Perception of pitch is primarily dependent upon the frequency of the sound wave. For a harmonic spectrum pitch perception depends on the fundamental frequency while for an inharmonic spectrum it is a function of the amplitude-weighted mean of the spectral components. The audible range for most humans is from 20 to 15000 Hz. The smallest degree of pitch discrimination between two pitches depends on their intensity and frequency range. Experiments have shown that the human ear is more sensitive to frequency changes at the mid-frequency region between 1 and 4 kHz. The *jnd* for pitch is typically about 1/30th of the critical bandwidth at a particular frequency. The perception of pitch is also dependent on the duration of the sound. A short duration sound will be heard as a click rather than a pure tone. On average, sound should have duration of at least 13 ms to be ascribed as a definite pitch. Although the human ear has sensitivity up to around 20 kHz, sensitivity of the human ear drops significantly at higher frequencies. Thus, it is reasonable to use frequencies in the middle of the audible range, i.e. 100-5000 Hz, so that the sound is audible in most circumstances.

Those characteristics of sound which enable the human auditory system to distinguish between sounds of similar pitch and loudness are, by definition timbre. Timbre perception depends upon the harmonic content, temporal evolution, and the vibrato and tremolo properties of the sound waves. Timbre may be useful to represent multiple data streams simultaneously.

2.2.2 Sound synthesis

A number of sound synthesis methods such as additive synthesis, subtractive synthesis, frequency modulation (FM) synthesis, and granular synthesis can be used to generate sound [28]. For any given application, there is no preferred technique, as each has its own merits and demerits. In our sonification, FM synthesis was used, which has the advantage of generating a rich variety of sounds with the control of only a few parameters. The FM signal is described as $A\cos(2\pi f_c t + M \sin 2\pi f_m t)$ where f_c is the carrier frequency, f_m is the modulating frequency, A is the amplitude and M is the modulating index. In this technique, the carrier wave frequency f_c is modulated by the modulating wave frequency f_m. The FM modulated signal consists of a complex tone with frequency components separated from one another by the modulating frequency as shown in Fig. 3(a). However, if there are reflected side frequencies due to sidebands falling into the negative frequency domain of the spectrum, then the ratio f_c / f_m would determine the position of the components in the spectrum [29]. The amplitude of the components can be determined by Bessel functions, which would be a function of the modulating index M. For higher values of M, more spectral energy will be dispersed among the frequency components.

2.2.3 Parameter mapping

The parameters extracted from the OCT data can be mapped to any or all of the attributes. We selected as the significant parameters the slope of the A-scans and the spectral parameters corresponding to the low (*I*), middle (*II*) and high frequency (*III*) regions of the Fourier spectrum of the data as shown in Fig. 2(e). These parameters were mapped after appropriate scaling into the carrier frequency f_c, modulation index M, amplitude A, and modulating frequency f_m, respectively, where f_m = [(Energy in region *III*) x (f_c)].

Interpretation of the mapping is shown in Fig. 3(b). The slope of the A-scan is mapped to the pitch. The high frequency content determines the separation of the spectral components relative to the carrier frequency, while the low frequency content determines the spectral energy within these spectral components. The final synthesized sound is strongly influenced by the choice of carrier frequency. In our data sets, slope was the variable with the greatest discriminating power and hence was mapped into the carrier frequency. As a result of these mappings, the sonification of signals from adipose and tumor tissues had non-overlapping

#123990 - $15.00 USD Received 17 Feb 2010; revised 20 Apr 2010; accepted 21 Apr 2010; published 27 Apr 2010

audio spectra and the perceived sound of tumor had a higher pitch. This makes intuitive sense as the Fourier spectrum of tumor tissue has greater energy at higher frequencies compared to that of adipose tissue.

