ANALYSIS 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

In prior work, we examined the possibility of sound generation from colon tissue scan data using vocal synthesis models. In this work, we review key results and present extensions to the prior work. Sonification entails the mapping of data values to sound synthesis parameters such that informative sounds are produced by the chosen sound synthesis model. We review the physical equations and technical highlights of a vocal synthesis model developed by Cook. Next we present the colon tissue scan data gathered, and discuss processing steps applied to the data. Finally, we review preliminary results from a simple sonification map. New findings regarding perceptual distance of vowel sounds are presented[1,2].

1. INTRODUCTION

We seek an intuitive means of representing and analyzing highly dimensional and complex data. Sonification[1, 2] attempts to provide such a means by generating sound based on a selected set of data values. In a typical sonification scenario, data values (e.g. from a tissue scan, or financial trends) are mapped to the parameters of a sound synthesis model (e.g. a model of a wind instrument such as a flute). The sound produced is intended to aid the observer in analyzing and classifying the data observed.

In a prior work[3], we explored the possibility of generating sound from colon tissue scan data using vocal synthesis models. In this work, we review key results and present extensions.

[1]Ryan J. Cassidy supported by the Natural Sciences and Engineering Research Council of Canada and The Banff Centre (Banff, Canada).

[2]Supported by DARPA award F41624-03-1-7000.

2. CASCADED-TUBE-SECTION VOCAL SYNTHESIS MODEL

The synthesis model we present is a vocal synthesis model proposed by Cook[4], in which distinct vowel sounds are produced by varying the radii of a series of acoustic tube sections, a process intended to simulate the changing shape of the human vocal tract. When this tract model is driven by a generic glottal waveform, as shown in Figure 1, approximations to distinct vowel sounds are produced at the model output.

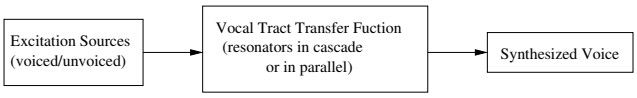

Fig. 1. Source-filter model diagram, wherein a glottal waveform (the source) is applied to a filter capturing the resonances of the vocal tract[5].

The aforementioned series of acoustic tube sections is approximated using a digital waveguide model[6]. To understand this model, we first review the physics of a single acoustic tube section, whose cross-section is illustrated in Figure 2. The quantities of interest are the pressure and volume velocity at position x in the tube and at time t, which we denote $p_1(x,t)$ and $u_1(x,t)$ (respectively).

We focus on an infinitesimal element of the tube (shown in Figure 2). By Newton's second law,

$$\underbrace{-\left(p_1(x+\Delta x,t)-p_1(x,t)\right)A_1}_{\text{Force}} = \underbrace{\rho_1 \Delta x A_1}_{\text{Mass}} \underbrace{\frac{\partial \frac{u_1(x,t)}{A_1}}{\partial t}}_{\text{Accel.}}, \tag{1}$$

where A_1 is the cross-sectional area, ρ_1 is the mass density of the fluid in the tube, and Δx is the width of the cross-section. Taking limits as $\Delta x \to 0$ and simplifying yields:

$$-A_1 \frac{\partial p_1(x,t)}{\partial x} = \rho_1 \frac{\partial u_1(x,t)}{\partial t}. \tag{2}$$

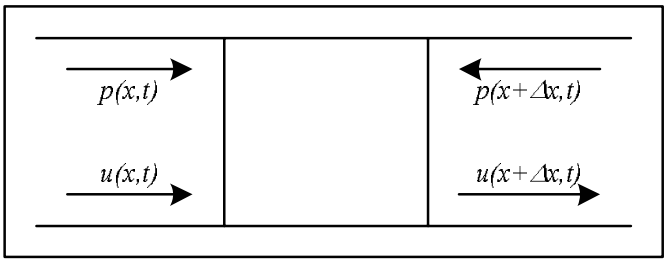

Fig. 2. Cross-section of an acoustic tube, showing an infinitesimal element. The functions $p_1(x,t)$ and $u_1(x,t)$ denote the pressure and volume velocity in the tube at position x and time t in the tube.

