

INVESTIGATIONS IN COARTICULATED PERFORMANCE GESTURES USING INTERACTIVE PARAMETER-MAPPING 3D SONIFICATION

Natasha Barrett
University of Oslo, Department of Musicology
natasha.barrett@imv.uio.no

Kristian Nymoen
University of Oslo, Department of Musicology
kristian.nymoen@imv.uio.no

ABSTRACT

Spatial imagery is one focus of electroacoustic music, more recently advanced by 3D audio furnishing new avenues for exploring spatio-musical structures and addressing what can be called a tangible acousmatic experience. In this paper we present new insights into spatial, temporal and sounding coarticulated (contextually smeared) gestures by applying interactive parameter-mapping sonification in three-dimensional high-order ambisonics, numerical analysis and spatial composition. 3D motion gestures and audio performance data are captured and then explored in sonification. Spatial motion combined with spatial sound is then numerically analyzed to isolate gestural objects and smaller coarticulated atoms in time, space and sound. The results are then used to explore the acousmatic coarticulated image and as building blocks for a composed dataset embodying the original gestural performance. This new data is then interactively sonified in 3D to create acousmatic compositions embodying tangible gestural imagery.

1. INTRODUCTION

In electroacoustic composition composers record a wealth of sounds and use these as sources in their work: dissecting, transforming spectra, time and space to create the building blocks of composition. Rather than being concerned with refined instrumental techniques, recording and its creative use are guided by physicality, acoustics and kinetic behavior. In this way, spatial imagery has developed hand in hand with electroacoustic composition and more recently, composers' interest in 3D sound.

To gain greater insight into the potential of gesture in the formation of spatial-temporal images, we propose a new approach relevant to a wide variety of performed sounds. 3D motion gestures and audio performance data are captured and first explored with interactive parameter-mapping sonification in three-dimensional high-order ambisonics as a way to identify significant features. Spatial motion and spatial sound are then numerically analyzed to isolate gestural objects, smaller coarticulated gestural atoms and their connectivity rules. The results are sonified to verify the results and then used in the composition of a new fictional dataset embodying the original spatial-gestural performance. This dataset is then explored in sonification as a performance and compositional tool. The work is sonified using 'Cheddar' [1], which has been developed over a number of projects in conjunction with both scientific and artistic sonification needs. The method, results and further work described in this paper apply an analytical and rigorous approach to some ad hoc assumptions suggested in [2].



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. METHODS

2.1. Source sounds and data recording

Instrumental performers acquire a refined control over motion that is less obvious in a non-musician's action-perception cycle. For this reason, our work focuses on 'non-instruments' more familiar in electroacoustic music, the performance of which also stimulates investigation and yields surprise. We chose a balloon as our non-instrument, where the action-sound language consists of a variety of spatial and spectral dynamics.

Audio was recorded with five DPA 4015 cardioid response microphones, four arranged in a rectangle with diagonal of 80 cm and one elevated above the centre. The balloon motion occurred mainly inside this microphone array. Motion data was captured using the Qualisys optical motion-capture system and eight Oqus 300 cameras at a rate of 200Hz. Six markers were placed on the balloon and 27 detailing various points on the fingers, hands and upper body. Two contrasting recordings were chosen in developing our analysis method and to provide the first results presented in this paper: (a) 'Bouncing': involving balloon and body large motorics; (b) 'Slip-Grip': involving balloon and fingers micro-movements.

2.2. Sonification

Cheddar is an interactive parameter-mapping 3D spatial sonification program built in MaxMSP and described in [1]. Cheddar sonifies multiple 3D spatial datasets in high-order ambisonics (HOA) where the virtual listening position can be freely moved to explore the spatial world in real-time. Sound is transformed by the data with a flexible, user-defined mapping. Parameter mapping sonification is important in our work: data acts as a layer of detachment from the original sounding event, thus avoiding any multi-modal inferences that may mislead the investigation, as well as allowing modulations in time and space which may clarify qualities hidden at the original tempo. In all sonification examples velocity is mapped to volume and vertical motion mapped to pitch shift. Accompanying examples are in binaural for headphones, originals are in 5th order 3D HOA (www.natashabarrett.org/ICAD2015/)

2.3. Data analysis

In our study, we consider *gestural-spatial images* consisting of sound, excitation action and other performance motions that proceed and follow the sound. From this combined motion and audio image we are interested in isolating phrases (a number of small sound-spatial objects linked together), sub-phrases (different phases in the phrase that may be separated in some way) and coarticulated elements (elements that contextually smear into the sub-phrase). [3] discusses coarticulation temporal frameworks and [4] analysis options. Although using these as a guide, our framework focuses on the temporal-spatial characteristics of our sound-source. Our phrases were selected aurally by evaluating the mix of data sonification with the original audio.

Sub-phrases and coarticulated elements were studied with numerical analysis of the audio and motion data: Audio from the five microphones was gated to remove low level noise, band filtered and the RMS at each microphone used to reveal general changes in the direction and location of the audio image. Motion data was reduced to the balloon, thumbs and middle fingers. Features extracted were absolute velocity and acceleration, 2nd derivative of distance between finger and balloon centre and the rotational speed of the balloon. Each feature vector was labelled positive when exceeding a threshold (set as mean value of the vector). When 50% or more of the feature vectors returned a positive value, this signified a segmentation point of some kind. The resulting boolean vector was filtered with a simple moving average spanning a temporal threshold of 130 ms. Thresholding the moving average allowed extracting sound-gesture units based on both spatial and temporal parameters. When the time between units was less than 130 ms we would assume the unit to be coarticulated. Values greater than this would mark a sub-phrase.

