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

The human ability to recognize, identify and compare sounds based on their approximation of particular vowels provides an intuitive, easily learned representation for complex data. We describe implementations of vocal tract models specifically designed for sonification purposes. The models described are based on classical models including Klatt[1] and Cook[2]. Implementation of these models in MatLab, STK[3], and PD[4] is presented. Various sonification methods were tested and evaluated using data sets of hyperspectral images of colon cells[1,2].

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 formant filter model based on Klatt[1], and a physical model based on [2]. In this paper we describe the model, its implementation and its implications in the sonification of complex data.

2. FORMANT-FILTER-BASED VOCAL SYNTHESIS

When identifying dissimilar sounds such as human vowels, the human auditory system is most sensitive to peaks in the signal spectrum. These resonant peaks in the spectrum are called formants (Figure 1). Formant frequencies for human vowels vary according to the speaker type (man, woman, or child) and sex (male, female), as well as the type of vowel uttered. In 1952, Peterson and Barney[5] measured formant frequencies of human vowels and suggested control methods.

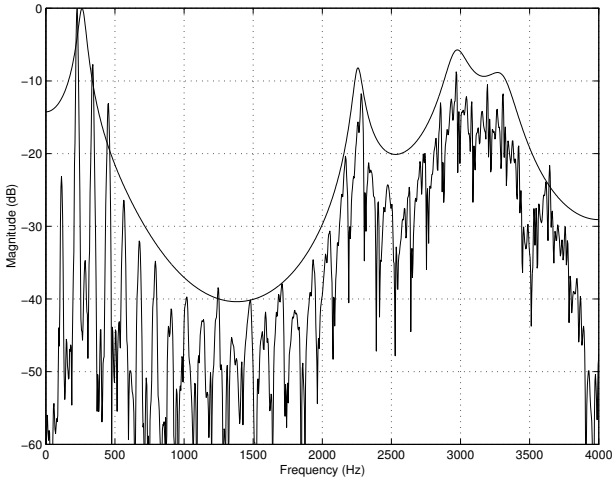

Figure 1: *Spectrum of vocal utterance of the vowel /i/ as in* team. *The smooth line enveloping the lower spectrum corresponds to the vocal tract transfer function. The resonant peaks of this curve are called formants.*

In 1980, Klatt[1] presented a method for vocal synthesis that uses the first few formant peaks as resonant peaks in a source-filter model[3](Figure 2). In his model, he used spectrally rich sounds as excitation sources. The sources are then filtered by a bank of resonators (configured either in series or parallel), whose resonant peaks correspond to formant peaks of the vowel to be synthesized. The amplitude, bandwidth, and center frequency of each formant peak play an important role in the simulation of human vowels.

Like Klatt, we have chosen a band-limited impulse train as an excitation source since, in voiced phonation, glottal

[1]Ryan J. Cassidy supported by the Natural Sciences and Engineering Research Council of Canada.

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

[3]In the context of vocal sound production, a source-filter model assumes, roughly speaking, that the time-varying air pressure waveform produced at the glottis is filtered by the vocal tract in order to produce distinct sounds. The source-filter concept was introduced by Fant[6].

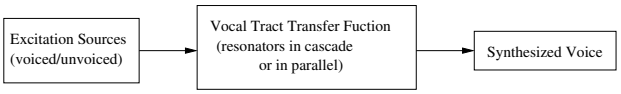

Figure 2: *Klatt's source-filter model for vocal synthesis. Radiation of the output pressure at the nose and lips to the environment is not shown.*

folds open and close periodically, producing a pulse-like excitation.

As has been mentioned earlier, in addition to formant frequencies, the amplitude and the bandwidth of each formant peak help determine the quality of the output sound. This is where formant-filter-based synthesis may be used in the sonification of highly-dimensional data. For example, by mapping the data to the amplitudes and bandwidths of the formant peaks in a vowel sound, distinct vowel-like sounds will be produced for different data points. Furthermore, by selecting vowel types whose sonorities are as perceptually distant as possible, and mapping the data between these sounds, we can maximize the possibility of data classification via sonification.

