SONIFICATION USING DIGITAL WAVEGUIDES AND 2- AND 3-DIMENSIONAL DIGITAL WAVEGUIDE MESH

Kyogu Lee, Gregory Sell, Jonathan Berger

CCRMA
Department of Music, Stanford University
Stanford, California, USA
brg@ccrma.stanford.edu

ABSTRACT

We describe a method of auditory display of complex data, object identification, and classification using digital waveguides and waveguide mesh. Our overall goal is to distinguish highly dimensional data sets from one another in such a way that reveals meaningful differences in a particular context. In this paper we provide a summary of the application of waveguide and waveguide mesh architectures to sonification, and demonstrate the digital waveguide, 2- and 3-dimensional mesh in a variety of sonification tasks.

1. INTRODUCTION

The recent and continuing emergence of new and improved data acquisition technologies in science, medicine and industry poses unprecedented demands on data analysis, data exploration, classification, navigation and interpretation. Applications such as medical imaging, or data mining often demand real-time diagnosis and interpretation of enormous data sets. Auditory representation of data can provide beneficial enhancements and in some instances new opportunities for data exploration. In some applications such as medical diagnostics and assistive technologies for blind navigation sonification can be a critical tool.

While directly mapping a data stream to an auditory parameter such as pitch or loudness is sometimes a simple and effective sonification method, high dimensionality generally deems direct mapping of data to synthesis parameters impractical. Indeed, in many routine situations of auditory analysis such as medical auscultation and automotive engine analysis, interpretation and diagnosis involves comparing the gestalt timbral result of integrated parameters rather than contrapunctally tracking after individual parameters. An alternative to direct parameter mapping is to view the data as part of a physical construction which, when excited by an impulse, provides an auditory result in which the emergent sound is the result of the physics of the system [1].

In this paper we describe a series of experiments in physical model based auditory display. Specifically, we consider the construction and implications of using digital waveguides as a vehicle to represent complex data.

The primary goals of the approach is to provide methods of auditory display that provide intuitive and easily-learned sound representations. The research path thus first explores the use of physical models of the human vocal tract as a means of producing auditory representations that answer these needs.

2. PHONEME BASED SONIFICATION

Phoneme perception is one of the most widely studied areas in auditory processing and cognition. Evidence of phoneme perception in neonates suggest categorical perception of phonemes along acoustic dimensions during infancy [2], irregardless of variation in vocal timbres and fundamental frequency [3]. While these attributes are modified by the statistical prevalence of particular sounds of the native language heard, this capacity to similarly categorize a given phoneme across a widely varying range of acoustical dimensions remains a critical aspect of the human ability to comprehend speech.

The seemingly effortless task of phoneme perception, provides a basis to address our essential goals of sonification - readily discerned and categorized.

Preliminary experiments with frequency modulation based formant synthesis for sonification [4] supported the feasibility of vocal-like synthesis in auditory display. While the approach proved promising in terms of intuitive mappings with relatively clear categorical boundaries, the limited number of control parameters made FM formant synthesis correspondingly limited in the number of dimensions the technique could represent.

2.1. Source-filter model based formant synthesis

Further experiments with formant synthesis using a source-filter model [5] extended this approach to representation of data-sets with higher dimensionality. To generate vowel-like sounds a model of the characteristics of the vowel sound production mechanisms is constructed using a band-limited impulse train as a glottal source subsequently filtered by a parallel or cascaded series of resonators with appropriate corresponding formant frequencies and bandwidths.

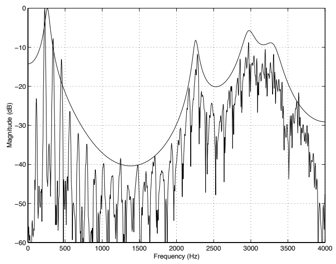

Figure 1: Spectrum of vocal utterance of the vowel /i/ as in team. The smooth line enveloping the lower spectrum corresponds to the vocal tract transfer function. The resonant peaks of this curve are called formants.

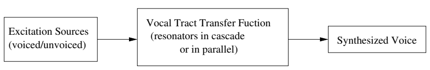

Figure 2: Klatt's source-filter model for vocal synthesis. Radiation of the output pressure at the nose and lips to the environment is not shown.

This classic approach to formant synthesis was implemented in Matlab using Peterson and Barney's formant tables [6] to create a formant matrix, which contains the first three formant frequency values from 10 American-English monophthong vowels as spoken by 76 speakers (33 men, 28 women and 15 children). Although it has only three control parameters - pitch, gender (male, female, or child), and vowel type - it is far more suitable for hyperspectral data sonification because if we map data values to the amplitudes and the bandwidths of formant peaks, we could obtain vowel sounds with different sonority.

