{knum})$
imode	2
inyx	0.5

To introduce the planet's impact factor (*kb*) in the generator's fundamental frequency control of *buzz* opcode, *kcps* variable was defined as $\text{kx_norm} * 90 * (2 + \text{kb})$. *Table 4* illustrates how planet's *Radius* (*kRp*), also modifies this frequency.

Table 4. Conversion used in *d planet* sonification.

buzz opcode arguments	Mapping value
xamp	$2 * \text{kamp} * \text{kv} * \text{kMast}$
xcps	$\text{kcps} * 2 * \text{kRp}$
knh	$20 * \text{ky_norm} * (2 + \text{kb})$
ifn	<i>gisine</i>

Planets *e* and *f* use identical mapping so their sound differences in the Trappist-1 system simulation are exclusively due to the values of the planets' astronomical data. As can be experimented in the prototype soloing these planets, this kind of approach is highly recommended if auditory display properties differentiation is a must.

Table 5. Conversion used in *e* and *f planets*.

poscil3 arguments	Mapping value
kamp	$\text{kamp} * \text{kv} * \text{kMast}$
kcps	$50 * (2 + \text{kb}) * \text{kRp} / \text{ky_norm}$
ifn	-1
iphs	<i>ftgenonce</i> 1,0,4096,10,1

In the sonification of *g planet*, the same *poscil3* opcode was used. A multiplication factor was introduced to obtain one octave higher representation.

Table 6. Conversion used in *g planet* sonification.

poscil3 arguments	Mapping value
kamp	$\text{kamp} * \text{kv} * \text{kMast}$
kcps	$100 * (2 + \text{kb}) * \text{kRp} / \text{ky_norm}$
ifn	-1
iphs	<i>ftgenonce</i> 1,0,4096,10,1

The tone (*ktone*) and brightness (*kbrite*) of the last planet sonification needed to be scaled to assure sound stability.

Table 7. Conversion used in *h planet* sonification.

hsboscil arguments	Mapping value
kamp	$\text{kamp} * \text{kv} * \text{kMast}$
ktone	Between -4 and $40 * \text{kb}$
kbrite	Between $-2 * 1.6 / \text{kRp}$ and 14
ibasfreq	100
ioctfn	0
iphs	2

Table 8. Conversion used for MIDI keyboard.

foscil arguments	Mapping value
xamp	$\text{iam} * \text{kenv}$ (MIDI, p5)
kcps	icps (MIDI, p4)
xcar	1
xmod	1.414
knxd	2
ifn	1

7. SIMULATED VS MEASURED PLANETARY SEQUENCE

Once the prototype was completely implemented, the sonification of the Trappist-1 system was recorded and analyzed to test its behavior. The spectrograms of *Figure 5* obtained with Sonic Visualizer software [26], show the audio file content in frequency (vertical axis) and time (horizontal), with time in seconds and a frequency range of 80Hz to 22Kz in the upper graph, and 10Hz to 1.5KHz in the low frequency detail spectrogram. The transit moments of each planet around Trappist-1 star are also shown in the lower graph of *Figure 5*. Here, horizontal axis corresponds to time in days and every planet is represented in different colors and lines, suggesting the correspondence of each transit with the generation of a sound in a clear analogy to musical notation. This information was obtained by the NASA Spitzer Space Telescope (SST) in 2016 [27], measuring the star light intensity variations, almost continuously for 500 hours.

Figure 5: Comparison between the simulated eclipse sequence and the Trappist-1 system telescope observations published by Guillon, M. et al., 2017. Both spectrograms (top and middle) are related to same audio file and its frequency axis was split for a complete recognition of sound events.

The comparative image shown on *Figure 5* allows us to appreciate how the simulated sequence matches the behavior of the observed exoplanets system -assuming no precision- a fact that suggests the possibility of predicting transit moments from the simulation prototype beyond the observed data, which might be useful for informative purposes.

Analyzing the spectrograms, planet *h*, with a single transit and a concave curve spectral representation can be easily distinguished. It only appears in the upper spectrogram as its fundamental frequency exceeds the limit used in the lower representation. With two transits, Planet *g* is also easily identifiable as two narrow horizontal straight lines in the upper area of the low frequency graph. The length of these lines, equivalent to sound duration, corresponds approximately to half of its orbital period (sounds have an envelope of amplitude with a certain time of attack and relaxation that can reduce accuracy). Since the sonifications of planet *f* and *e*, with three and four transits respectively, were performed with the same synthesizer as *g*, their representations also appear as straight lines whose vertical position depends on the fundamental frequency assigned by radius and impact factor parameters. Likewise, the five transits of planet *d* can be recognized as columns of spaced horizontal lines corresponding to its synthesized harmonics. On the other hand,

sounds generated by planets *c* and *b* are difficult to identify in the spectrogram due to their time coincidence and their similar spectral representation. Both sonifications generate sinusoidal pulses, represented as vertical columns that occupy the entire frequency range. The two planets are, however, easily distinguishable by listening to their sound representation.

Figure 6: Trappist-1 sonification detail showing pitch cues of planets *e*, *f* and *g*. Respectively 157.968 Hz (D#3+27c), 236.685Hz (A#3+27c) and 487.241Hz (B4-23c). Sonic Visualizer software.

Finally, it's important to mention that the intense orange shapes appearing in the spectrogram correspond to sound energy accumulation and should not be interpreted as any additional event. It is also necessary to mention how the activation of the sounds doesn't match that of maximum amplitude for reasons of concordance with the visual representation, so these areas have no meaning outside the prototype.

8. APPLYING THE MODEL TO OTHER PLANETARY SYSTEMS

In order to verify the usefulness of the prototype in the representation of planetary systems, and in addition to the default configuration of the Trappist-1 system, two presets have been incorporated representing respectively the Kepler planetary systems 444 and K2-72. Both planetary systems were observed by the Kepler space telescope using the photometric transit detection method and have been selected from the NASA's official file of confirmed exoplanets [28].

The Kepler-444 system is formed by five Earth-like planets orbiting around its star with periods comprised between 3,6 and 9,74 days. It is estimated that it was formed when the universe was 20% of its current age so is the oldest known Earth-like planetary system. All the information about this system can be found in the article by Campante et al. [29]. K2-72 is a system formed by four Earth-like planets orbiting EPIC 206209135. It is estimated that the planet K2-72d is 10% warmer than Earth while K2-72e could be 6% colder than our planet [30].

This evaluation of the prototype's possibilities revealed some considerations to take into account. In order to represent any planetary system, the scale of the variables involved in the simulation must be reconfigured attending to their astronomical dimensions. The above-mentioned planetary systems were chosen from NASA's file to fit the initial simulation model, as some other planetary systems were out of range. To solve this issue, an end-user's interactive variable range control is recommended for future versions or designs. The time-delay based eclipse sequence generator proposed in the prototype was validated in all cases. The sound representation proposed allows us to establish clear comparison between orbital periods and semi-axes of different planetary systems, but to satisfy any end-user's comparative

requirements, an interactive synthesizer selection control should be implemented in future versions or designs.

9. PERCEPTUAL ASPECTS

According to Vogt's approaches [31], the task of human auditory system consists in segregating and fusing the frequency components in correspondence with real-world phenomena. This fusion can occur due to the proximity of sounds' amplitude, time, frequency, periodicity or spatial localization and generates masking effects, widely studied in Psychoacoustics reference literature listed by De Campo [32]. Other aspects, such as context in which a sound is generated or timbre similarity of sound elements, can also produce associations in the auditory system. These groups can sometimes be desirable, like in the case of generating musical chords, or generate confusion and loss of information, as in the case of a conversation masked by an ambulance's siren. While analyzing and designing sonification tools, these associations may cause loss of information. In addition, subjectivity in sound perception is presented as one of the greatest obstacles to overcome in the development of sonification tools and its importance as a cognitive process is often overlooked due to its everydayness.

Planthesizer's sonification design was, therefore, oriented to provide a clear auditory distinction of each periodic sound event represented in the simulation, with the exception of planets *b* and *c*, whose intentional similarity was focused to test their graphical recognition through spectrogram analysis. Despite using similar oscillators -sinusoid pulses- in both cases, whose spectrograms are difficult to differentiate in time coincidence situations, auditory recognition is possible thanks to the perception of their fundamental frequency, demonstrating that auditory representations offer a useful complementary source of information to graphical display. As suggested by Walker and Brewster [33], the spatialization of sound sources also facilitates the interpretation of information. Attending to this idea, a stereophonic representation was chosen for the prototype, in which the recreated stereo image width varies as function of the planets' orbital radius.

With the intention of providing a didactic example for masking effects both in graphic and auditory displays, the three identical oscillators of planets *e*, *f* and *g* can be tuned to desired frequencies modifying the corresponding radius or impact factor of each planet. In this way, a chord of three notes can be generated, allowing a 'desirable' fusion of the sounds. Starting from identical time configurations for the three planets, a synchronized sound block is generated allowing the observation of how the graphic display loses its usefulness, despite the auditory display still providing a clear differentiation of the sounds that conform the chord. In a similar way, the 'Cocktail party effect' can be experimented representing the entire system or focusing attention on a specific planet. The prototype also allows 'Haas effect' monitoring acting over initial delay on two unison sounding planets. *Chords*, *Arpeggio*, *Unison*, *Cocktail* and *Haas* presets were included in the prototype for this application. As a creative musical add-on, it's also possible in the prototype to generate real time arpeggios introducing an initial delay or switching the *On/Off* button of the desired modules. Musical figures can also be made by controlling the planets' speed through orbital period or global tempo alterations.

10. CONCLUSION AND PROSPECTIVE

Throughout the development of this research work, an intense approximation to the techniques and processes related to astronomical data sonification was made. The analysis of several works carried out in this field resulted crucial to understanding the problem of sound representation and its application in scientific research. However, the comparison and evaluation of projects of this nature is complex due to the amplitude and heterogeneity of the parameters involved. Under this perspective, and despite the fact that sonification processes are in full swing, it is necessary to invest resources, time and effort in the development of new tools that encourage the expansion of this field and its countless applications. Frauenberger, C. [34] also claims the need to demystify the process of designing sound representations to establish the auditory channel as an appropriate interactive medium in the transfer of knowledge.

One of the greatest difficulties found in the development of sonification tools lies in the conjunction of multiple interdisciplinary aspects required to ensure the final usefulness of the prototype. From this point of view, the possibility of using *opcodes*, *frameworks* and accessible documentation for their use alleviated the project burden in an important way. As Stephen Barras expresses in his golden rules for the design of sound representations [35], through the experimental development of the prototype, it was verified how most direct representations require less psychological descriptions. In this sense, vector synthesis from a graphic representation's coordinates, offered a direct and complete way of representation for the basic description of the model under study. Likewise, the multimodal representation achieved complies with the 'level principle' (Cleveland WS, 1985, cited in [35], chapter 13, p.7), which expresses how the potential of a representation is a function of its ability to summarize a general behavior, while allowing examination in detail.

Although it has been necessary to delimit the ranges of values of some parameters to maintain the internal coherence of the simulation, these restrictions are easily eliminated or adaptable, if a new relation of input data is required. Several compromise solutions that distance the simulation from reality were made in order to provide an intuitive interface for non-expert users and to add creative facilities to the prototype. It is worth mentioning here the case of the orbital inclination, oversized to magnify its response or the impact factor sign interpretation. These deviations should be considered descriptive purpose licenses that can also be easily eliminated if needed. Giving final-users the ability to choose the type of synthesis and range of values of each sound parameter, would imply an interesting improvement of the prototype providing the synthesizer with total interactivity.

Planthesizer virtual instrument development has finally served to confirm the possibilities of multimodal representations and sonification as a discipline capable of actively contributing to the development and expansion of interdisciplinary knowledge and creativity. It provides a range of functionalities that extend its range of application beyond the simulation of planetary systems and the generation of musical sequences, allowing the creation of micro-tonal chords and even its use for educational or informational purposes related to astronomy, exoplanet detection, sonification processes, musical perception or audio synthesis.

11. ACKNOWLEDGMENT

Special thanks to Edith Alonso Sánchez for the support and guidance in the research work on which this document is based [36], and to Carmen Enrique and Helen Gräwert for their clinical eye.

12. REFERENCES

- [1] García Martín, R. (2009). La teoría de la armonía de las esferas en el libro quinto de *Harmonices Mundi* de Johannes Kepler. (Doctoral Thesis). Universidad de Salamanca.
- [2] Twyman, L. (2010). Retrieved on 05/01/2017 from: <http://whitevinyl.design.com/solarbeat/2010>
- [3] Ballora, M. (June, 2014). Sonification strategies for the film Rhythms of the Universe. *Proceedings of the 20th International Conference on Auditory Display*, New York, USA.
- [4] Quinton, M., McGregor, I. and Benyon, D. (July, 2016). Sonifying the Solar System. *Proceedings of the 22nd International Conference on Auditory Display*, Canberra, Australia.
- [5] Diaz Merced, W. L. (2013). *Sound for the exploration of space physics data*. (Doctoral Thesis). University of Glasgow.
- [6] Diaz Merced, W. L. (2014). *Supernova Sonification*. Retrieved on 05/25/2017 from: <http://www.npr.org/sections/thetwo-way/2014/01/10/261397236/dying-stars-write-their-own-swan-songs>
- [7] Diaz-Merced, W. L., Candey, R.M., Brickhouse, N., Schneps, M., Mannone, J.C., Brewster, S. and Kolenberg, K. (2012). Sonification of Astronomical Data. *New Horizons in Time-Domain Astronomy Proceedings IAU Symposium No. 285, 2011. R.E.M. Griffin, R.J. Hanisch & R. Seaman, eds*.
- [8] Alexander, R., Zurbuchen, T. H., Gilbert, J., Lepri, S. and Raines, J. (2010). Sonification of AC Level 2 Solar Wind Data. *Proceedings of the 16th International Conference on Auditory Display*, Washington D.C., USA.
- [9] Winton, R., Thomas M. Gable, T. M., Jonathan Schuett, J. y Walker, B. N. (June 2012). A sonification of Kepler space telescope star data. *Proceedings of the 18th International Conference on Auditory Display*, Atlanta, GA, USA.
- [10] Walker, B. N. (2010). Sonification Sandbox (Version 6.1) [Software]. Atlanta, GA: Georgia Institute of Technology. Retrieved from: http://sonify.psych.gatech.edu/research/sonification_sandbox
- [11] Quinn, M. Retrieved on 05/01/2017 from: <http://drsrl.com>
- [12] Ferguson, J. (2015). *From Hipparchus to Hipparcos: A Sonification of Stellar Catalogues*. Retrieved from: <http://jfergusoncompsci.co.uk/research>
- [13] Pyle, T. (2017). *Trappist Transits*. Retrieved on May 24 from: <http://www.spitzer.caltech.edu/explore/blog/371-Making-Music-from-Exoplanets>
- [14] Gillon, M., Jehin, E., Lederer, S. M., Delrez, L., De Wit, J., Burdanov, A., Van Grootel, V., Burgasser, A. J., Triaud, A. H. M. J., Opitom, C., Demory, B., Sahu, D. K., Gagliuffi, D. B., Magain, P. and Queloz, D. (2016). *Temperate Earth-sized planets transiting a nearby ultra-cool dwarf star*. *Nature*, Vol 533, pp. 221-223
- [15] Trappist one, 2017, accessed June 2017, <http://www.trappist.one/#about>
- [16] Gillon, M., Jehin, E., Lederer, S. M., Delrez, L., De Wit, J., Burdanov, A., Van Grootel, V., Burgasser, A. J., Triaud, A. H. M. J., Opitom, C., Demory, B., Sahu, D. K., Gagliuffi, D. B., Magain, P. and Queloz, D. (February 2017). *Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1*. *Nature*, Vol 542, pp. 456-460
- [17] Waldmann, I. 2013, *Finding exoplanets, the Transit technique*. Accessed June 2017, <http://studylib.net/doc/8188000/111--finding-exoplanets-the-transit-technique>
- [18] *Planet Hunters Guide*, NASA, accessed May 2017, <https://www.planethunters.org/#/education>
- [19] Iain McCurdy, accessed March 2017, <http://iainmccurdy.org/>
- [20] CSound software, accessed February 2017, <http://www.csounds.com>
- [21] Cabbage software, accessed January 2017, <http://cabbageaudio.com>
- [22] Blender software, accessed March 2017, <https://www.blender.org>
- [23] Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J. H., Miner, N. and Neuhoff, J. (2010). Sonification Report: Status of the Field and Research Agenda. *Faculty Publications, Department of Psychology. Paper 444*. Retrieved from: <http://digitalcommons.unl.edu/psychfacpub/444>
- [24] Xenakis, I. (1992). *Formalized Music. Thought and Mathematics in Music*. Hillsdale, NY: Pendragon Press.
- [25] Vercoe, B. (May, 2017). *The Canonical Csound Reference Manual*. Retrieved on 06/22/2017 from: <http://csound.github.io/docs/manual/index.html>
- [26] Cannam, C., Landone, C. and Sandler, M. (2010). Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files. Presented in *Proceedings of the ACM Multimedia International Conference*, Firenze, Italia.
- [27] <http://www.trappist.one/#plots>
- [28] NASA. (2017). *Exoplanets archive*. Retrieved on 07/20/2017 from: <https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=planets>

- [29] Campante, T. L. y Barclay, T., Swift⁵, J. J., Huber, D., Adibekyan, V. Zh., Cochran, W., Burke, C. J., Isaacson, H., Quintana, E. V., Davies, G. R., Silva Aguirre, V., Ragozzine, D., Riddle, R., Baranec, C., Basu, S., Chaplin, W. J., Christensen-Dalsgaard, J., Metcalfe, T. S., Bedding, T. R., Handberg, R., Stello, D., Brewer, J. M., Hekker, S., Karo, C., Kolbl, R., Law, N. M., Lundkvist, M., Miglio, A., Rowe, J. F., Santos, N. C., Van Laerhoven, C., Arentoft, T., Elsworth, Y. P., Fischer, D. A., Kawaler, S. D., Kjeldsen, H., Lund, M. N., Marcy, G. W., Sousa, S. G., Sozzetti, A., White T. R. (2015). An ancient extrasolar system with five sub-Earth-size planets. *Astrophysical Journal*. Retrieved on 07/16/2017 from: <https://arxiv.org/abs/1501.06227>
- [30] Dressing, C. D., Schlieder, J.E., Crossfield, I.J.M., Newton, E. R., Ciardi, D.R., Fulton, B. J., Gonzales, E.J., Howard, A. W., Isaacson, H., Livingston, J., Petigura, E. A., Sinukoff, E., Everett, M., Horch, E. and Howell, S.B. (2017). Characterizing K2 candidate planetary systems orbiting low-mass stars II: Planetary systems observed during campaigns 1-7. *Cornell University Library*. Retrieved on 07/20/2017 from: <https://arxiv.org/abs/1703.07416>
- [31] Vogt, K. (June, 2010). *Sonification of Simulations in Computational Physics*. (Thesis). Institute for Electronic Music and Acoustics, University of Music and Performing Arts, Graz, Austria.
- [32] De Campo, A. (February, 2009). *Science by ear. An interdisciplinary Approach to Sonifying Scientific Data*. (Thesis). University for Music and Dramatic Arts, Graz.
- [33] Walker, A. y Brewster, S. (2000). *Spatial Audio in small screen device displays*. University of Glasgow, Scotland. Retrieved on 06/24/2017 from: http://eprints.gla.ac.uk/3205/1/spatial_audio1.pdf
- [34] Frauenberger, C. (July, 2009). *Auditory Display Design*. (Thesis). Queen Mary University of London.
- [35] Boulanger, R. (Ed.) (2000). *The Csound Book: Perspectives in Software Synthesis, Sound Design, Signal Processing, and Programming*. Cambridge, MA, USA: MIT Press.
- [36] García Riber, A. (October, 2017). *Planthesizer: Aproximación a la sonificación de exoplanetas*. (Master's Thesis). International University of La Rioja.

TOXSAMPLER: LOCATIVE SOUND ART EXPLORATION OF THE TOXIC RELEASE INVENTORY

Michael V. Blandino

Center for Computation and Technology
School of Music
Louisiana State University
mblandi@lsu.edu

ABSTRACT

Regulatory geographic datasets that inform citizen's lives are, in general, responsive to engaged search and visual, attentive browsing, but are not designed for directly informing the lived context. The density of sensors and software interfaces present in mobile devices allows for integration of these resources with contextual applications. ToxSampler is an iOS application that modifies the immediate environmental audio scene with associated data from the Toxic Release Inventory (TRI) of the United States Environmental Protection Agency. The application applies digital signal processing (DSP) to the microphone signal based upon the location of the participant and associated TRI data releases. The system, as a result, affords an informed awareness of the datascape through an immediate augmentation of the sensed setting.

1. INTRODUCTION

Extremely large data sources describing comprehensive, annual releases of toxic industrial pollution for the United States are difficult to comprehend or to utilize due to their size, dimensionality, ranges of specificity, and the complex methodologies of their collection and manipulation. Accordingly, one must consult separately stored column descriptions or indexes in order to know what data are presented in particular comma-delimited zones of a set. Further, data are significantly abstracted from the setting in which people live, work, and play, by associating values with coordinates of facility identification numbers or in referenced data grids improved by sophisticated, yet rearranged, GIS projections as back-end databases for maps. The Toxic Release Inventory (TRI) is such a data source. Helpful visualizations of its data as maps are available directly from governmental institutional sources[1], but even visual information is removed from experience if presented in an aerial, illustrated view that is both uncommon for the average person and difficult to connect to lived experience in a personal, narrative scope. These maps are primarily available in browser-based applets, but the National Institutes of Health (NIH) has recently provided an iOS application to find facilities and release information, bringing the data in closer relation to the user location.

Beyond informational presentations, the increasing ubiquity of mobile devices provides an opportunity for the diffusion of sound

art through a locative system that may inform participants of their proximity to toxic release activity documented in the historical record of this federal regulatory program. Mobile devices offer high quality components for the realization of such a system: GPS, microphone, audio output, and software interfaces for accessing data feeds from these components. Beyond the halls of a museum or concert hall, an interactive engagement with art informed by these data is possible at locations chosen by or incidental to the participant. ToxSampler is a realization of such a system.

2. MOTIVATION

Initially, the intention was to provide a processed mediation between music played by applications on the device and line output, affecting playback. However, limitations preventing access to the music stream of proprietary applications block the routing necessary for this signal processing design. These limitations, though, pointed a way toward processing the available microphone audio stream, an approach which creates an interaction with the actual sounds of the landscape and opens possibilities for augmented audio using the toxic release data.

ToxSampler engages the participant with the TRI release data proximal to their location at a resolution of 2000m² by applying audio effects to the microphone input, thereby immersing participants in the data space. There are contours to the release data by location, and salient indicators of their values are not otherwise available while traversing the landscape without referring to visual cartographic resources that display facility location markers and associated release quantities or, at best, shade regions by entire counties to display health risk. Signal processing of the sounds of the present environment based upon values in these data informs the *perception* of the participant, providing information, but also providing an augmented experience within the senses that can invoke an emotional response. As part of a human/computer system, participants perceive, through their augmented cyborg senses[2], the estimated health risks of their environment that cannot otherwise be perceived.

Use of the personal mobile device empowers the individual to explore spaces where they may incidentally be active. If one were to arrive at a new location for some purpose, querying the risk factors through existing resources on an informative map may give a reference and expectation, but to hear features of the data and experience its contours directly connects one's perception and the choice of one's location to what would not be discernible otherwise. Readily accessible, the mobile device and the application together allow one to *incorporate* awareness of the underlying en-

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

vironmental situation of a place.

3. BACKGROUND

Processing an audio signal with control parameters affected by data is a form of sonification or auditory display. The state of this practice lies between poles of tool-making in the service of scientific inquiry and the making of music or sound art informed by data. Much discourse about the mapping of data to sound features emphasizes discernibility, and suggests using high and low pitch levels as preferred options.[3] These recommendations implicitly assume that the primary goal of sonification is primarily for fidelity of display.

Applications of auditory display to data collected in monitoring environmental quality have been realized before. Sounds mapped to air quality data have been synthesized with an orientation toward clarity or even basic representation, using metaphors of sensor instrumentation, such as the Geiger counter or heart monitor[4]. Andrea Polli's *New York Sound Map* (2006) provided an interactive interface for listening to sounds of the city in an exploration of the sound space. This work includes an engagement with the city as it has manifested and with the features of its soundscape. Her *Airflight Taipei* [5, 203-250] (2006) project featured air quality sensor readings in a visual and sonic art display, although available only in a fixed location as an exhibition. Similarly, Knowbotic Research's *Dialogue with the Knowbotic South* [6, 83-86] (1994-1997), which was a seminal work of visual and auditory display of ecological data with impressive visual clarity for the computational capabilities available at the time, could only be explored in the exhibition setting.

A more variable sampling of air quality was realized in the *PigeonBlog* [7, 192-199] (2006) of Beatriz da Costa, where the sampling was initially controlled by site selection for the release of sensor-equipped pigeons, but with distribution patterns left to the animals from that point. Air quality sensors equipped with GPS components were attached to pigeons who were released to collect data at sites. The website of the project distributed the interface, which included data readings and maps in textual and visual displays.

There is a strong tradition in new media art of making unseen forces or phenomena perceivable through display of gathered data. Greg Niemeyer and Chris Chafe's *Ping* [6, 136] (2001) was a sonic art work that made packets traversing the network of the world-wide-web audible within a fixed outdoor location at a museum. An event-based work, Usman Haque's *Sky Ear* [6, 222] (2004), made electromagnetic fields visible via a construction of helium balloons configured in connected rings with light emitting electronics to form a large mass with discrete sensors and flashing displays. The *Sky Ear* detected and gave feedback of electromagnetic activity in the vicinity of the site. Participants could interactively call parts of the apparatus to enable effects. Christina Kubisch's *Cloud 2017* [8] features unseen magnetic fields as composed presentations recorded from actual locations. Participants hear unseen yet physically present patterns through a direct sensor interface within the custom headphone apparatus that detects her sculpted playback interface. A more locative and situationally exploratory project, Kubisch's *Electrical Walks of San Francisco* (2016) brings unseen electromagnetic fields that are active in the cityscape to participants' ears. ToxSampler brings unseen toxic release data together with the soundscape, unifying several threads present in these historical examples, but utilizing the personal, mo-

Figure 1: A subsection of the prepared ToxSampler release grid that controls sound parameters

bile device.

4. TECHNICAL REALIZATION

The software of the system is programmed as an iOS application with data derived from, at the time of its development, the most recent comprehensive TRI data set from 2015. The participant's location is accessed by the software through the Location Services function persistent within the iOS, which draws upon GPS, cellular, and wireless internet location data. The application will still work when outside of cellular or wireless networks if a GPS chip is present on the device.

ToxSampler is built with the Swift programming language. The release data grid is stored within the application and is imported on build from a gridded summation of release data convolved with a Gaussian kernel to extend meaningful values from release coordinate points. Audio processing makes use of AVAudioEngine effects with control parameters guided by the grid data according to coordinate locations from the location services utility.

Out of respect for privacy, and for conformance to development guidelines enforced by the operating system manufacturer, the participant must grant permission to utilize the microphone as well as to access the location. In the ToxSampler system, these data are strictly used only for the audio signal processing and are not stored persistently beyond the application session, nor are they remotely accessible or transmitted in any way.

4.1. Augment Reality

ToxSampler does not provide *instrumentation* or visual feedback in units on a dial. While the intention is to convey the measured risk of pollution's effects in the locale of the listener, it is through an experiential and representative deformation of the directly sampled audio presence of the locale itself that this risk is presented. One can proceed to explore the joined physical space informed by the processed audio and data space (grounded in actual release data). Periodic location checks update the TRI reference value and change the attendant signal processing values accordingly. Thus, the virtual locale presented by ToxSampler augments reality with the mapped sound design parameters.

Due to its derivation from release data, the application is *informative* and augments with truthful observations as reported through the policy regime. In that it informs through manipulation of sensation of the world, the augmentation provided by the application is *experiential*, changing the sensory system in which the participant perceives the acoustic energy of the immediate sound field. In a perceptual sense, sensation is purposefully degraded by the application to connect the toxic release to the sensory apparatus. Disorientation is a result — one intended to inform, not only by discernible quantified values, but by the qualitative effect on experiencing the world through the mediated deformation.

4.2. Sound Design and Parameter Mapping

As opposed to synthesized or direct realization of sampled data into audio signals as is often found in instrumentation-oriented auditory displays, ToxSampler uses the information values as control signals to affect the audio signal from the microphone. The signal processing effects and the data are discernibly related in changes of direction and magnitude.

The echo is a phenomenon of recognizance and of returning sound of perceived events or, in the case of sounds originating from an agent, actions. The sounds return, but encoded within these sounds is *information* about the event or action and about the resonating structure. The amount of time spanning a natural reverberation or echo is information itself, informing a witness of the dimensional features found in a space. Echo added through analog or digital signal processing delay embeds information about a virtual space. ToxSampler adds a digitally realized delay to the microphone signal in units of time (maximum 2 seconds) to partially encode the level of toxic releases in a location. For however long the delay time, the higher the documented release amount is for the given data region. Because the range of release data has large outliers and because significant values are intended to be salient in ToxSampler, a logarithmic scaling is used to set the control parameters.

The gain of the delayed signal (on a scale of 0-100) is set by a mapping, (1),

$$4\log_2(r + 1) \quad (1)$$

where r is the release sum for the associated location grid.

The delay time in seconds is mapped according to (2).

$$0.08\log_2(r + 1) \quad (2)$$

The impulse response¹ of the ToxSampler system to a noise impulse in a recording studio setting is shown in figure 2. With a local toxic release value yielding a delay of 1.478 seconds, the delay and feedback gain values affect the signal accordingly. Impulse decay from -3dB to -60dB occurs over a span of 29.49 seconds.

Another effect, distortion, causes noise-like influences on the quality of audio signals, introducing textures that deform the sound significantly. Distortion is added to the ToxSampler signal at a scaled level that is increased as the TRI data increases. One cannot compare or recall levels of distortion as clearly as one can a delay time value. Distortion, however, is more suggestive of a degradation of the sound environment which mirrors the degradation of the life-sustaining environment by toxic releases.

¹The displayed impulse response recording of figure 2 may be accessed at <http://blandino.art/ToxSamplerImpulse.wav>, along with other audio examples.

Figure 2: Impulse response of ToxSampler at a location with resulting delay time of 1.478s.

Figure 3: The audio signal flow of the ToxSampler application.

The distortion level (0-100) applied to the incoming signal is mapped according to (3).

$$0.5\log_2(r + 1) \quad (3)$$

4.3. User Interface

Informing participants through sound as opposed to visual cartographic and quantitative display modes, the ToxSampler application does not provide a graphic control surface beyond enabling the app and deactivating it. A demonstration mode is available for presentation and testing settings in order to compare the sound results from an area with high levels of releases vs. another with none or to engage participants who only are allowed a limited interaction with the application by selecting their home location or a favorite (or known release) site.

5. DISCUSSION

5.1. Blending the Real and the Virtual

Modification of the *immediate* casts the perception of the user into a virtual, mediated context. In the case of ToxSampler, the context is a virtual one informed by the measurement of personal risk factors to the best level of knowledge available. While virtual, recognizability of the real is preserved, in that sounds may still be

discerned within the level of distortion set and within the delayed returns. The making unreal of the real and making real of the obscure are operations that support resulting reflection, but they simultaneously dull the impact of the health risk and effects of the toxic discharges on the community of the user or that of the participant's temporary locale as synthetic constructions. On the other hand, the release data are declarations of physical, located releases that were real and are likely to continue in the location alongside the continued, associated industrial activity. The manifestation of the representations of these releases in our perception, however virtual, is sourced in the real.

5.2. Ethics and Manipulation

The scale and quality of sonification signal processing must be informed by considerations of design to ensure salience of the data but also by considerations of ethics — out of concern for participant perception, for representing facility actors, and for representation of health risks. It is possible to process an audio signal to such a degree that no sounds are audibly recognizable or at a volume that is directly painful. To do so in this case would sever the link between present environmental experience that connects events and actions in the environment to what is heard through the system. Such an extreme influence of the data space would suggest commensurate extremes of degradation of the environment that would imply that an area is altogether uninhabitable or even instantly, fatally toxic. Thus, recognizability of environmental sounds captured by the microphone is intentionally preserved in the system design.