3. 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[2]. 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

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

and

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

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

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

and similarly, for the left-traveling wave component,

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

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

$$R_1 = \frac{\rho c}{A_1}, \tag{5}$$

then we have the following relations:

$$p_1^+ = R_1 u_1^+, \tag{6}$$

and

$$p_1^- = -R_1 u_1^-, \tag{7}$$

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

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

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 [7].

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[4] 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:

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

and

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

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 3. The input to the first tube section (the left-most section in Figure 3) 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).

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

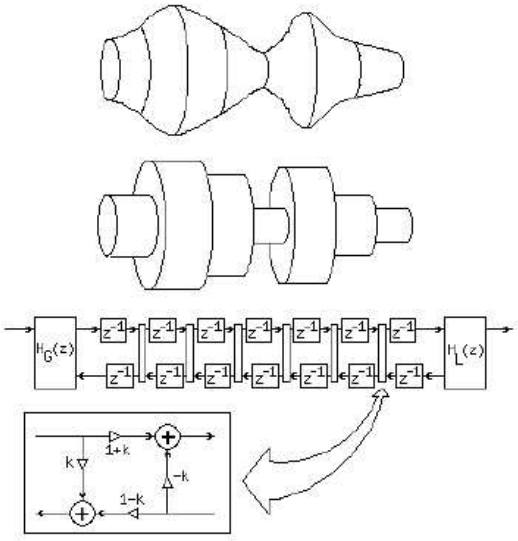

Figure 3: *Digital waveguide model of the human vocal tract, with scattering junctions between adjacent tube sections to account for changing radii (adapted from [2]).*

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 [2].

4. 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 Applied Mathematics at Yale University. First, a series of slides, each slide containing more than 300 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 mag-

nified image is recorded with a Sensovation 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 roughly 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 [8], [9], and [10].

In the study of [10], 15 datacubes of normal colonic tissue and 46 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. Figure 4 shows a grayscale image of the benign tissue, and Figure 5 shows a grayscale image of the malignant tissue.

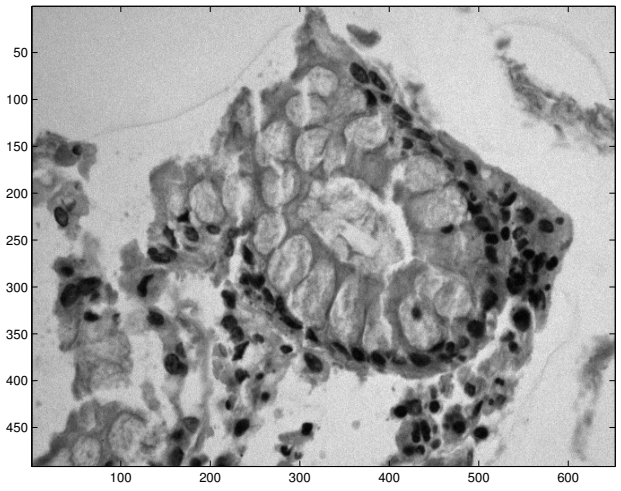

Figure 4: *Grayscale image of benign tissue used for sonification experiments.*

5. 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 [11] for a general treatment and [10] for details in this specific application).

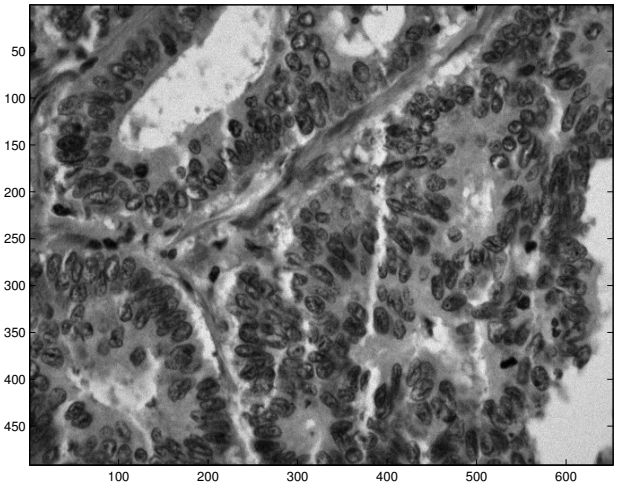

Figure 5: *Grayscale image of malignant tissue used for sonification experiments.*

3. Finally, the 5 dimensions are reduced to 3 using a Nearest-Neighbour Classification technique, again discussed in [10].