In addition to Matlab implementation, the STK (Synthesis Tool Kit), a CCRMA-created collection of C++ classes for the synthesis and processing of musical instrument sounds, contains a C++ class VoicForm for the synthesis of vowel sounds based on formant filtering of a band-limited impulse train.

2.2. Digital waveguide vocal tract model

Kelly and Lochbaum's [7] model of the vocal tract uses a time-variant circular one-dimensional acoustic tube of variable cross-sectional area. When excited at one end by noise, speech-like formants are produced according to the shape of the tube at a given time. An extension of this model by Cook [8], was used as a basis for data sonification [9]. Subsequently, the waveguide vocal model was implemented using PD, a real-time audio optimized synthesis framework which, combined with Open Sound Control (OSC), allows for real-time sonification of virtually any type of data received over a network.

The method involves approximating the vocal tract by a series of acoustic tube sections, each with a radius that varies from one

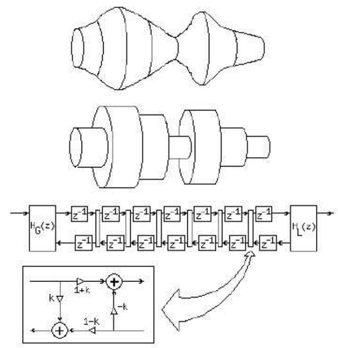

Figure 3: Spectrum of vocal utterance of the vowel /i/ as in team. The smooth line enveloping the lower spectrum corresponds to the vocal tract transfer function. The resonant peaks of this curve are called formants.

vowel sound to the next. The radii of adjacent tube sections govern the transmission and reflection of acoustic energy at the junction between such sections. For each tube section, discrete-time delay elements are used to model the forward- and reverse-traveling wave components of the digital waveguide simulation [10]. Between the delay elements, a scattering junction is used to handle the change in radius from one tube section to the next. In addition to the convincing sounds that it generates, this physical model can have as many tube sections as possible, which makes it perfect for sonification of very high-dimensional data. The model used four control parameters,

- frequency of glottal pulse train,
- the shape of tract radii for a desired phoneme
- radii is a vector that sets radii of N-tube sections
- tube length of sections

2.3. Summary

The driving premise for sonification with a vocal-tract model is that phonemes provide intuitive and easily learned auditory cues. Although this approach proved useful for sonification of certain types of data sets, the approach seems limited in use with highly dimensional data in a number of respects, not the least being that although categorization is readily done with phoneme-like sounds, the absence of an effective metric of the timbral distance between sounds limits this approach.

3. SONIFICATION WITH WAVEGUIDE MESH

3.1. Overview

The generalization of the digital waveguide network described above, in which paired delay lines propogate wave signals and scattering junctions scatter the waves has been extended to a rectilinear mesh which we explore as a basis of sonifying complex data.

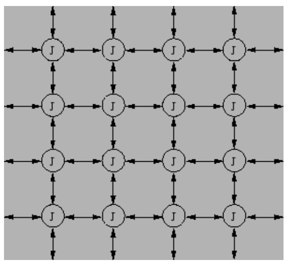

Figure 4: Spectrum of vocal utterance of the vowel /i/ as in team. The smooth line enveloping the lower spectrum corresponds to the vocal tract transfer function. The resonant peaks of this curve are called formants.

The digital waveguide mesh provides an approach for simulating two- and three-dimensional wave propagation in musical instruments and acoustic spaces by modeling the behavior of physical wave motion. The two-dimensional waveguide can be thought of as a membrane over which an impulse driven wave propogates.

The lossless scattering equations of a two-dimensional rectilinear waveguide mesh with a one-unit delay between junctions is:

$$v_j = \frac{2\sum_i R_i v_i^+}{\sum_i R_i}, and v_i^- = v_j - v_i^+ \tag{1}$$

where v_j is the junction velocity and v_i^+, v_i^- are the inbound and outbound waves at a junction.

Waveguide mesh have been applied to a wide range of musical instrument simulations and synthesis models including piano, bowed strings, winds, brass, percussion, and room reverberation.

The model of physical wave traversal over, within, or between waveguide structures provides a strong conceptual framework for applying waveguide mesh models to sonification of complex data, object identification and classification.