At the other end of the data domain, where there is zero risk of exposure to toxic releases, the audio signal is passed through with no influence, not even to compensate for the attendant distortions or artifacts inherent to listening to the environment through the small microphone of a mobile device. In the contour of salience that runs between the maximum and minimum, one could err in either underrepresenting or overrepresenting risk in the rate of increase in level of influence caused by the data at various points of the mapping curve.

5.3. Scenario

Walking a landscape, be it urban, suburban, rural, or industrial, with headphones and either the built-in microphone or an accessory microphone, the user of ToxSampler will, as a basis, hear the sounds before them. The attendant attributes of listening through a small microphone will be present: everyday sounds are amplified such that the resonance of a space is prominently noticeable and small sounds near the microphone, especially those associated with the user's movement, are noticeable in the soundscape. Dependent on the proximity to documented facility locations with releases in the TRI, the sounds will be colored by the signal processing representative to the data space.

The intention of the listener to sample the data space, and to choose how to move within it, gives agency to introduce a potential for manipulation of the signal and explore the landscape with this manipulation added and, thereby, experiencing it through the mediating "lens" of the dataspace. Unattached to a gallery or fixed terminal or kiosk, the work is mobile, and control of its realization is distributed.

5.4. Discovery

ToxSampler does not provide a reference back to what industrial entities are releasing toxic chemicals or any municipal zoning or regulatory influences that allow or encourage their existence. Situationist exploration of the landscape[9] alongside the annotated sound experience is, thereby, an exploration of recorded release events in actual space and of the artifacts and expressions of desire by investors, consumers, and political actors who bring them about. One need not travel excessively far to perceive a gradient increase or decrease in release values due to the resolution of data, if near a release site. Multiple ToxSampler listening experiences in locations that are far apart will provide contrasting experiences in remembered relationship across time.

5.5. Tactical Action

There are limitations to the implementation of ToxSampler as a vector for tactical action. It is prepared more readily as a means of informing individual experience with environmental risk factors. Wearing headphones is an isolating gesture not given to community interaction and sharing, but a group could certainly go on a ToxSampling excursion together. To use ToxSampler, one must be in a setting where care can be taken to hear well enough via headphones or amplification, and either take a static ToxSample or move across the static data space. One may record the output of the application quite readily with an auxiliary cable and recording device and share the results or incorporate into derivative art. Listening in a car to traverse more rapidly across the dataspace allows one to hear contours in the parameter shifts and introduces the possibility of feedback from loudspeakers. Intervention at a public location could certainly be realized with preloaded devices.

5.6. Participant Feedback

Participants who have used ToxSampler report an increase of knowledge about the releases in their locale as a result of engaging with the system. In one instance, an area that a participant assumed would have high release data was discovered to have a null release value. Participants who live and work in an area surrounded by multiple sites with relatively high release levels, the near-continuous alteration of the audio gathered from the environment elicited concern about health risks and about complacency regarding the continuous presence of environmental health risks. Transacting mundane activity such as pedestrian travel while listening to the ToxSampler signal in such an area was reported to be disorienting but informative in an enveloping way.

5.7. Complicity

Participating in a human/computer system that includes a highly-sophisticated, mass-produced electronic device is an inherent engagement and complicity in toxic releases resulting from the creation of the device and emissions from its energy sources. It is important to note that participants can generally be considered complicit as consumers and, to the extent that their political power may be exercised, *willing* as citizens to accept the distribution of risks involved with the creation of products that range from the arguably necessary to the disposable. Although ToxSampler does not include global data for toxic releases and, therefore, releases associated with the manufacture of iOS devices outside of the United

States, the device and its environmental costs as part of the human/computer system must be acknowledged as part of the meaning of the system.

6. FUTURE WORK

The ToxSampler makes use of direct release information and does not incorporate water flow, toxicity risk factors, prevailing wind directions, and other important contributing influences on health factors for a given location. Informing the ToxSampler with the Risk-Screening Environmental Indicators (RSEI) data grid is the next phase of development for this project as this data is more directly related to health factors and includes much analytical expertise dedicated to the public interest.

7. REFERENCES

- [1] *TRI and Superfund Environmental Maps — TOXMAP*. National Institutes of Health. [Online]. Available: <https://toxmap.nlm.nih.gov/toxmap/>
- [2] N. Wiener, “Men, machines, and the world about,” *Medicine and Science*, pp. 13–29., 1954.
- [3] A. De Campo, “A data sonification design space map,” in *Proceedings of the 11th International Conference on Auditory Display (ICAD-2007)*, 06 2007, pp. 342–347.
- [4] M. S. Pierre and M. Droumeva, “Sonifying for public engagement: A context-based model for sonifying air pollution data,” in *Proceedings of the 22nd International Conference on Auditory Display (ICAD-2016)*, 07 2016, pp. 60–64.
- [5] A. Polli, *Toxic Airs: Body, Place, Planet in Historical Perspective.*, A. Johnson and J. R. Fleming, Eds. Pittsburgh, Pa. : University of Pittsburgh Press, 2014.
- [6] C. Paul, *Digital Art*. Thames and Hudson, 2015.
- [7] X. Burrough, *Net Works : Case Studies in Web Art and Design*. New York ; London : Routledge, 2012.
- [8] R. Frieling, *Christina Kubisch*. SFMOMA. [Online]. Available: <https://www.sfmoma.org/publication/soundtracks/christina-kubisch/#cloud>
- [9] M. d. Certeau, *The Practice of Everyday Life*. Berkeley : University of California Press, 1984.

SONIFYING STOCHASTIC WALKS ON BIOMOLECULAR ENERGY LANDSCAPES

Robert E. Arbon^{1*}, Alex J. Jones^{1,2,3,*}, Lars A. Bratholm^{1,4}, Tom Mitchell³, and David R. Glowacki^{1,2}^{*}These authors contributed equally to this manuscript¹School of Chemistry, University of Bristol, Bristol, BS8 1TS, UK²Dept. of Computer Science, University of Bristol, Bristol, BS8 1UB, UK³Dept. of Computer Science and Creative Technologies, University of the West of England, Bristol, BS16 1QY, UK⁴School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK{robert.arbon, alex.j.jones, lars.bratholm, glowacki}@bristol.ac.uk
tom.mitchell@uwe.ac.uk

ABSTRACT

Translating the complex, multi-dimensional data produced by simulations of biomolecules into an intelligible form is a major challenge in computational chemistry and biology. The so-called “free energy landscape” is amongst the most fundamental concepts used by scientists to understand both static and dynamic properties of biomolecular systems. In this paper we use Markov models to design a strategy for mapping features of this landscape to sonic parameters, for use in conjunction with visual display techniques such as structural animations and free energy diagrams. This allows for concurrent visual display of the physical configuration of a biomolecule and auditory display of characteristics of the corresponding free energy landscape. The resulting sonification provides information about the relative free energy features of a given configuration including its stability.

1. INTRODUCTION

1.1. Project context

Richard Feynman famously stated [1] that “everything that living things do can be understood in terms of the jiggings and wiggings of atoms”. A complete understanding of how these atomic jiggings and wiggings give rise to the structure, dynamics and function of biomolecules remains an outstanding scientific challenge with implications across a wide range of disciplines. For example, dynamical processes like protein folding are implicated in neurological diseases (e.g. Alzheimer’s) [2] and conformational changes in enzymes are linked to their biological function [3].

Computer simulation is an important tool to understand biomolecular dynamics because of its ability to reveal chemical information at the atomic level with a high degree of temporal and spatial resolution [4, 5]. Its popularity is associated with three developments: accurate and computationally efficient ways of modeling the interactions between atoms (the atomic ‘force-field’)[6], the increasing availability of highly parallel computer architectures such as general purpose graphical processing units (GP-GPUs)[7], and a variety of user friendly software packages which exploit both these developments [8, 9]. This has enabled the study of bigger systems at longer timescales, moving the dynamics of biomolecular systems into the ‘big-data’ era [10].

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

Extracting scientific information from the output of computer simulations is difficult owing to the quantity of data available. Output from molecular dynamics (MD) simulations [11] include time series of atomic positions (known as *trajectories*) and associated data (e.g., system energy, volume, pressure, etc.). Making sense of this data requires reducing the dimensionality by removing irrelevant features and producing an accurate but understandable model of the process being investigated. Analysing trajectories is commonly performed through visual display using animations of the molecules (often with atoms rendered as balls and chemical bonds as sticks). However, there is typically far too much data for a researcher to understand. Dimensionality reduction is achieved by only displaying certain atoms while features of the data can be calculated and mapped to visual aesthetics. For example, common structural motifs in proteins, such as alpha-helices, can be drawn as a cartoon helix on top of the molecular structure to highlight their presence. A particularly important feature of the system is its *free energy*. Any given molecular configuration has an associated free energy, from which several important properties can be calculated [12]. Often described as a “landscape”, free energy is commonly represented as a topographical contour map. “Mapping” the free energy landscape of biomolecular systems remains a significant challenge. Nevertheless, understanding how a molecule’s 3D structure relates to its free energy landscape is crucial for scientists to gain an understanding of biomolecular dynamics. The software package MolPX [13] has attempted this by linking two separate visual objects - molecular animations and free energy diagrams. However this strategy has two drawbacks: (1) the free energy landscape is limited to two dimensions and (2) the researcher’s focus is split across the two visual objects. Display of higher dimensional landscapes is possible by combinations of two dimensional projections but this only exacerbates problem (2). Sonification has the ability to overcome problems with displaying high dimensional free energy landscapes: the topography and important features of the landscape can be heard concurrently with visual structural information. However, creating a sonic representation of features of the free energy landscape presents a number of technical and design related challenges which are explored in this paper.

1.2. Sonification techniques for molecular data

Two major techniques for approaching an auditory display challenge are *model-based* and *parameter mapping sonification* (PM-Son). Model-based techniques aim to transform a dataset into a dynamic model, which one can interact with and aurally examine [14]. In contrast, PMSon exposes features that describe the

data and maps these to sonic parameters. Previous sonifications of molecular simulation data seemed to have favoured PMSon and *auditory icons/earcons* rather than model-based techniques. This is likely because MD simulations already represent a physical model (see section 2.2) and adapting this dynamical system for the purposes of model-based sonification is challenging.

Rau et. al. [15] demonstrate a PMSon for interrogating features of a static molecule in the *Megamol* [16] visualization framework. Their approach was to create audio representations of features that are known to be chemically interesting, such as the forming and breaking of hydrogen bonds. *Sumo* is a plug-in for the Python based molecular simulation environment *PyMol* [17]. This project had the fairly broad aim of providing a general framework for implementing various sonifications within *PyMol*. A relevant example application created parameterized earcons by mapping the pairwise distances of all atoms of an amino acid onto the parameters of a set of resonant filters. These were offset in both time and space according to their deviation from a reference amino acid. Although no formal studies were undertaken, the designers observed that it was possible to distinguish between them and thus perceive conformational differences in the molecule [18]. Grond et al show a technique for mapping structural information of secondary RNA structures to sonic parameters with a view to aiding browsing and classification; however, there was no attempt to sonically render dynamical walks along free energy surfaces, which is our focus here. [19]. Hermann describes some of the critical issues that arise when designing a PMSon, observing that mappings are not necessarily transparent to a first time listener without some kind of “code book” [20]. This point is reiterated by Wishart when describing the design of his piece, *Supernova*, which sonified astronomical observations: “...there is no particular reason to use one mapping rather than another. As a result, the sonic outcome would be entirely dependent on the mapping chosen.” [21].

1.3. Molecular representation

Representational arbitrariness is a particularly interesting issue when it comes to atomistic and molecular representations, owing to the fact that they are too small to be experienced directly, either visually or aurally. There is an inevitable degree of flexibility when designing representations of imperceptible phenomenon, which is evident in the range of available molecular visualisation approaches, e.g., Pauling’s paper protein helices [22], Kendrew’s metal, wood and plastic protein structures [23], and the increasingly common digital ‘ribbon renderings’.[24]. When it comes to sonification, rendering conventions are far from established. Depicting an atom as a sphere is, in some sense, an arbitrary decision, but makes some sense insofar as both are spatially delimited. Attempting to define such a clearly delimited object in the audio realm is not as straightforward, neither spatially nor compositionally. It is difficult to assert what constitutes a single atomistic object in a piece of sound design.

However, there are a wide range of “non-local” properties important in biomolecular science (e.g. potential energy, free energy, electrostatic energy, temperature, conformational state membership, etc.). Such properties are extremely difficult to visualize using conventional rendering strategies (and even if there were effective strategies, would lead to significant visual congestion) owing to their non-locality. We believe that such properties are the most interesting to explore in a sonification context: hence our focus on free energy in this work.

There is a question of the level of intervention that the sonification designer should take. If a dataset is rendered as directly as possible (i.e. converted to audio), then perhaps any audible features present must be features of the data. But this rule may depend on the source and type of data (plus artifacts of the transform). For example, a set of measurements of how temperature changes over time might be treated differently to a non-local parameter that represents the overall instability of a system. In the latter case, it may be necessary to map the data in a less direct way to convey its provenance. Scaletti [25] categorises the directness of mappings through the idea of different orders: for 0th order, the data is directly read as an audio waveform, for 1st order the data is used to modulate a carrier signal. The approach presented here uses many 1st order, one-to-many [26] mappings that attempt to create a certain perceptual effect related to the significance of a given feature. A problem that may be encountered with this approach is that it is *atheroetical*; the decisions made are based on some subjective sound design process and the results often represent the designer’s sensibilities just as much as the underlying data set [27]. The techniques used in this project are primarily parameter mappings, which certainly do encounter some of the issues raised above. Acknowledging these issues is important although addressing them all in detail is out of scope for this paper.

This work extends the practice of molecular sonification by seeking an auditory display of how a molecule walks along a free energy landscape and its relation to fundamental dynamic processes of biomolecules, something that builds on our previous work developing real-time sonification strategies for molecular dynamics simulations ([28, 29, 30]). In these prior works, we sonified atomistic systems and events including atomic collisions, atomic clustering, and vibrational spectra. This work focuses on the features of molecular systems. A simulation of the biomolecule Alanine dipeptide (AD) is analysed using both hidden and observed Markov models to extract features of the underlying free energy landscape. These features are then mapped to sonic parameters to generate coupled visual and auditory display of structural and dynamic information respectively.

This paper is organized as follows: section 2 explains some of the underlying physical ideas and the modeling of biomolecular dynamics, section 3 explains our sonification strategy, some details of the implementation are given in section 4, and our conclusions and outlook for further work are given in section 6. An example of the sonification described in this paper can be found at <https://vimeo.com/255391814>.

2. BIOMOLECULAR CONFORMATIONAL DYNAMICS

2.1. Free energy landscape

Biomolecules such as proteins and nucleic acids are dynamic objects comprised of n atoms, each of which interacts with other atoms in the same molecule and the cellular environment. A molecular system has $3n$ degrees of freedom: each atom moves in the x , y , and z direction. Typically biomolecules are comprised of thousands of atoms, leading to high-dimensional dynamics.

At any given time, a molecule adopts a particular shape, or *conformational* state. Researchers are typically interested in understanding the networks of conformational states that characterize a particular molecule. Networks of highly connected states in which the system has a relatively long residence time are called *metastable states*. Of particular interest in many applications is

understanding how long it takes a molecule to travel between different metastable states. Conformational states are of interest because they are directly linked to the molecule's function, insight [31] which has been verified extensively through experiments [32] and computational studies [33, 34].

Any given conformational state has an associated *free energy*. Highly probable conformations have a lower free energy than improbable conformations. Rises and falls in free energy as a function of molecular shape (atomic coordinates) define a *free energy landscape*, which is illustrated in further detail in what follows.

2.2. Molecular dynamics simulation

The output of MD simulations are a series of regularly timed snapshots (*frames*) of atomic configurations as the system evolves, called *trajectories*. An animated example trajectory can be viewed at <https://vimeo.com/255526473>. With enough trajectories it is possible to understand the probability that a molecule occupies certain states, and develop a corresponding map of the free energy landscape. The primary challenge in constructing such maps arises from the system's high dimensionality. Researchers are therefore focussing on dimensionality reduction techniques for identifying only those coordinates (or combinations of coordinates) which take the system from one metastable state to another.

2.3. Markov Models

Markov state models have found widespread use in recent years as a dimensionality reduction technique to analyze the metastable dynamics of biomolecules [35]. Their popularity stems from their ability to produce predictive and easy to understand results as well as their ability to parallelize the problem of resolving very long timescale processes. Markov models transform a trajectory into a *chain* of n discrete states. These states are called *observed states* (or sometimes *microstates*) and form the data from which both types of Markov model can be estimated. In general we refer to a chain as x_t and a specific element by its position in the chain: $x_2 = 3$ denotes that the second frame of the chain is in state 3. We refer to the set of all possible discrete states as $\mathbf{x}$. There are two related and widespread approaches: observed Markov state models [36] and hidden Markov models [37]. This work uses both.

An *observed Markov state model* (or simply Markov state model, MSM) assumes the probability of transitioning to observed state b in a time τ given we are in state a , $P(x_{t+\tau} = b | x_t = a)$, only depends on the states a and b and not on the states visited at times $t - 1, t - 2, \dots, 0$. This property is known as the *Markov property* and any chain that satisfies this is known as Markovian. The dynamical information of the MSM is contained within a *transition matrix*, $\mathbf{T}(\tau)$, whose elements are the conditional transition probabilities, i.e. $T(\tau)_{a,b} = P(x_{t+\tau} = b | x_t = a)$.

The primary problem with observed Markov state models is that they contain too much information: the observed states need clustering into a smaller number of metastable states in order to make quantitative predictions about their dynamics. A *hidden Markov model* (HMM) represents a sort of *fuzzy* clustering of the observed states into a set of metastable states (or hidden states, $\mathbf{X}$), i.e. instead of describing a particular conformation (observed state) as unambiguously belonging to a metastable state, a probability of membership is given. A HMM consists of a transition matrix for the metastable states and a membership matrix $\mathbf{M}$ whose elements are the conditional probability of being in

Figure 1: Atomic structure of Alanine dipeptide (AD). The cylinders represent chemical bonds and their intersections represent atoms. Grey, blue, red and white colors are carbon, nitrogen, oxygen and hydrogen atoms respectively. The atoms involved in the ϕ, ψ dihedral angles are labeled and highlighted as spheres. The ϕ angle is formed from the intersection of the planes formed by the atoms (C_1, N, CA) and (N, CA, C_2). The ψ angle is formed from the planes formed by the atoms (N, CA, C_2) and (CA, C_2, O).

a metastable state (A) given a particular observed state (a), i.e. $M_{A,a} = P(X_t = A | x_t = a)$. HMMs work well in describing biomolecular dynamics in the regime where the underlying dynamics are metastable. In other words the proportion of observed states with membership probabilities intermediate between 0 and 1 are small in comparison with the total number of observed states.

2.4. Alanine Dipeptide Model

The “hello world” example of a biomolecule exhibiting metastable dynamics is Alanine dipeptide (AD), as shown in figure 1. The metastable dynamics of AD are reasonably well described with reference to two dihedral angles made by atoms in the peptide bonds [38], the ϕ and ψ angles, also shown figure 1. The free energy landscape of AD projected onto these two dimensions is shown in figure 2A. The light yellow colour denotes free energy wells, i.e. regions with a low value of free energy which define the metastable states. The lighter purple regions are those which are visited only briefly on the way to a metastable well, known as transition regions. As the dihedral angles are periodic, conformations with $\phi/\psi = 180^\circ$ are equal to those with $\phi/\psi = -180^\circ$. This means that rather than a 2D plane, the free energy landscape actually resides on a *torus*, i.e. each edge of the chart should be wrapped around to meet the opposite side. For the sake of simplicity we show it here in the form in which it is typically rendered by practitioners in the field.

For the purposes of this paper two models were created - an observed MSM and a HMM. Details of the data and calculations used to generate the models can be found in section 4. The estimated transition matrix for the MSM results in 500 eigenvectors which describe the various relaxation modes of the dynamics. The first eigenvector $\mathbf{q}^1$ is equal to the stationary distribution, $\mu(\mathbf{x})$. The next three eigenvectors are slow relaxation modes ($\mathbf{q}^{2,3,4}$) which define population transfer *between* metastable states. The next five eigenvectors ($\mathbf{q}^{5-9}$) are fast relaxation modes which define population transfer *within* metastable states. Each relaxation mode has an associated timescale (the corresponding eigenvalue). The remaining eigenvectors were discarded as the associated timescales for these modes was faster than the time resolution of the data used to estimate the model and so were not considered statistically robust. The HMM was estimated by assuming four metastable states. The results are shown in figure 2. Figure 2D shows the metastable state transition matrix, $\mathbf{T}$. Each circle repre-

Figure 2: The hidden Markov model for AD. **A**: The free energy landscape of AD projected onto the peptide ϕ and ψ dihedral angles. The lighter the colour the lower the free energy and hence the more stable those atomic configurations. The yellow regions define four metastable basins centered on coordinates $(60^\circ, 60^\circ)$, $(60^\circ, -120^\circ)$, $(-120^\circ, -60^\circ)$, $(-120^\circ, 120^\circ)$. The purple regions are not visited during the simulations used as input for the model. The black circles are the centers of 500 observed states for both the observed and hidden Markov models. Each frame of the trajectory is assigned to the nearest observed state. **B1-4**: Each panel represents rows 1 - 4 of the membership matrix. Each circle represents an observed state coloured according to its membership probability to each metastable state. **C**: Ten sampled conformations of each metastable state of AD overlaid over the center of their free energy well. The hydrogen atoms have been removed for clarity. **D**: The transition matrix and stationary distribution of metastable states. Each circle represents a metastable state with the area of each circle proportional to the stationary distribution. The arrows between each state show possible transitions with the width proportional to the conditional transition probability (also shown in the white boxes on each arrow).

sents a metastable state with the area of the circle proportional to its stationary distribution, $\mu(X_{1,\dots,4})$. The arrows show the conditional probability of transitioning to each state. For example, the probability of transitioning from state 1 to state 4 is 6.79%. State 4 is by far the most stable, followed by 3, 2 and then 1. As there are no transition regions between state 1 and 3 and between 2 and 4, the probability of transitions between these pairs of states is zero. Figure 2C shows an overlay of ten characteristic structures for each metastable state, overlaid over their respective free energy wells. Figures 2B1-4 show the rows of the membership matrix. Each circle represents one of the observed states, coloured according to the membership probability to each metastable state. The partitioning of the basins is clearly shown by the regions of black circles (high probability of metastable membership) vs. white circles (low probability of metastable membership).

3. SONIFICATION

The sonification is designed to convey features of the free energy landscape concurrently alongside visual display of structural information, i.e., a molecular animation from an example trajectory. It therefore requires three objects: a model of AD dynamics, an example trajectory and an animation of the example trajectory. In this work the data used to estimate the model of AD (the input trajectories) are different to the example trajectory although in principle trajectories from the input data could be used as example trajectories. The following section provides a description of the features derived from the data followed by a description of the components of the sonification and how features of the data are mapped to synthesis parameters.

3.1. Free Energy Landscape Features

The features of the data are split into two categories: *static* and *dynamic*. Static features are derived from the properties of the AD

Type	Scope	Parameter	Layer
Dynamic	Intra	Free Energy ($F(a)$)	B
		Fast Mode ($[q^i]_a$)	C
	Inter	Shannon Entropy (H_a)	A
Static		Membership probability ($[M]_a$)	A
		Well min ($UL[h]$)	A
		Well max ($UB[h]$)	A
		Histogram area ($A[h]$)	A

Table 1: Simulation features exposed via OSC. **Type** refers to whether the features are fixed at time 0 (static) or change at each trajectory frame (dynamic). **Scope** refers to whether the features relate to transitions between states (inter) or within a state (intra).

molecule and do not change after they are initialised at time 0. Dynamic features are those derived from each frame (i.e. from each observed state) of the example trajectory as well as the model features. As well as categorizing the features as static or dynamic, a further distinction is drawn between those that relate to changes within the current metastable state (intrastate) and those that pertain to changes between the metastable states (interstate). Table 1 summarizes these two categorizations. This second classification is useful because it draws the distinction between features based on characteristics of the physical dynamics rather than how they were generated. The interstate features represent the most physically important characteristics of the dynamics and free energy landscape and so form the core of the mapping strategy.

Firstly, we discuss static features: these are derived from the shape of the free energy wells associated with each metastable state. Each part of the free energy landscape is assigned a probability of membership to a metastable state so the limits of each well are not well defined. In order to overcome this problem each observed state is assigned to the metastable state for which it has

the highest probability of membership. The three static features for each metastable state, $UB[h]$, $LB[h]$, $A[h]$ are interstate features. They are principally related to two physical characteristics: the relative stability and conformational flexibility of each metastable state. The focus of the mapping strategy is to find a way to describe them aurally. The free energy for each observed state a , $F(a) = -kT \ln(\mu([x]_a))$ was calculated and scaled to lie in the range (0, 1). Here k refers to the *Boltzmann* constant and T to the temperature. The static features for each metastable state were derived from a histogram of the free energies of observed states assigned to that metastable state. We denote the histogram for metastable state A as $h_A(F)$ (or h in general). The properties we derive from $h(F)$ are (1) its upper bound, $UB[h]$, (2) its lower bound $LB[h]$ (3) its area $A[h] = \int_0^1 dF h(F)$. The histograms for states 1 (blue) and 4 (pink) are shown in figure 5. The model has four metastable states meaning there are $4 \times 3 = 12$ static features. The upper and lower bounds are related to the free energy well minima and maxima for each metastable state. The area of the histograms is proportional to the overall volume of the free energy well.

Secondly, we discuss dynamic features: these change with each observed state in the example trajectory. For each observed state they are: (1) its *probability of membership* to each of the four metastable states, (2) the *Shannon entropy* of its metastable state assignments (3) its absolute *free energy*, (4) its *projection* into the five fast relaxation modes. For each observed state there are $4 + 1 + 1 + 5 = 11$ dynamic features. The membership probability describes the probability that a given observed state can be assigned to a given metastable state. For observed state a there are four membership probabilities given by $([M]_{1,a}, [M]_{2,a}, [M]_{3,a}, [M]_{4,a})$. The information or Shannon entropy is a measure of the degree of certainty with which the assignment of an observed state to a particular metastable state can be made. The Shannon entropy for an observed state a , H_a is given by $H_a = -\sum_{i=1}^4 [M]_{i,a} \ln([M]_{i,a})$. $H_a = 0$ indicates the observed state is definitely in one metastable state. $H_a = \ln(4) \approx 1.4$ indicates it is equally likely to be in any of the four metastable states. The free energy of observed state a , $F(a)$ measures the observed state's global stability. These were the same free energies used in the calculation of the static features in section 3.1. The projection of observed state a onto the i 'th fast relaxation modes is given by $[q^i]_a$. Large oscillations in a projection indicate that the system is relaxing along that mode. The values of these projections were scaled to lie in the range $(-1, 1)$. Figure 3 shows the dynamic features for a section of an example trajectory.

3.2. Sonification Layers

As the molecule traverses different regions of the free energy landscape the sonification conveys information about its properties with the following layers in the audio stream: (A) a continuous pad sound representing membership of different metastable states and their properties, (B) a pulse sound representing the stability of the system and (C) a set of synthesized tones that represent how the system is changing within each metastable state.

The **Pad Sound (A)** was constructed as follows:

1 Representation: The pad sound represents features of the well in the free energy landscape corresponding to the current metastable state assignment.

Figure 3: The dynamic features of the model. Each panel shows how a model features varies over 0.1ns (100 frames) of the example trajectory. **State 1-4** show the membership probability of each frame's observed state to each metastable state. The values in each frame across all four panels sum to 1. **Shannon entropy** measures the uncertainty of the assignment of each observed state to the metastable states. High entropy indicates an observed state could be considered to belong to more than one metastable state. **Free energy** is the free energy of each observed state, the lower the free energy the more stable the state. **Mode 5** is the first fast relaxation mode which redistributes population within a metastable state. The other four fast modes are not shown.

2 Synthesis: Created using a subtractive, polyphonic synthesiser with noise as a source signal. The basic signal flow of a single voice is shown in Figure 4. The main parameters are the Q value of the resonant filter, the depth and rate of frequency modulations for each voice and the amplitude of each voice.

3 Mapping: Each metastable state is characterised by a set of notes referred to as a note cluster. This design arises from the desire to represent metastable state transitions by tonal groupings, a musical device that non-trained listeners should be able to perceive. The maximum note range is predefined to three octaves. The relative values of $A[h]$ for each metastable state defines the number of notes in its corresponding note cluster, these notes are then evenly distributed between the lowest and highest notes. The values of $LB[h]$ and $UB[h]$ define the lowest and highest notes of each cluster respectively. This is shown in figure 5 for metastable states 1 and 4. State 4 (pink) has a smaller range, but a larger area resulting in a dense, tightly spaced note cluster at the lower extent of the note range. State 1 (blue) has a large range but small area resulting in a more sparse note cluster. State 4 has its lower bound below that of state 1 and so

Figure 4: Single synthesis voice as used for each note of a metastable state note cluster. The set of multiple instances (one per note) form the pad sound. The inputs are the frequency of each note in the note cluster and the entropy. The note cluster is determined by the membership probability values at the current frame.

the lowest note of the cluster is below that of state 1. An accepted limitation of this mapping is that it does not attempt to classify the clusters in terms of their harmonic relationships. Instead, it deals with them as distributions of notes within a range, with a given extent and density. It may be possible to create a hierarchy of tonal groupings from which to choose but this is highly genre specific. This is an outstanding issue in that listeners may interpret the harmonic relationship of two note clusters as significant when this is not intended as part of the mapping (e.g. stacked whole tones vs. stacked fourths). The membership probabilities (M) control the choice of note cluster by linearly interpolating between the values defined by the static parameters of each state. This means that if $[M]_{1,a} = 1$ then the cluster defined by the static properties of metastable state 1 will be used. In the case that $[M]_{1,a} = [M]_{2,a} = 0.5$ then the lowest and highest notes will lie halfway between those defined by the static parameters of state 1 and 2. Large values of the Shannon entropy (H) represent observed states which could be assigned to more than one metastable state, physically this means they are in transition regions between two metastable states. H is mapped to the width and rate of frequency modulations of the voices for the note clusters as well as the bandwidth of the filter such that it tends toward noise for higher values of H . This is a one-to-many mapping that is designed to create a perceptual effect of instability. H tends to remain at 0 with occasional spikes as the system transitions. This is shown in figure 3 where there are spikes between 80.05 and 80.06 nanoseconds as transitions occur between all four states. These spikes, although fleeting, are important features. To emphasise transitions an asymmetric smoothing function is employed such that H values increase quickly but decrease slowly.

The Pulse sound (B) was constructed as follows:

1 Representation: The overall stability of the current observed state in relation to the global minimum. The pulse sound is designed to underpin the entire sonification. The rationale here is that a rhythmic kick drum sound is commonly used as a rhythmic marker upon which all other elements are constructed. This mirrors the way in which the free energy landscape underpins the dynamics of the physical system.

Figure 5: The mapping of static properties of the metastable states to note clusters. Each of the observed states was first assigned to exactly one metastable state. The bottom chart shows the distribution of free energies of observed states which have been assigned to the metastable states 1 (blue) and 4 (pink). States 2 and 3 are not shown for clarity. The coloured notes of the keyboard are the note clusters used to represent the metastable states. The relative upper and lower bounds of the distributions determine the highest and lowest notes in the cluster (as shown by the vertical connectors). The relative area of each distribution determines the number of notes in the note cluster. The ratio of the areas of state 4's distribution to state 1's distribution is approximately 3:1. This determines the 9:3 ratio of the number of notes in each note cluster.

- 2 Synthesis:** Generated by sweeping the frequency of a sine wave from approximately 130Hz to 20Hz over a short period (100-550ms) and passing it through a hard clipping function and stereo delay line. The relevant parameters are the rate at which the sweep is triggered, the length of time the sweep takes, the amount of drive in the hard clipping function and the amount of feedback in the delay path.
- 3 Mapping:** The synthesis parameters are mapped to the absolute free energy $F(x_t)$. The variable is inverted and filtered logarithmically. The filter settings are slightly different for each synthesis parameter but, in general terms, act to allow fast increases in its values and slow decreases to ensure the temporal alignment between visual and sonic representations. The $(F(x_t))$ of each observed state measures its stability relative to the global free energy minimum. This feature allows the sonification to draw a distinction between an observed state being globally unstable and yet part of a relatively stable metastable state (or vice versa).