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 [10].

6. IMPLEMENTATION AND RESULTS

6.1. 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++" [3]. STK classes have been successfully integrated into prior sonification projects (see, for example, [12]). The STK already contains C++ classes for vocal synthesis, namely FM synthesis[13] and formant-filter-based synthesis[1]. However, no class is available for vocal synthesis using Cook's cascaded-tube-section tract model described in Section 3. 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.

6.2. Parameter Mapping and Distance Preservation

There are various possibilities for choosing the sonification mapping $\mathcal{S}$ from the data X to the synthesis parameters P.

In general we expect this choice to be data- and application-dependent. However, all good sonification maps should preserve distances between the data; that is, the perceptual distances between the sounds produced should be proportional to the distances between the corresponding data points. Mathematically, we require $d_X(x, y) \approx d_S(\mathcal{S}(x), \mathcal{S}(y))$, for any two data points x, y, where d_X is a distance in data space and d_S is the perceptual distance in sound space. This relationship ensures that distant data points will be mapped to sounds that are perceived as being very different, while nearby data points will be mapped to sounds perceived as similar. While finding a quantitative distance between data points is usually straightforward (though often data- and application-driven), the distance in perceptual space has most often to be studied empirically[14]. However, once the two distances are given, the problem of finding a map satisfying the condition above can be solved mathematically and algorithmically in efficient and optimal ways (e.g. [15], and references therein). Another criterion a good sonification map should satisfy is that it should be easily learned. This constraint should drive the selection of the sonification space and its parameters. As mentioned in the introduction, the vowel sounds are very natural to the human auditory system and easily recognizable, which makes them good candidates for a target sonification space.

In our situation, the data space X is three dimensional (the coordinates being the three probabilities $(p_1, \ldots, p_3)$ described above), and we could naturally map it to the space of vocal tracts with three segments. We then map the coordinate p_i (appropriately rescaled) in X to the radius of the ith segment. This mapping should preserve the discrimination between various sets of probabilities, so that differences between the sounds corresponding to various tissue types can be heard when sonifying different portions of the slide. Other sonification mappings are possible (for example, mapping the ith coordinate in X to the length of the ith tract, or each coordinate in X onto a different vowel, with intensity proportional to the value of the coordinate, etc.), and while the choice described may give a good result for this particular data, further investigation of the perceptual distance properties of the sonification space needs to be performed to hopefully discover optimal mappings.

6.3. Preliminary Sonification Using Cook's Model

In the examples of [2], Cook assumes a vocal tract with 8 tube sections, each a single sample of delay in length. Then an algorithm (described in [2]) is used to determine the section radii required to produce various vowel sounds (e.g. the vowel /iː/ heard in the word *team*, or /ʊ/ of the word *took*). For our preliminary sonification attempts, we focus on two vowel sounds: let

$$r_{eee} = \begin{bmatrix} r_{eee,1} & r_{eee,2} & \cdots & r_{eee,8} \end{bmatrix}^T \tag{25}$$

denote the radii required to produce the sound /iː/ of the word *team* and let

$$r_{aah} = \begin{bmatrix} r_{aah,1} & r_{aah,2} & \cdots & r_{aah,8} \end{bmatrix}^T \tag{26}$$

denote the radii required to produce the sound /a/ as in *father*. Finally, let

$$r = \begin{bmatrix} r_1 & r_2 & \cdots & r_8 \end{bmatrix}^T \tag{27}$$

denote the radii chosen for an image point selected for sonification.