3.2. Data sonification with a 2-D waveguide mesh

Application of the waveguide mesh to distinguish highly dimensional data sets from one another in such a way that reveals meaningful differences in a particular context provided promising results. Using a 2-dimensional mesh of impedance matrices we explore various methods in which data controls or effects wave dispersion. We first consider situations in which the data effects mesh size, and proceed to describe mapping data to points or regions of impedance within the mesh. In preliminary experiments data parameters of a high-dimensional data set were mapped as mesh control parameters. Since there is no physical limitation to the size of the mesh N-dimensional data, no matter how large N may be, can, for example, be mapped to the initial excitation condition of each junction in the mesh. In these initial experiments three mapping strategies were tested. These include:

1. A rectilinear 2-D mesh is created with the number of points equals the (even) number of dimensions in a particular data set. The initial condition may be any type of wave variable - displacement, velocity, or force.
2. a mesh of size NxM is created, where N corresponds to the dimension of the data to be sonified, and M can be arbitrary. A planewave whose initial conditions are determined by the data is used as an excitation along the axis of N points.
3. a mesh of size NxM is created, where N corresponds to the dimension of the data to be sonified, and M can be arbitrary. Instead of mapping the data to an initial excitation condition as before, they are mapped to the boundary condition of the mesh. Since one pole filters are used at the boundaries the data can be mapped to control the gain or to change the pole location of the filters. In this case, the initial excitation can be anything - impulse, planewave, or a set of impulses. Since the resulting timbre of the sound produced by the mesh is largely determined by its size, success in producing sounds with salient differentiation in timbre was limited. We thus proceeded to describe a new technique using impedance is applied with the aim of establishing a sense of perceptual distance and character.

3.3. Data clustering with a 2-D waveguide mesh

In the previous sonification method using a 2-D digital waveguide mesh, the data was mapped to the control parameters such as the initial point wise excitation, the initial plane wave excitation, or the boundary condition of the mesh. While this approach enabled us to map very high-dimensional data without sacrificing any dimensions, the resulting sounds were barely distinguishable because with the pseudo random initial condition represented by the data, the size of the mesh would dominate in resulting sounds. Furthermore, it was point wise sonification; that is, a single point in the data sets corresponds to one mesh, thus not only making the computational costs very high, but also failing to provide a good data clustering scheme.

We took a new approach to focus on the data clustering this time instead of sonifying every single data point. We create an N-point mesh (with proper width X and height Y where $XxY = N$) from N data points, and thus one mesh can now represent a number of data points, or a data cluster. The major drawback of this method, however, is that now we must reduce the dimension of the data down to a few in the case of the 2-D rectlinear mesh since a junction in a 2-D mesh may have only a couple of control parameters. We have chosen four most significant dimensions in the 128-D data, and have mapped the data to the wave impedances of the 2-D mesh, where one junction at time n has four wave impedances in each branch i.e., $Rx[n], Ry[n], Rx[n+1], and Ry[n+1]$.

Using the test data set - we know which part is benign or malignant - we created two meshes, one of which contains only the benign cell data, and the other contains only the malignant cell data, and used them as the references. The two resulting sounds were very easily distinguishable. The next step is to create a composite mesh containing both benign and malignant cell data points. In our first approach, however, the data were manipulated in such a way that the left half of the mesh should include only benign cell data, and the right half should contain only malignant cell data (Figure x).

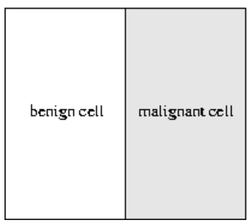

Figure 5: Composite mesh with both benign and malignant cell data.

3.4. The game of Battleship - object identification

As an example of classification and identification consider a version of the popular game *Battleship* in which a player tries to locate objects (ships) on the opponent's hidden grid (the sea) by guessing coordinates. In the variant described auditory cues provide more information than the standard response of 'hit' or 'miss'. In the following examples the ocean surface is represented as a two-dimensional rectilinear waveguide mesh, and a second 2-D mesh of equal size represents the ocean at a particular depth. Timbral segregation between surface and submerged regions are created by setting distinct boundary conditions for each mesh. In the surface level the pole location was set to 0.05 (the default setting for a metallic plate model (Smith, 1999), while the one-pole filters at the boundaries of the submerged mesh were set at 0.8, the default settings for a wood block. An object of arbitrary size can be simulated by setting wave impedance along areas within the mesh, These virtual objects can be characterized according to the amount of impedance imposed upon the regions. Since wave propagation is constrained in proportion to the amount of impedance in these regions, these characteristics can naturally represent physical attributes of the particular object.

A simulation of the sonified game of battleship was configured as follows: A $32x32$ 2-D mesh corresponding to the ocean surface. Two 'ships' of arbitrary sizes were placed at arbitrary locations in the mesh. Each ship was characterized in terms of 'durability' by setting setting the wave impedance with distinct values. The higher the impedance, the less a wave will propagate in the region. higher wave impedance, which makes sense because wave will not propagate very well if the impedance is high, thus having little effect on the ship.

