

A FORMULATION OF HOLOPHONIC DISPLAY

Conor O'Sullivan, Mikael Fernström

Interaction Design Centre,
Department of Computer Science and Information Systems,
University of Limerick, Ireland
conor.osullivan@ul.ie , mikael.fernstrom@ul.ie

ABSTRACT

This paper presents the results of a project that aims to develop an algorithm to synthesise a natural auditory phenomenon. The aspiration was the construction of a listener's holophonic experience of these auditory events, to develop a sonification that was both realistic and accurate, as a reflection of how we hear our own environment [1].

The natural aural stimulation in the virtual environment is that of rain hitting leaves. This sound is synthesised and then processed to create a type of holophonic display so that physical, psychoacoustical and perceptual aspects of the spatial localisation are accounted for. Research that has been subsequently accomplished in the areas of digital waveguide techniques, head-related transfer functions (HRTFs), mathematics, signal processing and digital audio is assembled and applied to the formulation of this algorithm. An implementation in the Pd (Pure-Data) software package is also offered [2].

1. INTRODUCTION

Pushing the boundaries of sound synthesis has long been the aim of composers, music technologists, scientists and anyone with an interest in the sonic domain. For thousands of years, music makers have created and explored new musical instruments, using their craft to excite the harmonics of countless unheard timbres. As the previous century entered the Digital Age, the field of sound creation extended itself naturally to this new realm. In the new realm, however, the instrument remains physically static, while its virtual shape evolves in a Heraclitean manner. The computer has become both obstacle and springboard- it defines the limitations yet provides the opportunity for infinite creation. The advent of digital signal processing (DSP) has witnessed impressive milestones in the world of science. With speeds currently permitting millions of operations every second, increasingly the only real limitations of digital audio are those imposed by the creativity of the mind. This creativity consistently proves to be unyielding.

There are various different approaches taken toward the advancement of sound synthesis. Some people look to achieve the most realistic sound, whilst others seek to invert reality and let the computer be an instrument of previously impossible timbres. It is possible to manipulate the auditory display so that there is a correlating effect on the listener's perception of reality. To create a sense of physical surroundings by means of digital audio is as useful as it is intriguing. It is a goal aspired to by these authors and shall continue to be so for many others.

2. OVERVIEW

The goal is the formulation of an algorithm for holophonic, or three-dimensional audio, synthesis. The technique is driven by the necessity for a digital representation of real-world effects, as seen by the authors. The supporting research in the fields of physics, mathematics and DSP physical modeling, is drawn upon because of the particular relevance in its quantification of natural phenomena.

Waveguide modeling has proven to be an extremely effective method of physical modeling synthesis. A digital representation of a leaf can be modeled using a waveguide mesh. Thus, by taking an impulse response from the mesh, a resulting modeling of the sound of rain hitting a leaf is obtained. The timing of the impulse response firings is controlled by an element of chaos theory called the logistic equation.

The filtering effect of the upper body and shape of the ear can be captured and imposed on sounds to give the impression of aural placement. Localisation is achieved in this algorithm by making use of HRTF theory. Other factors that promote the sense of realism for the listener are also addressed. Physical and perceptual issues, such as reverberation and head motion, are addressed to this end. An implementation of this synthesis algorithm is offered as programmed external objects for the Pd (Pure-data) software package [2].

3. MICRO-MEMBRANE DESIGN

Basically, digital waveguides model the behaviour of physical wave motion in the virtual domain. This means that the displacement caused by travelling sound waves can be effectively represented and their results used to synthesize waves occurring in physical space [3]. By considering the leaf as a type of membrane, a two-dimensional waveguide mesh can be used as a physical model of the membrane and so, the leaf.

Smith defines scattering equations for a two-dimensional rectilinear waveguide mesh with a one-unit delay between junctions as follows;

$$v_{l,m}(n) = \frac{v_{1l,m}^+(n) + v_{2l,m}^+(n) + v_{3l,m}^+(n) + v_{4l,m}^+(n)}{2} \quad (1)$$

$$v_{-il,m}^-(n) = v_{Jl,m}(n) - v_{il,m}^+(n) \quad (2)$$

where $v_{Jl,m}(n)$ is the velocity at a junction at position l,m on the mesh, at sample time n ,

$v_{il,m}^+(n)$ the four input velocities at the junction,

$v_{i,m}^{-}(n)$ the four output velocities at the junction.