As a preliminary sonification attempt, we adopt the following simple map: take the second element p_1 of the three-dimensional vector at a given point (this element corresponds to the probability that the selected point belongs to abnormal nucleic tissue). Then use this value to map linearly between the two vowel sounds:

$$r = (r_{eee} - r_{aah})p_1 + r_{aah}. \tag{28}$$

When a point with a high probability of belonging to abnormal nucleic tissue is selected, the vowel /iː/ will be produced, whereas when a point not having a high probability of belonging to abnormal nucleic tissue is selected, the vowel sound /a/ will dominate.

Figure 6 shows a colour image of benign colon tissue, with the three aforementioned probabilities represented by the colours red, green, and blue. Similarly, Figure 7 shows an image of malignant tissue, with the same colour scheme. Figure 8 shows a spectrogram of the vowel /iː/ produced by our implementation of Cook's vocal tract model, and Figure 9 shows a spectrogram of the vowel /a/ produced by the implementation. Note the difference in the intensity of the partials for each vowel. These sounds are produced when red and non-red points (respectively) in either image are selected.

6.4. Sonification via Formant-Filter-Based Vocal Synthesis

We used Klatt's formant vocal synthesis technique to sonify the datacubes with three dimensions per pixel (as discussed above). In order to generate sounds as perceptually distant as possible, we selected three different vowels that are most physically distant from each other in the vowel chart by the International Phonetic Association (Figure 10). In addition, we assigned each vowel to a distinct speaker type (man, woman, or child), to increase perceptual distance between the target sounds. Finally, we generated control parameters for vocal synthesis using a linear combination of three vowel sounds, whose weights are determined by the data to be sonified.

In our experiments, we used three formant frequencies and the fundamental vowel frequency as control parameters,

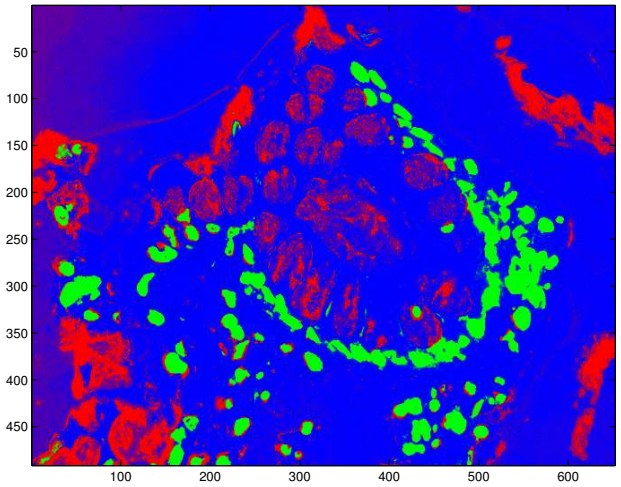

Figure 6: *Benign colon tissue image, with the 3 probability dimensions represented by the colours red, green, and blue. Note the red sections (indicating abnormal nucleic tissue) all correspond to non-nucleic tissue, so the misclassification in these patches of red is not serious.*

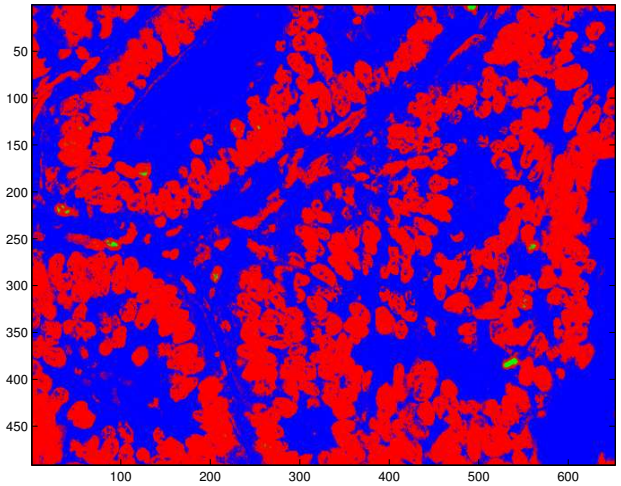

Figure 7: *Malignant colon tissue image, with the 3 probability dimensions represented by the colours red, green, and blue.*