3. RESULTS

3.1. Main phrase, sub-phrases and coarticulated elements

Sound example 1a is the Slip-Grip audio recording as an ambisonic image in front of the listener. Sound example 1b sonifies two balloon and two finger markers from Slip-Grip in the same spatial area as example 1a and mixes 1a into the image. The same is repeated for Bouncing in example 2a and 2b. We hear the relationship of the physical motion in relation to the balloon sound, yet also how perceptual segmentation of the total image is different from that when assessing the solo balloon image.

Figure 1 shows the analysis from Slip-Grip phrase 1, for a reduced set of parameters. In the forth section, blocks indicate identified units of sound or motion information. If we impose the 130ms threshold temporal value we can extract coarticulated atoms and sub-phrases, as shown in the sixth section. When the spacing of atoms was greater than 130ms, this would mark the start of a new sub-phrase (emphasised with grey top line).

3.2. Macro and micro-movements analysis

We saw that when large movements of the balloon produced loud sounds, large body movements were involved; small movements of the balloon producing loud sounds involved micro body movements. This suggests different sets of markers are appropriate for different types of sound-motion correlation - a trend that can also be used as a rule influencing the choice of sonification scaling, especially volume and spatial scaling.

3.3. Connectivity rules and composing a new dataset

By looking at the spatial-temporal displacement between the atoms and sub-phrases we can calculate connectivity rules. Connectivity rules would normally be derived from the complete recording, but as an example we will focus on just phrase 1 shown in figure 1. Consecutive atoms are spaced a maximum of 0.22cm with a maximum duration of 1.165 seconds; sub-phrases are spaced a maximum of 1.28cm with a maximum duration of 4.575. Following these rules we can make a fictional data set capturing aspects from the original performance. Example 3a sonifies Slip-Grip phrase 1, repeating and slowing down for two balloon and two hand markers. The listening location is placed central to the motion activity as a way to enhance the spatial-gestural image. Connectivity rules are used to create a new dataset, sonified in example 4. When atoms are

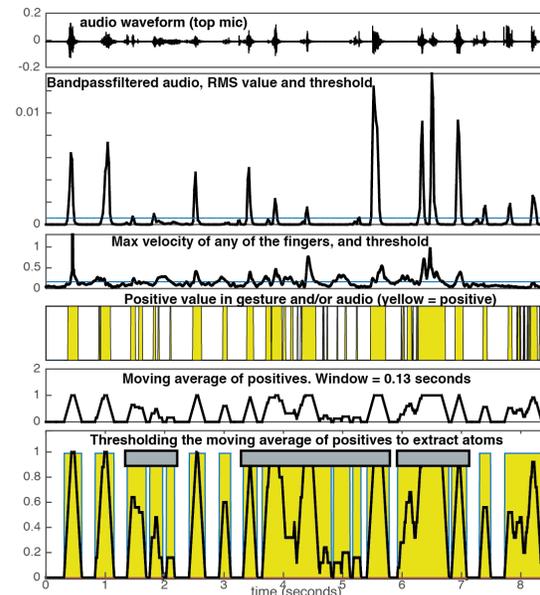


Figure 1. Slip-grip phrase 1. See text for details

further apart than the spacing rule they are translated within the proximity threshold. Also, some milliseconds of ‘padding’ are allowed as a spatial cross-fade between atoms. A rule controlling this padding duration is being investigated.

4. DISCUSSIONS AND ONGOING WORK

We saw no correlation between the balloon’s spatial audio analysis and the motion data, but this is an interesting area to investigate. We are also considering analysis methods such as canonical correlation analysis for marker and parameter selection, phrase-transition parameters and threshold values. Creative work will focus on how sonification can be used to explore spatial-gestural coarticulation, how temporal and spatial scaling influences our perception of each type of gestural unit, and how the results can be used in the composition of 3D spatial imagery along side other sonification and 3D audio techniques. Most importantly we need to undertake listener tests to establish salient features in spatial-temporal gesture, the degree of connection or abstraction from the original source performance and a general understanding of 3D motion-gesture imagery.

5. REFERENCES

- [1] Barrett, N. Design and application of ‘Cheddar’: an interactive 3D parameter mapping sonification programme. In review. www.natashabarrett.org/Cheddar.pdf 2015.
- [2] Barrett, N. Creating tangible spatial-musical images from physical performance gestures. *NIME’15, Louisiana State University, May 31-June 3, 2015*. In Press.
- [3] Godøy, R. *Understanding Coarticulation in Musical Experience*. IN Sound, Music and Motion. M. Aramaki et al. (Eds.): CMMR 2013, LNCS 8905, pp. 535–547, 2014. Springer 2014.
- [4] Caramiaux, B., Wanderley, M., Bevilacqua, F. Segmenting and Parsing Instrumentalist’s Gestures. *Journal of New Music Research* 41(1): 13-29. Taylor & Francis 2012.