Based on these equations, and the fact that the junctions are connected by a unit sample delay, the following equation for calculation of update sample values at each junction can be derived;

$$v_{J_i,m}(n) = \frac{1}{2} \sum_{x=1}^4 v_{J_x}(n-1) - v_{J_i,m}(n-2) \quad (3)$$

where $v_{J_x}(n-1)$ is the velocity of the four neighbouring junctions at sample time $n-1$, and $v_{J_i,m}(n-2)$ the velocity at the update junction two samples ago.

3.1. Design Optimization

The main concern regarding the membrane algorithm was to keep the computational cost as low as possible. The design is of a digital waveguide mesh as a square membrane of 5 x 5 nodes. The sound of the rain hitting the leaf would be synthesised by firing an impulse response into the centre node, at position (3, 3). The sound travels outward from the centre and this is achieved by updating all neighbouring nodes according to the update formula (3) above. The wave is reflected back at the boundary. The formula for updating the boundary nodes can be altered to suit the required conditions. A difference equation that describes the boundary node update, using a reflection coefficient is given by;

$$v_{J_b}(n) = (1+r)v_{J_n}(n-1) - rv_{J_b}(n-2) \quad (4)$$

where $v_{J_b}(n)$ is the boundary node to be updated, $v_{J_n}(n-1)$ is the (only) neighbouring node at sample time $n-1$, $v_{J_b}(n-2)$ is the boundary node at time $n-2$, and r is the reflection coefficient [4].

The output sound is taken by sampling the value at the centre node at the speed of the sampling rate. Obviously, this adds a large burden onto the computational cost of the algorithm and this increases with every operation or with the addition of new nodes on a larger mesh.

The above equations are given for an ideal isotropic membrane. It is also possible provide built-in impedances to the mesh algorithm so that each node has reflective impedance. This slows down the travelling speed of the wave, which would be the case for a real-life leaf.

The rectilinear digital waveguide mesh does provide a reasonably effective representation of wave dispersion throughout the membrane. The direction of the wave propagation in a rectilinear mesh is generally north-south, east-west, with diagonal propagation not accounted for [4]. In the ideal case, waves should travel in all directions across the membrane, exciting all possible modes. There have been several effective attempts made at improving the representation of wave dispersion throughout the membrane. The triangular mesh structure, where the nodes are connected in the shape of a triangle, or an interpolated rectangular mesh structure, are two examples of such [5]. An interpolated rectangular mesh connects eight or more node neighbours for every node to represent diagonal dispersion.

4. FORMULATING HOLOPHONICS

4.1. Three-dimensional Audio

In attempting to model a virtual environment, it is necessary to consider how we experience sound in an everyday context. We hear sound as it happens in the real world; in three-dimensional space. For digital audio to represent and delineate 'real world' sound effectively, these spatial characteristics must be accounted for. The goal of a perfectly realistic digital sonic representation can seem unattainable, certainly to date unattained, however every attempt we make to understand and quantify sound brings us one step closer.

There are a number of areas of acoustic and psychoacoustic theories that accurately explain our perception of sound and as such must be addressed in an approach to three-dimensional audio. These include localisation issues such as duplex theory, a consideration of the motion of the head, distance perception, the role played by reverberation and the effect of filtering by the pinnae.

Duplex theory refers to the use of both interaural time difference (ITD) and interaural intensity difference (IID) in determining sound source position. Although relatively concise and lucid, duplex theory does not provide us with all the answers to how we localise sound. The theory can only explain our perception of azimuthal location. Spatial positions are not fully accounted for according to this theory alone. Our ability to localise sound increases immensely when we are given the power to move our head. Instinctively we move our head when we hear a sound as our body naturally attempts to extract all the information from the interaural differences and their change over time. The motion of the head is known to disambiguate the process involved in front - back localisation. If our head moves toward the source, the intensity level at the ear nearest the source increases, thus aiding our ability to place that sound. Head motion cannot therefore be ignored when it comes to the implementation of a three-dimensional audio system. Even from a psychological point of view, if head movement directly effects our perception of the sound source, our ability to localise the sound will undoubtedly increase.