Within-Metastable State Tones (C) were constructed as follows:

- 1 Representation:** The extent to which the system is relaxing along one of the fast modes. This gives an impression of high frequency movements of the molecule that are not represented by the membership probabilities.
- 2 Synthesis:** The fast mode projections are oscillatory signals between -1 and 1 and are exported as PCM wav files (a 0th order mapping in Scaletti's terminology [25]). At the 20Hz frame-rate being used for the animation the oscillations are subsonic. *Scanned synthesis* is used in order to render the content as audible [39]. The relevant parameter is the audio file that is used as the wavetable to be scanned. Other parameters that affect the sound are the window size and the scanning frequency.

3 Mapping: Scanned synthesis was developed to allow for the direct manipulation of synthesis timbre using a physical model and represents an extension of wavetable synthesis. In this case it allows for a rolling window of the audio buffer to be scanned at a given frequency, with the effect that increased frequency and amplitude of oscillation in the window results in a brighter timbre. These can be heard panned from hard left (mode q^5) to hard right (mode q^9). The scanning frequency is defined by the lowest note of the set of note clusters (effectively the root of the most stable state).

4. IMPLEMENTATION

The simulation data was taken from a publicly available repository which accompanies the paper [38] found at <https://simtk.org/projects/alanine-dipeptide/>. A full explanation of the methods used to generate the data are given in the paper. All modelling was done using the Python 3.5 programming language, the Markov models were generated by PyEMMA 2.4 [40]. Both the HMM and MSM used the same set of 500 observed states and each were estimated with a lag time of $\tau = 1.0$ ps. The HMM was estimated by specifying four metastable states. This number was chosen so that the approximation of metastability was most accurate at the lag time used. The molecular animations were generated using VMD [24]. A 500ns long example trajectory was used in the sonification. This was generated from the HMM, rather than using an input trajectory as they were all of insufficient length to sample each metastable state regularly. While this trajectory does not strictly obey the original equations of motion used to generate the input data it reproduces all the modelled features (transition probabilities, relaxation time-scales, stationary distribution) and so is indistinguishable from a trajectory generated using the original equations of motion. A Python script, using the package OSC 1.6.4, was used to create a client which sent the static and dynamic parameters as messages to the audio processing software. The messages were sent using the OSC protocol. The dynamic parameters were sent at a rate of 20 trajectory frames per second (corresponding to a ratio of 20ps to 1s simulation to physical time). The audio processing was implemented in Max/MSP.

Everything required to reproduce this work (except the input data which can be downloaded separately) can be found at https://osf.io/rzp3k/?view_only=b5802dfce6da4dd59dfb6b406ae033f0.

5. EXAMPLE SONIFICATION

The example provided (link in sec. 1) shows an animation of an excerpt of trajectory data for Alanine Dipeptide accompanied by our sonification. This section provides a description of the audible features and how they relate to the mapping strategy. The pad layer (A) provides information about the structure of the free energy landscape for the configuration at a given frame; it can be heard ramping between the note clusters defined by the static features of the four metastable states. It stays primarily in state 4, as this is the most stable. At 00:13 the animation shows a significant change in the dihedral angles; this is linked to a change in the metastable state and a shift to a different well on the free energy landscape, which is depicted by the change in note cluster of the pad sound. The noise/wind effect that can be heard at 00:10 is caused by an increase in the Shannon entropy and gives an indication that there is some uncertainty in the assignment of

the current observed state. The pulse/kick layer (B) is correlated to the absolute free energy; at 00:10, the current metastable state has a lower absolute free energy and the rate, level of distortion and delay feedback are increased resulting in the sound becoming more prominent. The layer representing the fast relaxation modes (C) can be heard as the metallic, buzzing sounds panned across the sound-field. The scanned synthesis method means that amplitude and frequency of activity in the data is correlated to density of spectra and amplitude of the synthesiser. So an increase in brightness of this sound on the left of the sound-field indicates that the molecule is relaxing along the first of the five modes used.

6. CONCLUSIONS AND FUTURE WORK

We have presented a strategy for representing features of the dynamics and the free energy landscape of Alanine Dipeptide with sound. This auditory display can be used in tandem with visual display techniques to help build an understanding of how the physical structure of AD relates to the underlying free energy landscape and the resulting dynamic processes. Future user studies with domain experts will allow an assessment of the efficacy of the sonification; a sketch for this would entail participants viewing an animation of a simulation trajectory with and without the various layers of the sonification to assess if and how they show an increased appreciation of the free energy dynamics. In the future we hope to extend this implementation to allow a degree of interactivity in manipulating the example trajectory. Initially, this would take the form of allowing the user to manipulate the playback position, speed and loop points of the trajectory. This would allow them to focus on regions of interest in the dynamics. In the long term, allowing the user to manipulate the example trajectory using interactive molecular dynamics (for example, NanoSimbox [41]) and hear the resultant sonic effects is an exciting prospect and opens up the possibility of using the system as an instrument for musical expression as well as data exploration.

7. ACKNOWLEDGMENT

David Glowacki acknowledges funding from the Royal Society (UF120381) and the Leverhulme Trust. Robert Arbon is funded by the Royal Society (RG130510). Alex Jones is supported by a studentship from the EPSRC and Interactive Scientific Ltd. Lars Bratholm acknowledges EPSRC grant EP/P021123/1.

8. REFERENCES

- [1] R. P. Feynman, *et al.*, *The Feynman lectures on physics*. Addison-Wesley, 1963.
- [2] M. Hashimoto, *et al.*, "Role of Protein Aggregation in Mitochondrial Dysfunction and Neurodegeneration in Alzheimer's and Parkinson's Diseases," *NeuroMolecular Medicine*, vol. 4, pp. 21–35, 2003.
- [3] P. Singh, *et al.*, "Linking Protein Motion to Enzyme Catalysis," *Molecules*, vol. 20, no. 1, pp. 1192–1209, 2015.
- [4] S. A. Adcock and J. A. McCammon, "Molecular dynamics: Survey of methods for simulating the activity of proteins," *Chemical Reviews*, vol. 106, no. 5, pp. 1589–1615, 2006.
- [5] D. Antoniou, *et al.*, "Computational and Theoretical Methods to Explore the Relation between Enzyme Dynamics and Catalysis," *Chemical Reviews*, vol. 106, no. 8, p. 3170, 2006.

- [6] M. W. van der Kamp, *et al.*, “Biomolecular simulation and modelling: status, progress and prospects,” *J. Royal Society Interface*, vol. 5, no. Suppl 3, p. 173, 2008.
- [7] J. E. Stone, *et al.*, “Gpu-accelerated molecular modeling coming of age,” *J. Molecular Graphics and Modelling*, vol. 29, no. 2, p. 116, 2010.
- [8] P. Eastman, *et al.*, “Openmm 4: A reusable, extensible, hardware independent library for high performance molecular simulation,” *J. Chemical Theory and Computation*, vol. 9, no. 1, p. 461, 2013.
- [9] D. Case, *et al.*, “Amber 2017,” University of California, San Francisco, 2017.
- [10] M. Álvarez-Moreno, *et al.*, “Managing the computational chemistry big data problem: The iochem-bd platform,” *J. Chemical Information and Modeling*, vol. 55, p. 95, 2015.
- [11] B. J. Alder and T. E. Wainwright, “Studies in molecular dynamics. i. general method,” *J. Chemical Physics*, vol. 31, no. 2, p. 459, 1959.
- [12] P. W. Atkins and J. de Paula., *Atkins’ physical chemistry*, 10th ed., 2014.
- [13] G. Prez-Hernandez, *et al.*, “markovmodel/molPX: More PyEMMA integration through handling of TICA objects,” Mar. 2017. [Online]. Available: <https://doi.org/10.5281/zenodo.437707>
- [14] T. Hermann and H. Ritter, “Listen to your Data: Model-Based Sonification for Data Analysis,” in *Advances in intelligent computing and multimedia systems*. Inst. for Advanced Studies in System research and cybernetics, 1999, pp. 189–194.
- [15] B. Rau, *et al.*, “Enhancing visualization of molecular simulations using sonification,” in *1st International Workshop on Virtual and Augmented Reality for Molecular Science*. IEEE, March 2016, pp. 25–30.
- [16] S. Grottel, *et al.*, “MegaMol-A prototyping framework for particle-based visualization,” *IEEE Transactions on Visualization and Graphics*, vol. 21, pp. 201–214, February 2015.
- [17] “Pymol,” 2017. [Online]. Available: <https://pymol.org/2/>
- [18] F. Grond and F. Dall Antonia, “Sumo. A Sonification Utility for Molecules,” in *Proceedings of ICAD 2008*, no. 14. International Community for Auditory Display, 2008, pp. 1–7.
- [19] F. Grond, *et al.*, “Browsing RNA Structures by Interactive Sonification,” in *ISon, 3rd Interactive Sonification Workshop*, Stockholm, 2010, pp. 11–16.
- [20] T. Hermann, *et al.*, “Sonification of markov chain monte carlo simulations,” in *ICAD 2001*. Espoo, Finland: Georgia Institute of Technology, 2001, pp. 208–216.
- [21] T. Wishart, “From Sound Morphing to the Synthesis of Starlight. Musical experiences with the Phase Vocoder over 25 years,” *Musica/Tecnologia*, vol. 7, 2013.
- [22] L. Pauling, *et al.*, “The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain,” *P. National Academy of Sciences*, vol. 37, no. 4, p. 205, 1951.
- [23] J. C. Kendrew, *et al.*, “A three-dimensional model of the myoglobin molecule obtained by x-ray analysis,” *Nature*, vol. 181, p. 662, 03 1958.
- [24] W. Humphrey, *et al.*, “VMD – Visual Molecular Dynamics,” *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 1996.
- [25] C. Scaletti, “Sound synthesis algorithms for auditory data representations,” in *Auditory Display: Sonification, Audification, And Auditory Interfaces*, G. Kramer, Ed. Reading: Addison-Wesley, 1994.
- [26] A. Hunt, *et al.*, “The Importance of Parameter Mapping in Electronic Instrument Design,” *Journal of New Music Research*, vol. 32, no. 4, pp. 429–440, 2003.
- [27] E. Murphy, *et al.*, “A Semiotic Approach to the Design of Non-speech Sounds,” in *Haptic and Audio Interaction Design*, D. McGookin and S. Brewster, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 121–132.
- [28] D. R. Glowacki, *et al.*, “A gpu-accelerated immersive audio-visual framework for interaction with molecular dynamics using consumer depth sensors,” *Faraday Discuss.*, vol. 169, pp. 63–87, 2014.
- [29] T. Mitchell, *et al.*, “danceroom spectroscopy: At the frontiers of physics, performance, interactive art and technology,” *Leonardo*, vol. 49, no. 2, pp. 138–147, 2016.
- [30] D. Glowacki, “Sculpting molecular dynamics in real-time using human energy fields,” in *Molecular Aesthetics*, P. Weibel and L. Fruk, Eds. Cambridge, MA: The MIT Press, 2013, pp. 248–257.
- [31] H. Frauenfelder, *et al.*, “The energy landscapes and motions of proteins,” *Science*, vol. 254, no. 5038, p. 1598, 1991.
- [32] Y. Santoso, *et al.*, “Conformational transitions in dna polymerase i revealed by single-molecule fret,” *P. National Academy of Sciences*, vol. 107, no. 2, p. 715, 2010.
- [33] D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*, ser. Cambridge Molecular Science. Cambridge University Press, 2004.
- [34] N.-V. Buchete and G. Hummer, “Coarse master equations for peptide folding dynamics,” *J. Physical Chemistry B*, vol. 112, no. 19, p. 6057, 2008, PMID: 18232681.
- [35] J. D. Chodera and F. Noé, “Markov state models of biomolecular conformational dynamics,” *Current Opinion in Structural Biology*, vol. 25, pp. 135 – 144, 2014.
- [36] J.-H. Prinz, *et al.*, “Markov models of molecular kinetics: Generation and validation,” *The Journal of chemical physics*, vol. 134, no. 17, p. 174105, 2011.
- [37] F. Noé, *et al.*, “Projected and hidden markov models for calculating kinetics and metastable states of complex molecules,” *The Journal of chemical physics*, vol. 139, no. 18, p. 184114, 2013.
- [38] J. D. Chodera, *et al.*, “Automatic discovery of metastable states for the construction of markov models of macromolecular conformational dynamics,” *J. Chemical Physics*, vol. 126, no. 15, p. 155101, 2007.
- [39] B. Verplank and R. Shaw, “Scanned Synthesis,” *The Journal of the Acoustical Society of America*, vol. 109, no. 5, 2001.
- [40] M. K. Scherer, *et al.*, “PyEMMA 2: A Software Package for Estimation, Validation, & Analysis of Markov Models,” *J. Chemical Theory & Computation*, vol. 11, p. 5525, 2015.
- [41] M. O. Connor, *et al.*, “Sampling molecular conformations and dynamics in a multi-user virtual reality framework,” *ArXiv e-prints*, Jan. 2018, ArXiv:1801.02884.

EXTENDED ABSTRACTS

REAL-TIME EMG SONIFICATION SYSTEM FOR GAIT

¹Ryosuke Nakagawa, ²Hideki Kadone, ³Masaki Matsubara, ⁴Kenji Suzuki, and ⁵Hiroko Terasawa

¹Graduate school of Library Information and Media Studies

^{3,5} Faculty of Library Information and Media Science

²Faculty of Medicine, ⁴Faculty of Engineering, Information, and Systems

University of Tsukuba

1-2 Kasuga, Tsukuba, Ibaraki, Japan

¹ryosuke@slis.tsukuba.ac.jp, ²kadone@ccr.tsukuba.ac.jp, ³masaki@slis.tsukuba.ac.jp

⁴kenji@ieee.org, ⁵terasawa@slis.tsukuba.ac.jp

ABSTRACT

We developed a system that sonifies muscle activity during walking in real time. We aim at improving the efficiency and quality of walking rehabilitation by exhibiting the coordination of multiple muscles during gait. Our initial goal is to sonify the coordination of multiple muscles in real time, and we conducted two pilot tests: (1) gait experiment to monitor the gait pattern change with auditory biofeedback, and (2) listening test to verify if the individual gait patterns are successfully sonified. Three healthy adults participated in the gait experiment. They did not show any significant change in walking speed and stride length depending in the biofeedback condition. The listening test showed that listeners can distinguish individual patterns of muscle activities during gait with the sonification at a statistically significant level.

1. INTRODUCTION

Walking is one of the most familiar movements to us, often conducted unconsciously. However, people with walking disabilities due to illness may walk slowly, with small or awkward steps. When walking rehabilitation programs with physical therapists are chosen, the therapists help them improve their gait through observation, palpation, and verbal feedback, etc., in order to accomplish appropriate muscle coordination during walking. Yet, concurrent observation of multiple muscles is uneasy. Therefore, we constructed a real-time system to sonify multichannel electromyographic (EMG) signals that facilitates the observation of the coordination of multiple muscles. We are particularly interested (1) if the presence of the auditory biofeedback would change the gait patterns, and (2) if the individual patterns in the muscular coordination can be sonified.

2. GAIT AND MUSCLE COORDINATION

Human gait is conducted in cycles: A gait cycle is one cycle from the grounding of the foot on one side of the body to its next grounding. The gait cycle is divided into the stance phase (i.e., the period when the foot is on the ground) and the swing phase (i.e., the period when the foot is away from the ground and carried forward).

Figure 1: Muscles of the lower limbs.

The lower limbs have three representative joints: the hip joint, the knee joint, and the ankle joint. The extension (i.e., dorsiflexion) and flexion (i.e., plantarflexion) of these joints involves the activities of four major muscles as described in Figure 1: The gluteus maximus (GM) extends the hip joint; The rectus femoris (RF), which is one of quadriceps femoris muscles, flexes the hip joint and extends the knee joint; The gastrocnemius (GA) plantarflexes the ankle joint; And the tibialis anterior (TA) dorsiflexes the ankle joint.

3. RELATED RESEARCH

Much research has been conducted on rehabilitation using EMG biofeedback. Intiso et al. used EMG biofeedback in a group of patients to improve foot-drop after stroke, who showed recovery of foot-drop in the swing phase [1]. Morris et al. applied EMG biofeedback combined with contemporary physical therapy to stroke patients with genu recurvatum and determined that the combination enhanced the effectiveness of the treatment [2]. Matsubara et al. implemented auditory feedback that helped to understand the movement of multiple muscles [3]. Yang et al. conducted a study to improve the quality of exercise by giving auditory feedback to the exercise

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

Figure 2: System overview.

[4, 5]. In addition, many studies have been made to sonify EMG signals [6, 7, 8].

As for the other methods of auditory biofeedback rehabilitation than EMG, Aruin et al. conducted an experiment on patients with narrow width between their legs and showed that patients who used audio feedback were able to widen the width, compared with those who did not receive audio feedback [9]. One study showed sufficient rehabilitation of the ankle joint performed on blind persons using auditory feedback [10]. Claude et al. reported feedback of joint motion with sound may be a substitute for an ability to sense positions, locations, orientations, and movements of the body and its parts [11].

4. DEVELOPMENT OF REAL-TIME SYSTEM

4.1. Devices

Our sonification system consists of the following devices: a wireless EMG sensor kit made by Delsys (Natick, MA, USA), two laptop PCs (Windows and Mac), and two loudspeakers. Figure 2 shows the system overview.

The Windows PC receives EMG signals via Bluetooth, conducts data-preprocessing, and sends them to the Mac, which is connected with a LAN cable. The Mac normalizes the signals and converts them to sound. Two loudspeakers are used to play back the sonified sound to the participants.

During system development, we tried to use Bluetooth earphones instead of loudspeakers. However, the delay time in the transmission of sound from the Mac to the Bluetooth earphones was too long (approximately 1 second), and we decided that Bluetooth earphones were not suitable for the real-time sonification system.

4.2. Data pre-processing, normalization, and synthesis

In the preprocessing, the original EMG signal at 2000 Hz sampling rate is processed by windowing with length of 66 samples. Within a window, we first calculate the absolute values of the signal, then a mean value to represent the window. (Note this is essentially same as taking the RMS value and moving average, which is the standard way to process EMG signals.) The pre-processed EMG signals are sent via the LAN cable using UDP protocol to the Mac.

MATLAB on the Mac normalizes the received data. A walking exercise was measured beforehand to calculate the appropriate thresholds within a channel. The following equations were used for calibration:

$$y_i = \max_{10i \leq t \leq 10i+100} x(t) \quad (1)$$

$$z_i = \min_{10i \leq t \leq 10i+100} x(t) \quad (2)$$

$$Y = \text{median } y_i \quad (3)$$

$$Z = \text{median } z_i \quad (4)$$

where $x(t)$ is the pre-processed EMG signal at t and i denotes the index number for the window. This normalization uses the median of the maximum and minimum values of muscle activities observed in the sliding windows. Typically, maximum muscle activity is used to normalize the EMG signal. However, in this walking exercise, the maximum muscle activity appears as a sharp spike in a gait cycle. In order to sonify the state between spikes, it is desirable to set the upper threshold to a value that is observed more frequently. In addition, owing to the ground noise level, it is not desirable to set the lower threshold to 0. If we want to hear the states between sharp spikes, normalization should be based on a threshold calculated from middle-values but not from maximum values.

The thresholds were calculated for each person with the EMG data of 5 seconds of steady gait, excluding the start and end, where the walking speed and stride were unstable. In this calibration stage, the maximum and minimum values are detected in the window of 100 samples for each channel with a shifting-step width of 10 samples. The values in the window are stored in arrays. The median of the values of the two arrays was taken as the threshold. In the normalization, the values between the thresholds are linearly normalized. The values larger than threshold (Y) were rounded to 1, and the values smaller than threshold (Z) were rounded to 0, resulting in the normalized EMG signals ranging from 0 to 1.

The normalized EMG signals are then sent to SuperCollider, a sound synthesis software. It maps each channel of EMG to a pitch in a harmony, and the amplitude of the normalized signal of the channel is mapped to the volume of the tone, as described in the next section. Synthesized sound is played with the two speakers.

5. SONIFICATION DESIGN

We aimed to produce a sound design that can distinguish muscle coordination patterns based on the overtone-mapping technique [12]. Each channel of EMG signals is mapped to the amplitude of a sinusoid, and the pitches of multiple sinusoids comprise a harmony. When a certain muscle is used, the

Figure 3. Relative size of EMG during gait cycle.

amplitude of the corresponding sine wave increases, making that tone louder than the other in the harmony. When multiple muscles are active in a synchronized manner, the corresponding tones in a harmony overlap and comprise a harmonic sound object.

In our implementation, we sonified the activities of four muscles (GM, RF, TA, and GA) on the right side of the body. Table 1 shows the mapping between the muscles and the pitch. During a gait cycle, a coordinated transition of the muscle activity across the four muscles exists, resulting in a rhythmic pattern in the harmony. At the beginning and end of the gait cycle, GM, RF, and TA are active. GA becomes active at 40% to 60% of the gait cycle. Figure 3 shows the activities of the muscles during the gait cycle. These envelopes of muscle activity become the envelope of the harmonic sounds. We only sonified four channels because adding another channel can lead to difficulty in comprehending the feedback and would require more practice to the users.

Table 1: Mapping between muscle and sound's frequency.

Channel	Muscle	Note	Frequency (Hz)
1	GA	C5	523
2	TA	G4	392
3	RF	E4	330
4	GM	C4	262

6. GAIT EXPERIMENT

We conducted a gait experiment to monitor if the presence of sonification changes the gait behavior of healthy people.

6.1. Condition

Three participants (two men and one woman; aged 23-24 years) participated in the gait experiment. It was conducted in the measurement room of the Center for Innovative Medicine and Engineering in the University of Tsukuba Hospital, with the approval by university IRB. Twelve wireless EMG sensors were attached to the lower limbs of the participants. Four of them were used for sonification, and the remaining channels were recorded for analytic purposes. We also used a motion capture system and ground sensors to measure the walking speed, stride, and heel grounding timing.

Figure 4: Experimental setup.

6.2. Method

In the experiment, they walked approximately 10 m under two conditions: sound and no-sound conditions. Figure 4 shows the experimental setup.

Under the no-sound condition, the system provides no sound and the participants only had to walk. They walked three times under this condition. In the sound condition, four channels (right GM, TA, GA, and QF) of the lower limb muscles are sonified in real time. Participants walked ten times while listening to the sound, which is played from the speakers.

In both conditions, participants were asked to walk normally, without paying particular attention to the sounds.

6.3. Results

Figures 5 and 6 show the average walking speed and stride length, respectively, of the three participants for each condition. There were no significant differences in walking speed and stride length between the two conditions.

Figure 5: Average walking speed under sound and no-sound conditions.

Figure 6: Average stride length under sound and no-sound conditions.

7. LISTENING TEST

We conducted a listening test to verify whether individual walking patterns can be identified from gait sonification.

7.1. Method

An online listening test using Google Form was conducted. Ten participants took the test. They first listened to three sample audio files. The three files are sonification of three walkers from the gait experiment. Then the participants listened to nine audio files, three sonifications for the three walkers randomly presented. They were asked to identify the walker by listening to the sonification. Participants can listen to the three sample audio files as many times as they want.

7.2. Results

The results of the listening test are shown in Table 2. The numbers in each element shows the distribution of the correct and incorrect responses to the identification task. Figure 7 shows the correct answer rates of each respondent (listener). The overall correct answer rate was 0.67. When χ^2 test was conducted with the correct answers of each walker and the expected number from the chance level, they were significantly different ($p = 0.00001$).

Table 2: Result of the listening test.

	Walker ID	Source data of sonification		
		1	2	3
Listener's response in the identification task	1	0.67	0.07	0.33
	2	0.20	0.77	0.10
	3	0.13	0.17	0.57

8. DISCUSSION

The present system did not change walking speed or step speed, but the sound feedback was considered not to interfere with their walking.

One participant's hip joint extension and knee joint extension increased under the sound condition. The walking distance in this test was as short as 10 m in one trial. The trial most likely ended before the influence of the sound appears. We did not ask to pay attention to the sonification either. A longer walking distance in one trial, and more specific instruction may cause some changes on the gait pattern.

Based on the results of the listening test, individual gait patterns can be identified from the sound. Some listening test respondents had difficulties in identifying the walker, but they can recognize the difference in sounds. From the post-test interview, we determined that the speed of the attack and the timing of the sound appearance within one cycle were the clues for the identification.

9. CONCLUSION

This study developed a real-time EMG sonification system during walking. The activity of four major muscles was mapped to the amplitude of the sine wave of a different pitch to comprise a harmonic chord. Three healthy adults walked with auditory biofeedback. They walked with no significant difference in the walking speed and stride with and without the biofeedback, showing that the sonification did not interfere

Figure 7: Correct answer rate of the listening test by respondent.

with their walking at least. One of three participants showed more extended joint with auditory biofeedback.

A listening test was conducted to investigate whether it is possible to identify the individual from the sound. Ten people tried to identify the three walkers. The overall average correct answer rate was 0.67. The possibility of identifying the individual from the sound has been shown.

10. REFERENCES

- [1] M. D. Intiso *et al.*, Rehabilitation of walking with electromyographic biofeedback in foot-drop after stroke, *Stroke*, vol. 18, pp. 665-669, 1987.
- [2] M. E. Morris *et al.*, Electrogonometric feedback: its effect on genu recurvatum in stroke, *Arch. Phys. Med. Rehab.*, vol. 73, pp. 1147-1154, 1992.
- [3] M. Matsubara *et al.*, Sonification of, muscular activity in human movements using the temporal patterns in EMG, in *Proc. of APSIPA ASC, 2012*.
- [4] J. Yang and A. Hunt, Real-Time Sonification of Biceps Curl Exercise Using Muscular Activity and Kinematics, in *Proc. of 21st ICAD, 2015*.
- [5] J. Yang and A. Hunt, Sonic trainer: real-time sonification of muscular activity and limb positions in general physical exercise, in *Proc. of ISON 2013, 4th Interactive Sonification Workshop, 2013*.
- [6] P. Sandra and A. Hunt, The Sonification of EMG Data, in *Proc. of 12th ICAD, 2006*.
- [7] R. V. Croce, The effects of EMG biofeedback on strength acquisition, *Biofeedback Self-Regul.*, vol. 11, no. 4, pp. 299-310, 1986.
- [8] Y. Nakayama *et al.*, Real-time smile sonification using surface EMG signal and the evaluation of its usability, in *Proc. of the 21st ICAD, 2015*.
- [9] A. S. Aruin *et al.*, Base of support feedback in gait rehabilitation, *Int. J. Rehab. Res.*, vol. 26, pp. 309-312, 2003.
- [10] M. Matsubara *et al.*, Wearable auditory biofeedback device for blind and sighted individuals, *IEEE MultiMedia*, vol. 22, no. 1, pp. 68-73, 2015.
- [11] G. Claude *et al.*, An auditory display system for aiding interjoint coordination, in *Proc. of 6th ICAD, 2000*.
- [12] H. Terasawa *et al.*, Sonifying ECOG seizure data with overtone mapping: a strategy for creating auditory gestalt from correlated multichannel data, in *Proc. of the 18th ICAD, 2012*.

M.U.S.I.C. (MUSICAL UTILITY SOFTWARE FOR INTERACTIVE CREATIONS): THE MULTIPURPOSE MOTION-SOUND MAPPING PLATFORM

Joseph Ryan, Steven Landry, Jaclyn Barnes,
Myounghoon Jeon

Mind Music Machine Lab,
Michigan Technological University,
Houghton, MI 49931 USA
{jdryan; sglandry; jaclynb; mjeon}@mtu.edu

Chung Hyuk Park

Assistive Robotics & Tele-Medicine Lab,
The George Washington University,
Washington, DC 20052 USA
chpark@gwu.edu

ABSTRACT

With the advent of embodied interaction, research on interactive sonification has rapidly been increasing. There have been consistent efforts on the enhancement of motion-sound mappings, but still more improvements can be made. The goal of the Musical Utility Software for Interactive Creations (M.U.S.I.C.) platform is to provide a flexible, robust, and easy to use interactive GUI for taking motion data from other devices or software and mapping them to sound or music. In the M.U.S.I.C. platform, users can perform actions when the data meet user-generated conditions. Taking inspiration from the previous motion-sound mapping software, Interactive Sonification Markup Language (ISML), the M.U.S.I.C. structure allows for changes to be both dynamic and static, be it to load the program from launch or to make changes in real-time. Furthermore, any device's data can be used on-the-fly. After introducing the design and implementation of M.U.S.I.C., we discuss its comparisons with ISML and future plans.

1. INTRODUCTION

By supplementing or replacing traditional cognitivism, embodied cognition/interaction has emerged as a new theoretical paradigm in Cognitive Sciences [1] and Human-Computer Interaction (HCI) [2] and so does in interactive sonification. In sports, aerobics [3] and rowing in a boat [4] were translated into sound. In arts, movement of performing artists [5] and dancers [6] made aesthetic melody lines. Also, in Human-Robot Interaction gestures of children and robots were sonified to reflect their interactions [7]. In these various scenarios, auditory display researchers are specifically interested in motion-sound mappings. In the present paper, we introduce Musical Utility Software for Interactive Creations (M.U.S.I.C., <https://github.com/trimlab/iISO-P-GUI-Interface>) to offer a more flexible, robust, and easy to use graphical user interface for making mappings between motion data from external devices (e.g., sensors, cameras, robots, etc.) to sound or music. We expect that this work can help non-programmers easily configure and design the interactive sonification in real-time without hard coding. Also, other researchers and developers can update and

contribute to making M.U.S.I.C. more powerful and accessible.

2. INTERACTIVE SONIFICATION MARKUP LANGUAGE (ISML)

ISML (ISHMAEL) GENERATOR V0.5.4

Figure 1: ISML example.

To make efficient motion-sound mappings, we first developed the Interactive Sonification Markup Language (ISML) [8] (Figure 1). The primary goal of ISML was to enable researchers with non-programming backgrounds (e.g., cognitive scientists, artists, musicians, etc.) to configure mappings in an accessible way. ISML was designed for the immersive Interactive Sonification Platform (iISO-P) where artists and technologists collaborate with each other [9]. The project focused on how researchers and artists could use sound in a 3D virtual environment to interactively communicate with users. The iISO-P project utilized motion capture data via a Vicon tracking system and thus, ISML mainly communicated with the Vicon tracking system. This system allowed users to use their entire body as a controller. The motion capture data were taken in as input and mapped onto both visual and auditory displays in a number of ways.

ISML files can be considered as an activity flowchart or state machine defining application behavior in response to external inputs. ISML bears similarity to scripting languages since it is designed to automate the behavior of iISO-P's sound generation application. Specifically, it is an example of an audio synthesis scripting language. Languages of this type already exist. To name a few, examples include the ChuckK

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

audio programming language [10], the Csound [11], the commercial Reaktor software, and the MPEG-4 Structured Audio standard [12]. Some of these systems use graphical user interfaces for their scripting languages, while others are textual. Additionally, some of them support live coding, the ability to change the program's behavior while it is running. By comparison, ISML scripts can be created in a graphical or textual manner. To the best of our knowledge, ISML was the only scripting language specifically designed to translate physical movements into dynamically generated sound, which makes it very different from other audio scripting languages.

Using ISML to the mappings is much more efficient and accessible to non-technical researchers than attempting to reprogram the system for every new experiment with new specifications. Through our empirical experiences, we have found that using ISML could save 20-40% of time, compared to hard coding after 3-4 specification developmental cycles. For more details of ISML, see [8].

The ISML platform allowed researchers to create items that had both conditions and actions. Using this platform, researchers were able to do some basic object comparisons and interactions. ISML allowed the user to create simple 1:1 mappings between tracked objects and pre-defined MIDI notes and melody lines. However, unfortunately, developers and users were still required to hard code some features into the audio code and manually adjust ISML scripts. It was then realized that development was almost entirely being done in code or via text file manipulation, which was the opposite of the original intention of ISML. Moreover, the manipulations had to be made offline. The formation of the audio program was a static load up. In other words, users had to choose a file that they made at the compile time and could make no changes while the program ran. Features were program specific and could not be shared between users. This made it very hard for more than one person to use the platform effectively. The ISML script generator was also separated from the audio program. Therefore, users had to use multiple programs just to produce any sound. Another problem was the lack of data feedback that was received from the program. Data were calculated and used by the system, but the values were not displayed to users. This made it unnecessarily difficult to accurately assess system actions and make useful changes. The software lacked transparency and provided little feedback for development and testing of sonification scenarios. Finally, even though ISML's structure was very logical, it was somewhat confusing to new users, specifically for non-programmers. These issues led to many complications and difficulties when developing motion-sound mappings for the iISoP project.