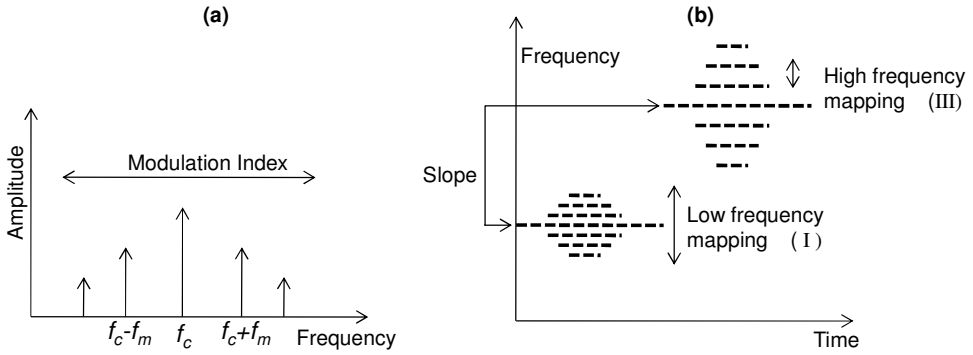

Fig. 3. Frequency Modulation (FM) synthesis. (a) Spectral components in FM synthesis. (b) Mapping of parameters for sonification via FM synthesis.

2.2.4 Sound rendering modes

The sonification of OCT data has been organized into two modes: A-scan sonification and image-mode sonification. In the A-scan sonification mode each individual A-scan (or a group of A-scans for faster playback) is sonified. Although this mode has high resolution, it has the limitation of being non-real-time as the typical A-scan acquisition rate (~0.1 ms for an A-scan rate of 10 kHz) will be much higher than the playback time (~100 ms) of the sound. A playback time of 100 ms was chosen based on the tone perception of the human ear.

Image-mode sonification may be used for real-time sonification of the data. In the image-mode, each frame is played for the duration of the playback time of the sound, and is therefore much faster than the A-scan sonification mode. In this mode, each frame is divided into a certain number of blocks and for each block the average value of the parameters are calculated and mapped into sound as shown in Fig. 4. The final synthesized sound consists of the summation of the waveforms from each individual block. The sonification (parameter calculation + sound synthesis) of each block is independent of all the other blocks. Hence, these calculations can be done in parallel for each block, which can significantly decrease the computational time for each frame. However, this mode will have a lower resolution than the A-scan sonification mode (where the resolution depends on the number of divisions of each frame).

Sound was synthesized using Matlab and played at a sample rate of 10 kHz. The final synthesized sound from each of these modes contained a clicking sound due to appending of the sound waveforms (~100 ms). These artifacts were minimized by multiplying each of the 100 ms sound waveforms with an envelope having linearly rising and decaying slopes at the edges.

#123990 - $15.00 USD Received 17 Feb 2010; revised 20 Apr 2010; accepted 21 Apr 2010; published 27 Apr 2010

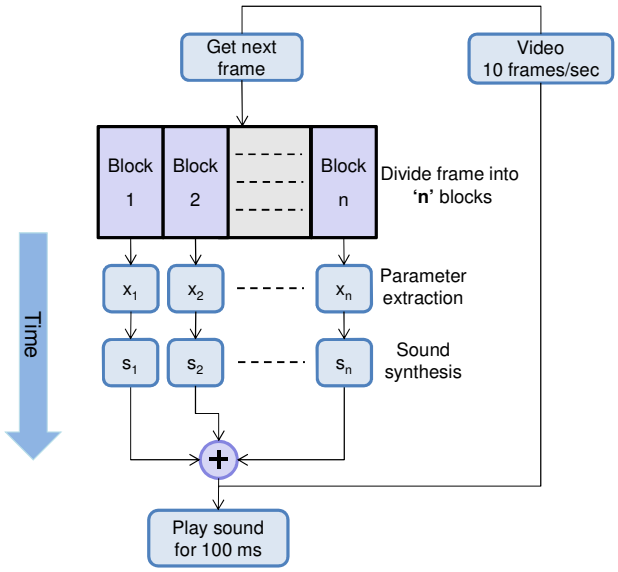

Fig. 4. Block diagram for image-mode sonification.

