AUDITORY DISPLAY OF HYPERSPECTRAL COLON TISSUE IMAGES USING VOCAL SYNTHESIS MODELS

*Ryan J. Cassidy**, *Jonathan Berger, Kyogu Lee*

The Center for
Computer Research in Music and Acoustics
Stanford University
Stanford, CA 94305

Mauro Maggioni, Ronald R. Coifman

Department of Mathematics
Yale University
New Haven, CT 06520

ABSTRACT

Vowel approximation is an intuitive, easily learned representation for complex data. We describe implementation of a physical model of the vocal tract and its application in sonifying multi-dimensional data. The model, based on [1], was implemented using STK [2] and PD [3] and is applied to hyperspectral images of colon cells.

1. INTRODUCTION

Sonification of highly dimensional data by directly mapping data to synthesis parameters is often limited by the lack of an auditory model that will ensure coherent and intuitive sonic results. Our exploration of alternatives have focused upon vocal tract models with the rationale that recognition and categorization of vowels is a highly developed feature of human auditory perception. To this end we have implemented a number of vocal tract models including a physical model based on [1]. In this paper we describe the model, its implementation and its implications in sonification of complex data.

2. CASCADED-TUBE-SECTION VOCAL SYNTHESIS MODEL

The model for sound synthesis that we have chosen is based on a vocal tract model developed by Perry Cook[1]. The vocal tract is approximated by a series of acoustic tubes, each with a distinct length and radius.

The model may be derived as follows:

First, consider a single cylindrical tube with constant cross-section $A_1 = 2\pi r_1^2$. With $p_1(x,t)$ and $u_1(x,t)$ denoting the pressure (in Newtons per square meter) and the volume velocity (in cubic meters per second) in the tube (respectively) at position x and time t, it can be shown that the tube supports wave propagation, with the solutions to the wave equation yielding

$$p_1(x,t) = p_1^+(t - \frac{x}{c}) + p_1^-(t + \frac{x}{c}), \tag{1}$$

and

$$u_1(x,t) = u_1^+(t - \frac{x}{c}) + u_1^-(t + \frac{x}{c}), \tag{2}$$

where p_1^+, p_1^-, u_1^+, and u_1^- denote the left- and right-going components of the traveling pressure and volume-velocity waves (respectively).

It can further be shown that

$$p_1^+(t - \frac{x}{c}) = \frac{\rho c}{A_1} u_1^+(t - \frac{x}{c}), \tag{3}$$

and similarly, for the left-traveling wave component,

$$p_1^-(t + \frac{x}{c}) = -\frac{\rho c}{A_1} u_1^-(t + \frac{x}{c}), \tag{4}$$

where ρ is the density of fluid in the tube (in kilograms per cubic meter), c is the speed of sound in the tube (in meters per second), and A_1 is the aforementioned cross-sectional tube area.

If we define the acoustic impedance of the tube as

$$R = \frac{\rho c}{A}, \tag{5}$$

then we have the following relations:

$$p_1^+ = R u_1^+, \tag{6}$$

and

$$p_1^- = -R u_1^-, \tag{7}$$

where we have omitted the arguments of the pressure and volume-velocity functions for convenience of notation.

Thus, we can model wave propagation in a single tube as a pair of delays (one for the left-traveling wave component and one for the right-traveling wave component). The length of each delay, in seconds, is given by

$$T_{d1} = \frac{l_1}{c}, \tag{8}$$

*Supported by the Natural Sciences and Engineering Research Council of Canada.

where l_1 is the length of the tube (in meters). For a discrete-time simulation, the number of samples of delay necessary is thus given by

$$N_{d1} = f_s T_{d1}, \tag{9}$$

where f_s is the sampling rate (in Hz). Note that this value may well be a non-integer; techniques for simulating a non-integer delay are discussed in [4].

When a second tube section with different cross-sectional area $A_2 = 2\pi r_2^2$ is joined to the end of the first tube, the mathematical relationships between the left- and right-traveling wave components may be derived using the results stated above:

By conservation of mass, the volume velocity of fluid in the first tube (flowing into the junction) must be equal to the volume velocity of fluid in the second tube (flowing out of the junction). Thus we may further derive

