Sonification of optical coherence tomography data and images

Adeel Ahmad[1], Steven G. Adie[1], Morgan Wang[1], Stephen A. Boppart[1,2,*]

[1] *Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Biophotonics Imaging Laboratory, Beckman Institute for Advanced Science and Technology, 405 N. Mathews Avenue, Urbana, IL 61801, USA*

[2] *Departments of Bioengineering and Internal Medicine, University of Illinois at Urbana-Champaign, Biophotonics Imaging Laboratory, Beckman Institute for Advanced Science and Technology, 405 N. Mathews Avenue, Urbana, IL 61801, USA*

**boppart@illinois.edu*

Abstract: Sonification is the process of representing data as non-speech audio signals. In this manuscript, we describe the auditory presentation of OCT data and images. OCT acquisition rates frequently exceed our ability to visually analyze image-based data, and multi-sensory input may therefore facilitate rapid interpretation. This conversion will be especially valuable in time-sensitive surgical or diagnostic procedures. In these scenarios, auditory feedback can complement visual data without requiring the surgeon to constantly monitor the screen, or provide additional feedback in non-imaging procedures such as guided needle biopsies which use only axial-scan data. In this paper we present techniques to translate OCT data and images into sound based on the spatial and spatial frequency properties of the OCT data. Results obtained from parameter-mapped sonification of human adipose and tumor tissues are presented, indicating that audio feedback of OCT data may be useful for the interpretation of OCT images.

OCIS codes: (170.4500) Optical coherence tomography; (100.2960) Image analysis; (100.3008) Image recognition, algorithms and filters

References and links

1. F. T. Nguyen, A. M. Zysk, E. J. Chaney, J. G. Kotynek, U. J. Oliphant, F. J. Bellafiore, K. M. Rowland, P. A. Johnson, and S. A. Boppart, "Intraoperative evaluation of breast tumor margins with optical coherence tomography," Cancer Res. **69**(22), 8790–8796 (2009).
2. F. T. Nguyen, A. M. Zysk, E. J. Chaney, S. G. Adie, J. G. Kotynek, U. J. Oliphant, F. J. Bellafiore, K. M. Rowland, P. A. Johnson, and S. A. Boppart, "Optical Coherence Tomography: The Intraoperative Assessment of Lymph Nodes in Breast Cancer," IEEE Eng. Med. Biol. Mag. **29**(2), 63–70 (2010).
3. A. M. Zysk, F. T. Nguyen, E. J. Chaney, J. G. Kotynek, U. J. Oliphant, F. J. Bellafiore, P. A. Johnson, K. M. Rowland, and S. A. Boppart, "Clinical feasibility of microscopically-guided breast needle biopsy using a fiber-optic probe with computer-aided detection," Technol. Cancer Res. Treat. **8**(5), 315–321 (2009).
4. W. Drexler and J. G. Fujimoto, *Optical Coherence Tomography: Technology and Applications* (Springer, New York, 2008).
5. T. Hermann, "Taxonomy and definitions for sonifications and auditory display," in *Proceedings of the 14th International Conference on Auditory Display* (*Paris, France* 2008).
6. E. S. Yeung, "Pattern recognition by audio representation of multivariate analytical data," Anal. Chem. **52**(7), 1120–1123 (1980).
7. S. Barrass and G. Kramer, "Using sonification," Multimedia Syst. **7**(1), 23–31 (1999).
8. E. Jovanov, K. Wegner, V. Radivojević, D. Starcević, M. S. Quinn, and D. B. Karron, "Tactical audio and acoustic rendering in biomedical applications," IEEE Trans. Inf. Technol. Biomed. **3**(2), 109–118 (1999).
9. G. Baier, T. Hermann, and U. Stephani, "Event-based sonification of EEG rhythms in real time," Clin. Neurophysiol. **118**(6), 1377–1386 (2007).
10. M. Ballora, B. Pennycook, P. C. Ivanov, A. Goldberger, and L. Glass, "Detection of obstructive sleep apnea through auditory display of heart rate variability," in *Computers in Cardiology* (2000), pp. 739–740.
11. A. D. N. Edwards, G. Hines, and A. Hunt, "Segmentation of Biological Cell Images for Sonification," in *Congress on Image and Signal Processing, cisp* (2008), pp. 128–132.