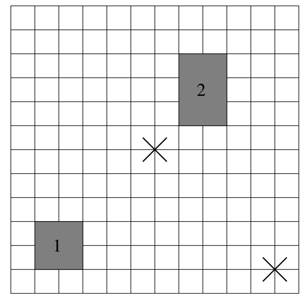

Figure 6: surface.

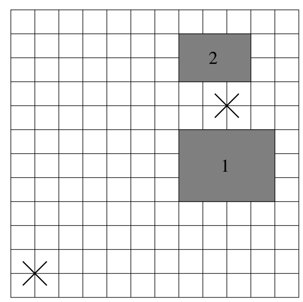

Figure 7: Basic configuration for (a) sea surface level and (b) submerged level. Dark areas indicate ships, and 'X' marks show attack positions. Ships are not scaled correclty.

A second mesh with different boundary conditions was then created to represent a particular underwater depth. A 'submerged' object was placed on this mesh, again designated with characteristic size, location, and durability (figure 4).

An impulse to the mesh simulated the firing of an artillery projectile whose location of hit was represented by the location of excitation, and whose relative force was represented by the impulse amplitude.

Because of the isolated area of higher impedance in a virtual object, the sounds of a 'hit' and a 'miss' are distinctive from one another. Decoupling of the ships from the sea, when a ship is hit, would have the similar effect of having smaller size of mesh, thus making its impulse response easily distinguishable from that of missing since the original mesh, which represents sea, is much bigger in size than the ships, thus making the effect much less noticeable. Secondly, the ship variables - size, location, and durability (wave impedance) - will affect the resulting sounds because wave propagation will vary as they change. Lastly, the attack position and power will also have an effect on the overall wave propagation.

3.5. Sonification of data clusters using impedance to control resonance

The reflective behavior of a wave as it encounters greater impedance can be harnessed to direct the wave to certain regions within the mesh. If a shell of high impedance is wrapped around a region of low impedance, the interior becomes a zone of resonance producing a unique and identifiable timbre and frequency when responding to an impulse. A promising aspect of altering the impedance within the shell to reflect resonant properties is the audible percept of proximity to the region from the impulse source. Because of the nature of the dispersion of a wave, a point further from the impulse will receive a smaller proportion of the wave's original energy. Thus, an impulse placed closer, but not within, a resonance zone will filter more of its energy through the shell, creating a significantly lower degree of resonance resulting in a noticeable alteration in timbre.

The ability to completely control the impedance parameters of the waveguide mesh offers an intriguing possibility for data mapping. Any wave, when traveling from one impedance level to another, reflects on the order of $\frac{Z_2}{Z_1+Z_2}$, where Z_2 is the impedance that the wave is traveling toward. So, it is clear that moving from low impedance to high impedance causes most of the wave to be reflected, while high impedance to low impedance movement allows most of the wave to be absorbed.

This quality allows for an interesting form of data separation. Any set of data (n x m) can be mapped to a mesh of size $n-1xm-1$ so that the impedances of any point on the mesh map to the differences between the corresponding data points. In other words, the x-direction impedance of (a,b) is connected to the value $(a+1,b)$ - (a,b) and the y-direction impedance at the same point is connected to $(a,b+1)$ - (a,b).

The advantages gained by this means of mapping are that zones of like data all exist in regions of low impedance, because the differences between those data points will be minimal. However, the differences between unlike data points will be significantly greater (and this can be amplified through data manipulation), therefore creating walls of high impedance separating the low impedance zones.

The end result of this is that resonance zones are created within the waveguide mesh. Then, the nature of any given region can be determined by creating an impulse. If the excited point is located within a small region of like data, then a high frequency resonance is created. If the excitation is within a large region of like data, then a low frequency resonance is created. If the excitation is not within a region of like data, then no resonance is created (except possibly that of the entire mesh). In this case, various frequencies can be created from the many possible styles of reflection, but no audible resonance is found.

This approach to data mapping creates a very natural orientation within the mesh. The location of the clusters in the mesh will indicate the location of the clusters in the data matrix. In the case where the data was collected and organized based on physical location, the mesh becomes a model of the very subject for the data collection. In the case of the 2D mesh, this can be extremely appropriate for data gathered from slides, surfaces, or any other two dimensional plane. However, in many cases, data is collected from a three dimensional space, creating a need for a 3-dimensional method for mapping in order to preserve the relevance of location.

3.6. Expansion of impedance mapping to 3 dimensions

With the same concepts in mind, we expanded the 2-dimensional mesh to include a third impedance matrix and to employ 6-port scattering junctions instead of only 4-port. This allows for the creation of a three dimensional mesh that can, in turn, map data taken from a three dimensional space with meaningful location.