$$\begin{aligned}
0 &= u_1 - u_2 && (10)\\
&= u_1^+ + u_1^- - u_2^+ - u_2^- && (11)\\
&= G_1(p_1^+ - p_1^-) + G_2(p_2^- - p_2^+) && (12)\\
&= G_1(2p_1^+ - p_J) + G_2(2p_2^- - p_J) && (13)\\
\Rightarrow p_J &= \frac{2p_1^+ G_1 + 2p_2^- G_2}{G_1 + G_2} && (14)\\
\Rightarrow p_1^- &= p_J - p_1^+ && (15)\\
&= \frac{G_1 - G_2}{G_1 + G_2} p_1^+ + \frac{2G_2}{G_1 + G_2} p_2^- && (16)\\
&= \frac{R_2 - R_1}{R_1 + R_2} p_1^+ + \frac{2R_1}{R_1 + R_2} p_2^-, && (17)\\
\Rightarrow p_2^+ &= p_J - p_2^- && (18)\\
&= \frac{G_2 - G_1}{G_1 + G_2} p_2^- + \frac{2G_1}{G_1 + G_2} p_1^+ && (19)\\
&= \frac{R_1 - R_2}{R_1 + R_2} p_2^- + \frac{2R_2}{R_1 + R_2} p_1^+, && (20)
\end{aligned}$$

where u_2 is the volume velocity in the second tube, u_2^+ and u_2^- are the corresponding left- and right-traveling wave components, p_J is the common pressure at the junction, and G_1 and G_2 are the acoustic conductances[1] in the first and second tube section (respectively).

If we now define the scattering coefficient k as

$$k = \frac{R_2 - R_1}{R_1 + R_2}, \tag{21}$$

then we can re-write the scattering relations Equation (17) and Equation (20) as follows:

$$p_1^- = kp_1^+ + (1-k)p_2^-, \tag{22}$$

and

$$p_2^+ = -kp_2^- + (1+k)p_1^+. \tag{23}$$

[1]Conductance G is simply the inverse of impedance R, i.e. $G = \frac{1}{R}$.

Note the scattering coefficient k may also be written as a function of the radii of the tube sections:

$$k = \frac{r_1^2 - r_2^2}{r_1^2 + r_2^2}. \tag{24}$$

By joining additional tube sections to the end of the pair to form a cascade, we obtain an approximate vocal tract model. The process is illustrated in Figure 1. The input to the first tube section (the left-most section in Figure 1) represents the back of the vocal tract (nearest the glottal folds), and the output of the last tube section represents the point where the tract ends (at the lips). The blocks between the delay elements implement the relations of Equation (22) and Equation (23), and are known as Kelly-Lochbaum scattering junctions (they occur in wave propagation whenever there is an impedance discontinuity).

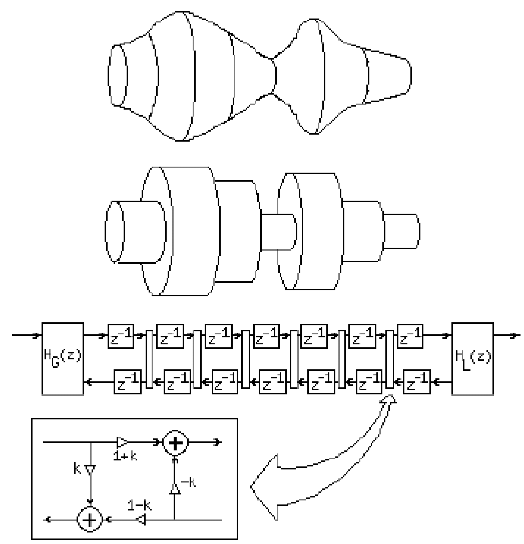

Fig. 1. Digital waveguide model of the human vocal tract, with scattering junctions between adjacent tube sections to account for changing radii (adapted from [1]).

By modifying the radii of (or equivalently, the scattering coefficients between) the tube sections, the quality of the output sound that results when a waveform is applied to the glottal input, will vary. It is this effect that we hope to take advantage of: by mapping the data we wish to sonify to the radii or scattering coefficients of the vocal tract sections, we obtain a new sonification technique.

One important detail that remains to be discussed is the nature of the waveform applied to the input of vocal tract model, or the glottal waveform. In short, the waveform is well approximated by an impulse train to which a lowpass filter is applied. More details may be found in [1].

3. HYPERSPECTRAL TISSUE IMAGE DATA

We aim to apply our sonification technique to hyperspectral images of colon tissue, the collection of which is discussed in this section.

The tissue images have been collected in cooperation with the Department of Pathology at Yale University. First, a series of slides, each slide containing several hundred microdots, each microdot corresponding to a slice of colon tissue (roughly 0.5 mm by 0.5 mm in size) from a distinct patient, is prepared. Each microdot may correspond to either normal or malignant colon tissue. Next a slide is chosen and illuminated with a tuned light source (capable of emitting any combination of light frequencies in the range of 450–850 nm), and the transmitted image is magnified 400X by a Nikon Biophot microscope. The magnified image is recorded with a Sensovision CCD camera. Several images are taken, each using a different combination of light frequencies, consistent with a technique known as Hadamard spectroscopy. The result is an image with 128 values per pixel, each value corresponding to a frequency of light in the range 440–700 nm (the frequencies are linearly spaced across the range). Each hyperspectral image may be referred to as a *datacube*, with dimensions x by y by 128, where x and y are the number of horizontal and vertical pixels in the image. More details may be found in [5], [6], and [7].