To solve these issues, the Musical Utility Software for Interactive Creations (M.U.S.I.C.) platform was created. It serves as an interactive platform intended not only for the iISoP project, but also other interactive sonification projects in order to increase the robustness and usability of the entire system so that anybody can develop future applications in a more convenient way.

3. M.U.S.I.C. STRUCTURE

The structure of the M.U.S.I.C. platform is similar to the structure of the ISML platform. However, it has more

Figure 2: M.U.S.I.C. platform's basic structure.

```

Procedure conditionalChecker:
  variable doActions;
  begin
    repeat
      set doActions value to true;
      repeat
        if condition is not met:
          set doActions value to false;
          go outside of repeat
        end if
      until conditionList has no condition left

      if doActions is true:
        perform all actions
      end if

    until conditionList has no conditions left
  end
  
```

Figure 3: Pseudocode of the conditional checker.

features to bridge the gaps of execution and evaluation. The structure of M.U.S.I.C. is depicted in Figure 2. The platform is still built around the ability to create conditional statements for motion-to-sonification. A conditional is a combined set of conditions and actions. Conditions are user-defined criteria that are constantly checked against the stream of incoming data. Different actions, such as playing a specific note or melody, are triggered when specific conditions are met. When all conditions are met, all actions are performed in separate threads. This allows for multiple conditionals to be run concurrently without slowing the speed of execution for assigned actions. Conditionals are built in a block of code, called "conditional builder". The conditional builder has a reference to all of the current objects, patterns, and sections as well as a list of previously made conditionals. All of these references act independently of the conditional builder, which allows changes to occur at any point in the development process.

Each conditional has a flag variable, the value of which determines if the conditional is enabled or disabled. The conditional checker is a separate thread of program that has a reference to the conditional list and checks all conditionals if enabled or not. This process is shown in Figure 3, a pseudocode of the conditional checker. It checks for every enabled conditional, and then performs all actions within that condition. The flag variable doActions is set to true so that all actions can be performed even when no conditions are enabled. This in turn allows for chosen actions to be executed continuously, if desired.

Figure 4: The Quick Start View.

Figure 5: The Objects View.

Figure 6: The Room View.

Figure 7: The Pattern Builder View.

4. M.U.S.I.C. GRAPHICAL USER INTERFACE

Following M.U.S.I.C.'s structure, the platform uses an interactive interface that allows non-programmers to "program" their modules without touching code. The M.U.S.I.C. GUI is split up into sub-sections called tabs, each of which has specific functions. Upon launching the program, the user will see the "Quick Start" tab (Figure 4). This tab is for starting up a previously created module. On this page, users can also learn whether they have a connection to the motion tracking devices (e.g., Vicon sender) and they are receiving the live data. The "Objects View" (Figure 5) allows users to view and access all of the different objects' data that the M.U.S.I.C. platform is tracking. Users can also toggle listening to certain objects, as well as adjust their color for the "Conditional View". In the "Room View" (Figure 6), users can divide the space into sections for later condition checking in the conditional builder, e.g., dividing the room into left and right halves so that conditional such as "play middle C when left hand is in the left half of the room" can be created. In the "Pattern Builder" tab (Figure 7), users can create their own MIDI patterns to use as actions of the conditionals. Users can simply click a box on the grid marking a note to play a pattern in a right timing. The simplicity of this view makes it easy for users to make their own patterns and sounds without additional music programs. It can also be easily connected/extended to external sound programs. The "Conditional View" (Figure 8) is our conditional builder. This is where users take all of their sections, patterns, and objects to make conditionals. As the current program stands, there are two types of conditions: 'in-section' and 'value'. The in-section condition checks to see if the object is within a created section. The value condition lets users choose from the list of data points and make a comparison between those numbers. The currently

Figure 8: The Conditional View.

available actions are to play a single note or to play a previously defined pattern.

5. THE COMPARISONS OF ISML AND M.U.S.I.C.

Although the main conceptual structures between ISML and M.U.S.I.C. are similar to one another, there are major differences between the two platforms (Table 1).

Table 1: Comparison between ISML and M.U.S.I.C.

Functionality	ISML	M.U.S.I.C.
No programming knowledge required		✓
Interfaces with Vicon tracking System	✓	✓
Graphical user interface	✓	✓
Faster than text file editing		✓
Can view tracked data live		✓
Dynamic adjustments		✓
Can build conditionals	✓	✓
Easy to use		✓
Can be utilized internet free		✓

There were difficulties faced when using ISML for extended development. One was simply the speed of script development. In most cases, developers had to start with ISML and then go to the generated text file and edit it manually. M.U.S.I.C. solves this problem by having an enhanced graphical user interface (GUI), which not only gives users direct visual/audio feedback for the work that they have done, but also enables them to simply make complex features without writing a single text file. In our experience, it is faster for most users to create conditionals in the GUI than in text files. In addition, the M.U.S.I.C. platform provides the users with feedback on the objects being tracked (the Objects View in Figure 5), which is particularly helpful for debugging. By allowing users to dynamically adjust patterns, sections, and conditionals while sonifying data in real-time, M.U.S.I.C. solves the static manipulation problem that ISML has had. Most importantly, any user can use this platform without having a programming background.

6. THE FUTURE OF M.U.S.I.C.

M.U.S.I.C. is currently in its infancy. There are considerable potential applications and future directions that this platform can go. First and foremost, we will add more conditions and more actions that can be utilized. The structure makes it very convenient to add new conditions and actions. In addition, when the algorithms for gesture recognition are created (e.g., drawing a circle in 3D space), it is just a matter of adding a new condition option for users to utilize. Next, adding more objects to interface with the platform, such as a muscle sensor (e.g., Myo) or brain sensor (e.g., EPOC), can add a new dynamic of biological data tracking. Code optimization to further improve performance as well as running usability tests to improve user experience are also planned. The subsection GUI, such as the Room View tab, could be greatly improved by a better graphics engine, such as the Unity Engine. Since this application is currently built in Java, developers are exploring the option of moving this application to Android as well. Conditionals are very useful concepts and could be generalized to other tasks and domains that would benefit from interactive sonification. For example, in the driving domain, additional auditory feedback could be provided based on driving performance. Similar applications are possible for athletic training, physical rehabilitation, and human-robot interaction.

7. REFERENCES

- [1] Davis, J. I. and Markman, A. B. (2012). The future of embodied cognition. *Topics in Cognitive Science*, 4, 4, 685-793.
- [2] Marshall, P., Antle, A., Van Den Hoven, E., and Rogers, Y. (2013). Special issue on the theory and practice of embodied interaction in HCI and interaction design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(1).
- [3] Hermann, T. and Zehe, S. (2011). Sonified aerobics: Interactive sonification of coordinated body movements. *Proceedings of the 17th International Conference on Auditory Display*, Budapest, Hungary.
- [4] Schaffert, N., Mattes, K. and Effenberg, A. O. (2011). The sound of rowing stroke cycles as acoustic feedback. *Proceedings of the 17th International Conference on Auditory Display*, Budapest, Hungary.
- [5] Jeon, M., Landry, S., Ryan, D. J., & Walker, J. W. (2015). Technologies expand aesthetic dimensions: Visualization and sonification of embodied penwald drawings. In A. L. Brooks, E. Ayiter, and O. Yazicigil (Eds.), *Arts and Technology. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (LNICST)* vol. 145. (pp. 69-76). Springer International Publishing.
- [6] Landry, S. and Jeon, M. (2017). Participatory design research methodologies: A case study in dancer sonification. *Proceedings of the International Conference on Auditory Displays (ICAD2017)*, PA, USA, June 20-23.
- [7] Zhang, R., Barnes, J., Ryan, J., Jeon, M., Park, C. H., and Howard, A. (2016). Musical robots for children with ASD using a client-server architecture. *Proceedings of the 22nd International Conference on Auditory Display (ICAD2016)*, Canberra, Australia, July 2-8.
- [8] Walker, J. W., Smith, M. T., and Jeon, M. (2015). Interactive sonification markup language (ISML) for efficient motion-sound mappings. *Proceedings of 17th International Conference on Human-Computer Interaction (HCII2015)*, Los Angeles, CA, August, 2-7, 2015.
- [9] Jeon, M., Smith, M. T., Walker, J. W., and Kuhl, A. S. (2014). Constructing the immersive interactive sonification platform (iISoP). In N. Streitz and P. Markopoulos (Eds.), *Distributed, Ambient, and Pervasive Interactions (DAPI)*, LNCS 8530, pp. 337-348. Switzerland: Springer International Publishing, 2014.
- [10] Wang, G. (2008). *The ChucK Audio Programming Language*. Dissertation. Princeton University, New Jersey.
- [11] Boulanger. R. (2000). *The Csound Book*. MIT Press.
- [12] Scheirer. E.D. (1998). The MPEG-4 structured audio standard. *Acoustics, Speech and Signal Pro-cessing*, 1998. *Proceedings of the 1998 IEEE International Conference on*. Vol. 6. IEEE.

REPRESENTING TWITTER USERS' ENGAGEMENT BY SONIFICATION

Robert Wheeler

Columbia College Chicago,
Audio Arts and Acoustics Department
600 S Michigan Ave, Chicago, IL 60605
USA
robert.wheeler@loop.colum.edu

David Worrall

Columbia College Chicago,
Audio Arts and Acoustics Department
600 S Michigan Ave, Chicago, IL 60605
USA
dworrall@colum.edu

ABSTRACT

Social Media has become one of the main ways that people engage with each other and companies engage with their customers and potential customers, so the statistical analysis of this engagement has become common, and a useful tool to track social trends and preferences. This project's aim was to create a tool to sonify and analyze Twitter users' engagement with a user-defined topic over time.

1. INTRODUCTION

This project was undertaken for the purpose of quantifying users' engagement with a topic or keyword on Twitter [1] over time, and presenting this data using sonification so that that it can be absorbed in a compressed amount of time. There have been several previous reports of the sonification of Twitter data, one notable example being a 2012 report by Thomas Hermann et al.[2]. The majority of previous work seems to focus on real-time data.

This project ("Mike Twypyson (MT)") takes each tweet (a post by a user of Twitter), converts it into a discrete data-point and sonifies the data-points individually along a user-scalable timeline. Unlike other approaches, this project deals with historical not realtime data in the hope that this would be more useful for observing longer term trends in engagement.

2. DATA COLLECTION

A data set is drawn by the user from Twitter in non-real time. When the program starts, the user is prompted to enter

- (a) Search term (a name, keyword, or hashtag).
- (b) The length of time they wish the sonification to take place over.
- (c) How many tweets should be sonified, and
- (d) A "temporal resolution threshold" (see later).

In order to access their database, Twitter requires access credentials which it grants upon registration. In order to access Twitter via Twython [3], a Twitter API, requires access credentials from Twitter, and so the application "Mike Twypyson" (MT) was registered through the site's developer portal to obtain these credentials.. For the Twitter database access, only the search term and the number of

tweets to sonify are required. MT retrieves chunks of tweets from the Twitter database for the given user query until the tweets collected meets the number specified by the user, or Twitter runs out of tweets to return. If the number of user-requested tweets is large, accumulating the data may take a considerable time, and/or the search for tweets for a given search-term is exhausted. A list of retrieved tweets is parsed for useful information and turned into `Tweetts`, a Python Class object, for ease of processing.

3. DATA PROCESSING

Once the collation of `Tweetts` is complete, their timestamps are used to order them temporally in preparation for the time-warping necessary for sonification according to the user's requirement for a sonification of a particular duration. A scalable temporal gamut in which all of the sonified tweets will be placed is then calculated. Due to the potentially high density of tweets per minute, however, it proved necessary for some periods of time to be further sub-divided. If, for example, in one minute of "tweet-time" (i.e. real time), there was more than one tweet, we sub-divide the portion of the temporal gamut devoted to that minute into an equal number of divisions, according to the number of tweets in that minute. This creates an arpeggio-like effect. The `Tweetts` in this arpeggio's timbres are increasingly "brightened" by the addition of more harmonics based on their position in the arpeggio.

With the temporal division of the sonification settled, and mechanisms to sonify the density of `Tweetts` in place, the `Tweetts` are analyzed for user engagement, which is quantified using four main criteria:

- (a) Retweets,
- (b) Favorites,
- (c) User Mentions, and
- (d) The Follower Count of the user that authored the tweet.

The number of retweets, favorites, and mentions of other users (through authors tagging their handles) are summed to obtain one "engagement value", with each element of this value being weighted based on how important these criteria are when judging engagement. For example, retweets are weighted x4, user mentions are weighted x2, and favorites are weighted x1). This value is converted to the pitch of the tone that will be created from this tweet. If the tweet's author has 10,000 or more followers, this is seen as a potentially influential tweet in its own way, and so, in order to distinguish it, the resulting sonified data-point's fundamental frequency is lowered by three octaves.

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

When the density of the data over the user-specified sonification time was very high, a user-defined “temporal resolution threshold” mechanism was implemented to make the sonification more listenable. This mechanism lets the user decide how quickly they are able to absorb the data points, combs through the times that these sonified points are set to play, and if the time between points is smaller than what the user has specified for their temporal resolution threshold the Tweets are consolidated into a single sonification event; the timbre brightness adjusted and the duration increased accordingly. In this way, the sonification of tweet density over the allotted playback time is not lost, and the user is made aware that temporal separation of events have been lost. This separation loss is also indicated in the statistics provided to that the user before and after the playing the sonification.

4. PLAYBACK

After the main processing is complete, each Tweet is converted into a Csound event and appended to the sonification event list which is then passed to Csound for sound for rendering in real-time or to soundfile. The user is provided with some basic statistics about the sonification they are about to hear, and then prompted to initiate the sound synthesis. The Csound instrument used for the sonification is a simple fixed-waveform wave-table tone generator, which, as discussed earlier, adds additional harmonics based on the density of tweets within time.

5. EXAMPLES

Examples of the project’s sonifications are publicly available on Soundcloud [4] are provided in order to illustrate what the project does well, and what the project fails to do well, as well as show the range of results the program can yield:

- (a) *monster energy* [5] is the sonification produced by the program with the settings “monster energy” for query term, “120” for the sonification length in seconds, 5,000 tweets being sonified, and 5ms set as the resolution threshold. The example shows a change in users’ engagement over time. The sonification begins with sparse tweets, many with little engagement, but with small bursts of both tweet density (sonified with brighter timbres), and engagement with those tweets (sonified by pitch), and with the occasional tweet by a user with over 10,000 followers (sonified by a note 3 octaves lower than usual). As the sonification proceeds, we observe more frequent bursts of engagement, suggesting a trend of increased Twitter user engagement with the keywords over the period of the tweets sonified, which, in this case, was 20 hours prior to the enquiry (and as reported by the program’s user interface).
- (b) *switch* [6] is the sonification produced by the program with the settings “switch” for query term, “240” seconds for the desired sonification length, 5,000 tweets to be sonified, and 30 milliseconds as the resolution threshold. This sonification highlights the resolution threshold in action well. Around the 6 second mark, we hear the density of

sonified data points decrease, but the length and timbre of these points increases. The increased length is the signifier that points have been consolidated, and from the bright timbre we get an indication of the density of the consolidated points (the point we hear takes on the timbre of that of the “brightest” note absorbed).

6. EVALUATION

While this project does provide the user with accurate information about users’ engagement with keywords on Twitter, there are many possible improvements. For example, one of the main aims for several of the features of the sonification was to make it more listenable and easily absorbed by the user. Unfortunately, the data sonified can be so temporally dense that it becomes hard to get a sense of trends in the sonified data. High density of tweets over a short period of time can also be a problem in drawing a large enough data set to be able to discern gradual trends in the data over time. Twitter’s publicly searchable database of tweets may only extend a day or two into the past when querying for even thousands of tweets, but if the query is extended into the ten-thousands of tweets the time it takes to process them all becomes significant. The lack of information about how many tweets on a specific keyword the public database contains makes the data acquisition process cumbersome: Queries for thousands of tweets often fall short, and more testing is required to determine if this is because of a lack of tweets by users or if the database simply doesn’t return some search results. Also, it is difficult for the listener to discern how much time has passed in the sonification in regards to the relation of the timestamp of the current tweet being sonified to the timestamp of any other tweet being sonified. A system to differentiate the different kinds of engagement with the tweets (eg. retweets, favorites, user mentions, etc.) as opposed to combining all of them into one value for general engagement might improve the user’s understanding of the complexity of the tweetscape: High numbers of followers for the author of a tweet is identified, but having each type of engagement control different parameters of the sonification’s timbre may yield a result with more richness in conveyed data. The use of a more complex sound-synthesis model would make listening to the flurry of sine tones generated by the project easier.

7. REFERENCES

- [1] “Twitter. It’s What’s Happening.” *Twitter*, Twitter, <http://www.twitter.com/>. Accessed May 2, 2018.
- [2] Hermann, T., A. V. Nehls, F. Eitel, T. Barri, and M. Gammel, “Tweetscapes: Real-time sonification of twitter data streams for radio broadcasting” in *Proceedings of the 18th International Conference on Auditory Display, Atlanta, GA, USA. June 18-21, 2012*
- [3] “Twython.” *Twython-Twython 3.6.0 Documentation* <http://twython.readthedocs.io/en/latest/>. Accessed May 2, 2018.
- [4] *Mike Twipyson* SoundCloud. <http://soundcloud.com/twipyson> Accessed May 2, 2018.

- [5] *monster energy* SoundCloud,
<http://soundcloud.com/twipyson/monster-energy>.
Accessed May 2, 2018.
- [6] *switch* SoundCloud.
<http://soundcloud.com/twipyson/switch>. Accessed May
2, 2018.

AUDIOGRAPH: AUDITORY GRAPHS FOR GOOGLE SHEETS WITH FAUST

Andrés Aparicio

Center for Research in Audio Technologies
Music Institute
Faculty of Arts

Pontificia Universidad Católica de Chile
aaparicio@uc.cl

Rodrigo F. Cádiz

Center for Research in Audio Technologies
Music Institute, Faculty of Arts
Department of Electrical Engineering
School of Engineering
Pontificia Universidad Católica de Chile
rcadiz@uc.cl

ABSTRACT

In this article we present Audiograph, a web-based auditory graph application that allows users to sonify numerical data directly from a spreadsheet, using Google Sheets. One or more data series can be converted into sound at the same time. Our tool is implemented as an add-on inside Google Sheets using the powerful sound synthesis engine FAUST. We have chosen FAUST as the sonification platform due to its highly efficient rendering of digital sound processing algorithms and its integration with WebAssembly, which allows to use any modern web browser in a much more powerful way than with previous JavaScript-based technologies.

1. INTRODUCTION

People with visual impairments or blindness have serious difficulties to access and comprehend the information contained in graphs, which are widely utilized to describe numerical data, especially in scientific or mathematical fields. Text-to-speech or screen-reader software available in most modern operating systems today may help them to understand the context of a graph, including information such as the title, axis labels, and what the visual display is about, but not the actual data itself. This is especially the case in spreadsheets environments such as Microsoft Excel or Google Sheets.

Understanding data in visual graphs is not an easy task for sighted people and, in the case of visually-impaired people, the difficulty increases. As data becomes more complex, multi-dimensional or bigger in size, the issue becomes more difficult [1]. In the United States, coordinate graphs and number lines are an essential part of mathematics education [2], and are often part of required standards from kindergarten to high school, while in Chile, one of the objectives of the primary and secondary school curriculum is the usage of graphs to interpret and access different

kinds of information¹. Graphical data is used from 1st to 10th grades, starting first with pictographs, and later introducing other kinds, such as bar, line or pie graphs. In the case of blind children, currently there are no established methodologies for teaching the understanding of graphs except via tactile-graphs, when available. This turns out to be a problem later in life for visually-impaired people pursuing careers in the natural or social sciences, where the use of graphs is essential and implementation of accessible options is not widespread.

There is significant evidence that auditory graphs allow for a better understanding of numerical data contained in graphs, especially for visually-impaired people [3] [4] [5] [6]. This approach has many advantages over the standard method of tactile-graphs as they are easy to implement with any general purpose computer equipment available today for educational purposes. Also, in the specific case of educational settings, auditory graphs are a satisfactory alternative to tactile-graphs, in terms of curriculum compatibility, students performance and teachers satisfaction [7].

Our main goal is to implement graphical data sonification strategies that would allow blind or visually-impaired people to access quantitative information contained in visual graphs, especially in educational, professional or academic contexts. This is the reason why we have focused on a free widely used tool for numeric calculation and plotting on a spreadsheet, such as Google Sheets. Spreadsheets are a very common and useful tool for the storage and analysis of data, and they are used throughout industry, commerce, government and education [8]. We would like to add the capability of not only seeing the data in a spreadsheet environment, but also to hear it.

This article is structured as follows. In section 2, the general sonification strategy is presented. Section 3 discusses the usage of FAUST as the engine for sound generation and processing and provided one example of the type of synthesis methods that can be used. Section 4 presents our application in detail. Finally, conclusions and future work are discussed.

2. SONIFICATION STRATEGY

There are two main ways of using auditory graphs: to estimate tendencies in the data or to estimate exact values. In the first case, the idea is to determine patterns in the data while the latter tries to

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

¹Retrieved on 4/20/2018 from <http://www.curriculumnacional.cl>

determine with the most possible precision the sound of a particular point in the data space [6]. At the current state of our research, we are focusing in pattern estimation for the case of line graphs of one of more series.

In the particular case of spreadsheet sonification, there has been significant work that has been reported in the literature. Stockman [8] proposed an integrated approach to sound-based access to spreadsheets. For this purpose, he developed a macro for Microsoft's Excel written in Visual Basic, while the sound rendering was done in CSound. This macro first identifies important features of the data such as global maximum and minimum numeric values, offset and scaling parameters. The system is able to sonify a selected range of cells either row-wise or column-wise or partial portions of the selected range.

Hetzler and Tardiff [9] developed a spreadsheet-based tool that uses sonification to support and reinforce calculus instruction, by means of graphical and numeric representations of mathematical functions. Their objective was to dispose of a tool that "students could use to create auditory graphs, to explore and learn about limits, derivatives, concavity, and extreme values". The implementation was also embedded into Microsoft's Excel using Visual Basic and ActiveX.

In the case of line plots, auditory graphs are typically sonified by converting values in the y-axis to pitches or frequencies and the x-axis to time, as done by [10] [11]. The way that the y-axis data is converted into sound may vary, such as in the case of [9] where they used a discrete number scale, mapping each value logarithmically into the fundamental frequency of a chord. Major chords were used to represent non-negative values, and minor chords to represent negative values.

The handling of time can also vary, by means of a continuous or discrete presentation of sound events. Harrar and Stockman conducted an experiment where the continuous presentation mode generated more accurate results when the participants were asked to draw the audio graph in comparison to the discrete mode of presentation. But for point estimation tasks, discrete sound presentation generated a better accuracy than the continuous sound presentation [12].

In our approach, data series are converted to an audio stream in a sequential manner, and we map the value of each cell to a particular pitch. There are two fundamental modes of presentation: as a discrete series of audible events and as a continuous frequency-varying sound. Intensity is coupled with frequency, as they have been proven to be integral dimensions [13]. The duration of the whole sonification process is in the range of 3-7 seconds, as it has been shown to be the most effective time range [12].

It is important to emphasize that studies of auditory graphs with more than one data stream are fewer than the ones that focus on only one data series. This is the reason why we designed our tool to be able to sonify more than one data stream at the same time, which leads to interesting research questions such as how many sound parameters should be used, which ones should be used in all series and which ones separately [13]. In concrete, when more than one data series are present we also incorporate stereo mixing as a strategy to separate the different data series in the stereo field.

Our tool can be run on any modern web browser that supports WebAssembly and sonification parameters can be easily customized and changed in real-time and provide almost immediate feedback on the visual information to the listener.

3. SOUND SYNTHESIS

We have chosen FAUST [14] as the sound synthesis platform for our tool. FAUST is a powerful and expressive programming language that provides a purely functional approach to audio signal processing while offering a high level of performance comparable to C/C++ based audio libraries.

The semantic of FAUST is simple and well defined, allowing the FAUST compiler to be semantically driven. This means that instead of compiling a program literally, it compiles the mathematical function it denotes. A FAUST program denotes a signal processor, a mathematical function that transforms input signals into output signals.

Contrary to most JavaScript-based audio libraries, which provide usually a very simple and limited subset of sound synthesis algorithms and data control, FAUST is a complete digital audio processing engine, comprising hundreds of sound analyzers, generators, filters and effects².

However, one of the most important facts of using FAUST for web-based auditory graphs, is its recent support for WebAssembly. [15]. According to the official website, "WebAssembly (abbreviated Wasm) is a binary instruction format for a stack-based virtual machine. Wasm is designed as a portable target for compilation of high-level languages like C/C++/Rust, enabling deployment on the web for client and server applications". In practice, audio that runs via WebAssembly on a web browser is far more accurate and efficient than JavaScript-based solutions.

Figure 1 presents an example of the type of sound sources that can be generated in FAUST. One of the nice features of FAUST is that it provides block diagrams of all the digital signal processing subroutines being generated in the code. In this example, the instrument consists on three sawtooth oscillators connected to an amplitude envelope packed in a group called voice.

We now proceed to describe our web-based auditory graph solution in more detail.

4. AUDITORY GRAPHS IN GOOGLE SHEETS

We believe that the most important factor that differentiates our proposal from the previous ones [8] [16] [9], is the usage of open-source and state of the art platforms such as Google sheets, Faust and WebAssembly, that provide a much better performance and flexibility for the design of different sonifications. Anyone with a Google account can create a spreadsheet, install our Audio-graph tool and use it. The previous proposals were based on Microsoft Excel, relying on obscure and not well-maintained development environments such as Visual Basic (deprecated in 2008), or CSound, which is not being used actively in the computer music community. Excel is surely still widely used, but it remains a commercial tool. Our add-on tool is free and open-source. While FAUST is a relatively new tool, it is gaining momentum and it is very powerful. Our solution is also available as a JavaScript API, meaning that it could be incorporated into any web-based application.

To accomplish our goal, we designed a web system with three components³ (Figure 2). First, the *FAUST Layer* provides an instrument DSP (Figure 1) with a set of parameters that is compiled into a WebAssembly module. This module, in turn, is exposed

²<http://faust.grame.fr/library.html>

³Available for download at <https://github.com/rcadiz/web-audiographs> and as a add-on inside Google Sheets.

Figure 1: FAUST block diagrams of a simple organ-like instrument.

via a Javascript API packaged as a dynamic module that may be loaded asynchronously. This abstracts the sound synthesis details and allows changing the sonification strategy by updating only the WebAssembly module when the instrument DSP is changed. Second, the *Audiograph Module* provides a set of operations that are used to both to handle incoming data and its sonification process. The sonification process may be controlled either programmatically by code, or through a standard set of controls that may be manipulated by users to change sonification parameters. Finally, the *Client Interface* uses API calls to the *Audiograph Module* both to render a user interface (UI) and to manage the sonification process. The implementation of this component depends on where the system is to be deployed: as part of a webpage, an extension of another web system, or as part of a Node application, for example.

We implemented a *Client Interface* for Google Sheets, a part of Google suite of office web applications. These applications may be enhanced with add-ons implemented in Google App Scripts, an scripting language based on JavaScript. These add-ons can be used to extend the application UI via menus or sidebars and, in this case,

Figure 2: Architecture of the proposed solution.

operate on numeric data stored in the spreadsheet. Each add-on requires a UI and a script that handles communication between the spreadsheet and the add-on UI. In our case, the add-on UI contains the user controls provided by the *Audiograph Module* (Figure 3).

Once installed and enabled, the add-on is used by selecting a group of cells in the spreadsheet and activating the add-on. This loads the *Audiograph Module*, which renders the UI, connects the user controls to the appropriate browser events, then reads the selected data from the spreadsheets and sends it to the sonification process, starting automatically as soon as possible. Initially, the sonification uses its default settings but they may be changed with the user controls, and the process restarted. Available settings are:

1. Mode: whether the sound is discrete or continuous.
2. Duration: time, in milliseconds, for the audiograph to play.
3. Frequency: possible range used when mapping values to frequencies.
4. Scale type: controls whether the mapping from values to frequencies maps the minimum and maximum values from the data series to the minimum and maximum frequencies (the relative type), or whether the mapping uses a fixed range of values provided by the user (the absolute type).
5. Absolute scale: minimum and maximum values used in an absolute scale.
6. Timbre: type of sounds and synthesis algorithms used for the sonification.

Figure 3: Example of Google Sheets Audigraph add-on sidebar with basic options displayed.

5. CONCLUSIONS AND FUTURE WORK

We have shown that it is possible to implement a web-based auditory graph application to sonify numerical data from Google Sheets that allows users to hear the data in addition to manipulating and seeing it. We have used FAUST, an advance digital sound processing engine, which allows the use of complex sonification strategies in a web-based context. Our architecture enables changing sonification strategies easily while keeping a standardized interface for the end users.

In terms of future work, we are planning to conduct tests with end-users in order to validate the design decisions we have made with the application. We are also considering to add sonification for context information (tick marks, for example), as a separate layer of sound that the user can select at will. We are looking into the addition of binaural sound or ambisonics in order to better position audio signals in the listening space. A final application that we think could be useful is the sonification of analytical data functions instead of data series.

6. ACKNOWLEDGMENTS

This research was funded by Google via a 2017 Latin American Research Award (LARA) and by Fondecyt Grant #1161328, Government of Chile.

7. REFERENCES

- [1] B. N. Walker and L. M. Mauney, "Universal design of auditory graphs: A comparison of sonification mappings for visually impaired and sighted listeners," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 2, no. 3, p. 12, 2010.
- [2] B. J. Tomlinson, J. Batterman, Y. Chieh Chew, A. Henry, and B. N. Walker, "Exploring Auditory Graphing Software in

Classroom: The Effect of Auditory Graphs on the Classroom Environment," *ACM Trans. Access. Comput. Article*, vol. 9, no. 3, 2016.

- [3] M. A. Alonso-Arevalo, S. Shelley, D. Hermes, J. Hollowood, M. Pettitt, S. Sharples, and A. Kohlrausch, "Curve shape and curvature perception through interactive sonification," *ACM Trans. Appl. Percept.*, vol. 9, no. 4, pp. 17:1–17:19, Oct. 2012.
- [4] B. Brown, L.M., Brewster, S.A., Ramlohl, S.A., Burton, R., and Riedel, "Design Guidelines for Audio Presentation of Graphs and Tables," *9th International Conference on Auditory Display*, no. July, pp. 284–287, 2003.
- [5] F. Delogu, M. Palmiero, S. Federici, C. Plaisant, H. Zhao, and O. Belardinelli, "Non-visual exploration of geographic maps: does sonification help?" *Disability & Rehabilitation: Assistive Technology*, vol. 5, no. 3, pp. 164–174, 2010.
- [6] B. N. Walker, L. M. Mauney, and J. Godfrey, "The Audio Abacus: Representing a Wide Range of Values with Accuracy and Precision." in *Proceedings of the International Conference on Auditory Display*, 2004.
- [7] B. K. Davison, "Evaluating auditory graphs with blind students in a classroom," *ACM SIGACCESS Accessibility and Computing*, no. 102, pp. 4–7, 2012.
- [8] T. Stockman, "The design and evaluation of auditory access to spreadsheets," in *Proceedings of the 10th International Conference on Auditory Display (ICAD2004)*, 2004.
- [9] S. M. Hetzler and R. M. Tardiff, "Two Tools for Integrating Sonification into Calculus Instruction." in *International Conference on Auditory Display*. Georgia Institute of Technology, 2006, pp. 281–284.
- [10] D. L. Mansur, M. M. Blattner, and K. I. Joy, "Sound graphs: A numerical data analysis method for the blind," *Journal of Medical Systems*, vol. 9, no. 3, pp. 163–174, jun 1985.
- [11] L. M. Brown and S. A. Brewster, "Drawing by ear: Interpreting sonified line graphs," in *Proceedings of the International Conference on Auditory Display*, 2003.
- [12] L. Harrar and T. Stockman, "Designing Auditory Graph Overviews: An Examination of Discrete vs. Continuous Sound and the Influence of Presentation Speed," *Proceedings of the 13th International Conference on Auditory Display (ICAD2007)*, pp. 299–305, 2007.
- [13] S. C. Peres and D. M. Lane, "Auditory graphs: The effects of redundant dimensions and divided attention," in *Proceedings of the International Conference on Auditory Display*, 2005, pp. 169–174.
- [14] Y. Orlarey, D. Fober, and S. Letz, "Faust: an efficient functional approach to dsp programming," *New Computational Paradigms for Computer Music*, vol. 290, p. 14, 2009.
- [15] S. Letz, Y. Orlarey, and D. Fober, "Compiling faust audio dsp code to webassembly," in *Web Audio Conference*, London, UK, 2017.
- [16] T. Stockman, L. Nickerson, and G. Hind, "Auditory graphs: A summary of current experience and towards a research agenda," in *International Conference on Auditory Display*. Georgia Institute of Technology, 2005, pp. 420–422.