Next, by the law of conservation of matter,

$$-\left(u_1(x+\Delta x,t)-u_1(x,t)\right)\rho_1 = \frac{\partial \rho_1}{\partial t} A_1 \Delta x. \quad (3)$$

Simplifying and taking limits (again as $\Delta x \to 0$) yields

$$-\frac{\partial u_1(x,t)}{\partial x} = \frac{A_1}{\rho_1}\frac{\partial \rho_1}{\partial t}. \quad (4)$$

If we assume a constant proportion between excess density ρ_e (i.e. the difference between the instantaneous fluid density ρ_1 and the initial fluid density ρ_0) and pressure $p_1(x,t)$, e.g.

$$\rho_e = \frac{p_1(x,t)}{c^2}, \quad (5)$$

one obtains

$$-\frac{\partial u_1(x,t)}{\partial x} = \frac{1}{\rho_1}\frac{\partial \rho_e}{\partial t} = \frac{A_1}{\rho c^2}\frac{\partial p_1(x,t)}{\partial t} \quad (6)$$

Combining Equation (2) and Equation (6) yields

$$\frac{\partial^2 p_1(x,t)}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 p_1(x,t)}{\partial t^2}, \quad (7)$$

which is the well-known one-dimensional wave equation.

As shown by d'Alembert[7], the pressure and volume velocity above, each adhering to the wave equation, may be decomposed into left- and right-traveling wave components as follows:

$$p_1(x,t) = p_1^+\left(t-\frac{x}{c}\right) + p_1^-\left(t+\frac{x}{c}\right), \quad (8)$$

and

$$u_1(x,t) = u_1^+\left(t-\frac{x}{c}\right) + u_1^-\left(t+\frac{x}{c}\right). \quad (9)$$

Combining Equation (2), Equation (8), and Equation (9) yields

$$\begin{aligned}\frac{\partial}{\partial x}\left[p_1^+ + p_1^-\right] &= \frac{-\rho_1}{A_1}\left[u_1^{+'}\left(t-\frac{x}{c}\right) + u_1^{-'}\left(t-\frac{x}{c}\right)\right] \\ &= \frac{\rho_1}{A_1}\left[\frac{\partial u_1^+\left(t-\frac{x}{c}\right)}{\partial x}c - \frac{\partial u_1^-\left(t+\frac{x}{c}\right)}{\partial x}c\right] \\ &= \frac{\rho_1 c}{A_1}\frac{\partial}{\partial x}\left[u_1^+ - u_1^-\right],\end{aligned} \quad (10)$$

where $f'(x)$ denotes differentiation of a function with respect to its argument, and we have omitted the arguments of the right- and left-traveling wave components in certain parts of the equation above for convenience. We further observe that

$$p_1^+\left(t-\frac{x}{c}\right) = \frac{\rho c}{A_1}u_1^+\left(t-\frac{x}{c}\right), \quad (11)$$

and

$$p_1^-\left(t+\frac{x}{c}\right) = -\frac{\rho c}{A_1}u_1^-\left(t+\frac{x}{c}\right). \quad (12)$$

If we further define the characteristic tube impedance

$$R_1 = \frac{\rho c}{A_1}, \quad (13)$$

then we have

$$p_1^+ = R_1 u_1^+ \quad (14)$$

and

$$p_1^- = -R_1 u_1^-. \quad (15)$$

The traveling disturbances represented by Equation (8) and Equation (9) propagate at a speed c, and thus a section of tube of length l_1 may be simulated using a pair of continuous delays (one for each traveling wave component) of size

$$T_{d1} = \frac{l_1}{c}. \quad (16)$$

In a discrete-time system with sampling rate f_s, we obtain a delay of

$$N_{d1} = f_s T_{d1} \quad (17)$$

samples[8].