In addition to our directional assessment of the sound source, another key to accurate spatial placement is our perception of distance. It is important that we get a sense of how far away a sound is and also whether or not that distance remains constant. This distance cue must be denoted in holophonic audio synthesis, to help establish a physical relationship between sound source and listener. It is apparent from everyday life that a sound that is further away generally has a loudness lower than it would if the sound travelled closer to us. It can also be influenced by other environmental factors, although loudness is the primary cue in distance perception. It is described physically by means of logarithmic scales, an example of which is based on the inverse square law. The inverse square law states that the intensity, I , of a sound at any distance r , is inversely proportional to r^2 . Using this law, the decibel (dB) change can be implemented by using the following expression;

$$\text{dB Change} = 20 \log_{10} \left[\frac{1}{d} \right] \quad (5)$$

where d is the relative change in distance [6]. Other methods that have been proposed include the use of the *sones scale* of

relative loudness and the use of the inverse cube law. Investigations into this area have attempted to provide a more realistic auditory and spatial range, incorporating the various parameters that play their role in our perception of space that goes beyond that of azimuth and elevation [7][8]. Effective use can be made of a lowpass filter to model the effect of air absorption that takes place in a diffuse-field environment.

Reverberation is another natural phenomenon that is familiar to us in an everyday context. Walking into a tunnel, church or concert hall, we quickly develop an awareness of our changing physical surroundings. Reverberation is important for the simulation of spatiality because often 'raw' digital audio synthesis appears to the listener to be emanating from within the head. It therefore helps in the perception of externalisation, which is the elusive key to holophonic synthesis. The algorithm includes a representation of reverberant effects by incorporating a virtual ground plane into the auditory space. Reverberation can be implemented digitally by means of filters and delay lines. Early reflections can be modeled using tap-delay lines, where the signal is delayed by a small time duration, scaled and then output. More complex algorithms for the generation of the diffused-spectra of later reflections have been developed. These generally make use of infinite impulse response (IIR) filters, such as a comb filter and also allpass filters. These filters are effective because they colour the sound, by altering the frequency spectrum, producing results that are similar to the natural colour of late reflections.

4.2. Head Related Transfer Function

In order to understand what role the ear and body play in our ability to localise sound, it is possible to measure the spectral changes they make to the sound. These spectral changes, characterised by the head-related transfer function (HRTF) are known to directly influence how we perceive directional features of three-dimensional space [9]. The head, torso and pinnae affect tiny delays and frequency-dependant resonances that are enough to significantly alter the spectral content of the sound; altered as a function of direction. This means that the ear asserts a unique spectral filter for every direction and frequency of the arbitrary sound source.

Synthesis of digital audio with spatial characteristics can be achieved using HRTFs. This is performed by filtering the HRTF with the desired signal. Specifically, convolution can be used to impose directional attributes on the input source. An equation for the calculation of the output signal is given by;

$$y_l(n) = x(n - n_{ITD_{\theta,\phi}}) * d_{\min_{l,\theta,\phi}}(n) \quad (6)$$

$$y_r(n) = x(n) * d_{\min_{r,\theta,\phi}}(n) \quad (7)$$

where $y_l(n)$, $y_r(n)$ are the left and right outputs respectively, $x(n)$ is the input signal,

$n_{ITD_{\theta,\phi}}$ is the interaural time difference, and $d_{\min_{l,\theta,\phi}}(n)$ and $d_{\min_{r,\theta,\phi}}(n)$ are the (minimum phase) impulse responses measured at azimuth θ , elevation ϕ , of the left and right ears respectively and * denotes convolution.

4.3. Convolution

In order to impose the spectrum of the finite impulse response filter, generally given by $h(n)$, with the input signal, $x(n)$, it is possible to convolve the two signals to produce the output signal, $y(n)$. Direct convolution is computationally very expensive to implement. It has been seen that the order of required multiplies for such an implementation is $M(N - M + 1)$, where N is the size of $x(n)$ and M is the size of $h(n)$. Clearly, as the lengths of the input sequences increase, the demands outstretch the available speeds of modern-day processors. A better method is required for a realistic implementation. One such method is called *fast convolution* and makes use of the convolution theorem, which states that convolution in the time domain is equal to multiplication in the frequency domain.

This implementation employs the use of the Fast Fourier Transform (FFT) to perform a method of convolution known as overlap-and-add [10].

4.4. Chaotic Control

Although no single universally-agreed definition of a chaotic system exists, a system can be said to be chaotic if it exhibits chaotic behaviour. Chaotic behaviour is exhibited when complex results are obtained from apparently simple systems. Chaotic systems exhibit sensitivity to initial conditions.