3. Results

The human tissue used in this study was acquired and handled under protocols approved by the Institutional Review Boards at the University of Illinois at Urbana-Champaign and Carle Foundation Hospital (Urbana, IL). The results obtained by sonification in the A-scan mode and the image-mode are shown below.

3.1 A-scan mode

In the A-scan mode the acquired A-scans are grouped together in bins, each 10 A-scans wide. The data parameters are calculated for each A-scan and averaged together for each bin. Each bin is played for a duration of 100 ms based on the tone perception of the human auditory system.

The mapping of the A-scans parameters obtained from human breast adipose and tumor tissues are shown in Table 1. These results show that adipose tissue has a sound of lower pitch with the spectral components more closely spaced to each other, and energy more widely dispersed among them. In contrast, tumor tissue has a sound of a higher pitch with relatively large spacing between the spectral components, and with most of the energy concentrated within the carrier frequency (due to the low modulation index *M*).

Table 1. A-scan parameter mapping for FM synthesis.

FM synthesis parameters	A-scan parameters	Adipose	Tumor
Carrier frequency (f_c)	Slope	Low	High
Modulation index (*M*)	Low frequency content (*I*)	High	Low
Amplitude (*A*)	Middle frequency content (*II*)	Moderate	Moderate
Modulation frequency (f_m)	High frequency content (*III*)	Low	High

Note: Roman numerals refer to frequency bands shown in Fig. 2.

Figure 5 (Media 1 – both video and audio) shows the sonification of a two-dimensional OCT image containing a tumor margin (boundary between normal adipose tissue and tumor). The data set in Fig. 5(a) was acquired using a spectral-domain OCT system with 800 nm center wavelength and 70 nm bandwidth, providing an axial resolution of 4 μm [22]. The audio spectrogram of the output sound is shown in Fig. 5(b). The audio spectrogram (computed using the short time Fourier transform) displays the frequency components of the

#123990 - $15.00 USD Received 17 Feb 2010; revised 20 Apr 2010; accepted 21 Apr 2010; published 27 Apr 2010

sound at each time instant and is helpful in visualizing the sonification results. Results demonstrate that tumor and adipose tissues have distinct sounds.

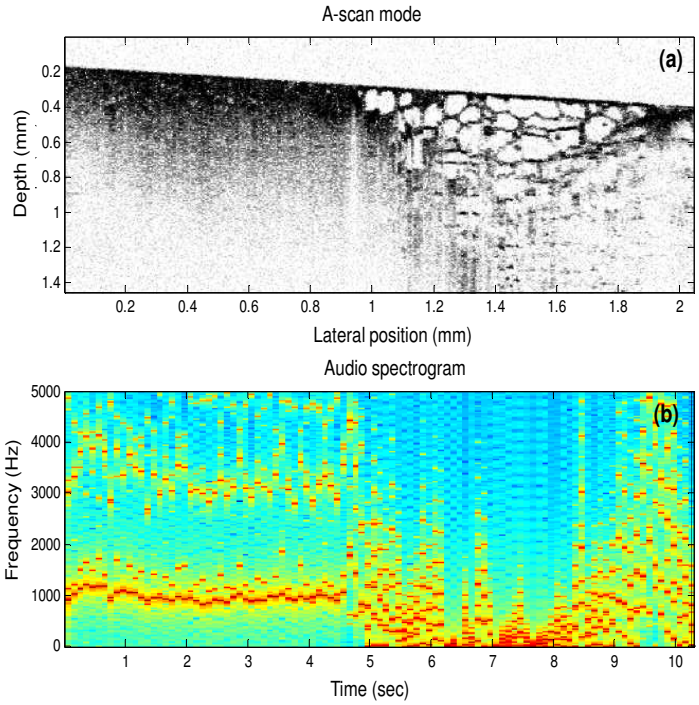

Fig. 5. Sonification using the A-scan mode (Media 1 – both video and audio). (a) Human breast tissue containing a tumor margin with tumor (left side of the image) and adipose (right side of the image). (b) Audio spectrogram of the output sound, where each column in the spectrogram corresponds to 10 A-scans in the OCT image in (a).