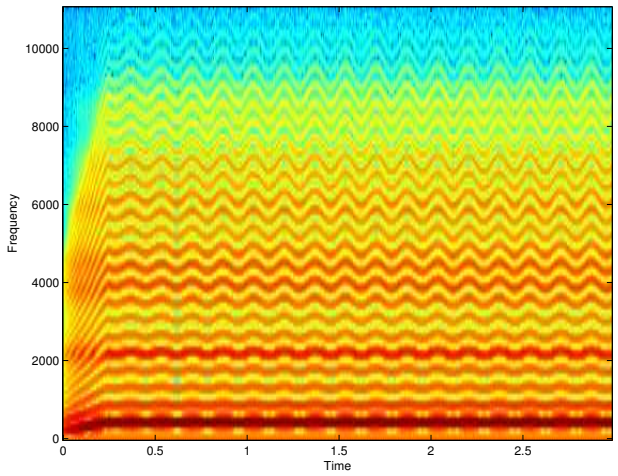

Figure 8: *Spectrogram of the vowel /iː/ (as in* team*), produced when a point with high probability of belonging to abnormal nucleic tissue is selected.*

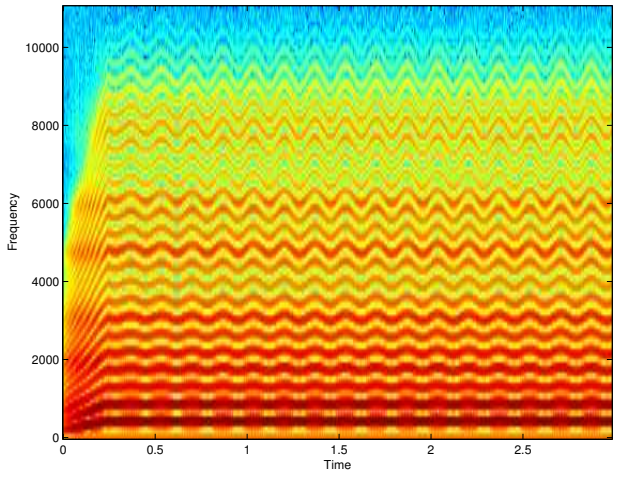

Figure 9: *Spectrogram of the vowel /a/ (as in* father*), produced when a point with low probability of belonging to abnormal nucleic tissue is selected.*

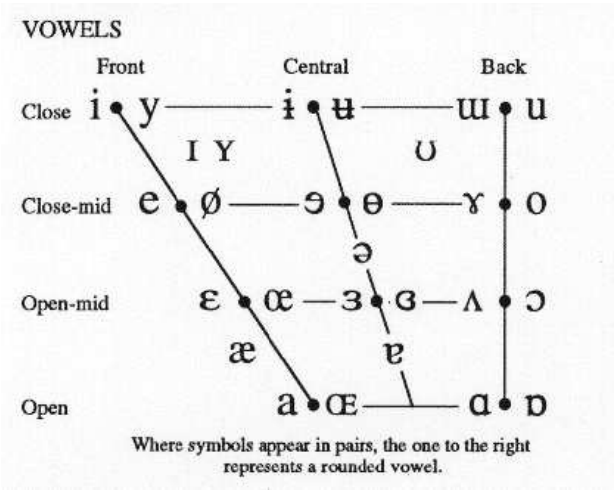

Figure 10: *Vowels plotted according to the point of primary obstruction by the tongue and the mouth during the articulation of the vowel sound (from the International Phonetic Association).*