The logistic equation is an iterative equation used to model population growth. The function definition is given by;

$$x_{n+1} = rx_n(1 - x_n) \quad (8)$$

where r is the growth rate, also called the fecundity. When $0 < r \leq 4$ the results of the equation are given by;

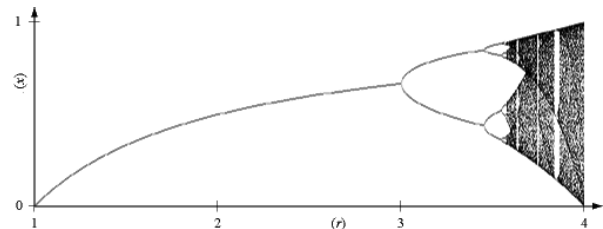


Figure 1: Graph of the logistic equation

Depending on the initial conditions, the behaviour of this system is characterised by one of three types; fixed ($0 < r \leq 1$), periodic ($1 < r < 3.5699457$), or chaotic [11]. The chaotic behaviour begins at $r = 3.5699457$. This is the iterative function used in the holophonic synthesis algorithm. It is used to time the 'firings' of impulse responses to the micro-membranes and as such is a control parameter to the algorithm.

5. AN ALGORITHM FOR SYNTHESIS

The aim of the algorithm is to collate the fields of research that were presented and expounded. These areas are applied to produce a formulation that is intended to be useful, possibly novel and certainly feasible in implementation. Conceptually, the idea is that the listener stands in front of a virtual wall of leaves, hearing the rain hit off them. The listener can move around, to a limited degree, and hear the difference in the spatial direction of the sound. The implementation is given through programmed objects in Pd [2], with the spatial position of the listener being controlled initially by sliders in the Pd

patch. This permits the perception of virtual motion by the user and a resulting alteration of immersive spatial characteristics.

The algorithm is illustrated in the following diagram;

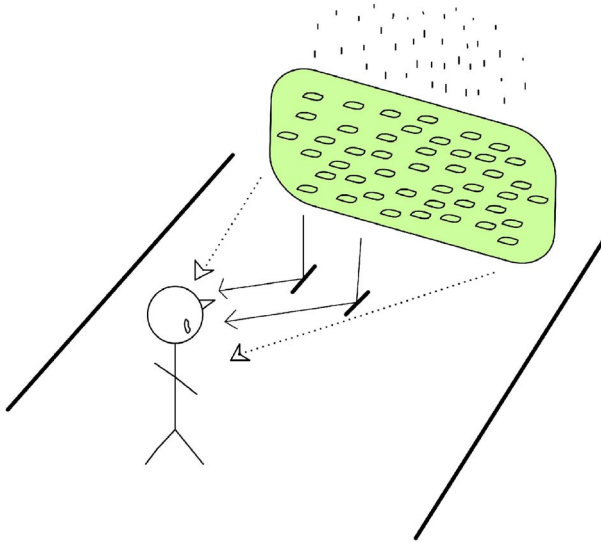


Figure 2: Illustration of the Algorithm

The ground plane is introduced to heighten the sense of realism for the listener. This allows for the reverberant effect of the sound source to reach the listener, increasing the fidelity of the otherwise open-air acoustics scene.

6. IMPLEMENTATION

The algorithm is implemented as a module in Pd, specifically by two programmed external objects, called leaf~ and holophon~. Pd (Pure-data) is a real-time, Open-Source, digital audio software system, as an ongoing development, by Miller Puckette [2]. It can be used for various music or interactive media performances, as a synthesis or control tool. The model for this sound synthesis algorithm was implemented under Linux in Pd.

The advantage of such an open-source software implementation allows for research and development to occur on a large-scale basis. This encourages the type of resource sharing and consultation that facilitates better adaptation and application of research.

A screenshot of the interface patch is shown here;