However, two central problems arise with the introduction of the third dimension: energy dispersion and computation time. The computation time is obviously a problem, as the number of scattering junctions increases with the third dimension, as does the computational complexity at each junction. Energy dispersion becomes an issue as well because, with the vastly increased space within which the wave spreads, the output soundfiles are almost completely inaudible. Both of these problems are fixed by reducing the size of the mesh, both to reduce the computation time and to reduce the volume that the impulse must fill. Then, both to reduce the computation time further and to counteract the audible effects of the size reduction, the sampling rate is decreased. The reduction in sampling rate allows for fewer samples to occupy a larger virtual space, because the sampling interval is inversely proportional to the sampling rate. Once the class is updated to include the third impedance matrix, the 6-port scattering junction, the smaller size, and the sampling rate reduction, the results are effective.

4. SUMMARY

Physical modeling provides potentially powerful methods for sonification of high dimensional data. In this paper we presented two approaches; first, using a waveguide vocal model to map data to easily recognizeable phoneme-like sounds, and second, a general approach to representing multi dimensional data using waveguide mesh architectures. Our future goals include harnessing the technique to sonify data distances and characteristics of data and data clusters in multi-dimensional space.

5. REFERENCES

[1] Hermann, T. and H. Ritter, Listen to your Data: Model-Based Sonification for Data Analysis,1999

[2] Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. Speech perception in infants. Science, 171, 303306, 1971.

[3] Jusczyk, P. W., Pisoni, D. B., & Mullennix, J.. Some consequences of stimulus variability on speech processing by 2-month-old infants. Cognition, 43, 253291, 1992.

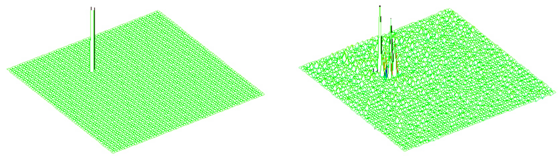

Figure 8: An impulse within the high impedance shell under initial conditions and after propagation.

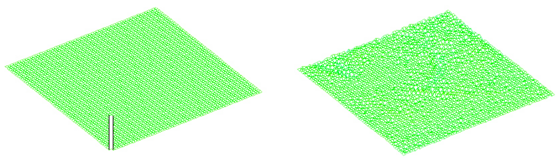

Figure 9: An impulse outside the high impedance shell under initial conditions and after propagation.

[4] Ben Tal, O, J. Berger and M. Daniels. De natura sonoris: Sonification of complex data. in Mathematics and Simulation with Biological, Economical and Musioacoustical Applications; Mathematics and Computers in Science and Engineering, p. 330. WSEAS Press. (Attellis, Kluv and Mastorakis, ed.) 2001.

[5] D. H. Klatt and L. C. Klatt, Analysis, synthesis and perception of voice quality variation among female and male talkers, Journal of the Acoustical Society of America, vol. 87, no. 2, pp. 820857, February 1990.

[6] Peterson, G.E. and H.L. Barney. Control methods used in a study of the identification of vowels. Journal of the Acoustical Society of America, 1952, 24, 175-184.

[7] Kelly, K.L. Jr and C. C. Lochbaum, Speech synthesis, in Proc. Fourth ICA, Copenhagen, Denmark, September 1962.

[8] Cook, Perry, R., Synthesis of the Singing Voice Using a Physically Parameterized Model of the Human Vocal Tract, Proc. of the International Computer Music Conference, pp. 69-72, Columbus, OH, 1989.

[9] Cassidy, R., J. Berger, K. Lee, M. Magioni and R. Coifman. Auditory display of hyperspectral colon tissue images using vocal synthesis models. in Proceedings of the 2004 International Conference on Auditory Display, Sydney, Australia, 2004.

[10] Smith, Julius ”Digital Waveguide Modeling of Musical Instruments”. http://www-ccrma.stanford.edu/ jos/waveguide/.

[11] J. O. Smith, Physical Modeling using Digital Waveguides, Computer Music Journal, vol. 16, no.4, pp.74 - 91, 1992.

[12] Van Duyne, S.A. and J.O. Smith. ”The 2-D digital waveguide mesh”. In Proceedings IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, New York. IEEE Press, 1993.

[13] Hermann, T and H. Ritter, Crystallization Sonification of High Dimensional Datasets, Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan, 2002.

[15] Peterson, G.E., and Barney, H.L., Control methods used in a study of the vowels. J. Acoust. Soc. Am. 24, 175-184 (1952).