SONIFICATION OF 3D POINT CLOUDS FOR SUBSTITUTION OF VISION BY AUDITION FOR BLIND USERS

Louis Commère

Université de Sherbrooke
NECOTIS GEGI, CIRMMT
Sherbrooke, Canada
louis.commere@usherbrooke.ca

Florian Grond

McGill University
CIRMMT
Montréal, Canada

François Côté

1296 Avenue Maguire
Québec, Canada
G1T1Z3

Jean Rouat

Université de Sherbrooke
NECOTIS GEGI, CIRMMT
Sherbrooke, Canada
jean.rouat@usherbrooke.ca

ABSTRACT

A low level sonification prototype of 3D point clouds for the sensory substitution of vision by audition for the visually impaired is investigated. The aim of this work is to study which point cloud features can be understood through the sonification of raw 3D data without the extraction of high-level features through algorithms. Preliminary results show the possibility for the user to localize objects and estimate their sizes but not to understand shapes of objects.

1. INTRODUCTION

Sensory substitution refers to the use of one sensory modality to access information usually available from another sensory modality. Sensory substitution devices are designed to help people with sensory loss to compensate with other remaining sensory modalities. This work focus on systems which substitute vision by audition for blind people. The first such device is the vOICE [1] which was originally developed by P. Meijer in 1992. The vOICE sonifies grayscale images by mapping the position and value of pixels to sound properties of short sine tones. Many other systems have subsequently been conceived for various purposes like text reading, locomotion, color recognition, obstacles detection, etc.

Systems like the ones by Zhong *et al.* [2] or some commercial applications [3, 4, 5] use sophisticated image recognition algorithms to give the user an easily understandable access to visual information. However, this strategy can produce false detections, often requires an internet connection for cloud computations and provides abstract information, which can lead to a lack of generalization if the system is used in a new environment. On the other hand, systems like the vOICE sonify raw data without the use of recognition or detection algorithms. This strategy has the advantage of providing direct information to the user. The work of Kim and Zatorre [6] shows that blindfolded sighted participants can learn to perceive tactile shapes with the vOICE. Studies have also shown that thanks to the great plasticity, the brain can use visual areas to process auditory stimuli [7]. The brain can interpret new sound stimuli to represent the visual space, and it does not seem necessary to use complex recognition algorithms.

We investigate which features of basic 3D point clouds can be perceived by sonifying the raw data. For this purpose, we propose a raw 3D point cloud data sonification prototype inspired by the work of coauthor F. Grond [8], where each point acts as a sound source. We use 3D data as inputs to the sonification prototype. Depth information is beneficial to plan actions and get a good understanding of the surroundings. Instead of using point clouds captured from depth sensors, in this work we generate artificial 3D points to ease the evaluation with reproducible stimuli.

The closest work we found related to ours [9] sonifies the interior contour of 3D points clouds captured by depth sensors to allow the user to localize and recognize the 3D real objects. Compared to this work we do not use any algorithm to extract contours or shapes. We sonify each point as describe above to study how far the brain can interpret them without complex processing.

2. INPUT DATA

We use four 3D artificial point clouds as illustrated in figure 1 to simulate virtual objects. Each object is initially composed of 5000 points. Points that are non-visible from the user's point of view are then deleted to mimic occlusion. The use of artificial points instead of real ones coming from 3D sensors gives us a complete control on the experimental conditions. Distances in the artificial space are arbitrary. To simulate pseudo-scenes (comprising one or two objects for now), point clouds are translated, rotated, rescaled and combined. Issues faced with real scenes are not addressed in this abstract.

3. SONIFICATION STRATEGY

3.1. Parameter mapping

As stated before we choose to sonify the raw data without pre-processing to study which features of a pseudo-scene, the subject can perceive. Each point acts as a short spatialized sound source. The emitted sound is a sinusoidal signal of 100 ms duration multiplied by a percussive envelope. The sinusoidal characteristics depend on the position of the point in the virtual space. The mapping between the sound characteristics and the point position is designed to be as natural as possible for the human perception to make the sound easier to interpret.

We linearly map the characteristics of the spatialized sound to the point position in the geometric space as follows :

This work is licensed under Creative Commons Attribution-Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

Figure 1: Illustration of the four point clouds used for the sonification: a cube, a sphere, a pyramid and a cylinder. Units in the figure are arbitrary. We choose for our preliminary test (section 4) that one arbitrary unit is approximately 10 cm.

$$\left\{ \begin{array}{l} Freq [200Hz ; 2000Hz] \Leftrightarrow El[-70^\circ ; 60^\circ] \\ Amp [0dB ; -20dB] \Leftrightarrow Dist[0.5 ; 5] \\ Delay [0.5s ; 5s] \Leftrightarrow Dist[0.5 ; 5] \\ Sound Az [-62^\circ ; 62^\circ] \Leftrightarrow Point Az [-62^\circ ; 62^\circ] \end{array} \right. \quad (1)$$

where *Freq*, *Amp*, *El*, *Az* and *Dist* are respectively the sound's frequency, amplitude, the point's elevation, the sound's and point's azimuth, and the distance between the user and the point. The distance has no units as we work with artificial data. The limit of human horizontal binocular vision is used to define the mapping between the sound azimuthal position and the point azimuthal position. The limit of the human vertical field of view and the frequency sensitivity of the human hearing are used to define the mapping between the sound's frequency and the point's elevation position. We spatialize the sound only on the azimuthal plane with binaural first order ambisonic techniques [10], as the elevation clue is not accurate without an individual tuning of HRTF filters.

3.2. Insights about the sonification

The sonification process resembles granular synthesis where each point produces a grain which features depend on the position of the point. The result¹ is a whistling sound with the following features:

- Larger point clouds give longer sounds with wider frequency and amplitude coverage;
- Farther point clouds give weaker and delayed sounds;

¹sound examples available under:
<https://www.gel.usherbrooke.ca/rouat/soundsICAD2018.zip>

- Elevation of point clouds is encoded into the pitch of sounds;
- Azimuthal position of the point clouds are encoded with sounds with the same azimuth;
- Denser point clouds generate louder sounds.

Density and distance of point clouds both impact the loudness of sounds. Although the distance is also encoded within the delay, this cross dependence can be confusing for the user. To address this issue, we currently work on a method to reduce the number of points to be sonified. We do not discuss in this abstract the problem of perceptual limit when two sounds produced by two close points cannot be differentiated by the human hearing system.

3.3. Technical details

We use python with the libraries PyQT [11], Numpy and Transformation [12] to generate and manipulate the point clouds shown in figure 1. We then use Supercollider to read the point cloud data (stored as csv files) and generate sounds. We use the ambisonic toolkit [13] available in Supercollider (first order ambisonic) to spatialize the sound on the azimuthal plane. As there are thousands of microsounds to generate, we use non real-time synthesis techniques in Supercollider.

4. PRELIMINARY TESTS

We have so far tested our sonification strategy with blind coauthor François Coté (F.C.). The goal was to see which features of pseudo-scenes the subject could understand with the 3D raw data sonification prototype. Examples of pseudo-scenes presented to the subject are illustrated in figure 2. F.C. had no prior knowledge of the sonification strategy we used.

We first tested the object localization potential of the sonification strategy. For this purpose, F.C. had to explore by hand few pseudo-scenes comprising real objects. At the same time, we were playing the corresponding sonified artificial point clouds. After this short training time (2 to 4 minutes), we played the sound corresponding to a pseudo-scene F.C. didn't know and asked him to point towards the object. F.C. was able to perfectly point towards the object almost always on the first trial. We then put two objects in the pseudo-scene and F.C. was also able to perfectly point to them at the first or second trial while listening to the sound.

We then tested the potential of the sonification to evaluate the size of objects. This time, F.C. had to sort different objects by size. During training, we place in the pseudo-scene two real objects having different sizes. F.C. touched them while we were playing the corresponding sounds. F.C. then had to sort four objects of different sizes based on the sonified point clouds corresponding to the objects. F.C. was able to perform this task perfectly after listening one by one to the four objects.

Finally, we tested the shape comprehension potential of the sonification. F.C. had to figure out whether the shape was a cylinder, a sphere, a cube or a pyramid. This time F.C. was unable to succeed with this task. F.C. told us that he did not find a natural link between the sound and the shape he was touching.

Although we still have to conduct quantitative experiments with more subjects to confirm these results, this test gives us an indication of what we can accomplish with this 3D raw data sonification strategy.

Figure 2: a) example of two different pseudo-scenes seen from above b) corresponding artificial point clouds (visualized with python and PyQT) that were used to generate the sounds.

5. CONCLUSION

This extended abstract presents an ongoing work to design a 3D point cloud sonification prototype which is still in development. The initial tests lead us to believe that we are going in the right direction to encode the position and the size of objects with the sound. For the shape recognition task, a pre-processing step using more elaborated algorithms seems to be needed. Feedback from the first study has provided avenues we are currently exploring to better encode shapes through sounds. Further work also includes the development of an interactive near real-time system. This poses the challenge of point cloud data reduction without losing too much information in order to have a smaller number of points to sonify.

6. ACKNOWLEDGMENT

We thank the financial support from Québec FRQNT. We also thank the CIRMMT for the infrastructure support, Franco Lepore for the stimulating discussions and Etienne Richan and Luca Celotti for a partial proofreading. We thank the 2 anonymous reviewers for their constructive comments.

L.C. and J.R. conceived the prototype and conducted the experiments; L.C. made the programming, built the prototype and wrote the manuscript; J.R. and F.G. proofread the abstract; F.G. contributed to the sonification based on his PhD thesis and his artistic installation: *the Haptophone*; F.C. was the testing subject.

7. REFERENCES

- [1] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 2, pp. 112–121, Feb 1992.
- [2] Y. Zhong, P. J. Garrigues, and J. P. Bigham, "Real time object scanning using a mobile phone and cloud-based visual search engine," in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 2013, p. 20.
- [3] <http://taptapseeapp.com>.
- [4] <https://camfindapp.com>.
- [5] <https://www.aipoly.com>.
- [6] J.-K. Kim and R. J. Zatorre, "Can you hear shapes you touch?" *Experimental Brain Research*, vol. 202, no. 4, pp. 747–754, May 2010. [Online]. Available: <https://doi.org/10.1007/s00221-010-2178-6>
- [7] H. Duffau and C. Duchatelet, *L'erreur de Broca: exploration d'un cerveau éveillé*. Michel Lafon, 2016.
- [8] F. Grond, "Listening-Mode-Centered Sonification Design for Data Exploration," Ph.D. dissertation, 2013, Bielefeld University.
- [9] H. Pourghaemi, T. Gholamalizadeh, A. Mhaish, G. İnce, and D. J. Duff, "Real-time shape-based sensory substitution for object localization and recognition," in *Proceedings of the Eleventh International Conference on Advances in Computer-Human Interactions*, 2018, pp. 45–50.
- [10] M. Noisternig, T. Musil, A. Sontacchi, and R. Holdrich, "3D binaural sound reproduction using a virtual ambisonic approach," in *Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2003. VECIMS '03. 2003 IEEE International Symposium on*, July 2003, pp. 174–178.
- [11] <http://www.pyqtgraph.org>.
- [12] <https://github.com/davheld/tf/blob/master/src/tf/transformations.py>.
- [13] J. Anderson, "Introducing... the ambisonic toolkit," in *Ambiosnics Symposium 2009*, 2009.

THE AUDIO GAME LABORATORY: BUILDING MAPS FROM GAMES

Brandon Biggs

Lena Yusim

Peter Coppin

OCAD University, M5T 1W1
Toronto, ON
Canada
3164451@student.ocadu.ca

OCAD University, M5T 1W1
Toronto, ON
Canada
3164470@student.ocadu.ca

OCAD University, M5T 1W1
Toronto, ON
Canada
pcoppin@faculty.ocadu.ca

ABSTRACT

Audio games demonstrate an emergence of interactive parameter mapping sonifications that potentially optimally display geographical information and a large number of simultaneous data variables. Our preliminary investigation of audio games is in response to a call for more research on parameter mapping sonifications, such as the best way of presenting auditory legends for representations, effectiveness of spatial audio, map comprehension techniques, and finding optimal sonic variable mappings. We also present a proposed set of auditory map interfaces observed in audio games. Commercially available interactive interfaces and audio games – that have been shaped and informally “tested” by the selection pressures of a demanding consumer market – can serve as examples of potentially effective conventions informing future work in the auditory display research community.

1. INTRODUCTION

Audio games are computer games that can be played entirely with an audio interface. These games have existed since 1972 [1] and span the spectrum from completely speech-based to being almost exclusively based on non-speech sounds. Within the context of this paper, audio games are what we refer to as a “natural laboratory” – a site of investigation where extensive iteration in culture, driven by selection pressures, has refined a set of artifacts and/or conventions to what are likely to be effective states. Elsewhere, this has been referred to as “artifact evolution” [2]. Here, we focus on audio games because they demonstrate an emergence of interactive parameter mapping sonifications with the potential to optimally display geographical information and a large number of simultaneous data variables. Our preliminary investigation of audio games is in response to a call for more research on parameter mapping sonifications, such as from Walker & Mauney [3], Krygier [4], and Flowers, Turnage, and Buhman [5]. Commercially available interactive interfaces and audio games – that have been shaped and informally “tested” by the selection pressures of a demanding consumer market – can serve as examples of potentially effective conventions that can inform future work in the auditory display research community. Past research on audio games has been limited, focusing mainly on their applicability to navigation for visually impaired users and little focus has been given to how the affordances of audio games can be applied to data

representation. Through our preliminary investigation of audio games presented in this paper, examples are provided to suggest solutions to some of the more difficult problems in sonification. This exploration of audio game interfaces aims to expand the research presented in chapter 20 of the *Sonification Handbook* by Brazil & Fernstrom [6], and to encourage greater investigation into the potential value of audio games to the auditory display community.

2. BACKGROUND

There has been little research on utilizing audio game interfaces for data representations and mapping. For example, researchers have demonstrated that utilizing some of the techniques in audio games have enabled blind people to improve their sense of orientation and cognitive mapping ability in the physical world [7], [8], [9]. However, the existing research has not acknowledged the variety of auditory interfaces presented in audio games. Such discoveries within the audio game community may expose new opportunities, such as strategic uses of text-to-speech labels that have not been the focus of auditory display researchers. For example, developments in sonification have particularly focused on non-speech sounds [10]. However, for auditory mapping displays, the strategic use of text-to-speech is fundamental. Edler & Lamnert-Siepmann [11], who have researched the addition of an auditory dimension to cartographic maps, recommend understanding what makes current map types successful in helping to structure auditory interfaces. In visual maps, there are several purposes for using text when an image will not suffice. Titles and labels provide context, and are examples where the text is more useful than an image. Similarly, there will be instances in sonification where text, in addition to non-speech sounds, may prove useful [12].

2.1. Call for research and clues from audio games

Researchers of the auditory display community have raised questions about representing complex data sets through sound that audio games may have an answer for. Here are three examples: Krygier [4] asked how one would best design a sonic legend for a map. Most audio games employ *auditory icons* – a type of representation that bears an analogical resemblance to the process or activity being represented – to reduce the need for a legend [13], [14], [8]. An example of this may be found in the audio game *Swamp* [15], in which sounds of buzzing flies represent corpses containing loot, and sounds of footsteps paired with groans represent zombies. If a map key is required, such as in *Adventure at C*, an ever-accessible tree-based menu presents a name of the different

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

items, and a keypress, such as enter, will play a sound corresponding to the selected item [16]. The second issue identified by Krygier [4], and Flowers, Turnage & Buhman [5], called for research on the ability for users to identify objects in spatial audio. The aforementioned *Swamp* utilizes 3D audio, requiring users to center objects or creatures in their headphones through rotating their orientation, before either shooting, avoiding, or running over the object. Users have become incredibly accomplished at this task, being able to quickly navigate through complex and varied terrain while defending against dozens of enemies in the game world, all while searching for a single item [17]. A third question posed by Krygier [4] concerns whether blind people can build up maps while only being able to view a single point at a time. Balan, Moldoveanu, & Moldoveanu [8], and Sanchez et al. [9], found the highly immersive and attractive nature of audio games enables blind people to create a spatial representation of their environment that can be assigned to aid performing real-life navigation tasks. For example, strategy games such as *Tactical Battle* are precisely based around the user being able to know exactly what is happening on all parts of the map at once [18]. Flowers, Turnage & Buhman [3], Walker & Nees [19], and Walker & Mauney [5], comment that it is difficult to determine appropriate mappings between data and displays for blind audiences for several reasons, including the fact that many sonification studies are conducted with sighted participants and these studies may not account for perceptual differences due to expertise. However, given that the audio game community is mostly blind, all successful audio games have been extensively – though informally – “tested” by blind individuals [1]. Therefore, observing the data-to-display mappings in audio games can suggest generally accepted practices in the blind population. New types of experiments could perhaps be developed and deployed using games within the audio game community.

3. INTERFACES

The authors propose that audio game interfaces fit into five distinct map types: MUD style, tree-based, grid-based, side-scroller, and first-person 3D. Each of these interfaces have completely different methods of conveying the three levels of a digital interface [21]. The three levels are: physical input and output, WIMP (window, icon, menu, and pointer), and the game world itself [21]. The output interface of audio games is almost exclusively through headphones [22]. Each map type utilizes its own basic conventions, such as the mouse in first-person 3D, and typing on a keyboard in MUD-based maps [8], [23]. The WIMP interface is present in audio games, but should be changed slightly to replace “pointer” with “message” (WIMM) because audio games generally do not utilize a mouse pointer [24], [7]. The WIMM elements, which contain information about the game, stats, health, and exact location are mostly presented through speech [8]. Every aspect of the audio game experience consists of either: auditory icons, earcons, and/or speech [8]. “The game world is represented as a landscape, topography, objects, and inhabitants” [21, p.55]. Each of these game world components and their interactions are represented by one or more audio elements which could be combined to represent different levels of realism, such as varying the tambour of footsteps based on size and ground surface characteristics [25]. The most significant attribute of auditory maps is the representation of location, either as a set

of coordinates or as parent-child relationships within a tree-based map structure. The five auditory maps are taken from an analysis of existing interface design literature, past studies, the author’s observations through a decade of audio game play, and a growing database of annotated games. During the design process for interactive parameter mapping sonifications, the third step, after identifying the data and the message the interface should convey, is the choice of what interface to use. Almost every data set can be represented with any interface, so it is the purpose of the sonification that dictates what interface should be used.

3.1. MUD-based

MUD stands for “multi-user dungeon” or “multi-user domain” and is a completely text-based game [23]. Users interact with the world by typing commands on a keyboard, such as “north” to move north, and access system options through typing instructions such as “set exits on” [23]. MUDs are considered audio games because blind users can access their content through their screen reader which reads the output window and players will often add in custom auditory elements to represent events in the game world [26], [27]. Game world components are represented by short text descriptions such as “A black iron pot is here” after the room’s description [28]. Rooms have a title attribute, such as “Shrine of St. Wiseheart” [29]. MUDs can represent any type of data that can be put into text. They enable a user to zoom in on elements through commands, such as “look pot” in the aforementioned example, to get a detailed description, and can represent a high number of variables simultaneously. A disadvantage is the high level of linguistic competency required to understand the content [8]. Interfaces where users need a strong analogical connection to the content do not work well in MUD interfaces. The content is conceptually specific, so it will say “a large dog”, rather than making the sound a large dog would produce. The command line on computers is very similar to the MUD interface.

3.2. Tree-based

Tree-based maps are a set of parent-child relationships that show hierarchical structure [30]. The input to tree-based maps is generally comprised of four parts: move forward, move backward, move up a level, and move down a level. The only analogical relationship is hierarchical; every other representation is conceptually specific. Browser-based games and list-based games such as *Stryth* and *Crafting Kingdom* are organized as tree structures [31], [32]. Game world components are represented as nodes, either pages or menu items, and often with accompanying speech and auditory icons that play when the user selects an object. Almost every game utilizes a tree-based menu structure to allow the user access to actions such as start, exit, and save [25], [33]. One advantage of tree-based maps is that users do not need to remember all the data points in a set; they simply need to remember the location of the data they are looking for. Menus allow users to explore data without needing to remember order to navigate through the interface. However, trees do not represent distances between two nodes very well. They also tend to have a steep learning curve, and as the size of the tree structure increases, browsing through it becomes impractical [34]. Representations showing groups of data and lists of

points are best shown with tree maps. File explorers and most websites are tree-based interfaces.

3.3. Grid-based

Grid-based maps are based on a set of coordinates representing squares placed together in a column–row relationship [35]. The most common example of a grid interface is the spreadsheet. The user navigates through a rectangular array of tiles or cells, and components are placed in different cells around the rectangular array to represent the map content. The input interface generally utilizes the arrow keys to allow moving between cells, and keystroke combinations, such as Ctrl+Up arrow to jump upward through the tiles [33]. Game messages are frequently accessed through menus and by pressing letter keys on different tiles. Each game world component is situated in a tile and the user can move around the squares and view component locations in relationship to one another. Auditory elements of a tile’s components play when the user enters a new tile. Strategy games (such as *Tactical Battle*), where showing an overview of a map is important, employ grid-based maps [18]. Zhao et al. [7] found that grid-based maps allow blind users to easily explore unfamiliar geographical data and quickly understand adjacent relationships. Grids are ideally suited to allowing the user to get an overview of the data being represented [34]. Grids, however, do not show irregular shapes; users do not like the number of keystrokes it takes to move through the map; and it takes a while for users to understand everything that the map has to offer [7]. Many data sets, in comma-separated values (CSV) format, can be represented using text in a grid-based spreadsheet. However, adding non-linguistic sonic cues and brief speech messages allow multiple variables to be represented in a single cell.

3.4. Side-scroller

Side-scroller maps utilize a side view of the map where the user primarily moves left and right and analogical object sounds are represented using stereo panning [36], [37]. Games such as *Adventure at C* utilize a side-scroller map. Users mostly use the arrow keys to move through the map and access conceptually specific status information through menus and letter key presses, such as “h” for health [33]. Oren performed a study designing a complex side-scroller game with multiple levels and different heights of objects [37]. He demonstrated that three variables – height, position, and type – could be effectively represented simultaneously using pitch, stereo panning, and texture respectively. The conclusion was that side-scroller interfaces allow for quick recognition of several variables at once, and complexity beyond that leads to confusion. The interface is optimal for quick navigation through a high level of analogical detail.

3.5. First-person 3D

First-person 3D maps are characterized by the use of 3D audio and are the most studied of the audio game map types [8], [38], [25], [39], [9], [40]. The input interface to first-person 3D maps span the range of input devices, from “keyboard only” (as in *Top Speed 3*), to gyroscope as seen in *Swamp* with the “see monkey”, to GPS location as presented by Rober & Masuch [39], [41], [15]. WIMM elements are

both inside and outside the game world. The game world is incredibly detailed and tries to mirror reality [38]. The analogical game world components are positioned around the user and use short looping sound samples to let the user know where they are [25], [39]. Terrain and topological features are represented by audible thuds as the user hits impassible obstacles, and changes in footstep tambour or tone quality as the user moves over different surfaces [25], [15]. First-person 3D provides the most detailed and analogical interface out of the five map types, and can allow for many different variables to be represented at once. This makes first-person 3D perfect for data sets with many variables. The disadvantages of first-person 3D are the easy disorientation of the users and the length it takes to explore an interface.

4. CONCLUSION

Audio games provide a large set of potentially effective sonic interfaces that can be used in designing data sonifications. Researchers can shape their hypotheses based off what seems to work in audio games. If one has access to the data in an audio game, it would be possible to perform empirical studies on user tendencies. Currently most sighted users have had no experience with these interfaces and it may require some training before they become proficient. This means that there may be factors other than good design that may lead to a successful interface and there may be interfaces beyond what has been mentioned that may be employed in audio games. Designing data sonifications using the framework set forth by Jørgensen will allow a more dynamic and intuitive interface and expand the capabilities of existing sonification practices, such as those presented in Brazil & Fernstrom [21], [6]. The use of auditory elements, such as auditory icons, earcons, and speech should be ubiquitous in most sonic interfaces. Future research needs to validate and clarify the systematic application of the five audio game map interfaces to data sonifications. A larger set of audio games also needs to be evaluated to insure validity.

5. ACKNOWLEDGMENT

We would like to thank the members of our research group – Richard C. Windeyer, Uttara Ghodke, Annie Levy, Marta Wnuczko, Dan MacDonald, Thangam Natarajan, and David Steinman. We would also like to thank our industrial partners at Esri Canada for their support, specifically Dr. Brent Hall, and Jonathan Van Dusen. Thank you to NSERC, SSHRC, and Scotiabank for their generous support towards this research.

6. REFERENCES

- [1] J. Swarts, “A look at audio games, video games for the blind,” March 2016. [Online]. Available: <https://inverse.com/article/13331>.
- [2] D. Kirsh, “Explaining artifact evolution,” in *The cognitive life of things: recasting the boundaries of the mind*, L. Malafouris and C. Renfrew, Eds. Cambridge, UK: McDonald Institute for Archaeological Research, 2009.
- [3] B. N. Walker and L. M. Mauney, “Universal design of auditory graphs: A comparison of sonification mappings for

- visually impaired and sighted listeners,” *ACM Transactions on Accessible Computing*, vol. 2, no. 3, pp. 12:1–12:16, March 2010.
- [4] J. Krygier, “Sound and geographic visualization,” in *Visualization in modern cartography*, 1994. [Online]. Available: <https://makingmaps.net/2008/03/25/making-maps-with-sound/>.
- [5] J. H. Flowers, K. D. Turnage, and D. C. Buhman, “Desktop data sonification: Comments on Flowers et al., ICAD 1996,” *ACM Transactions on Applied Perception*, vol. 2, pp. 473–476, 2005.
- [6] E. Brazil and M. Fernstrom, “Navigation of data,” in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany: Logos, 2011, pp. 509–523. [Online]. Available: <http://sonification.de/handbook/>.
- [7] Zhao, C. Plaisant, B. Shneiderman, and J. Lazar, “Data sonification for users with visual impairment: A case study with georeferenced data,” *ACM Transactions on Computer-Human Interaction*, vol. 15, pp. 1–28, 2008.
- [8] O. Balan, A. Moldoveanu, and F. Moldoveanu, “Navigational audio games: An effective approach toward improving spatial contextual learning for blind people,” *International Journal on Disability and Human Development*, vol. 14, pp. 109–118, 2015.
- [9] J. Sánchez, M. Espinoza, M. de Borja Campos, and L. B. Merabet, “Enhancing orientation and mobility skills in learners who are blind through video gaming,” in *Proceedings of the 9th ACM Conference on Creativity & Cognition*, 2013, pp. 353–356.
- [10] T. Hermann and A. Hunt, “Guest editors’ introduction: An introduction to interactive sonification,” *IEEE Multimedia*, vol. 12, no. 2, pp. 20–24, 2005.
- [11] D. Edler and N. Lammert-Siepmann, “Approaching the acoustic dimension in cartographic theory and practice: An example of mapping Estonia(n),” *Meta-Carto-Semiotics*, vol. 3, no. 1, pp. 1–15, 2010.
- [12] B. N. Walker, J. Lindsay, A. Nance, Y. Nakano, D. K. Palladino, T. Dingler, and M. Jeon, “Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus,” *Human Factors*, vol. 55, pp. 157–182, 2013.
- [13] C. Peirce, “Logic as semiotic: The theory of signs,” in *Philosophical Writings*. Dover, 1902. [Online]. Available: <http://theory.theasintreas.org/wp-content/uploads/2013/02/Peirce-C-S-Logic-Semiotic.pdf>.
- [14] W. W. Gaver, “Auditory icons: Using sound in computer interfaces,” *Human-Computer Interaction*, vol. 2, no. 2, p. 167–177, 1986.
- [15] J. Kaldobsky. Swamp. [Online]. Available: <http://www.kaldobsky.com/audiogames/>.
- [16] VGStorm. Adventure at C. [Online]. Available: <http://www.vgstorm.com/aac/about.php>.
- [17] Hannibal, “Pattern for the strip mall,” in *Black Screen Gaming*. [Online]. Available: <http://www.blackcreengaming.com/swamp/warehouses/strip-mall/pattern/index.php>
- [18] I. Reed, “Tactical Battle,” 2013.
- [19] B. N. Walker and M. A. Nees, “Theory of sonification,” in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany: Logos, 2011, pp. 9–39. [Online]. Available: <http://sonification.de/handbook/>.
- [20] F. Grond and J. Berger, “Parameter mapping sonification,” in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany: Logos, 2011, pp. 363–397. [Online]. Available: <http://sonification.de/handbook/>.
- [21] K. Jørgensen, *Gameworld interfaces*. Cambridge: MIT Press, 2013. [Online]. Available: <https://ebookcentral.proquest.com/lib/oculocad-ebooks/detail.action?docID=3339717>.
- [22] audiogames.net. Audiogames, your resource for audiogames, games for the blind, games for the visually impaired! [Online]. Available: <http://audiogames.net/>
- [23] The Mud Connector. Getting started: Welcome to your first MUD adventure. [Online]. Available: <http://www.mudconnect.com/mud-intro.html>.
- [24] P. Christensson. WIMP definition. [Online]. Available: <https://techterms.com/definition/wimp>.
- [25] F. J. Friberg and D. Gärdenfors, “Audio games: New perspectives on game audio,” in *Proceedings of the 2004 ACM SIGCHI ACE*, 2004, pp. 148–154.
- [26] CX2. VIP MUD. [Online]. Available: <https://audiogames.net/db.php?id=VIP+MUD>.
- [27] I. Williams. (2016, Sep) How blind players made a text-only RPG more accessible. [Online]. Available: <https://waypoint.vice.com/enus/article/kwz77n/how-blind-players-made-a-text-only-rpg-more-accessible>
- [28] Rumble. The builder’s tutorial. [Online]. Available: <https://www.tbamud.com/getting-started-menu>.
- [29] Materia Magica. A world of text adventure is waiting for you. [Online]. Available: <https://www.materiamagica.com/>.
- [30] E. W. Weisstein, “Tree,” in *MathWorld*. [Online]. Available: <http://mathworld.wolfram.com/Tree.html>.
- [31] M. H. Yarrows. Sryth the age of Igtheon. [Online]. Available: <http://www.sryth.com/>.
- [32] MetalPop LLC. Crafting Kingdom. [Online]. Available: <http://metalphopgames.com/forum/index.php?forum/11-crafting-kingdom/>.
- [33] I. Reed. (2013) User guide for Tactical Battle. [Online]. Available: <https://blindgamers.com/Home/IanReedsGames>.
- [34] J. Tidwell, *Designing Interfaces*, 2nd ed. Sebastopol, CA: O’Reilly Media, 2010.
- [35] M. Soegaard. The grid system: Building a solid design layout. [Online]. Available: <https://www.interaction-design.org/literature/article/the-grid-system-building-a-solid-design-layout>.
- [36] “What is a side scroller?” in *Techopedia*. [Online]. Available: <https://www.techopedia.com/definition/27153/side-scroller>.
- [37] M. A. Oren, “Speed sonic across the span: Building a platform audio game,” in *Proceeding CHI EA ’07 CHI ’07 Extended Abstracts on Human Factors in Computing Systems*, 2007, pp. 2231–2236.
- [38] I. Podkosova, M. Urbanek, and H. Kaufmann, “A hybrid sound model for 3D audio games with real walking,” in *Proceedings of the 29th International Conference on Computer Animation and Social Agents*. ACM, 2016.
- [39] N. Rober and M. Masuch, “Leaving the screen: New perspectives in audio-only gaming,” in *Proceedings of International Conference on Auditory Display*, 2005, 2005. [Online]. Available: <http://hdl.handle.net/1853/50168>.
- [40] E. Klein and O. G. Staadt, “Sonification of three-dimensional vector fields,” in *Proceedings of the SCS High Performance Computing Symposium*, vol. 8, 2004.
- [41] L. de Ruijter, P. de Ruijter, B. Duveigneau, and D. Loots, “Top Speed 3,” in *Playing in the dark*. [Online]. Available: <http://www.playinginthedark.net/topspeed3>

e.php.