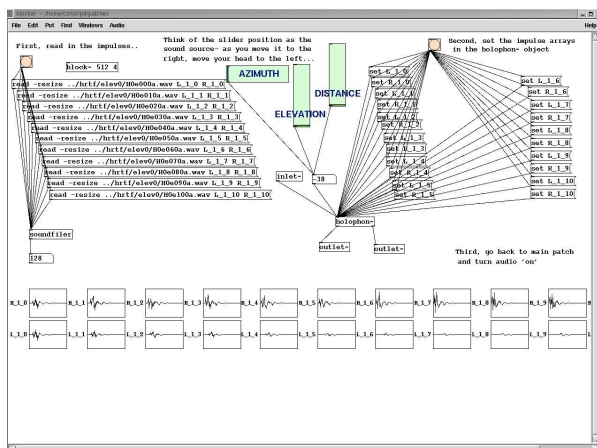


Figure 2: Screenshot of the Pd implementation

The excitation of the membranes is generated by the leaf~ object. A sub patch contains the holophon~ object, which has all appropriate impulse responses stored internally. This object is fed in the listener's spatial position, as well as the sound of the 'raindrops'. Thus the input sound is rendered with the apposite spatial characteristic. The HRTF set used were measured at the MIT Media Lab, from a KEMAR Mannequin [12]. It is important to also consider that this set is measured in a spherical anechoic chamber and therefore the distance parameter needs to be adjusted to account for the wall scenario. Headphones are therefore required to achieve the maximally desired spatial effect for the user.

In accordance with the nature of the implementation, a limited but effective real-time interactive scenario has been developed.

It has been found that an awareness of directional gesture of the head leads to a heightened perception of the change in spatial characteristics. The sliders of the pd patch, while not ideal, allow a point of focus for the user that therefore enhances the holophonic experience. The use of sliders as a control mechanism only affords a limited expansion of user-interface however. It is foreseen that a three-dimensional head-tracking device can be used to give a more natural control to the algorithm, so that the expressive intention of the user has a direct impact on the algorithm and the resulting output of sound.

7. CONCLUSIONS

This paper presents the results of a project that implemented the algorithmic modeling of a natural auditory phenomenon. Various stages of the realisation of this project are discussed here, such as the physically based modeling, perception issues, the interface, design and control of the implemented model.

The ultimate aim of the authors is a move toward a true holophonic environment. One area in which this work has particular relevance is in the development of the Digital Foley industry, such as the recent PHISM project by Perry Cook [13]. The production of digital audio in the film industry has yet to reach the levels of realism and accuracy attained by those of digital graphics and accordingly work, such as this type of physical modeling, aims toward redressing this imbalance. The results achieved by this project are certainly encouraging; nevertheless they also highlight the need for further research in this area.

8. REFERENCES

- [1] R. McGrath, T. Waldmann and M. Fernström. "Listening to Rooms and Objects" presented at *Audio Engineering Society 16th International Conf. on Spatial Sound Reproduction*, Rovaniemi, Finland, April 1999.
- [2] Pure Data, <http://www.pure-data.org/> January 01 2002.
- [3] J. O. Smith, "Physical Modeling using Digital Waveguides," *Computer Music Journal*, vol. 16, no.4, pp.74 - 91, 1992.
- [4] L. Savioja, *Modelling Techniques for Virtual Acoustics*, Doctoral Thesis, Helsinki University of Technology, 1999, online <http://www.tml.hut.fi/~las/publications/thesis/>
- [5] S. Bilbao, *Wave and Scattering Methods for the Numerical Integration of Partial Differential Equations*, Doctor of Philosophy Thesis, Stanford University, 2001, online

<http://ccrma-www.stanford.edu/~bilbao/master/node124.html>

- [6] D.R. Begault, *3-D Sound for Virtual Reality and Multimedia*. Academic Press, Cambridge, MA, 1994.
- [7] W. L. Martens, "Psychophysical Calibration for Controlling the Range of a Virtual Sound Source: Multidimensional Complexity in Spatial Auditory Display," in *Proc. 2001 Int. Conf. on Auditory Display*, Espoo, Finland, July– August 2001, pp. 197 – 207.
- [8] A. Sontacchi and R Höldrich, "Further Investigations on 3D Sound Fields Using Distance Coding," in *Proc. COST G-6 Conf. on Digital Audio Effects (DAFX-01)*, Limerick, Ireland, December 2001, pp. 152 – 154.
- [9] C.I. Cheng and G.H. Wakefield, "Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space." *Journal of the Audio Engineering Society*, vol. 49 no. 4, p. 231, 2001.
- [10] P. M. Embree, *C Algorithms for Real-Time DSP*, Prentice-Hall. Upper Saddle River, NJ, 1995.
- [11] G. Elert, "Logistic Equation", *The Chaos Hypertextbook*, <http://hypertextbook.com/chaos/42.shtml> January 01 2002.
- [12] B. Gardner and K. Martin, "HRTF Measurements of a KEMAR Dummy-Head Microphone", *MIT Media Lab*, <http://sound.media.mit.edu/KEMAR.html> January 01 2002.
- [13] P. Cook, *New Work and Future Directions*, <http://www.cs.princeton.edu/~prc/NewWork.html> May 01 2002.