3.2 Image-mode

The results from image-mode sonification are shown in Fig. 6 (Media 2 (13 MB) – both video and audio). The sonification is applied to a three-dimensional volumetric data set of dimensions 1.7 mm x 3 mm x 5 mm containing both adipose and tumor tissues. This data set was acquired intraoperatively using a 1310 nm spectral-domain OCT system with 11 μm axial and 20 μm transverse resolution. Each frame was divided into 10 blocks and sonification was performed based on the scheme shown in Fig. 4. A portion of the data set (after 30 seconds) is played backwards to highlight the distinction in sonification of adipose and tumor tissues, and to mimic real-time intraoperative imaging back and forth across a tumor margin. The audio spectrogram in Fig. 6(b) demonstrates that the sound of tumor has higher frequency content than the sound of adipose tissue.

The first 190 images or frames contain adipose tissue, except for the 37th frame, which consists of tumor. This particular frame was artificially inserted between frames of adipose tissue to highlight the sensitivity of our sonification technique and the human ear at identifying subtle changes in the image data. If only image data is displayed, then the rapid transition of adipose-tumor-adipose may be missed if the user does not pay close attention to the visual display at that particular instant in time. However, the addition of another sensory information channel in the form of audio feedback in conjunction with the visual display may make this abrupt transition more easily recognized during high-speed image and data acquisition.

#123990 - $15.00 USD Received 17 Feb 2010; revised 20 Apr 2010; accepted 21 Apr 2010; published 27 Apr 2010

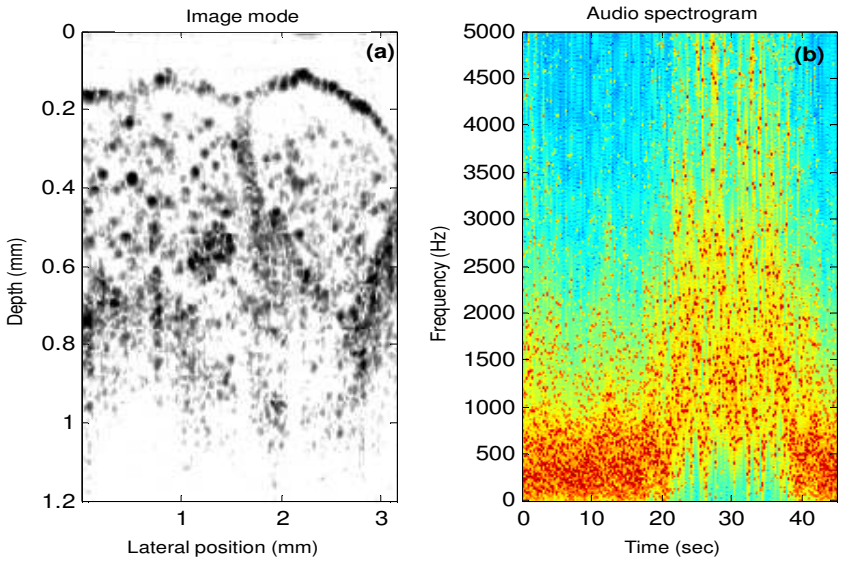

Fig. 6. Sonification using the image mode (Media 2 (13 MB) – both video and audio), (Media 3 (4 MB) – low display resolution video and audio). (a) A single frame from a three-dimensional volumetric data set, which consists of 450 frames played at 10 frames per second. (b) Audio spectrogram of the output sound where each frame in the three-dimensional volume now corresponds to a playback time of 100 ms, and the audio spectrum from each frame is represented by a single column in this spectrogram.

4. Discussion

Auditory representation of OCT images based on parameter-mapped sonification has been demonstrated in this study. The current method of sonification of OCT data may potentially be improved and be made aesthetically more pleasing by using more complex sound attributes such as vibrato and tremolo of the tones and by using dedicated hardware for sound manipulation and generation. Additional sound dimensionality such as stereo, where different parameters could be mapped to the left and right ear, may also be used. Moreover, depending on the tissue types and data sets employed, additional data parameters based on the histograms, A-scan peaks, standard deviation (for A-scan data), or textural parameters (for image-data) can be incorporated for sonification.