SIMPATY-HAND APP, A MUSICAL INSTRUMENT DESIGNED FOR PEOPLE WITH LIMITED MOTOR SKILLS

Carolina Valli

Electronic Music and New Technologies,
Conservatorio G. Verdi,
Milan, Italy
Carolina.valli@tiscali.it

ABSTRACT

Simpaty - hand App is an assistive music instrument realized with a smartphone that enables you to activate and modify audio sequences or rhythmic patterns with minimal movement of the arms or legs, suitable and designed for people with limited motor skills. Thanks to an internal accelerometer managed in Touch-OSC [1], the data on the movements of the limbs are transmitted and sent to a pc application developed in Max/Msp [2], via a wireless network. This app allows you to load audio files that will be reproduced based on the movement performed by simulating the use of a real instrument, with three different modes of operations: percussive sounds, riffs and loops.

was a lack of interaction and active participation, hampered by difficulties in movement and the impossibility of using traditional musical instruments. Inspired by the gestures used in interaction with a traditional instrument, such as a maraca, with the aim of maintaining the link between gesture and sound, the idea of using a common device such as the smartphone, able to take advantage of residual movements and adapt to most patients unable to make articulated movements. The realized instrument gives an extra possibility to those who, for the reason of accidents or illnesses, are not able to use any musical instrument but still want to express themselves and interact actively while participating to music performances. With this tool it easily becomes possible to control audio sequences or rhythmic patterns with a single and simple movement.

1. INTRODUCTION

Figure 1: Simpaty-hand App

The project was born as a bachelor degree thesis in collaboration with the center *Sim-patia* in Valmorea (Italy) [3], a structure that houses a day center rehabilitation and some housing units equipped for the needs of disabled people with severe disabilities. In the centre there were already music therapy activities that were often passively experienced by people with serious motor problems. There

2. CURRENT STATUS OF THE PROJECT

After initial theoretical research in the field of AMT (Assistive Music Technology), I started to design the Simpaty-hand app. I was inspired by a project of Perry Cook, the PhISEM Shaker Percussion that uses values of internal sensor to send MIDI commands to control a simple algorithmic jazz combo of bass, piano, and drums. In a simpler way I thought to use the internal sensor of the smartphone. The success with both adults and children, said Cook in his consideration about Principles for Designing Computer Music Controllers [4], in particular about the Frog Maraca, came from its simple interface (just shake it). Simpaty-hand App has been designed to help some of the members of the band present in the center in which I worked, to overcome their difficulties while performing with the band. They needed a musical instrument with an accessible and universal design, as Brendan McCloskey said when he describes design ethos in accessible design [5]. These are the main characteristics:

1. Wearable
2. Light
3. Easy and intuitive to use
4. Capable of using minimal movements
5. Which uses low-cost technologies

Hence the idea of using a smartphone which, thanks to its internal accelerometer, could easily detect data related to limb movements and send them to a PC application that has been developed with the Max/Msp software. It allows to load audio files and to play them according to the movements of the performer. The application provides three different playing modes for activating recorded music patterns such as percussive sounds, riffs and loops. An audio engine and a

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

calibration process are managed within the Max/Msp application on the pc. The prototype has been used by the band with positive results [6] and could yet be improved by developing an Android application that combines data acquisition and sound management on the same hardware and software platform in order to make a free use of an application for smartphone. It also remains to experiment it's use on a smartwatch with an Android operating system that would be even a lighter controller.

3. MAIN APP FEATURES

The main functions of the Simpaty-hand App are :

1. Data acquisition from the smartphone sensor;
2. System calibration according to the typology of gesture made by the performer with the device;
3. Data pre-processing (data filtering, onset detection, segmentation);
4. 3D acceleration data processing and mapping the final use and control of the audio engine.

To easily manage data acquisition, device calibration and audio synthesis, a graphical interface with pop-up menus has been developed.

4. GRAPHICAL INTERFACE

Figure 2: GUI with 3 main sections and remote controls on the smartphone

The graphical user interface is almost entirely based on pop-up menus to easily guide the user through a step by step configuration procedure. The interaction with the GUI requires fine motor control and it is usually managed by a caregiver or session organizer. It can be divided into three main sections.

4.1. Set up and calibration

In the first section you must set the communication port between the two devices that allows you to receive data from the smartphone and perform device calibration. Sensor values are displayed graphically and serve as feedback to verify whether the two systems are communicating correctly.

4.2. Choosing how to play and playing audio files

The application recognizes two different states based on the movements made on the device: an ON and OFF state. You can choose one of the three execution modes from a pop-up menu.

1. Solo mode: suitable for playing an audio fragment that will be activated (On) and stopped (Off) depending on the movement performed, such as when a key is pressed on the keyboard.

2. Sequence Mode: alternating the On/Off states it is possible to control the execution of an audio sequence (On) that can be interrupted (Off) and restarted from the point where it was previously stopped (On).

3. Duo mode: Alternating the two On/Off states, you can control two different sounds, suitable for different percussions .

For each mode you can use a default sound or upload an audio file in the window, dragging it with the mouse. Afterwards you can set the playback speed of the audio file and adjust its volume.

4.3. Additional Controls

Two additional controls are used to enable the remote control or to reset the application to its initial state. The remote control is possible thanks to the graphical interface developed in Touch-OSC that help to control from smartphone the volume, calibration or playback mode.

5. PRACTICAL CONSIDERATIONS AND USER EXPERIENCE

Simpaty-hand App was tested by Elena, a woman with paralysis of the right side of the body and limited left forearm movements. One of the problems encountered during the use of the device is the tendency to change movement during the execution of the same piece and therefore the need for new calibrations. For this reason, a motion filtering system was used that does not take into account sudden or undesired movements made within 240 ms. As far as latency is concerned, no particular problems were encountered using a hotspot connection between the two devices. As you can see in the video, the Solo mode was more suitable and in line with the needs of the band, used with the sound of maraca. The Duo mode has been experimented as an accompaniment to the rhythm of the drums and needs, as with any traditional instrument, Elena's learning of the rhythm to be performed according to the time marked by drummer. The Sequence mode was used as the basis on which the other band members improvised. From the feedback received and the tests carried out with the band has emerged the intention to continue using the Simpaty-hand both with Elena and to involve other guests, currently the app is available to the structure I worked with and it's still in the testing phase. For now, there aren't enough feedback or users who have tested it to carry out a survey or a more precise statistic of its use but remain the will to experiment and to make it known in the rehabilitation field as well as in music therapy.

6. SCALABILITY IN DIFFERENT CONTEXTS AND APPLICATION FIELDS

The instrument can be calibrated according to the motor skills and movement needs for each performer and can be used with a bracelet or without. In addition to being used as a musical instrument in groups or individually, as a

recreational activity, the application could be experienced as a rehabilitation aid. For example, to stimulate the accomplishment of painful or difficult movements. It could be less demanding for patients if rehabilitative activities were associated with sounds or audio tracks. Playing audio when the movement is correctly performed could particularly help children who would face rehabilitation as if it was a game. The main features of the tool such as simplicity of use, portability, adaptability, lightness, practicality and use of a common device such as the smartphone, makes it possible to use it in different contexts. In particular minor structures, with little money to invest in such as Schools, Rehabilitation Centers, Private, Associations, Hospitals, Leisure Centers, Health Residences for the disabled or Cooperatives.

7. POTENTIAL SOCIAL IMPACT

At a social level, the use of an instrument that can involve all the people who have more difficulties in music therapy contexts, concerts or during moments of recreation contributes to create a network of friendships, to promote communication and reciprocal knowledge while creating a framework of cooperation between patients, their families and the health staff. This type of aid can also be considered as a psychological support. Communities may create situations of interactions between all their actors which go beyond strict music therapy sessions. They can give birth to other musical realities, for example small bands that are involved in external and social activities such as concerts or events, increasing both self-esteem and well-being of all the involved people.

8. FUTURE DEVELOPMENTS AND ACHIEVEMENTS UP TO NOW

Simpaty-Hand App has been selected for the competition "Make to Care 2017" [7] organized by the Sanofi Genzyme Company among the top 10 finalists. It also has participated in the "Maker Faire" of Rome 2017 [8] and in the competition organized by the City of Milan "Call for Solutions: Open Innovation for Community Care" [9] within the Open Care project where it arrived among the top 3 winners. My future goal is to enhance the project and to develop an app for smartphone to be distributed in different facilities. I'm particularly interested in the rehabilitation field and I'd like to extend the original project by using more sensors to capture different and simultaneous movements, in order to control more sophisticated musical situations by introducing audio effects or acting on expressive parameters during performance.

9. REFERENCES

[1] <https://hexler.net/software/touchosc>

[2] <https://cycling74.com/products/max>

[3] <http://www.sim-patia.it/index.php>

[4] http://www.nime.org/proceedings/2001/nime2001_003.pdf

[5] Samuels K., (2014), "Enabling Creativity: Inclusive music interfaces and practices", in International Conference on Live Interfaces 2014, Queen's University Belfast, Northern Ireland

[6] https://drive.google.com/drive/folders/0B7_KtlzguBbUOU5xSFk4SGlsVGs

[7] <https://www.maketocare.it>

[8] <https://2018.makerfairerome.eu/it/espositori/?ids=1894>

[9] <http://opencare.cc>

A PRELIMINARY EVALUATION OF A SMART EXERCISE APPLICATION WITH WEARABLE SENSORS

Eric Vasey¹, Xingliang Li¹, Tejin Yoon², Myounghoon Jeon¹

Michigan Technological University
Mind Music Machine Lab¹, Neuromechanics Laboratory²
1400 Townsend Drive, Houghton, Michigan, United States
{emvasey, xli17, tyoon, mjeon}@mtu.edu

ABSTRACT

When doing lower leg exercises for strength training or to properly recover from a leg injury, people can encounter problems with boredom, lack of feedback on exercises, and potentially a worse patient outcome or incomplete recovery. We have previously developed and validated a Smart Exercise application with a wearable Bluetooth IMU, and we are now investigating the usability of the system and the usefulness, intuitiveness, and appropriateness of the auditory and visual feedback mechanisms. Our results indicate that the system has effective, intuitive, and appropriate auditory and visual feedback and shows potential for future improvement.

1. INTRODUCTION

The repetitive nature and lack of feedback while doing lower leg exercises for strength training can cause boredom and lower motivation to exercise. In the case of an individual with an injury, completing at-home physical therapy exercises is important for patients to make a full recovery. However, the boring, repetitive nature of these exercises and lack of feedback on performance can lower patient compliance and result in worse patient outcomes. In addition, physical therapists do not have access to any data on patient compliance or know if patients are performing their exercises correctly.

To address these issues, we have developed a Smart Exercise application which combines an Android phone and a low-cost Bluetooth IMU (Inertial Measurement Unit) from *Mbientlab* to track performance in forward lunges and squats while delivering real-time audiovisual feedback to improve a user's performance and motivation to exercise. Data are collected from the IMU and streamed to the phone at 100 Hz from the accelerometer and gyroscope, and at 30 Hz from the magnetometer. The data are passed through a threshold based filter that allows data through that exceeds 5% of at-rest calibrated values, then a Kalman filter is applied to remove sensor noise from the data.

Noh et al [1] conducted a pilot study of 13 participants to evaluate the functionality of the application and validate the data with a research grade accelerometer. The application performed as anticipated for detecting forward lunges and measuring participant

balance on a one-leg half squat task. Pearson's correlation coefficients were computed for the accelerometer and gyroscope data, and indicated a high correlation between our IMU and the research grade IMU ($r = 0.90$ to 0.99). A Bland-Altman plot was also utilized to validate the data, and indicated a low discrepancy between the IMUs.

1.1. Related Works

Spina et al [2] developed and evaluated an Android phone application to assist the fitness training and rehabilitation of patients with COPD (Chronic Obstructive Pulmonary Disease). The application utilizes the Android phone's on board accelerometer and magnetometer to generate an exercise model that personalizes the system to a patient and allows the patient to later complete exercises that are evaluated with regard to the prior model. The app provided auditory feedback for incorrect repetitions of an exercise, along with suggestions to improve. Evaluation indicated that exercise repetitions were classified at 96.7% accuracy, and the auditory feedback was effective in guiding participants to perform exercises correctly. However, this application required the phone to be on the user's arm, limiting the usefulness of visual feedback.

Rosati et al [3] investigated the impact of task-related, error-related, and visual-related feedback on a motion tracking exercise using a two degree of freedom joystick and touchpad. Their findings indicated that auditory feedback was more beneficial for the upper-body exercise task, visual task-related feedback improved performance significantly more than visual error-related feedback, and that auditory feedback on the motion of the target was better than feedback from moving the interactive device.

2. PROCEDURE

In this study, we conducted an evaluation of the usability of the Smart Exercise application and the appropriateness, usefulness, and intuitiveness of the application's auditory and visual feedback in its Lunge Piano/Lunge Sequence function. This function had users perform lunges at seven equally sized regions on a half-circle, with each region corresponding to a note on a C scale. Auditory feedback was given in the form of the corresponding note being played on the phone's speakers, and the region the user was facing was highlighted on a half-circle visualization (Figure 1). Lunge Sequences were series of notes that indicated the regions that the user needed to perform a lunge at in a particular order. These sequences were randomly generated and communicated to

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

the user via audio and visual feedback. For example, if the sequence was C-D-G-E, the user would need to face the region for C, perform a forward lunge, then rotate to face D, and perform a lunge, proceeding in this manner until they have performed lunges while facing all of the indicated directions in the correct order.

Audio feedback consisted of a text-to-speech voice instructing them which sequence notes/regions to perform lunges at. Performing a lunge at a correct region would play the corresponding note. Successfully completing a sequence was immediately followed by a preview of a new sequence. Performing a lunge at an incorrect region would play the corresponding note, then a series of three 100ms 311Hz tones that are 100ms apart, followed by a text-to-speech voice indicating that they were incorrect. Visual feedback was a half-circle visualization in the app which illuminated sections of the view corresponding to the notes/directions to perform lunges at. Performing a lunge at the correct region caused that region to fill with red, yellow, magenta, gray, cyan, light gray, or green depending on the region. Successfully completing a sequence would display 'Great Job! You completed the sequence correctly'. Performing a lunge at an incorrect region would display 'Sorry! That's not quite correct'.

Initially, participants were asked to complete the self-administered short version of the International Physical Activity Questionnaire (IPAQ). Then, they were asked to complete three sequences of four notes under three different feedback conditions - audio feedback only, visual feedback only, and both audio and visual feedback with order counterbalanced. In the audio feedback condition, the phone was taken from the participant, and they had to rely on their memory of the sequence and the ground tape markers (Figure 2) to indicate their direction. In the visual condition, the phone volume was muted, forcing the participants to utilize the visualization and messages to communicate the sequence to them and indicate they had completed a lunge correctly. If a participant performed a lunge at an incorrect note in the sequence, all subsequent notes were marked as incorrect, since the lunges needed to be performed in the correct order. Then, the error would be communicated to the participant, and a new sequence would be generated and provided to the participant. The app would log the sequence as well as the number of correct notes. After completing the three sequences under the three conditions, participants were asked to complete a questionnaire containing System Usability Scale (SUS) and the experimenter-designed questions.

Figure 1: Lunge Piano/Lunge Sequences Visualization.

Figure 2: Lunge Piano/Lunge Sequences Floor Markers.

Figure 3: Mean appropriateness, usefulness, and intuitiveness of audio and visual feedback.

2.1. Questionnaires

2.1.1. System Usability Scale (SUS)

The System Usability Scale consists of ten questions scored on a five-point Likert scale that aim to evaluate the usability of a system [4]. This is intended to evaluate the usability of the system as a whole.

2.1.2. International Physical Activity Questionnaire (IPAQ)

IPAQ is a questionnaire that is used to evaluate the physical activity level of an individual over a given time period and categorize it into low, moderate, or high levels [5]. We utilized the self-administered short version of IPAQ, which surveys physical activity over a seven day period.

2.1.3. Experimenter-designed Questionnaire

To investigate the app's feedback modalities, participants were asked to rate the audio and visual feedback mechanisms in terms of their usefulness, intuitiveness, and appropriateness for the given task.

Figure 4: Mean percent error for the Audio, Visual, and both Audio and Visual conditions.

3. RESULTS

A total of 18 participants (5 female, $M = 19.72$, $SD = 1.18$) completed the above procedure. The ten questions from the SUS were scored using the criteria in [4] ($M = 61.39$, $SD = 15.86$), and IPAQ was scored using the criteria at [6]. Five participants were excluded from IPAQ scoring due to incomplete responses. Of the 13 remaining, 6 were classified as having high levels of physical activity, and 7 as having moderate levels of physical activity. A paired samples t-test did not indicate any significant differences between the perceived appropriateness ($t = -0.809$, $p < 0.43$), usefulness ($t = 2.051$, $p < 0.056$), or intuitiveness ($t = 1.686$, $p < 0.11$) of the audio or visual feedback (Figure 3). A two-way repeated measures ANOVA on the mean percent error (Figure 4) did not indicate any significant differences ($F(2,30) = 2.355$, $p < 0.112$) between the audio only ($M = 52.78$, $SD = 28.01$), visual only ($M = 63.89$, $SD = 26.04$), or audio with visual conditions ($M = 56.48$, $SD = 27.79$). No significant difference in the mean percent error was found ($F(1,30) = 0.081$, $p < 0.778$) between individuals at a moderate physical activity level vs. individuals at a high activity level (Figure 5). The interaction effect for activity level and condition (Figure 6) was near significance ($F(2,30) = 3.187$, $p < 0.056$).

4. DISCUSSION

Utilizing the adjective rating scale developed by Bangor, Kortum, and Miller [7], the application's SUS score is between *OK* and *Good*. This indicates that we are on the right track in terms of the usability of the system, but additional improvements are needed.

While the results are not at the level of statistical significance, they are promising. The mean percent error for the audio only condition was lower overall than visual only, with audio and visual in-between. This may be due to the improved effectiveness of the auditory feedback in indicating a correctly performed lunge. In addition, the policy for errors was harsh - if a participant performed a lunge at any incorrect note or a note in the wrong order, all subsequent notes in the sequence were marked as incorrect, the

Figure 5: Mean percent error for Moderate and High activity individuals.

Figure 6: Interaction between IPAQ Physical Activity Level and Condition.

visual and audio feedback for an incorrect sequence would occur, and a new sequence would be given to the participant. This might also have influenced the SUS scores - participants might have rated the system lower due to its harsh policy on performance.

Participants rated both the visual and audio feedback as appropriate, useful, and intuitive (around 4 out of 5). The differences between auditory and visual feedback in usefulness and intuitiveness are approaching significance. Since the feedback was good overall, this could be due to the ceiling effect. Participants who were moderately physically active performed better with audio only feedback than those who were highly physically active. The more physically active participants may have been more confident in their ability to perform forward lunges correctly using their own technique. This might have impacted their compliance with what the application expected to be an appropriate lunge, and possibly proceeding with the sequence ahead of the application.

5. CONCLUSION

Based on our results here, we have a solid base from which we can improve the overall design and usability of the system. By adding additional tutorials to the application and iterating on the auditory and visual feedback, we can better guide users to perform their exercises more accurately and make them more enjoyable. We intend to conduct a similar evaluation to this one in the future with an updated version of the application and more forgiving failure conditions, and ultimately an in-home extended trial of the system.

6. REFERENCES

- [1] B. Noh, E. Vasey, K. Phillips, D. Verbrigghe, M. Jeon, and T. Yoon, "Cost-effective personal training aid to improve leg function using smart exercise application: Pilot study," in *Medicine and Science in Sports and Exercise*, vol. 50, no. 5 Supplement, 2018.
- [2] G. Spina, G. Huang, A. Vaes, M. Spruit, and O. Amft, "Coptdrainer: A smartphone-based motion rehabilitation training system with real-time acoustic feedback," in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '13. New York, NY, USA: ACM, 2013, pp. 597–606. [Online]. Available: <http://doi.acm.org/10.1145/2493432.2493454>
- [3] G. Rosati, F. Oscari, S. Spagnol, F. Avanzini, and S. Masiero, "Effect of task-related continuous auditory feedback during learning of tracking motion exercises," in *Journal of Neuro-Engineering and Rehabilitation*, vol. 9, 2012, p. 79.
- [4] J. Brooke, "Sus: A quick and dirty usability scale," <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>, 1996.
- [5] C. L. Craig, A. L. Marshall, M. Sjstrm, A. E. Bauman, M. L. Booth, B. E. Ainsworth, M. Pratt, U. Ekelund, A. Yngve, J. F. Sallis, and P. Oja, "International physical activity questionnaire: 12-country reliability and validity," in *Medicine and Science in Sports and Exercise*, vol. 35, no. 8, August 2003, pp. 1381–1395.
- [6] T. I. group, "Ipaq scoring protocol," <https://sites.google.com/site/theipaq/scoring-protocol>, November 2005.
- [7] A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," *J. Usability Studies*, vol. 4, no. 3, pp. 114–123, May 2009. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2835587.2835589>

EMOTIVE VOICE ACCEPTANCE IN HUMAN-ROBOT INTERACTION

Jaclyn Barnes, Erin Richie, Qiyue Lin,
Myounghoon Jeon

Mind Music Machine Lab
Michigan Technological University
1400 Townsend Ave., Houghton, MI 49931 USA
{jaclynb, enrichie, qiyuel, mjeon}@mtu.edu

Chung Hyuk Park

Assistive Robotics and Tele-Medicine (ART-Med) Lab
The George Washington University
2121 Eye Street, NW Washington, DC 20052 USA
chpark@gwu.edu

ABSTRACT

Social robots are expected to interact with people in a natural and socially appropriate manner. Often, this will involve communicating by speech. An experiment with two social robots investigated how participants accepted different voices used by the robots and how they interpreted emotions expressed in those voices. A humanoid and a dinosaur robot used recorded human voices and synthesized voices to express seven different emotions while commenting on fairy tales read by the participants. The preliminary results of this study suggest that monotone synthesized voices are not well-suited to emotion-rich interactions, while regular speaking voices and expressive character voices both have relative strengths. We hope that this preliminary study can spark lively discussions on emotive robot voices in the ICAD community.

1. INTRODUCTION

As robots develop more prominent roles as assistants and companions, the nuances of communication between robots and the people they interact with deserve careful consideration. While body language, gestures, expression, and other social cues can convey a wealth of information in social interactions, here we focus particularly on voice and emotion. The ability of a person to understand not only what a robot is saying, but what is meant by a statement based on context, word usage, and tone of voice is critical. Emotion is ever present in human interactions and, whether intended or not, will necessarily also be present in human-robot interactions. This research investigates the capability of participants to discern emotional responses in context from a robot, their understandings of these utterances, and their opinions of vocalization the robot uses to respond to the participant as they read stories. By investigating these factors, we hope to find valuable correlations that can help future companion and assistive robots required to interact emotionally to do so in a more meaningful and easily interpretable way.

2. BACKGROUND

There have been some studies related to acceptance of the sounds used by robots, though they have not received as much attention as physical appearance has. In a study comparing participants' textual responses to a mechanical robot using a low affect human voice, a synthesized voice, and beeping sounds, participants gave significantly more commands to

complete the same tasks under the synthetic voice condition than either of the other conditions. The researchers suggested that the robot with the human voice was assumed to be more capable than the synthetically voiced robot and therefore believed to need less instruction, while participants may have assumed that the robot that did not itself speak could not understand the speech of others [1]. Further supporting the idea that people anthropomorphize robots with human voices and relate to them with human expectations, a mechanical robot with a human voice was approached more closely than when the same robot was using synthetic voices [2]. Human voiced robots of the same gender as participants were anthropomorphized more than robots of the opposite gender or with a synthesized voice. Robots with a human voice were also considered more likeable [3], [4]. All of this suggests a general preference to interacting with robots similar to oneself. This is further supported by a study showing New Zealanders preferred interacting with robots that had the local accent rather than an American or British accent [5].

Coeckelbergh argues that including the appearance of emotion in robots is necessary to create suitably moral agents [6]. While emotional expression is complex, detection of emotion based on voice alone is possible, though not entirely reliable. Scherer's review of vocal emotion research found accuracy recognizing emotion in vocal portrayals tends to be around 60%, well above chance [7]. Our study is unique in focusing on emotional expression of robots and for including voice actor designed character voices (similar to what a voice actor would create for an animated character, for instance), which have more exaggerated emotional expression than typical speaking voices.

3. METHODS

3.1. Participants

Nine students of Michigan Technological University were recruited from the university psychology study participant pool. The participants included two females and seven males with an average age of 19. All participants self-identified as White/Caucasian.

3.2. Robots

We utilized two robots: Nao (Figure 1) and Pleo (Figure 2). Nao is a humanoid robot with rounded toy-like styling. Pleo is a dinosaur shaped robot with a cute, pet-like or toy appearance. Nao and Pleo were chosen as the robots to conduct the experiment in order to contrast the effect that robotic appearance has on voice acceptance. It is possible that participants may consider one voice more suited to a humanoid

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

robot than to an animoid robot or vice versa. A TV remote was used to control Pleo with IR signals and Nao was controlled from a nearby laptop over Wi-Fi. Participants were advised of the researcher's control of the robot during the consent process as the activation of the robots was conspicuous and thus the Wizard of Oz method was not effective in our particular set up.

Figure 1: Nao

Figure 2: Pleo

3.3. Voices

Different vocalizations were utilized to determine what impact, if any, a different voice or speaking style brings to interpretation and acceptance of emotional responses. A pair of young adult male voice actors with Midwestern American accents provided six of the voices used. Each actor recorded their lines in their regular speaking voice, a character voice envisioned for Nao, and a character voice for Pleo. Two synthesized voices from popular text-to-speech engines were also used: the Microsoft David voice and the iOS Alex voice. These conditions were chosen because they were perceived to cover a spectrum of expressiveness and appropriateness. A regular human speaking voice is a baseline that may be both expected of a robot and also highly emotive. In contrast, a character voice is likely to be very expressive, but may not be expected from a robot. Meanwhile, a synthesized voice may be expected of a robot, but may not be expected to be as highly emotive. These considerations of expectation versus emotiveness suggest a variety of possible interpretations of what is a suitable voice for a robot. As the research is still in progress, we will presently only consider results for one of the voice actors, Thomas, and the Alex synthesized voice, hereafter referred to as the human voices and synthesized voice for simplicity

3.4. Emotions

Seven emotions were tested in the study: happiness sadness, anger, fear, disgust, surprise, and anticipation. Six of the seven emotions are Ekman's six basic emotions and were chosen for their prevalence in psychology [8]. The seventh condition, anticipation, was chosen for its similarity to fear and surprise. Its inclusion allows us the opportunity to see if participants can discern an emotion that is not traditionally regarded as basic and to gauge confusion between emotions with subtle differences. When participants were asked to identify emotions, the questions were open-ended to avoid participants using process of elimination to "guess" the emotion from a list over their own intuition or expectation.

3.5. Scripts

The stories chosen were adaptations of the traditional fairy tales "The Boy Who Cried Wolf" and "The Three Little Pigs"

crafted specifically for this experiment. Fairy tales were selected because they naturally evoke a wide range of emotions that would be difficult to contrive in a game or other activity. Adjustments were made to the stories to accommodate certain emotions (e.g., an evocative description of the wolf was used to prompt disgust). These particular fairy tales were chosen for their similar cadence and structure, so while the emotions may appear in different orders, the stories themselves are fairly similar to minimize confounding variables. The exact words spoken by the robots differed in each story to be contextually relevant. The utterances are listed in Table 1.

Table 1: Robot utterances and corresponding emotions

The Boy Who Cried Wolf	The Three Little Pigs
That sounds nice. <i>(Happiness)</i>	I wonder what's going to happen! <i>(Anticipation)</i>
That's not nice! <i>(Anger)</i>	They shouldn't tease him like that. <i>(Anger)</i>
This should be good. <i>(Anticipation)</i>	He can't want to EAT them! <i>(Disgust)</i>
Gross! <i>(Disgust)</i>	Woah, that's fast! <i>(Surprise)</i>
What? Why didn't they help? <i>(Surprise)</i>	He destroyed their homes. <i>(Sadness)</i>
He's going to eat the sheep! <i>(Fear)</i>	Good. <i>(Happiness)</i>
All his sheep are gone. <i>(Sadness)</i>	Oh no! <i>(Fear)</i>

3.6. Conditions

The experiment was designed so that each participant would interact with all of the eight voices, two stories, and two robots, but not all combinations thereof. Alternating robots and stories, each participant went through four conditions with questionnaires, then completed a group questionnaire for those four voices, before repeating the procedure for the final four conditions and another group questionnaire. Each group of four conditions included all three voices (regular speaking voice, character voices for both Nao and Pleo) from a single voice actor and one synthesized voice. Conditions were counterbalanced between participants to counteract order effects. While the character voices were created with a particular robot in mind, we used them with both robots during the experiment for thoroughness. A number of participants ran out of time during the experiment and did not encounter all of the intended conditions.

3.7. Experiment

To begin session, each participant was welcomed and given a consent form. The experiment was explained verbally in addition to in writing within the consent form. Participants were also introduced briefly to the robots. The participants were given scripts for the stories indicating their lines and pauses for them to wait before continuing the story. At these pauses, the researcher activated the robot's response to the point in the story.

After each of the robot's responses, participants were asked what emotion they believed the robot was trying to convey. They were then asked to rate on a linear scale from 1 to 7 the clarity with which they felt the emotion was expressed and the suitability of the response coming from the robot. Each of the seven emotions appeared once in each story. After a story was completed, there was a set of linear scale questions evaluating

Table 2: Participant ratings of voice characteristics; ratings on a 7-point linear scale

	Likability		Attractiveness		Warmth		Honesty		Trustworthiness		Naturalness		Humanness		Roboticism	
Voice	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Nao Voice	4.10	1.60	3.50	1.43	3.40	1.07	4.40	0.52	4.30	0.95	3.20	1.48	4.00	1.25	3.10	1.10
Pleo Voice	4.10	1.66	3.50	1.27	4.00	1.63	4.70	1.16	4.70	1.34	3.00	1.33	3.50	1.08	3.50	1.08
Regular Voice	4.10	1.37	4.30	1.25	3.90	0.99	4.30	0.95	4.40	0.84	4.40	1.26	4.80	0.92	3.10	1.52
Synthesized Voice	3.18	1.60	2.73	1.10	1.91	1.04	3.82	1.17	3.45	1.29	1.64	1.03	1.36	0.92	6.55	1.21

properties of the voice overall. The process repeated for the next three conditions, with the robot and story alternated and the voice changed. At the end of four conditions, the participant completed a questionnaire that included basic impressions of one of the robots, evaluations of the voices, and demographics information. Voices were labeled A, B, C, and D respectively to avoid influencing participants. Finally, the procedure was repeated with four more conditions and the group questionnaire again.

The researcher conducting the experiment verbally asked all of the questions and noted the participant's responses in a form on a laptop rather than having the participant filling out the survey in writing, with one exception. The survey was coded with a participant ID and information about the setup of each iteration of the test (robot, voice, and story). With the researcher already at the computer to take notes and control robots, this design allowed researchers to record responses with key information or notes as needed. This design also required the questions be asked aloud, which created a more conversational atmosphere. We anticipated that participants would be more likely to give longer or more thorough answers when speaking as opposed to writing and could feel more comfortable asking clarifying questions.

There was a group of questions that were not done orally between the participant and the experimenter. Open-ended prompts for general thoughts on each of the voices were completed by the participants in writing when time permitted to allow us to directly quote participants.

4. PRELIMINARY STUDY RESULTS

On a seven point scale from strongly dislike to strongly like, participants rated the human regular speaking voice highest ($M = 5.11$; $SD = 1.45$), followed by the Pleo character voice ($M = 4$; $SD = 2.06$), the Nao character voice ($M = 3.78$; $SD = 1.64$), and finally the synthesized voice ($M = 3$; $SD = 1.41$). A single factor ANOVA comparing ratings approached statistical significance, but did not reach the conventional level ($F(3,32) = 2.48$, $p = 0.08$) due to the small sample size. Four participants listed the regular voice as their favorite, 3 listed the Pleo voice, 1 participant selected the Nao voice, and 1 the synthesized voice. Chosen least favorite voices were the synthesized voice (4), the Pleo voice (2), the Nao voice (2), and the regular speaking voice (1). In contrast, however, slightly more participants (4) believed the Pleo voice to be more emotionally expressive than the regular voice (3). Only 1 participant considered the Nao voice most expressive and no one chose the synthesized voice as most expressive.