When two acoustic tube sections are joined together, the relationship between right- and left-traveling wave components, noting that the pressure at the junction $p_J = p_1^+ + p_1^- = p_2^+ + p_2^-$, may be derived as follows:

$$\begin{aligned} 0 &= u_1 - u_2 && (18)\\ &= u_1^+ + u_1^- - u_2^+ - u_2^- && (19)\\ &= G_1(p_1^+ - p_1^-) + G_2(p_2^- - p_2^+) && (20)\\ &= G_1(2p_1^+ - p_J) + G_2(2p_2^- - p_J) && (21)\\ \Rightarrow p_J &= \frac{2p_1^+G_1 + 2p_2^-G_2}{G_1+G_2} && (22)\\ \Rightarrow p_1^- &= p_J - p_1^+ && (23)\\ &= \frac{G_1-G_2}{G_1+G_2}p_1^+ + \frac{2G_2}{G_1+G_2}p_2^- && (24)\\ &= \frac{R_2-R_1}{R_1+R_2}p_1^+ + \frac{2R_1}{R_1+R_2}p_2^-, && (25)\\ \Rightarrow p_2^+ &= p_J - p_2^- && (26)\\ &= \frac{G_2-G_1}{G_1+G_2}p_2^- + \frac{2G_1}{G_1+G_2}p_1^+ && (27)\\ &= \frac{R_1-R_2}{R_1+R_2}p_2^- + \frac{2R_2}{R_1+R_2}p_1^+, && (28)\end{aligned}$$

where G_1 and G_2 are the characteristic tube admittances ($G_1 = R_1^{-1}, G_2 = R_2^{-1}$). If we further define the junction scattering parameter as

$$k = \frac{R_2 - R_1}{R_1 + R_2}, \tag{29}$$

then we have

$$p_1^- = kp_1^+ + (1-k)p_2^-, \tag{30}$$

and

$$p_2^+ = -kp_2^- + (1+k)p_1^+. \tag{31}$$

Finally, using Equation (13), we can express the scattering parameter as a function of

$$k = \frac{r_1^2 - r_2^2}{r_1^2 + r_2^2}. \tag{32}$$

The relationships of Equation (30) and Equation (31) may be implemented by a block diagram shown at the bottom of Figure 4. Moreover, the right- and left-traveling wave components of a series of acoustic tubes may be modeled using the larger block diagram of Figure 4, with a scattering junction inserted between each pair of tubes to model the impedance discontinuity. The scattering junctions shown are known as Kelly-Lochbaum scattering junctions[9], and it is by modifying the scattering parameters of these junctions that different vocal tract shapes (and hence, different vocal sounds) result. By allowing our sonification data to modify the tract radii according to a mapping, different and intuitive vowel sounds result.

We have yet to discuss the excitation signal used as input to our tract model. This signal represents the pressure waveform at the glottal folds of the human vocal mechanism.

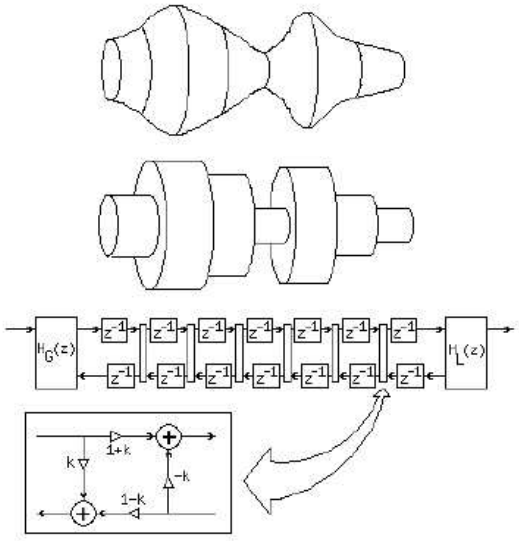

Fig. 4. Digital waveguide model of the human vocal tract, with scattering junctions between adjacent tube sections to account for changing tube impedance (adapted from [4]).