Sonification will be especially useful if done in conjunction with the acquisition of A-scans in real-time. For real-time performance, the calculation of the parameters and the subsequent mapping into sound attributes must be done faster than the data acquisition rate. A parallel implementation of the scheme presented in Fig. 4 can be used for real-time performance utilizing either commercially available sound synthesizers or parallel programming techniques [30]. For real-time sonification, the data must either be downsampled or averaged. This will not likely present a problem, as auditory feedback is intended to be a fast and efficient screening method for the identification of important data features that can alert the user to suspicious areas of tissue. For more detailed recognition and visualization, the user may look at the high-resolution image on the screen. The speed of real-time sonification can be increased by decreasing the playback time of sound (100 ms was used in this sonification). However, as mentioned previously, this will decrease the resolution of the sonification, producing audible clicks rather than sound tones.

One of the main challenges in sonification is finding the most efficient mapping of data parameters into sound attributes. Currently, there is no single optimized approach as the sonification technique will depend to a great extent on the type and form of the data, individual perception and preference of sound, and the computational requirements. With this

#123990 - $15.00 USD Received 17 Feb 2010; revised 20 Apr 2010; accepted 21 Apr 2010; published 27 Apr 2010

in mind, the current sonification scheme may not be optimal for every OCT data set. Data from different tissue types may have different distinguishing parameters and a sonification system would need experimentation with different mappings, synthesis techniques, and parameter tuning to customize it to the unique properties of the data sets employed. A versatile sonification system would likely have a calibration mode, where multiple parameters could be adjusted in real-time to optimize the sounds and sensitivity for identifying particular tissues of interest.

Sonification of data may also have certain fundamental drawbacks and limitations. Audio perception will vary between individual users and there could be potential interference from other sound sources such as speech and the environment. Another limiting factor is that sound attributes are not completely independent of each other. For example, loudness has frequency dependence while pitch perception also depends on the intensity levels, which may cause misinterpretation of mapped data features. The sound attributes must therefore be carefully chosen to compensate for these effects.

Future work will incorporate more tissue data from different and similar tissue types. The performance of human subjects at distinguishing between different tissue types based on audio feedback will also be evaluated. Experimentation with different mappings, different OCT data sets, and different variations in the scaling and polarity on the audio rendering is likely to further improve performance. Sonification of A-scans with multiple cell and tissue types present within a single A-scan will also be investigated.

5. Conclusion

In this paper we have demonstrated a new method to represent OCT data and images in the form of audio signals. This representation may complement the traditional visual display, and enable the user to utilize multi-sensory perception capabilities for the interpretation of OCT data under real-time imaging conditions, such as during surgical or diagnostic procedures. In the case of cancer surgery represented here, an estimate of the tumor location may first be gauged using audio feedback, with subsequent analysis of the image data from the suspect region made using tissue classification algorithms. Sonification is expected to be used as a complementary extension rather than a complete replacement of the traditional visual display. This multi-sensory approach has the potential to improve the real-time differentiation and interpretation of data during high-speed OCT imaging.

Acknowledgements

We thank Dr. Adam Zysk for providing the data sets used in Fig. 2 and Fig. 5. We thank our clinical collaborators at Carle Foundation Hospital and Carle Clinic Association, and their patients, for providing us with tissue for this study. We also thank Professor Sever Tipei from the School of Music and Professor Yongmei Michelle Wang from the Departments of Statistics, Psychology, and Bioengineering at the University of Illinois at Urbana-Champaign for their helpful discussion related to this work. This research was supported in part by grants from the National Institutes of Health, NIBIB, R01 EB005221, and NCI, RC1 CA147096. Additional information can be found at http://biophotonics.illinois.edu.

#123990 - $15.00 USD Received 17 Feb 2010; revised 20 Apr 2010; accepted 21 Apr 2010; published 27 Apr 2010