The average likability scores of all human voices were equal and were greater than the average score of the synthesized voice. The synthesized voice ranked highest in average roboticness and nothing else. The Pleo voice was highest in warmth, honesty, and trustworthiness. The regular speaking

voice had the greatest average score in attractiveness, naturalness, and humanness. See Table 2 for details.

Since participants used their own words to describe the emotions they heard, similar emotions were grouped for the purposes of data analysis. For example, curiosity and excitement were combined under "anticipation". While the actors had been asked to portray seven emotions (anger, anticipation, disgust, fear, happiness, sadness, and surprise), participants' responses included 24 emotions after consolidation. The most common unintended response was "Neutral" with 30 occurrences, followed by "Confusion" with 13. The vast majority (22/30) of the "Neutral" responses were for the relatively uninflected synthesized voice.

The emotions with the best recognition across conditions was fear (68.29% accurate recognition), followed by anticipation (65.85%), happiness (65.85%), sadness (56.10%), anger (53.66%), disgust (43.90%), and surprise (34.15%). However, the emotion with the highest average rated clarity was sadness (5.04). The average clarity ratings of the other emotions were: anger (4.86), surprise (4.86), anticipation (4.85), disgust (4.78), fear (4.75), and happiness (4.69). The emotions of the Nao voice were accurately recognized most reliably (67.14% of the time), then Pleo (61.43%), regular (60.00%), and synthesized (35.06%).

There were some robot effects. Utterances from the Pleo robot (irrespective of what voice was being used) were accurately recognized 62.86% of the time, while those from the Nao robot were recognized 48.30% of the time. Conversely, the average ratings for clarity and suitability of correctly identified emotions were higher for Nao (4.86 and 4.90, respectively) than for Pleo (4.81 and 4.33).

Participants remarked that the Nao voice was "a little weird for the robot...", "goofy like a bad cartoon", and "very creepy, like a puppet voice". The Pleo voice was called "childlike", "almost like a kid", "sounded like a child", "very high pitched", and "annoying and high pitch". The regular speaking voice was repeatedly noted as sounding older than the other two human voices: "Older tone (early 30s)", "sounds like a young adult", and "very mature". The apparent age of the voice conflicted with the appearance of the robot and the task of the experiment: "I thought this voice did not really go with both the context and the robot itself." and "I didn't feel like that voice should be listening to this type of story". Some also remarked that it was "very clear" and "showed emotion". Finally, the synthesized voice was also considered clear, but monotone and robotic. "Liked that it was very clear but it was monotone and unemotional[.]" "The voice was very monotonous and robotic, and made it feel cold and uninviting[.]"

5. DISCUSSION

Despite a number of negative comments, the human regular speaking voice was the highest rated. Many of the comments were related to the age of the voice being discordant with the appearance of the robots and the usual age for fairy tales. This suggests that when rating the voices, participants may have been considering them largely independent of the context of the experiment. However, when context was a factor, it was significant.

Unsurprisingly, the synthesized voice was poorly regarded. While many participants considered it robotic both in their comments and in the ratings of roboticness, they were unhappy with the monotonous tone and lack of expression. What is stereotypically suitable, such a robotic voice, apparently is not a good choice for social robots in emotional interactions.

The Pleo voice was highest in warmth, honesty, and trustworthiness, perhaps because the voice was the most child-like. Meanwhile, the regular speaking voice was highest in attractiveness, naturalness, and humanness. They were also the highest rated of the voices, the top two favorite voices, and equally likable. This suggests that the voices could be used to evoke different feelings or impressions without significantly undermining acceptance. For example, using the Pleo voice might be more effective than the regular speaking voice in a situation requiring warmth and trust, such as comforting a nervous pediatric patient.

The emotion recognition scores present a rather unclear picture. The differences in recognition between emotions suggests that some emotions can be adequately conveyed by voice, but not all. This is consistent with the findings of Scherer [7]. To improve emotional communication, it might make sense to focus on improving non-vocal emotional cues (e.g., body language) for emotions that are poorly recognized by voice. There is no clear pattern in the data between the recognition of particular emotions and their rated clarity. Analysis of the effects of the individual robots is also similarly unclear. Given the relatively small sample size at this stage, it is hard to come to any firm conclusions regarding these ambiguous factors.

6. FUTURE WORK

Some participants' responses were based on the wording of the utterance rather than the tone of voice. Some instances of respondent suitability scores may also be tied to this effect of phrasing over tone. While it is certainly a phenomenon that will be inescapable in naturalistic applications, it should be kept in mind while analyzing the results of experiments involving a relatively small number of interactions and be considered while designing future studies as well. Designing the study with different responses for certain emotional cues may work to mitigate this effect's impact on the data.

Given the homogeneity of our current participant group, similar work with older adults, children, and people of color should be conducted. Additionally, based on the findings of [3], [4], it would also be advisable to examine gender differences in more diverse pool of participants. It would also be worthwhile to follow up with a comparison of male and female voices in the same context as all of the voices in this study were male. Finally, future versions of this study should track the English language proficiency of participants.

7. CONCLUSION

These findings are preliminary, so we should be cautious of drawing conclusions too strongly. However, in designing this study and working through the preliminary data, it is apparent that the area of properly designing vocal emotional expression in robotics is filled with intriguing avenues of investigation and possible implications. The field of social robotics continues to grow and with it the necessity for robots to be not only functional but emotive as well.

8. ACKNOWLEDGMENT

Our thanks to our volunteer voice actors, Thomas Klonowski and Robert Mampe, and director Patricia Helsel for so generously creating the voices used in this study. This project was supported by the National Institutes of Health under grant No. 1 R01 HD082914-01.

9. REFERENCES

- [1] V. K. Sims, M. G. Chin, H. C. Lum, L. Upham-Ellis, T. Ballion, and N. C. Lagattuta, "Robots' Auditory Cues are Subject to Anthropomorphism," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 53, no. 18, pp. 1418–1421, 2009.
- [2] M. L. Walters, D. S. Syrdal, K. L. Koay, K. Dautenhahn, and R. Te Boekhorst, "Human approach distances to a mechanical-looking robot with different robot voice styles," *Proc. 17th IEEE Int. Symp. Robot Hum. Interact. Commun. RO-MAN*, pp. 707–712, 2008.
- [3] F. Eyssel, D. Kuchenbrandt, S. Bobinger, L. de Ruiter, and F. Hegel, "If you sound like me, you must be more human," *Proc. 7th ACM/IEEE Int. Conf. Human-Robot Interact.*, p. 125, 2012.
- [4] F. Eyssel, D. Kuchenbrandt, F. Hegel, and L. De Ruiter, "Activating elicited agent knowledge: How robot and user features shape the perception of social robots," *Proc. - IEEE Int. Work. Robot Hum. Interact. Commun.*, pp. 851–857, 2012.
- [5] R. Tamagawa, C. I. Watson, I. H. Kuo, B. A. Macdonald, and E. Broadbent, "The effects of synthesized voice accents on user perceptions of robots," *Int. J. Soc. Robot.*, vol. 3, no. 3, pp. 253–262, 2011.
- [6] M. Coeckelbergh, "Moral appearances: Emotions, robots, and human morality," *Ethics Inf. Technol.*, vol. 12, no. 3, pp. 235–241, 2010.
- [7] K. R. Scherer, "Expression of emotion in voice and music," *J. Voice*, vol. 9, no. 3, pp. 235–248, 1995.
- [8] P. Ekman and H. Oster, "Facial expressions of emotion," *Annu. Rev. Psychol.*, vol. 30, no. 1, pp. 527–554, 1979.

INSTALLATIONS

PHOTONE – INSTALLATION PROPOSAL

Niklas Rönnerberg

Linköping University
Media and Information Technology
SE-581 83 Linköping, Sweden
niklas.ronnerberg@liu.se

Jonas Löwgren

Linköping University
Media and Information Technology
SE-581 83 Linköping, Sweden
jonas.lowgren@liu.se

ABSTRACT

We present Photone, an installation comprising an interactive combination of photographic images and music. An image is displayed and a dynamic musical score is generated based on the overall color properties of the image and the color value of the pixel under the cursor. Hence, the music changes as the user moves the cursor, creating a modal synergy where visual and auditory output combine holistically with the chosen interaction technique.

1. INTRODUCTION

The starting point for Photone was the use of musical sonification to augment information visualization. In such application music serve as a complementary modality to enhance performance on primarily visual tasks, and the hierarchy between these modalities is clear. However, when applying a similar approach to photographic images for hedonic experience, the hierarchical relationship is not as clear. Photone was designed to explore this relationship, i.e. the **modal synergy** between photographic images and music, further.

1.1. The composition

The composition is made to be soft and relaxing, yet with interesting harmonic and melodic changes. The musical sonification is determined by the overall predominant hue in an image, and the musical expression differs in terms of harmonics and melodic components between these hues. The intention is to create different impressions of, for example, a whiter image compared to a greener one. The music within each hue is then determined by the pixel value for each color channel (red, green, blue) beneath the mouse cursor and the complexity of the harmonic ambience, the use of melodic components, and the timbre as well as the sound level are determined by the pixel value (see Figure 1). This means that the music changes in accordance to the user's interaction, from slow and soft changes to faster and more pronounced changes.

A short video demonstration of Photone can be found here:
<https://vimeo.com/246964768>

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

Figure 1: Each image is classified to a certain shade of color and the sonification is adjusted accordingly. The RGB color channels have different chords building up the ambient sound as well as different melodic components. The rest of the musical elements (the bass tones, the high chord, the bell-sound for white, and the synth sweep for black) are consistent within a certain main color shade, and are dependent on the intensity levels for each pixel.

Figure 2: Example setup for the Photone installation, with an area for exploring Photone and a neighbouring area for poster and discussions.

2. CONTRIBUTION OF PHOTONE

The combination of images and music are far from new. The history is full of excellent examples of composed music for images and motion pictures. There are also different approaches to visualizing music and for turning an image into music. However, we believe Photone is more than just audification of pixel values or the playback of precomposed music to an image. The combination of carefully composed musical elements and the user interaction, where different hues and image components are used, creates something more and we characterize the resulting experience as one of **modal synergy**.

Photone differs also in regards to meaning and connotation. Traditionally, music is composed to images based on their denotative and connotative meaning. Our artistic intention in Photone is rather the opposite: By building the sonification upon pixel values of hue and brightness, that is, syntactic rather than semantic properties of an image, we aim to cut through conventional ways of seeing to a more foundational level. Other syntactic image features that would be interesting to use in further iterations of Photone include edges, shapes and spatial frequency measures.

3. ON A TECHNICAL NOTE

The implementation of the interactive sonification is done in SuperCollider 3.9.1, a programming environment for real-time audio synthesis. Even though SuperCollider simultaneously runs over 70 synthesizers, each consisting of many oscillators, playing multiple tones, Photone runs smoothly on a modern laptop. Due to a bug in SuperCollider 3.8.0 the green color channel was not read properly, therefore, the first version of Photone used individual text files for the color values for each pixel. Reading these text files took some time, causing a delay when changing images in Photone. This, however, was solved in the new version of SuperCollider, resulting in a faster and even better user experience.

4. THE RECEPTION OF PHOTONE

Photone was exhibited and well received at the science center (Norrköping Visualiseringscenter C) in connection to Linköping University. At the science center some visitors stayed and interacted with Photone and exploring the music as well as the use of images for a long time. However, not all visitors seemed equally

interested in Photone. A tentative conclusion is that Photone attracts only some of the visitors in a public exhibition space, but the visitors engaging more deeply in the interaction are drawn into a more or less captivating state of exploration.

"First of all, [I] would like to say that the exhibition [i.e. Photone] was an interesting experience! Seems to be an exciting research area. On Saturday's visit, I expressed to my company that I would have liked to be able to sit for hours and explore this".

A further developed version of Photone will be included in a exhibition about visualization at the science center in the autumn of 018. In this version all interactions will be logged for more detailed analysis.

5. PHOTONE, THE INSTALLATION

Visitors to the Photone installation will be using headphones when exploring Photone, and hence the installation is quite robust to surrounding noise, ongoing conversations in the vicinity, et cetera. A pair of speakers at a low sound level could be used to attract new visitors and to provide a glimpse of the ongoing user exploration (see an example setup in Figure 2).

It would be preferable if the lighting could be reduced in the installation area, in order to make the visual saliency of the images match the intensity of the headphone-mediated music and in order to create an inviting atmosphere. However, this is not an essential requirement.

There are no safety hazards in this installation. Sound levels in the headphones will be monitored to be within safe levels, and sound levels in the speakers will be kept low enough to provide a sound ambience without being loud or disturbing.

Equipment asked for from the conference chairs, if possible, are a 24" LCD display with HDMI cable, a pair of speakers (active monitor speakers, or passive speakers with an external amplifier) with audio cables (6.3mm TRS out from the audio device), and a space for a poster in A1 format (594 x 841 mm). Additional equipment such as a laptop, an external audio device (Universal Audio Apollo Twin), headphones, and a USB-mouse will be brought to the demonstration.

For the installation we will also bring a poster. This poster will essentially be based on Figure 1, which we think will be good for discussions about Photone, the composition, and the interaction between music and image elements.

CERATONOVA SHASTA

Jason Fick

Oregon State University
School of Arts and Communication
1650 SW Pioneer Place, Corvallis, OR 97331
jason.fick@oregonstate.edu

ABSTRACT

Ceratonova Shasta is a two-channel fixed media composition featuring the sonification of microbiome data. All sound gestures made in the piece sonify measurements of a parasite spore sampled from 2008-2016. This parasite (*ceratonova shasta*) resulted in particularly high mortality rates at an index site located in the mainstem of Klamath River, where it intersects with Beaver Creek in Northern California. The sound content and formal procedures are influenced by this single data set on both a macro and micro level. Peaceful and distressed sound characters represent the life and death of the fish within this delicate microbiome. The composition can be found at: http://jasonfick.com/Jason_Fick/Ceratonova_Shasta.html.

1. GOALS AND AESTHETICS

This piece is the first in a series of collaborations with the department of Microbiology at Oregon State University. The initial goal of the project was to create an artistic commentary on the parasite data collected for a concert performance in 2017 but has extended to include future sonifications as a method to learn more about parasite infection: to uncover patterns through sound that might not be seen with traditional methods.

1.1. Motivations and Composition Choices

This composition is entirely driven by the sonification of nine years of microbiome data. The sampled data measures the amount of parasite infection (*ceratonova shasta*) at an index site located in the mainstem of Klamath River, where it intersects with Beaver Creek in Northern California. In some cases, this infection resulted in extremely high rates of mortality of salmon at the index site.

I desired to depict the varying levels of mortality through aggressive and calming musical gestures. Each of these gestures is devised of several layers of sound characters intended to mimic the many layers one would expect to find in a large body of water. While a majority of the sound characters are busy in nature, a conscious effort was made to combine gestures that have great variety in spectral content, spatial positioning, and amplitude balance.

In addition, the narrative of the piece represents the life cycle of these fish over the testing period. I choose to project a calming mood during the points of low infection and a more aggressive and uncomfortable state during high mortality rates.

2. TECHNICAL DISCUSSION

An extreme amount of variability exists in the data collected from this index site. There are long periods of low infection, in which the values stay at 3.0 or below and other shorter moments when the values jump to 900 and above (in which 90% of the fish died). There are 447 index values ranging from 0.0 to 1092, with a majority of the numbers occupying the extreme registers.

Ceratonova Shasta uses this single data stream as its sonification source (parasite infection level), but it is applied on both micro and macro levels. On the micro level, each individual sound event in the piece is influenced by the data series, whether the set is directly applied to a sound parameter or series of parameters, scaled, or augmented through additional processing. In many cases, the data stream is applied in a linear fashion to various effect parameters to influence change overtime (grain density, grain length, filter cut-offs, pitch variants, etc.). The audio source material is comprised of processed samples that are constructed into sound characters that are then shaped into intricate textures. Example sound characters include low rumbles, busy high frequency chatters, dense rattles and rustles, water-like droplets, among others. In many cases, these sounds are used as motives and return during characteristic formal moments, namely during points of high and low mortality. The sound samples used are all of things breaking (egg shells, aluminum cans, tiles, and a milk carton). I built a tool to read, edit, and alter playback speed of the data in Max/MSP, as well as patches to scale, augment, and map the values to meaningful parameters. In some cases, the playback speed was varied to maximize artistic result. The sound modification tools consist of granulators, filters, and pitch shifters all built by the composer in Max/MSP.

In addition to the localized gestures being influenced by the data stream, the activity intensity and emotional essence of the composition are related to a linear timeline derived from the index values. The piece is five minutes in length. All 447 index values are addressed, with each new data point occurring every .671141 seconds. Using an activity timeline in this manner provided creative opportunities to design aggressive and active sound during moments of high mortality and less active and calming music during points

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

of low infection. The seven instances of high mortality (0:00; 0:23; 0:49; 2:54; 3:25; 4:06; 4:49) are represented by musical extremes in amplitude, textural density, harsh frequency content, and emotional content.

3. REVIEW OF LITERATURE/RELEVANT WORKS

Ceratanova Shasta seeks to present a creative and narrative representation of a particular data set of microbes. There is a growing trend in the makers and circuit bending communities to build devices and generate music based on microbiome data, such as Energy Bending Lab's, *Non Human Rhythms* and the various musical projects made by BiotaBeats. In both of these examples, the music is commercial in nature with driving beats in 4/4, repetitive harmonic progressions, and homophonic-dominated textures. My composition falls into the computer music and electroacoustic genre. Its treatment of subject material connects more closely to Stephen Taylor's sonification of genetic data, specifically in those works that feature electronics, such as *Indian Hedgehog*, *Writhe*, and *Sonic Hedgehog*. In each of these compositions, the composer sonifies multiple streams of DNA sequences through vivid aural manipulations and presents intricate sound textures and goal-driven gestural progressions.

A review of recent ICAD conference compositions show composers have sonified data from the stock market, atoms, spatial ambisonics, weather, and various other environmental data. *Ceratanova Shasta* is a hybrid in musical style between the sound mass-based sonifications found in pieces such as *Atom Tone* by Jirí Suchánek and Stephen Roddy's *Sonification: The Good Ship Hibernia* with the more gesturally-dominated shapes seen in Natasha Barrett's *Topology Chamber 2*.

4. USER FEEDBACK

Because the purpose of this composition was to present creative interpretations of a single data set, composer reactions and feedback to the values, and the initial perceptions of their sonifications was a crucial part of the design process. In many cases, feedback from early improvisations informed the current shapes of the sound characters. Upon preliminary hearings/improvisations, alterations were made to the mapping or processing of the data set to better suit the artistic aims of the project. Example modifications include altering the scaling factor or algorithmic processing, changing the reading/playback rate of the set, or simply applying the pattern to different parameters. Furthermore, additional feedback from others was taken under consideration during initial draft, specifically in the granular processing during the more aggressive sections.

Further projects are in development for the microbiologist in which sonifications can broaden our understanding of infectious spores.

WAVEFORMS

Michele Zaccagnini

Brandeis University

michelez@brandeis.edu

ABSTRACT

WAVEFORMS is the latest installment of my ongoing compositional research which aims at creating pieces that truly exist at the boundary of aural and visual senses. To achieve this perceptual blurring of senses I implement algorithmic compositions that can be rendered as sound and images synchronously. My approach to composition is tied to the world of visual arts in three different ways. Firstly, my musical aesthetic draws heavily from abstract expressionism: my music is textural and lacks inner contrasts, i.e. my pieces are “sound panels” rather than musical narratives based on the ideas of variation or development. Secondly, I use visual stimuli and processes to generate rhythmic structures and manipulate the piece’s *gestalt*. Finally, I incorporate some of the visual elements I have used to compose the piece as part of the piece itself, as a multi-sensorial accompaniment.

1. CONCEPT

Generally speaking, when multi-media composers are faced with the issue of connecting aural and visual elements they can work on a continuum of a spectrum that spans between “high level” and a “low level” perceptual connection between aural and visual elements. The former is an abstract connection: a scene from a film is accompanied by a soundtrack that musically suggests what kind of emotion the viewer “should” feel. At this side of the spectrum, the multi-media creator can work in procedurally loose fashion as she is allowed to freely associate sounds and images based on a general aesthetic idea. On the other hand, when working at the “low level” end of the spectrum the multi-media creator aims at creating a tight knit perceptual relation between sounds and visuals so that virtually *all data represented in the aural realm find a correspondent in the visual realm (and vice versa)*.

This approach comes with a fair amount conceptual and technical issues of its own. Conceptually, a group of issues to grapple with is the idea of *mapping* data into sounds and visuals. If, for example the goal is to represent a musical pattern (rhythm, pitches, timbre) how does one map musical data into visuals? Exploring different solutions to such problems and matching the results to their perceptual outcomes is one of the more compelling aspects of this kind of research. But, as my ongoing efforts have shown me, it is equally crucial for the multi-media composer to carefully choose programming tools and practices that are best suited to achieve just that aesthetic goal. Not dissimilarly from a traditional composer who adjusts her writing according to the instru-

ment or combination of instruments she is writing for, creating audio-visual compositions of the kind I have just described requires the multi-media composer to program the pieces in a way that is idiomatic to the chosen programming platform so that she can achieve a certain quality of outcome, crucial for a certain idea to perceptually come through.

2. TECHNICAL DETAILS

This installation draws from a series of tools and practices I have developed and differs from other kinds of audio-visual works in fundamental ways. As I was *zeroing in* the issue of “low level” connection between music and image I realized how the common approach to audio-visualization, commonly defined as “audio-reactive,” was proving excessively cumbersome and computationally expensive: the use of audio-reactive algorithms can majorly limit the possibilities of representing musical events in a discrete manner especially in the case of polyphonic music. Since my approach to composition is algorithmic at its core, I am able to bypass the audio-reactive step. by feeding matrices of raw musical data that I gathered from my algorithms into rendering.

3. PRESENTER’S BACKGROUND

I graduated with a Ph.D. in Music Composition and Theory at Brandeis University in 2014. My research focuses on generative algorithmic systems, interactivity and music visualization. In my recent projects I challenge traditional ways in which music is experienced by creating interactive platforms for immersive multi-sensory experiences. Along these lines of research, I developed CINC (Computer Interface for Neuro-Composition) an audio-visual interface that reacts to a user’s Electroencephalogram data feed generating audio-visual textures. I have presented this work at various conferences in England, Brazil and more recently at the MIT Science Museum as part of a Dinosaur Annex Music Ensemble commission for the Cambridge Science Festival. I am also working on several projects as a free-lance software developer on different platforms such as MaxMSP, Python and Javascript. I recently joined the *Consciousness Hacking* team to work on Group Flow: a project aimed at creating open-source sensor sonification software for meditative practices. I am currently developing software tools and practices to write generative audio-visual pieces.

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

OFFBEAT NEWS

Matteo Castiglioni

Conservatorio di Milano,
Via Conservatorio 12, 20100, Milano, Italia
casty1098@gmail.com

ABSTRACT

"Offbeat news" is a sound installation made of a radio that always sounds the same channel. This channel consists of a continuous rap flow made of news taken in real time from the network, these are analyzed and through some algorithms new ones are created, generating a situation in the balance between the real and the false.

The installation is made of a "Boombox" radio that was used by hip hop groups to bring their language to the streets. In the same way in the installation the radio is a means with which it is possible to transport a textual language from the virtual to the real

The installation plays on the concepts of truth and post-truth, questioning the real. The analysis of the news posted in real time identifies which are the terms and the most used categories at that time on the news-network, this allows to generate new informations that are in constant balance between what really happens around the world and what is pure fiction.

This puts the listener in a constant questioning situation, simulating what happens in everyday life by placing many people in a situation of compulsive control of the news and of what happens in the world.

CONCERT

FLUOR SONSESCENCE (FOR THE SONIFICATION CONCERT)

Christopher Jette

Center for Computer Research in Music
and Acoustics (CCRMA)
Stanford University
Palo Alto, CA, USA
jette@ccrma.stanford.edu

Mark Broshinsky

Manhattan School of Music
New York, NY, USA
mbroschinsky@msmnyc.edu

ABSTRACT

Fluor Sonescence is a work for trombone, electronics and video, where sounds are translated to video and the videos are translated to sounds. The video documentation includes the video and sound components of the work. Please see <https://vimeo.com/255790972/> to hear and view Fluor Sonescence.

1. PROJECT OVERVIEW

Fluor Sonescence explores the creation of a piece of concert music for trombone, electronics and video using the visualization of air masses from a brass instrument as source material. Smoke is placed into the air column of brass instrument during the production of a tone(s). As the smoke exits the bell of the instrument a laser illuminates a slice of the cloud and a high frame-rate camera captures the visual data (see figure 1). This visual material is then translated into sound and combined with the video. The process of constructing this work is a series of translations from sounds to visuals and then back from visuals to sounds. The resulting artistic artifact is a musical composition for trombone, electronics and video projection.

The larger context of this project is the creation of a real time instrument for controlling and producing both auditory and visual elements. The first step in this larger plan is the work Fluor Sonescence, where the composer both develops the techniques that will be incorporated in the forthcoming instrument and develops what will turn into a formalized translation process. This work serves as an opportunity to investigate the sonic and visual palette of working with embodied sound and types of translation techniques. Fluor Sonescence serves as a first investigation of the sonic and visual material that both extends a compositional trajectory of the composer and combines research in the aesthetics of mixing sonic and visual material.

2. MOTIVATION

The interpenetration of the translated material provides temporal and gestural continuity. The translations are mediated and not

Figure 1: A still from the video, illustrating gray and white smoke in a black frame

meant to strictly synthesize across media but rather to balance compositional intuition and expressive data. The basic musical material is generated by translating image to sound and these musical phrases are edited to create a musical composition. This approach of working with recorded data in a studio setting in order to create a finished composition is elicits the approach of Musique Concrete composers working with recordings on magnetic tape in a studio environment. With Musique Concrete, the transformations of sound resulting from manipulations in playback speed, tape direction and tape splicing define a series of constraints. For composers working exclusively with tape, the source material and the operations present a set of constraints which gives rise to a range of sounds. While the transformation in Fluor Sonescence are less extreme, with only the reordering of material and minor temporal expansions and compressions. Like the constraints of tape music, in Fluor Sonescence the sonification process defines the sonic palette and the organization of the material gives rise to a particular sonic palette.

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

Figure 2: A still from the premiere of Fluor Sonescence

The composer will diffuse the work. A detailed technical description can be provided upon request.

3. AESTHETIC GOAL

The goal of this work is to present the audience with gestural information through multiple streams, video, acoustic instrument and electronics. The interplay of these streams of information is informed with notions of counterpoint. The individual elements combine sensorially in parallel and contrary motion with points of articulation created by gestures of punctuation. The translation across media imbues each stream with similar gestural archetypes. The challenge in creating the work became a matter of scaling the raw information to a perceptually relevant range. To achieve meaningful translations, each gesture in each domain needed to be identified and considered for similarities and differences in the rendering. With this information, the combination of the gestures becomes a puzzle where each media type contributes to sense of an archetypal gesture in a different way.

4. CONCERT PRESENTATION

Fluor Sonescence is presented as a concert work with live trombone performer, processing of the trombone, fixed electronics and video projection. For the concert presentation we require a stereo audio projection system, video projection surface and access to theatrical lighting in order to frame the live performer while not detracting from the video presentation. The work uses live processing of the trombone so an ideal mixing station enables the blending of both the electronic stereo signal and live trombone.

TEMPLE OF NO RELIGION

Alessio Rossato

Music Conservatory “S. Cecilia” in Rome, Italy
Electroacoustic Composition Teacher
Via San Barnaba 29, 00176 Rome
alessiorossato8@gmail.com

ABSTRACT

Piece created with material recorded at Villa Smilea of Montale Pistoiese (Tuscany-Italy), where Andrea Dami's metal and sheet metal sculptures were sampled in various recording sessions including short improvisations. The sculptures were played in different ways, from the naked hands to wooden sticks, metal, rubber, with different thickness chains etc... and different types of microphones were used with "active" recording mode, searching different sound and position in the sculptures. The piece consists of five sections, the composition process is based on the concept of musical sculpture: from section I, four repetitions/versions have been made in which each one is based on precise focused music. The title is inspired by a building *Iglesia sin Religion* by Colombian architect Simón Vélez, a cathedral-temple built only with bamboo canes, simple material and with Attention to eco- sustainability and that contradicts, at least in this work, the monumentality and functionality of the building itself...

INTRODUCTION

1. PLACE AND MATERIALS

Piece created with material recorded during a short residence at Villa Smilea of Montale Pistoiese (Tuscany - Italy), where Andrea Dami's metal and sheet metal sculptures were sampled in various recording sessions including short improvisations.

Figure 1: Villa Smilea, Montale Pistoiese (Tuscany – Italy)

Figure 2,3: Andrea Dami's metal sculptures

2. SCULPTURES: PLAY AND RECORDING

The metal sculptures were played in different ways, from the naked hands to wooden sticks, metal, rubber felt or cobblestones or with different thickness chains, different types of microphones were used with "active" recording mode, searching different sound and position in the sculptures.

Figure 4: Recording session

3. SOUND ELABORATIONS AND STRUCTURE

Recorded material were treated with the following typical processes of early concrete music:

- . editing, simple editing and overlapping
- . stereophonic space distribution
- . transpositions
- . different file speed reading
- . reverse file reading
- . reverberation

The acousmatic piece consists of five sections, the composition process is based on the concept of musical sculpture: from section A, four repetitions/variations have been made in which each one is based on precise focused music, mainly in the elimination of material.

Figure 5: Temple of No Religion Sonogram

Figure 6: Temple of No Religion Sections

4. SUGGESTION

The title *Temple of No Religion* is inspired by a building Colombian architect Simón Vélez, and in particular at *Iglesia sin Religion* (a church without religion in Cartagena, Colombia) a cathedral- temple built only with bamboo canes simple material and with Attention to eco-sustainability a that contradicts, at least in this work, the monumentality a functionality of the building itself...

Figure 7: *Iglesia sin Religion*, Cartagena (Colombia) by Simón Vélez

5. PERFORMANCE ADVICE

The beginning of the piece is the strongest (*sforzatissimo*) point, it requires sound power to create an epidermal experience and immediately impose, to the listener, a mass material condition of the sound, which gradually dissolves in various shades along the course of the piece.

During the performance is recommended to diffuse a feeble light, tending to the dark.

LOUDSPEAKERS ARRANGEMENT

6. SUMMARY

Title: Temple of No Religion
 Autor: Alessio Rossato
 Year of composition: 2016
 Duration: 10 min. 46 sec.
 Number of channels: 2 or multiples
 Kind of composition: acousmatic
 First performance: November 13, 2017
 OUA Electroacoustic Music Festival 2017
 (略称: OUA-EMF2017) Art Information Center –
 Osaka University Of Arts

Temple of No Religion audio:

<https://soundcloud.com/alessiorossato/temple-of-no-religion>

Website: <http://alessiorossato.wix.com/home>

PLANETHESIZER: SONIFICATION CONCERT

Adrián García Riber

AGR Image and Sound Art,
Music Research and Sound Engineering,
Francisco Martí Mora 1-B, 22-3, IB 07011, Spain
adrian@imageandsoundart.com

1. MOTIVATIONS AND CONTEXT

The exponential increase on virtual instruments development and their popularity and establishment as part of the current workflow on audio productions, both in home studios and professional environments, open a new opportunity in approaching science to general public and designing audio tools based on physical or mathematical models of undoubted creative and artistic utility. Under this context, the publication of the discovery of seven Earth-like planets orbiting around the Trappist-1 star on February 2017, inspired the creation of a model-based virtual instrument. As result, the Virtual Interactive Synthesizer prototype *Planethesizer* is presented, which default configurations display a multimodal Trappist-1, Kepler-444 and a K2-72 planetary systems simulation. With the development of this astronomical data interactive sonification VST plugin, this work also tries to suggest a possible line of interdisciplinary audio tools development.