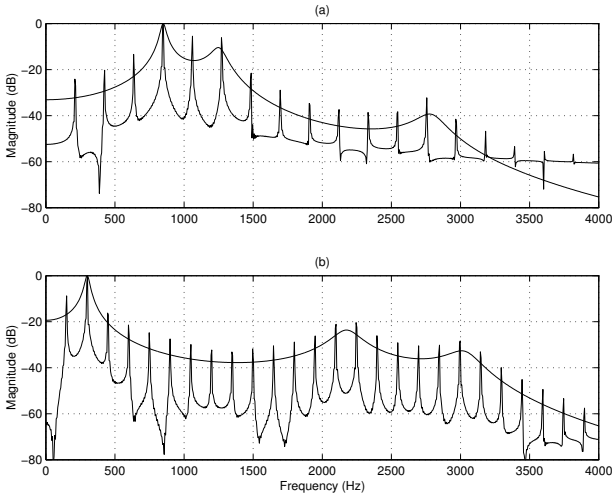

Figure 11: *(a) Spectrum of the vowel /a/ (as in* father*) as uttered by a woman, produced when a point with high probability of belonging to normal nucleic tissue is selected. (b) Spectrum of the vowel /i/ (as in* team*) as uttered by a man, produced when a point with high probability of belonging to abnormal nucleic tissue is selected.*

and we chose three vowel types: the male /i/, the female /a/, and a child's /u/. Then we have

$$f_0 = p_1 f_{0,man,/i/} + p_2 f_{0,woman,/a/} + p_3 f_{0,child,/u/} \quad (29)$$

$$f_1 = p_1 f_{1,man,/i/} + p_2 f_{1,woman,/a/} + p_3 f_{1,child,/u/} \quad (30)$$

$$f_2 = p_1 f_{2,man,/i/} + p_2 f_{2,woman,/a/} + p_3 f_{2,child,/u/} \quad (31)$$

$$f_3 = p_1 f_{3,man,/i/} + p_2 f_{3,woman,/a/} + p_3 f_{3,child,/u/}, \quad (32)$$

where f_0 is the fundamental frequency that determines the pitch, f_i is the ith formant frequency, $f_{i,speaker,vowel}$ is ith formant frequency for a speaker of type *speaker* uttering the vowel *vowel*, and the p_is are given by the aforementioned probabilities for a selected pixel. Using the above data mapping to control paramters for vocal synthesis, we could have three different criteria in the auditory domain that could help us classify multi-dimensional data. For example, with $p_1 = 1$ and $p_2 = p_3 = 0$, we would generate the pure sound of a man's /i/, and we could conclude with great confidence that the data belongs to the appropriate class. On the other hand, sonifying the data with $p_1 = 0, p_2 = 0.2$ and $p_3 = 0.8$ would result in a sound between a woman's /a/ and a child's /u/, but closer to the latter, from which we could classify the data point accordingly. It is also possible to control other parameters such as amplitudes and/or bandwidths of formant peaks.

The spectra of Figure 11 show some examples of synthesized vowels using the above data mapping method.

7. CONCLUSIONS

Initial experiments with a variety of vocal tract models suggest that human ability to easily identify vowel-like sounds is promising for intuitive sonification. Our experiments with formant-filters and a physical model of the vocal tract provide a rich domain for sonification with the principle benefit of producing sounds that the human auditory system is innately constituted to recognize and distinguish. This feature addresses the need for intuitive sonification paradigms in which many dimensions contribute to a particular categorical class of sounds. The human ability to recognize multiple simultaneously sounded vowels [16] further suggests the potency of vocal-like sounds for sonification [17]. The implications for sonification are broad and we continue to explore the benefits and challenges of this approach.

8. REFERENCES

[1] Dennis Klatt, "Software for a cascade/parallel formant synthesizer," *Journal of the Acoustical Society of America*, vol. 67, pp. 971–995, 1980.

[2] 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).

[3] 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.

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

[5] Gordon E. Peterson and Harold L. Barney, "Control methods used in a study of the vowels," *Journal of the Acoustical Society of America*, vol. 24, no. 2, pp. 175–184, Mar. 1952.

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

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

[8] 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.

[9] 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.

[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] 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).

[13] 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.

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

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

[16] U. T. Zwicker, "Auditory recognition of diotic and dichotic vowel pairs," *Speech Communication*, vol. 3, pp. 265–277.

[17] 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.