The period of this waveform gives the vocal sound its fundamental frequency, and also controls its volume. Though we could well allow our data to modify these parameters, we choose instead to keep the fundamental frequency and amplitude of the glottal waveform constant. More details regarding the waveform may be found in [4].

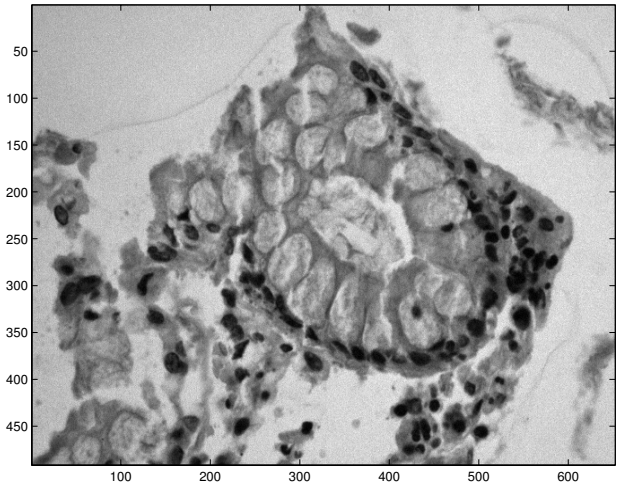

Fig. 3. Sample specimen showing normal colonic tissue (gray-scale image).

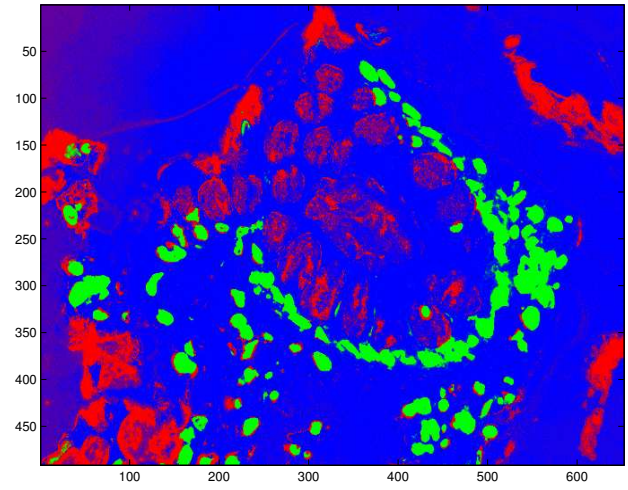

Fig. 5. Normal colon tissue image reduced from 128 to 3 dimensions per pixel via data preprocessing.

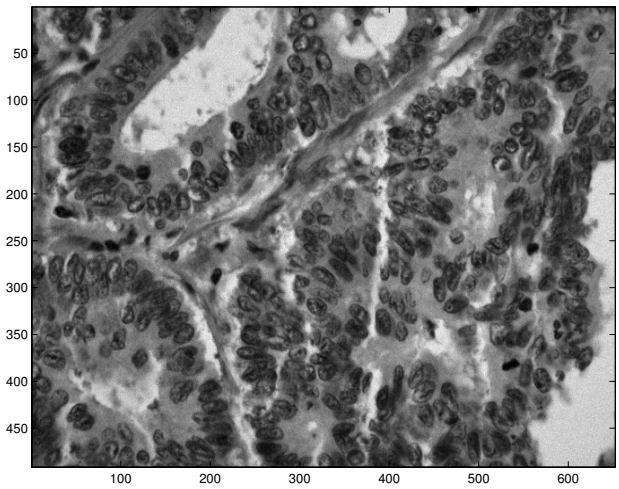

Fig. 6. Sample specimen showing abnormal colonic tissue (gray-scale image).