In an attempt to make a state of the art review in terms of astronomical data sonification and to analyze the technology, processes, tools and techniques that are currently involved in this field, results clear that Pythagoras (S. VI B.C.) and his *Music of the Spheres* as well as Kepler's *Harmonices Mundi* (1619), represent the crucial starting point of the relation between occidental music and astrophysics [1]. Highly influenced by this fundamental works, appear the animated sonifications of the solar system from Twyman (2010) [2], and Mark Ballora's sonifications for the film *Rhythms of the Universe* (2013) [3], a co-production of the ethno-musicologist and percussionist from the mythical band *Grateful Dead* and the cosmologist George Smoot III from *Lawrence Berkeley Labs*. Quinton, M. et al. [4] went a step further with their perceptual research on the solar system's sonifications created for the planetarium Explora (2016, island of Malta), proposing a planet properties through sound recognition test, realized on both specialized and non-specialized audience. In this sense of probing the effectiveness of the astronomical data sonification processes, also highlights the work of Diaz Merced, W. L. (2013) [5], the *Supernova Sonification* [6] and the prototype *xSonify* project [7], built in collaboration with NASA. Trying to equilibrate creative freedom and scientific accuracy to the original data sheet, it's also worth mentioning the solar wind data sonifications of Alexander, R. et al (2010) [8].

In a more creative context, Winton et al. [9] developed the sonifications of stellar data from Kepler Telescope (2012), using *Sonification Sandbox* software [10] and *Matlab's* interpolation functions to explore the aesthetical possibilities

of direct unprocessed conversions. Also interesting is the orchestral character coding sonifications of Quinn, M. for the exoplanets and stars information data base of the *European Southern Observatory* [11], and the mapping sonification work of Jamie Ferguson in collaboration with ESA *From Hipparchus to Hipparcos* (2014) [12], that translate several historic stars catalogues to sound in a way that reminds John Cage's *Etudes Australes*. Finally, it's impossible not to mention Tim Pyle's *Trappist Transits* musical representation (2017) [13], illustrating NASA's *Spitzer* Telescope data for informational purposes.

2. PLANETHESIZER CONCEPT

Planethesizer prototype was designed under the inspiration of Iain McCurdy's works [14], *LorentzSynth* (2015) and *Planet* (2012). It can be defined as a vector synthesis virtual instrument implemented from Csound's *Planet* opcode. This opcode or computing operation code, generates the coordinates (x , y , z) of a planet orbiting a binary star system simulation which outputs -calculated for each of the seven planets-, are used to control seven software synthesizers. The instrument also provides an additional eighth synthesizer, controlled through a real time MIDI keyboard that improves musical interaction. Csound [15] was used in combination with Cabbage front end [16] to implement a VST plugin and to maximize compatibility with any DAW (Digital Audio Workstation), both on Mac and PC computers. Blender software [17] was used for the graphic design of the star.

The prototype can be downloaded for testing from:
<https://archive.org/details/@agriver>

A musical composition using *Planethesizer* in a creative context is also available from:
https://www.youtube.com/watch?v=gUfA7kF_P0k&t=18s

In order to maximize intelligibility and simplicity of calculation, in *Planethesizer*, every sound appears linked to the graphic simulation of its planet, forming a multimodal representation that minimizes its learning curve.

Figure 1 shows the prototype interface that allow end-users to control sound properties as function of the radius, semi-axis, inclination, period and impact factor for the orbit of each planet or synthesizer, expressed in astronomical units. To reproduce the eclipse sequence cadence, each planet can be initial delayed introducing time differences in days. Level control is allowed in master output, keyboard's synthesizer as well as in every planet's module. The *Restart* button initializes the eclipses sequence and the *Init. Delay Off* button deactivate initial delay, allowing to synchronize the orbiting planets of the simulation for creative porpoises. The prototype also includes a global tempo, global zoom control and one *on/off*

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

button for each planet. Instrument default configuration represents a multimodal simulation of Trappist-1 system. Saving and recalling presets is also allowed.

Figure 1: *Planethesizer* interface illustrating the axes used in the simulation's graphic display.

3. SOUND DESIGN AND CAPABILITIES

Intensive testing with *opcodes* described in Csound's database [18] was made to achieve *Planethesizer*'s final design which required a compromise solution between processing CPU load and sound richness.

Table 1. Csound opcodes, synthesis types and initial delay (days), used for Trappist-1 System Sonification.

Planet	Opcode	Synthesis	Delay
b	fof	Granular	0,6
c	Vco2 Moogladder	Wavetable LPF	0,5
d	buzz	Harmonic sinusoids	0,9
e	poscil3	Hi resolution sinusoid	1,2
f	poscil3	Hi resolution sinusoid	0
g	poscil3	Hi resolution sinusoid	0
h	hsboscil	Wavetable	9,5
Keys	foscil	FM	-

The sonification process was oriented to provide a clear auditory distinction of each periodic sound event represented in the simulation with the exception of planets *b* and *c*, whose intended similarity was focused to test their graphical recognition through spectrogram analysis. Planets *e*, *f* and *g* use identical oscillators that can be tuned to desired frequencies modifying planets' radius or impact factor. In this way, a chord of three notes can be generated, allowing a "desirable" fusion of the sounds. Starting from identical time configurations for the three planets, a synchronized sound block is generated allowing to observe how the graphic representation of the prototype loses its usefulness despite the auditory display still provides a clear differentiation of the sounds that conform the chord. In a similar way, Cocktail party effect can be experimented representing the whole system or focusing attention to a specific planet. The prototype also allows Haas effect monitoring acting over initial delay on two unison sounding planets. *Chords*, *Arpeggio*, *Unison*, *Cocktail* and *Haas* presets were included in the prototype for these application. As a creative musical add-on, it's also possible in the prototype to generate real time arpeggios by introducing an initial delay or switching *On/Off* button on desired module. Musical figures can be made controlling planet's speed through orbital period parameters.

4. SIMULATED VS MEASURED PLANETARY SEQUENCE

Once the prototype was completely implemented, the sonification of the Trappist-1 system was recorded and analyzed to test its behavior. The spectrograms of *Figure 2* obtained with Sonic Visualizer software [19], show the audio file content in frequency (vertical axis) and time (horizontal), with time in seconds and a frequency range of 80Hz to 22Kz in the upper graph, and 10Hz to 1500Hz in the low frequency detail spectrogram.

Figure 2: Comparison between the simulated eclipses sequence and the Trappist-1 system telescope observations published by Guillon, M. et al., 2017 [20]. Both spectrograms (top and middle) are related to same audio file and its frequency axis was split for a complete recognition of sound events.

Analyzing *Figure 2*, it's possible to appreciate how the simulated sequence matches the behavior of the observed exoplanets system -assuming no precision-, fact that suggests the possibility of predicting transit moments from the simulation prototype beyond the observed data, which might be useful for informative purposes.

5. DEMONSTRATION PROPOSAL

For the public presentation of *Planethesizer*, a live creative Sonification concert is proposed. The performance would consist of a continuous musical piece of five movements that explore the prototype capabilities interacting with a drum set, piano and some effects. The music would be composed *ad hoc* for the conference. *Figure 3* shows the instrumental set up needed, initially planned to be reproduced in a quadraphonic environment with live video projection of the prototype.

Figure 3: Stage plot for *Planethesizer*'s demonstration concert.

6. REFERENCES

- [1] García Martín, R. (2009). La teoría de la armonía de las esferas en el libro quinto de *Harmonices Mundi* de Johannes Kepler. (Doctoral Thesis). Universidad de Salamanca.
- [2] Twyman, L. (2010). Retrieved on 05/01/2017 from: <http://whitevinyl.design.com/solarbeat/2010>
- [3] Ballora, M. (June, 2014). Sonification strategies for the film Rhythms of the Universe. *Proceedings of the 20th International Conference on Auditory Display*, New York, USA.
- [4] Quinton, M., McGregor, I. and Benyon, D. (July, 2016). Sonifying the Solar System. *Proceedings of the 22nd International Conference on Auditory Display*, Canberra, Australia.
- [5] Diaz Merced, W. L. (2013). *Sound for the exploration of space physics data*. (Doctoral Thesis). University of Glasgow.
- [6] Diaz Merced, W. L. (2014). *Supernova Sonification*. Retrieved on 0525/2017 from: <http://www.npr.org/sections/thetwo-way/2014/01/10/261397236/dying-stars-write-their-own-swan-songs>
- [7] Diaz-Merced, W. L., Candey, R.M., Brickhouse, N., Schneps, M., Mannone, J.C., Brewster, S. and Kolenberg, K. (2012). Sonification of Astronomical Data. *New Horizons in Time-Domain Astronomy Proceedings IAU Symposium No. 285, 2011. R.E.M. Griffin, R.J. Hanisch & R. Seaman, eds*.
- [8] Alexander, R., Zurbuchen, T. H., Gilbert, J., Lepri, S. and Raines, J. (2010). Sonification of AC Level 2 Solar Wind Data. *Proceedings of the 16th International Conference on Auditory Display*, Washington D.C., USA.
- [9] Winton, R., Thomas M. Gable, T. M., Jonathan Schuett, J. y Walker, B. N. (June 2012). A sonification of Kepler space telescope star data. *Proceedings of the 18th International Conference on Auditory Display*, Atlanta, GA, USA.
- [10] Walker, B. N. (2010). Sonification Sandbox (Version 6.1) [Software]. Atlanta, GA: Georgia Institute of Technology. Retrieved from: http://sonify.psych.gatech.edu/research/sonification_sandbox
- [11] Quinn, M. Retrieved on 05/01/2017 from: <http://drsrl.com>
- [12] Ferguson, J. (2015). *From Hipparchus to Hipparcos: A Sonification of Stellar Catalogues*. Retrieved from: <http://jfergusoncompsci.co.uk/research>
- [13] Pyle, T. (2017). *Trappist Transits*. Retrieved on May 24 from: <http://www.spitzer.caltech.edu/explore/blog/371-Making-Music-from-Exoplanets>
- [14] Iain McCurdy, accessed March 2017, <http://iainmccurdy.org/>
- [15] CSound software, accessed February 2017, <http://www.csounds.com>
- [16] Cabbage software, accessed January 2017, <http://cabbageaudio.com>
- [17] Blender software, accessed March 2017, <https://www.blender.org>
- [18] Vercoe, B. (May, 2017). *The Canonical Csound Reference Manual*. Retrieved on 06/22/2017 from: <http://csound.github.io/docs/manual/index.html>
- [19] Cannam, C., Landone, C. and Sandler, M. (2010). Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files. Presented in *Proceedings of the ACM Multimedia International Conference*, Firenze, Italia.
- [20] Gillon, M., Jehin, E., Lederer, S. M., Delrez, L., De Wit, J., Burdanov, A., Van Grootel, V., Burgasser, A. J., Triaud, A. H. M. J., Opitom, C., Demory, B., Sahu, D. K., Gagliuffi, D. B., Magain, P. and Queloz, D. (February 2017). *Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1*. *Nature*, Vol 542, pp. 456-460
- [21] <http://www.trappist.one/#plots>

AUDIOVISUAL COMPOSITION: THE CASE OF “ATLAS OF UNCERTAINTY”

Massimo Vito Avantaggiato

Conservatorio della Svizzera Italiana
Lugano (Switzerland)
massimo.avantaggiato@conservatorio.ch

ABSTRACT

This paper regards the interaction of electroacoustic music with visual elements, outside realtime hypotheses. In particular, this writing tries to suggest how analysis techniques for acousmatic music can help - without the presumption of being general or strictly objective - the creative process for achieving any audiovisual work.

In the following article we face the following issues:

- Audiovisual works, repetition and multilevel composition;
- The metaphorical relationships between video and audio;
- Semiotic Units, discretization and sense of form;
- On time management/ temporality;
- The spatial reevaluation of sonic and graphic details;
- The importance of coordination with analysis techniques for acousmatic music.

We identified some guidelines followed in the achievement of our recent work “Atlas of Uncertainty”, to exploit the potential of sound at different levels of structuration.

1. INTRODUCTION

Video offers the opportunity to create music in which the smallest sonic detail is important, turning seemingly ephemeral, or transitory details into something permanent and particularly meaningful. Outside real time, this allows us to reflect on our sounds, taking advantage of the possibilities of internal development diligently: we identify a sound, we experiment with it, researching its creative potential, innate and latent.

Video also allows composers to extend the meaning of a work, as it enriches a piece of music with extra-musical contents: poetic-literary; pictorial; biographical; descriptive and so on.

In this article we described the experiment of our audiovisual work “Atlas of Uncertainty”. The hyletic universe of this work is represented by heterogeneous inharmonic sound materials, in great part natural or electronically generated percussion sounds, explored through various techniques: granular, subtractive amongst the other. Four Classical elements, that typically refer to the Ancient Greece concepts of water, fire, earth and ether were instead used as presupposition and extra-musical factor around which the entire video was built: images were used indiscriminately and alternatively to handle very different sound phenomena.

The music that accompanies this computer - generated video is a sonic continuum, ranging from unaltered natural sounds

to entirely new sounds - or, more poetically - from the real world to the realm of the imagination.

The sounds and the images were combined in well-identifiable gestures.

2. A MULTILEVEL COMPOSITION

A first level of coordination was reached identifying repetition and variation processes at the musical level. But repetition, which is a major tool for bearing form in musical pieces, may be considered also at visual level. In this way repetition may concern and may lead to a contemporary exploration of the sound and visual detail.

The repetition processes in music can be identified at the level of the sound object (mini-structure), at level of rhythmic patterns (meso-structure) or at an higher organizational level (musical gesture and phrases, subsections, overall form):

Repetition at various levels, are often indicated implicitly by ; various analysis technique (Fig.1 and 2); only sometimes in a explicit way. Amongst these, we can cite: a) Temporal Semiotic Unit, which describes the Invariant Temporal Semiotic Unit by Repetition [1]; b) Roy's techniques [2], in his functional rhetorical analysis and c) the adaptation of Pierre Schaeffer's [3] Typo-morphology by Lasse Thoresen [4], amongst others.

Identifying repetition allowed us to create points of convergence or synchronization between visual moments and a particular sounds, between a movement of the image and sound morphologies. An example of synchronization can be a theme or a sound pattern that accompanies the same type of image. This sound creates a special atmosphere that recalls the image and reproduces itself each time this image appears. In the same way, the dynamics, the frequency, the speed of a sound can follow and reinforce the appearances and thematic variations of the image”.

2.1. Audiovisual composition: metaphorical relationships

Therefore the metaphorical relationship between video and audio in *Atlas of Uncertainty*, were not only relation of a) conformity or conformance [5], or Consistency [6]: in the sense that audio and image aim at having the same metaphorical relationship. In fact, in *Atlas of Uncertainty*, other channels, were activated: b) Complementarity (or complementation): one of the media expresses its potential dissimilar and complementary to the other; c) Conflict (or contest): the two media are in conflict with each other and compete to impose their meaning. In this way the "emerging" meaning of a certain section of the work, is to be understood as an interaction between the various components of the multimedia work [7].

This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

Some theories such as I-son theory by F. Bayle [8] suggests clear visual images - La nuit, L'eau, Le cri, Pulsation, L'air – and can be used to have a quick connection between images of graphic elements (water, air amongst other) and audio. A second level of coordination of sound material with images was considered for the making of Atlas of Uncertainty, and acts at the level of the following associations: a) on the basis of direct or indirect acoustic affinities - also derived sounds - of the starting audio material; b) on the basis of associations with the same class/genre of sound: electronic; natural/concrete; instrumental. c) on the basis of other formal or linguistic associations.

Analysis Technique	Repetition Process (Indexes)
Schaeffer (SCHAEFFER, 1998)	Redundant Sounds (Hx ; Hn ; Tn – Tx Special Weft)
Lasse Thoresen (THORESEN, 2007)	Accumulation; Iteration; Impulses
Vande Gorne – Figures d'Espace (VANDE GORNE, 2002)	Accumulation; Envahissement; Scintillement; Vague
M.I.M. – Temporal Semiotic Unit (http://www.labo-mim.org)	Invariant T.S.U. (Waves, Spinning, Obsessive)
Roy (Functional Analysis - Process) (ROY, 2003)	Intensification/Accumulation; Attenuation/Dispersion
Roy (Functional Analysis - Rhetoric)	Thème; Répétition; Affirmation; Articulation; Déviation
Roy (Functional Analysis - Stratification)	Mouvement
Roy (Functional Analysis – Orientation)	Extension; Prolongement; Engendrement
Bayle I-Son (BAYLE, 1999; 2000)	Le murmure; Le cri; Tapage; Pulsation

Figure 1: Analysis technique for acousmatic music and indexes of repetition. Sonic phenomena can have correspondents at the audio-visual level.

Analysis Techniques	Accumulation Processes (Indexes)
Lasse Thoresen (THORESEN, 2006)	Accumulation;
Vande Gorne – Spatial Figures (VANDE GORNE, 2002)	Accumulation; Envahissement
M.I.M. – Temporal Semiotic Unit (http://www.labo-mim.org)	Chaotic followed by suspension/ suspending; questioning
Roy (Functional Analysis - Process) (ROY, 2003)	Intensification/Accumulation followed by Attenuation/Dispersion
Bayle I-Son (BAYLE, 1999; 2000)	Le murmure; Le cri; Tapage; Pulsation

Figure 2: Accumulation Processes and Analysis technique for acousmatic music.

3. SOME GUIDELINES

Particularly, in the composition of "Atlas of Uncertainty", some guidelines have been followed:

1) Narrative and dramaturgical dimension prevail [9], avoiding trivial exact repetition on the two perceptive levels. That said, it was possible to manage the sound repetition phenomena, gradually changing the graphic element of the circumstance: the repetition (varied) at the audio level does not have a corresponding redundancy in the video material

2) The time has been managed in four different ways: a) Continuation: no change, stasis; b) Slow transition: moderate change; c) Quick transition; d) Juxtaposition: co-existence or simultaneous presence of elements. Particularly, in the second part of the video - from 4.00 to the end of the piece - fast transitions and the juxtaposition between different elements prevail. This obeys a precise narrative intention, indicating that we are going towards a stage of more development or

excitement;

3) Temporality is influenced by the way in which the sound sustains the scene, through its variations or irregularities; from the predictability of the sound on the image and the time. Timbre, frequency, amplitude, speed of sound had, in this context, an important role and a significant impact on perception. In this way we were able to enhance some internal characteristics of a sound;

4) We tried to balance structure and expression, knowing that discrete elements carry structure and continuous variation carries expression. We avoided to glide continuously through several dimensions of sounds and images, helping the listener to acquire much of a sense of form.

5) We tried to describe and emphasize spatial movement not only in the acousmatic piece, but also in the video, shifting the attention from general to details and vice-versa;

6) We tried to develop our video using suggestions derived from analyses techniques of acousmatic music. This effort of coordination suggested us new ways of interpretation of music, increasing the overall efficacy of the whole work.

3.1. Discretization and sense of form

Sculpting sonic material into gestures or phrases, which are the middle layers of musical structures - involves mediation between the nature of raw sound materials and the poetic of the composer: a mediation that is not always easy and it's not immediately successful. Things, moreover, become more complicated if we combine sound and images.

In our work, gestural organization of sound objects and of visual elements was intended as an intermediate level of "aggregation", which allowed us the creation of "discrete entities" that, sometimes, correspond to "semiotic units". This discretization process can guide the listenings and watching processes, helping the spectator acquiring much a sense of a form. To reaching this aim we particularly read the acoustic surface of the piece using techniques of segmentation proposed by MIM researchers, that, for the first time, talked about TSUs [10]: "TSUs are not Schaefferian sound objects because they are not isolated from their context according to the same segmentation criteria. The Schaefferian sound object is extracted from its context in accordance with purely gestalt laws and not with meaning units like TSUs, in which a minimal segment corresponding to a well-defined meaning is searched". These semiotic units, identified in the piece through MIM techniques, were coordinated, exalted or exaggerated by images, to be more easily recognizable and to allow a more immediate contemporary understanding of the sonic and visual fact. What J. McAdams [11] expresses for music, regarding discretization, is also valid in visual organization:

"We remember discrete entities easierly than continuous or unclearly demarcated ones, at least for the memory of structures. This does not mean that continuous variation is not important in the appreciation of musical form. It is certainly vital for expressive variation of musical gesture".

Discretization of sound and visual elements are aspects that allowed us to clearly divide the piece and the video into sections and sub-sections, also thanks to:

1) the aid of transitions 2) the aid of other sound phenomena suggested by the narrative evolution of the piece and labelled by analytical theories as «Apparition/Disapparition»; «Invasion»; «Fluctuation»; «Suspension/Resolution»; «Divergence»), favoring a clear reading of the acoustic surface [11] and of the visual one.

3.2. About temporality

Another important tool to combine sound object and video together, is a wise management of temporality. Sound has the ability to give a sense of linearity to what could, instead, be perceived as abstract on a visual level. This function can be easily found in many acousmatic videos, where the use of abstract images is very widespread. In *Atlas of Uncertainty* the management of temporality has followed different addresses [12]:

- the «temporal animation»: it's the ability of the sound to change the perception of the spectator, transforming it from "static" to "moving" and vice-versa;
- «temporal linearization»: sounds give a comprehensible logic to - apparently - disordered images and disrupted actions;
- «directional properties»: the sound has the property of leading towards a predefined objective, a plausible or comprehensible end. These three functions play an important role in the building of the weave; «directional properties», particularly, were used in *Atlas of Uncertainty* to mark the starting point and the end of panels/sections, with the aim to help the spectator to understand the structuring processes thought by the composer.

3.3. The shift of attention to details

The enhancement of sound detail is an element that contributes to produce added value in several visual works. The concentration on details is important for various reasons:

- to fill perceptive gaps of images;
- to enhance some internal characteristics of sound of particular interest, that is the way in which it evolves spectrally, or in amplitude, or in frequency;
- to give importance to spatial movement inside images, moving from the general to the local or vice-versa: in this way we avoid strict redundancy in a "spatial" sense: spatial movements and figures [13] are not only important in the musical piece, but can importantly regard the video section for the reason that we have told.

3.4. From acousmatic analyses techniques to visual structuring

In the management of various phenomena indicated by analytical techniques with terms such as «Suspension»; «Fluctuation»; «Invasion»; «Accumulation»; «Suspension/Resolution» we avoid one-to-one predictable relationship among sonic gesture and images. In this way the effect of the two elements – sounds and image combined together - is more powerful and expressive than that of each part taken together. In fact:

- 1) the phenomena of «Suspension» (Roy), of «Fluctuation» (TSU/MIM) are rendered through slow camera movements; sometimes, trying to give additional readings, through transitions and interferences, suggested by the internal movements of the acoustic substrate;
- 2) the chaotic phenomena of «Accumulation» (TSU/MIM) or «Invasion» («Envahissement» in «Figures d'Espace» analytical model), are rendered through faster camera movements;
- 3) the phenomena of «Appearance» and «Disapparition» («Apparition/Disapparition» in the Spectromorphology, and in Orientation Categories in Roy) are visually rendered through a suction of the images only in two points in the video, points that, obviously, have particular significance in

the story and on the dramaturgical plane. In other sections of the video, phenomena of «Apparition/ Disapparition» remain operative only at the audio level and do not have a visual correspondence, creating a clear conflict between sound elements and images;

Figure 3: Element Earth (excerpt): progressive exploration of sound and images details.

Figure 4: Element Earth and gliding moon (excerpt). In the second part of the “Atlas of Uncertainty”, fast transitions and juxtaposition between different elements prevail.

Figure 5: Acousmographie: screen shots in sequence. The study of recurrence of visual events help the memory of structures.

- 4) «Suspension» (TSU/MIM) or «Resolution» (Spectromorphology: structural functions) phenomena are, instead, rendered by means of slight disconnections or transitions, or even by means of "combinations" of sudden images, which represent an element of novelty in the visual path. Sometimes, they are the result of an editing operation, of a compositing process through which an image is superimposed on / in a previous image;
- 5) The phenomena of «Chaoticity» or «Divergence» (T.S.U./M.I.M.) can, moreover, be expressed through several levels of reading on a graphic level or through the deeper exploration of the image details.

4. CONCLUSIONS

In an era in which the representation/notation of electroacoustic pieces is effectively difficult, due to the lack of universally accepted notation systems that could surely help to understand a piece, well-built visuals could help to lead to a more complete discourse on a compositional work, highlighting and emphasizing many aspect of a composition. Video can explore details, concentrating on spectral evolution of certain sound-object or other sonic aspects at ministructural level; it can illuminate on gestural or phrase organization which are the middle layers of musical structure; video can clarify the subdivision of a piece, exalting transitions, or other sound phenomena, labeled by analytical theories as «Apparition/Disapparition»; «Suspension/Resolution»; «Divergence», just to name a few; it can awakens the attention of the spectator and ease or deny a certain directionality of music; it can also allow a more rapid storage of information by the spectator. In other words, well-built audiovisual works can help to have a better understatement of the overall structure of a piece and can highlight important aspects regarding mesostructural, ministructural and macrostructural organization.

The goal of our work “Atlas of Uncertainty” was to create a good balance between perceived structure and expression, exalting and coordinating both the directionality of music and images.

Audiovisual composition allows to increase the possibilities of communication of the music as it is thought, thanks to the novelty of keeping the original message of the composer intact. In this way, video is not a passive instrument for a possible re-presentation and integration of sound materials, but shows a potential organizational function.

5. REFERENCES

- [1] <http://www.labo-mim.org>
- [2] S. Roy, *L'analyse des musiques électroacoustique. Modèles et propositions*, Paris, France: L'Harmattan, Univers Musical, 2004, pp.1-590.
- [3] P. Schaeffer, *Traité des objets musicaux*, Paris, France: Editions du Seuil, 1996.
- [4] L. Thoresen, Spectromorphological, “Analysis of Sound Objects: an adaptation of Pierre Schaeffer’s Typomorphology”, *Organized Sound*, vol.12, n°2, pp.129-141
- [5] E. Carterette, R. Kendall, “The communication of musical expression”, *Music Perception*, vol. 8, n°2, 1990, pp. 129-164.
- [6] G. Lakoff, M. Johnson, *Metaphors we live by*, Chicago, U.S.A.: University of Chicago Press, 1980, pp.1-50.
- [7] N. Cook, *Analysing Musical Multimedia*, Oxford, UK: Oxford University Press, 1998, p.98.
- [8] F. Bayle, “Image-of-sound, or i-sound: Metaphor/metaform”, *Contemporary Music Review*, 4, 1989, pp. 165-170.
- [9] J.A. Garavaglia, “Music and technology: what impact does technology have on the dramaturgy of music?”, *Journal of Music and Meaning*, vol. 7, n°2, 2008, pp.1-8.
- [10] A. Bregman, S. McAdams, “Hearing Musical Streams”, *Computer Music Journal*, Vol. 3, n°4, 1979, pp. 26-43.
- [11] D. Smalley, “Space-Form and the Acousmatic Image”, *Organised Sound*, vol.12, n°1, 2007, pp. 33–58.
- [12] M. Chion, 2005, *Audio-Vision: Sound on Screen*. New York, U.S.A.: Columbia University Press, p. 230.
- [13] Vande Gorne, “L’interprétation Spatiale. Essai de formalisation Méthodologique”, *Revue DEMèter*, Université de Lille, pp.1-6

SNOWFALL

Daniel Grayvold

Composer, Sound Designer, Musician.

daniel@dgrayvold.com

1. INTRODUCTION

Snowfall is a meditation on the interpretation of winter weather as a depressive event. A hybrid musical composition/audio installation, its purpose is to sonify the snowfall of a winter season in the Keweenaw county of Michigan, portraying the sadness, isolation, and loneliness that can accompany the season for its residents.

2. FOUNDATION

The season of winter is often portrayed in a joyful or celebratory manner. Popular culture is rife with stories of Christmastime merriment when friends and family come together, of children playing in the snow, and of the landscape being made beautiful by a white blanket of precipitation. This is easily seen in such works as Claude Monet's painting *The Magpie*, Bing Crosby's song *White Christmas*, and classic films such as *A Christmas Story* and *National Lampoon's Christmas Vacation*. For many, this is the feeling that they most commonly associate with that time of year.

For many, however, the season of winter feels completely different. Inclement weather can cause accidents and close roadways, leaving stranded many who have limited means of transportation. The lack of sunlight can harm those with Seasonal Affective Disorder, suffered by many who live in climates—such as Michigan's Keweenaw county—where winter stays longer than it is welcome. The arrival of weather that brings snowfall is therefore frequently seen negatively, an occasion for anything but celebration.

I wanted to create a work that exemplifies the distinct feeling of dread that can come with the falling of snow in the winter. The goal was to give those who may

have never felt this desperate loneliness a sense of what it can feel like.

3. CREATION

The composition of *Snowfall* began with the decision to create an installation that used snowfall data from the Keweenaw county of Michigan—a location that receives a heavy amount of winter precipitation every year and my home for a number of years.

County officials record the staggering amount of accumulated snow every year and make it available online. I was able to gather an up-to-date dataset of snowfall from the 2017-2018 season to form the first version of the piece. This data included daily snowfall, ground accumulation, and to-date information from October of 2017 to April of 2018.

With a good sample of data to work with, I created a patch in Pure Data, a multimedia visual programming application. The first step I took was to create the base scale that the song would be played on. I intended for the piece to cause listeners to feel two emotions: a sense of isolating loneliness and a constant desire to get to the end of this somber feeling. The first emotion would be created by selecting notes from a minor scale - F melodic minor in my case. To achieve the second feeling, I started the scale on the dominant note (C4) and only allowed for the tonic note (a note which gives a feeling of stability because of its nature as the base of a scale) to be randomly chosen 20% of the time.

From there, I needed to create a way to choose notes from this scale randomly. This part was relatively straightforward; I used the built-in random object to choose a number up to 10—the number of notes in the used scale. This would select a value from a list of the ten MIDI note values and

3. The scale that *Snowfall* uses to attain a depressing, always-unresolved timbre. The tonic (F) appears only twice out of a possible ten notes.

trigger a MIDI note at velocity 30 and that value.

Next came the incorporation of the snowfall data. I created an array and populated it with the snowfall data I had gathered for a total of 174 data points. This involved converting HTML information into a CSV file (accomplished by using a standard text editor) and feeding it into Pure Data through its textfile object. This method allows for future expansion with more data as it becomes available.

With the data imported, the last step on the technical side of the composition was to determine how the data would be converted into music. As mentioned earlier, this piece needed to be depressing, giving one the feeling of isolation. There also needed to be contrast between days in which no snowfall was recorded (usually indicating better weather) and days in which snowfall was recorded. After some experimenting, I found that using a slow tempo and using the daily snowfall count in inches as the number of notes played worked well. I split each day into measures of 4/4 time lasting 4000 milliseconds, or 1 second per beat—translating into 60 beats per minute. The 4000ms were then divided by the snowfall accrual in inches rounded

up. This would give the time needed for that amount of notes to be played in a single measure. As an example, a data point of 2 inches of snow would mean that two notes would be played during that measure with a 2000-millisecond delay between the notes. If there was no snowfall during that day, there would be no notes played.

With that set up, it was a simple matter of using a metro (metronome) object to trigger the random note player whenever needed. I also added a single note to be triggered at the beginning of each new measure; this enabled the audience hear the transition to each new day. The triggered MIDI notes were routed out of Pure Data and into Logic Pro X, where the final step of the composition was to take place.

The last part of the puzzle was to create the sounds that the composition would use for the played notes. I decided to use a sampled harp played with a flageolet technique since its hollow, haunting timbre perfectly fit the emotion that I was aiming for. It needed a little more to it, so I also created a drum-like hollow bell synth sound that would play along with the harp. These together created a sound that was eerie and desolate, giving the impression that it was being played from a great distance.

Date	Snowfall	On Ground	Month	Season
2017-12-30	6	22	81	120
2017-12-31	7	24	88	127
2018-01-01	3	24	3	130
2018-01-02	1	22	4	131

2. A sample of snowfall data from the 2017-2018 season in the Keweenaw county of Michigan. This data includes the snow fallen on that day, how much was on the ground, and the running total for the month and entire season. [1]

Going forward, this piece could be expanded in a few different ways. For one, more snowfall data can be added as it is recorded. New instruments could be incorporated to increase the emotive impact of the piece; a visual element could even be added to aid the aural experience.

4. CONCLUSION

Snowfall is a composition that is simple but extremely effective. Its sonification of actual snowfall data allows the listener to get a sense of the depressive feeling that can often accompany winter weather. Its ability to be presented as an installation that passerby can listen to or as a musical piece performed by a solo harpist or keyboardist gives it flexibility in reaching new audiences. This flexibility allows audiences to get a new perspective of the season of winter that they may not have otherwise ever experienced.

4.1. Media

A sample recording of *Snowfall* can be found at dgrayvold.com/features/snowfall.

5. REFERENCES

1. Keweenaw County. "Keweenaw County Snowfall Amounts." Keweenaw County Online. Accessed 20 April 2018. <http://keweenawcountyonline.com/snowfall2.php>.