In the study of [7], 15 datacubes of normal colonic tissue and 42 datacubes of abnormal colonic tissue were collected. The dimensions of each datacube are 491 pixels by 652 pixels by 128. In the following sections of this paper, we focus on two such datacubes (one of benign tissue, the other of malignant tissue), and illustrate how vocal sonification techniques, after data preprocessing, may be applied to help distinguish the tissue analyzed.

4. DATA PREPROCESSING

Prior to sonification, several preprocessing steps are applied to the datacubes:

1. First the 128 dimensions for each point in the image are reduced to 16 via principal components analysis.

2. Next the 16 dimensions are reduced to 5 using a technique known as Local Discriminant Bases (LDB) (see [8] for a general treatment and [7] for details in this specific application).

3. Finally, the 5 dimensions are reduced to 3 using a Nearest-Neighbour Classification technique, again discussed in [7].

The final result is an image with 3 values per pixel: the first denoting the probability that the pixel belongs to an abnormal nucleus, the second being the probability that the pixel belongs to a normal nucleus, and the third being the probability that the pixel does not belong to nucleic tissue. More details may be found in [7].

5. IMPLEMENTATION AND RESULTS

5.1. STK Implementation of Cook's Vocal Synthesis

To quote the manual, "The Synthesis ToolKit (STK) is a set of open source audio signal processing and algorithmic synthesis classes written in C++" [2]. STK classes have been successfully integrated into prior sonification projects (see, for example, [9]). The STK already contains C++ classes for vocal synthesis, namely FM synthesis[10] and formant-filter-based synthesis[11]. However, no class is available for vocal synthesis using Cook's cascaded-tube-section tract model described in Section 2. For this reason, a C/C++ class, ready for integration into the STK, has been written from scratch. The class is called `VoicTract` (cf. the STK class `VoicForm`, which performs formant-filter-based synthesis), and has been designed with an interface similar to other STK instruments.

6. CONCLUSIONS

7. REFERENCES

[1] Perry R. Cook, *Identification of Control Parameters in an Articulatory Vocal Tract Model, with Applications to the Synthesis of Singing*, Ph.D. thesis, Elec. Engineering Dept., Stanford University (CCRMA), Dec. 1990, (CCRMA thesis).

[2] Perry R. Cook and Gary P. Scavone, "The Synthesis ToolKit," in *Proceedings of the 1999 International Computer Music Conference, Beijing*. 1999, pp. 164–166, Computer Music Association.

[3] M. Puckette, "Pure data," in *Proceedings of the 1996 International Computer Music Conference, Hong Kong*. 1996, pp. 269–272, Computer Music Association.

[4] Julius O. Smith III, *Digital Waveguide Modeling of Musical Instruments*, http://www-ccrma.stanford.edu/~jos/waveguide/, May 9, 2002.

[5] R. A. DeVerse, R. R. Coifman, A. C. Coppi, W. G. Fateley, F. Geshwind, R. M. Hammaker, S. Valenti, F. J. Warner, and G. L. Davis, "Application of spatial light modulators for new modalities in spectrometry and imaging," *Proceedings of the SPIE*, vol. 4959, pp. 12–22, 2003.

[6] R. M. Hammaker, A. N. Mortensen, E. A. Orr, M. K. Bellamy, J. V. Paukstelis, and W. G. Fateley, "Multi-dimensional Hadamard transform spectrometry," *Journal of Molecular Structure*, vol. 348, pp. 135–138, Mar. 1995.

[7] M. Maggioni, G. L. Davis, F. J. Warner, D. B. Geshwind, A. C. Coppi, and R. R. Coifman, "Spectral analysis of normal and malignant microarray tissue sections using a novel micro-optical electricalmechanical system," *Modern Pathology*, vol. 17 Suppl1:358A, 2004, (Abstract 1513).

[8] R. R. Coifman, "Local discriminant bases and their applications," *Journal of Mathematical and Imaging Vision*, vol. 5, pp. 337–358, 1995.

[9] Oded Ben-Tal, Jonathan Berger, Bryan Cook, Michelle Daniels, and Gary P. Scavone, "SonART: The sonification application research toolbox," in *Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan*. 2002, pp. 151–153, Advanced Telecommunications Research Institute (ATR).

[10] John M. Chowning, "Frequency modulation synthesis of the singing voice," in *Current Directions in Computer Music Research*, Max V. Mathews and John R. Pierce, Eds., pp. 57–63. MIT Press, Cambridge, MA, 1989.

[11] Dennis Klatt, "Software for a cascade/parallel formant synthesizer," *Journal of the Acoustical Society of America*, vol. 67, pp. 13–33, 1980.