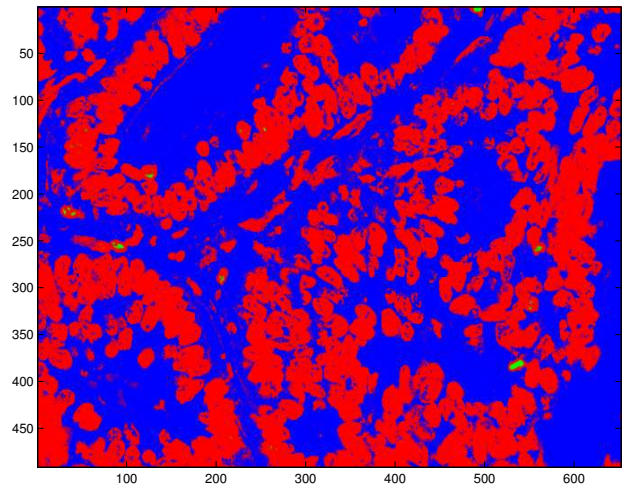

Fig. 7. Abnormal colon tissue image reduced from 128 to 3 dimensions per pixel via data preprocessing.

3. HYPERSPECTRAL TISSUE IMAGE DATA

The highly dimensional data we have chosen for sonification is obtained from hyperspectral scans of normal and abnormal colon tissue, the abnormal tissue being potentially cancerous. The data has been collected by our colleagues at Yale University[10]. For each spatial point in a specimen of colon tissue, light transmittances in each of 128 spectral bands are obtained using a Nikon Biophot microscope, with a technique known as Hadamard spectroscopy[10]. Each specimen measures roughly 0.5 mm by 0.5 mm in size, and gives an image of 491 by 652 pixels (128 values per pixel). In the study of [10], 15 normal and 46 abnormal specimens were examined, samples of which are shown in Figure 3 and Figure 6.

4. DATA PREPROCESSING

As discussed in [10], the 128 dimensions per pixel are reduced to 3 via a series of techniques including Principal Components Analysis (PCA), Local Discriminant Bases (LDB) [11], and a Nearest-Neighbour Classifier[10]. The final three dimensions give the probabilities that the point in the specimen belongs to abnormal nucleic tissue, normal nucleic tissue, and non-nucleic tissue (respectively). Examples of preprocessed specimens are shown in Figure 5 and Figure 7.

5. IMPLEMENTATION AND RESULTS

5.1. Parameter Mapping and Distance Preservation

In order to sonify the processed tissue data, we require a quantitative means of mapping data values to synthesis model parameters (in this case, these are the radii of adjacent tube sections in Cook's tract model). This map should be roughly distance-preserving, i.e. the map $\mathcal{S}$ should be such that $d_X(x,y) \approx d_S(\mathcal{S}(x),\mathcal{S}(y))$, where $d_X(x,y)$ is the distance between two data points x and y, and $d_S(\mathcal{S}(x),\mathcal{S}(y))$ is the distance between the sounds produced by the two points. While, in our current situation, the quantitative distance between data points is reasonably well understood (and provided implicitly), the quantitative distance between vowel sounds in perceptual space is less clear. A quantitative model for the perceptual distance between vowel sounds, however, has been investigated and provided by Pols et al.[12]. Such a metric has been derived empirically using Principal Components Analysis (PCA) and multidimensional scaling techniques. Further research on timbre perception is also available[13]. With such a metric in hand, the task of deriving a map becomes possible using multidimensional scaling techniques such as those discussed in [14].

5.2. Preliminary Sonification Using Cook's Model

As a preliminary sonification example, we adopt Cook's 8-section tract model, and tube section radii corresponding to two different vowel sounds:

$$r_{eee} = \begin{bmatrix} r_{eee,1} & r_{eee,2} & \cdots & r_{eee,8} \end{bmatrix}^T \tag{33}$$

for the sound /i/ as in *team*, and

$$r_{aah} = \begin{bmatrix} r_{aah,1} & r_{aah,2} & \cdots & r_{aah,8} \end{bmatrix}^T \tag{34}$$

for the sound /a/ as in *father*. Next we take the first element of the probability triple of a selected point in the specimen, and interpolate linearly between the two sounds:

$$r = (r_{eee} - r_{aah})p_1 + r_{aah}. \tag{35}$$

The result is the dominance of the sound /i/ when abnormal nucleic tissue is selected, and the dominance of the sound /a/ when normal nucleic tissue is selected. Spectrograms of the vowel sounds have been produced[3].

6. CONCLUSIONS

We have presented a vocal synthesis model and its application to sonification of hyperspectral colon tissue images. Opportunities for future work include the derivation of a distance-preserving map between data points and points in the perceptual space of vowel sounds, and the sonification of other complex data sets. The use of human vowel sounds provides an intuitive and easily-learned representation for the purpose of data analysis and classification, and the related implications of sonification are broad.

7. REFERENCES

[1] Oded Ben-Tal, Michelle Daniels, and Jonathan Berger, "De natura sonoris: Sonification of complex data," in *Mathematics and Simulation with Biological, Economical, and Musicoacoustical Applications*, C.E. D'Attellis, V.V. Kluev, and N.E. Mastorakis, Eds., p. 330. WSES Press, Cambridge, MA, 2001.

[2] Elizabeth M. Wenzel, "Spatial sound and sonification," 1992, Presented in the International Conference on Auditory Display (ICAD'92). Also published in "Auditory display: Sonification, Audification, and Auditory Interface, SFI Studies in the Sciences of Complexity", Proc. XVIII, edited by G. Kramer, Addison-Wesley, 1994).

[3] Ryan J. Cassidy, Jonathan Berger, Kyogu Lee, Mauro Maggioni, and Ronald R. Coifman, "Auditory display of hyperspectral colon tissue images using vocal synthesis models," *Proceedings of the 2004 International Conference on Auditory Display, Sydney, Australia*, 2004.

[4] 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).

[5] Gunnar Fant, *Acoustic Theory of Speech Production*, Mouton & Co., The Hague, 1960.

[6] Davide Rocchesso and Julius Smith, "Generalized digital waveguide networks," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 3, pp. 242–254, May 2003.

[7] Jean le Rond d'Alembert, "Investigation of the curve formed by a vibrating string, 1747," in *Acoustics: Historical and Philosophical Development*, R. Bruce Lindsay, Ed., pp. 119–123. Dowden, Hutchinson & Ross, Stroudsburg, 1973.

[8] Julius O. Smith, "Music applications of digital waveguides," Tech. Rep. STAN–M–39, CCRMA, Music Department, Stanford University, 1987, a compendium containing four related papers and presentation overheads on digital waveguide reverberation, synthesis, and filtering. CCRMA technical reports can be ordered by calling (650)723-4971 or by sending an email request to info@ccrma.stanford.edu.

[9] J. L. Kelly and C. C. Lochbaum, "Speech synthesis," *Proceedings of the Fourth International Congress on Acoustics, Copenhagen*, pp. 1–4, Sept. 1962, paper G42. Reprinted in [15, pp. 127–130].

[10] 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).

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

[12] L. C. W. Pols, L. J. van der Kamp, and R. Plomp, "Perceptual and physical space of vowel sounds," *Journal of the Acoustical Society of America*, vol. 46, no. 2, pp. 458–467, 1969.

[13] S. Handel, "Timbre perception and auditory object identification," in *Hearing*, B. C. J. Moore, Ed. Academic Press, San Diego, CA, 1995.

[14] Trevor F. Cox and Michael A. A. Cox, *Multidimensional Scaling*, Chapman & Hall, London, 1994.

[15] J. L. Flanagan and L. R. Rabiner, Eds., *Speech Synthesis*, Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Penn., 1973.