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

12. A. C. G. Martins, R. M. Rangayyan, and R. A. Ruschioni, "Audification and sonification of texture in images," J. Electron. Imaging **10**(3), 690–705 (2001).
13. H. F. Routh, "Doppler ultrasound," IEEE Eng. Med. Biol. Mag. **15**(6), 31–40 (1996).
14. V. X. D. Yang, M. L. Gordon, S. J. Tang, N. E. Marcon, G. Gardiner, B. Qi, S. Bisland, E. Seng-Yue, S. Lo, J. Pekar, B. C. Wilson, and I. A. Vitkin, "High speed, wide velocity dynamic range Doppler optical coherence tomography (Part III): *in vivo* endoscopic imaging of blood flow in the rat and human gastrointestinal tracts," Opt. Express **11**(19), 2416–2424 (2003).
15. E. Jovanov, D. Starcevic, V. Radivojevic, A. Samardzic, and V. Simeunovic, "Perceptualization of biomedical data. An experimental environment for visualization and sonification of brain electrical activity," IEEE Eng. Med. Biol. Mag. **18**(1), 50–55 (1999).
16. A. M. Zysk, D. L. Marks, D. Y. Liu, and S. A. Boppart, "Needle-based reflection refractometry of scattering samples using coherence-gated detection," Opt. Express **15**(8), 4787–4794 (2007).
17. B. D. Goldberg, N. V. Iftimia, J. E. Bressner, M. B. Pitman, E. Halpern, B. E. Bouma, and G. J. Tearney, "Automated algorithm for differentiation of human breast tissue using low coherence interferometry for fine needle aspiration biopsy guidance," J. Biomed. Opt. **13**(1), 014014–014018 (2008).
18. T. Hermann and H. Ritter, "Sound and meaning in auditory data display," Proc. IEEE **92**(4), 730–741 (2004).
19. G. Kramer, ed., *Auditory Display-Sonification, Audification, and Auditory Interfaces* (Reading, MA: Addison-Wesley, 1994).
20. S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "A detailed investigation into the effectiveness of earcons," in *Auditory Display*, G. Kramer, ed. (Reading, MA: Addison Wesley, 1994), pp. 471–498.
21. W. W. Gaver, "Synthesizing auditory icons," in *Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems*(ACM, Amsterdam, The Netherlands, 1993), pp. 228–235.
22. A. M. Zysk and S. A. Boppart, "Computational methods for analysis of human breast tumor tissue in optical coherence tomography images," J. Biomed. Opt. **11**(5), 054015 (2006).
23. X. Qi, M. V. Sivak, G. Isenberg, J. E. Willis, and A. M. Rollins, "Computer-aided diagnosis of dysplasia in Barrett's esophagus using endoscopic optical coherence tomography," J. Biomed. Opt. **11**(4), 044010 (2006).
24. K. W. Gossage, T. S. Tkaczyk, J. J. Rodriguez, and J. K. Barton, "Texture analysis of optical coherence tomography images: feasibility for tissue classification," J. Biomed. Opt. **8**(3), 570–575 (2003).
25. E. Zwicker and H. Fastl, *Psychoacoustics - Facts and Models* (Springer, Berlin, 1999).
26. H. F. Olson, *Music, Physics and Engineering* (Dover Publications, 1967).
27. B. Moore, "Psychoacoustics," in *Springer Handbook of Acoustics*, T. D. Rossing, ed., (Springer, 2007), pp. 459–501.
28. C. Scaletti, "Sound synthesis algorithms for auditory data representations," in *Auditory Display*, G. Kramer, ed., (Reading, MA: Addison Wesley, 1994), pp. 471–498.
29. J. M. Chowning, "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation," J. Audio Eng. Soc. **21**, 526–534 (1973).
30. H. G. Kaper, E. Wiebel, and S. Tipei, "Data sonification and sound visualization," Comput. Sci. Eng. **1**(4), 48–58 (1999).

1. Introduction

Optical Coherence Tomography (OCT) can intraoperatively provide valuable diagnostic feedback about microscopic tissue morphology [1–3]. Depth-resolved backscattered light from tissues or biological specimens can be obtained from a single A-scan in OCT. The tissue structure, morphology, and beam attenuation are encoded in the intensities of the back-scattered light. Conventionally, these A-scans are mapped to an image in synchronization with the transversely scanned position of the beam, where a single A-scan forms one column of an OCT image. To exploit the high-resolution, non-invasive and real-time subsurface imaging capabilities of OCT in the operating room environment, the technique should not only be simple to use and flexible under different operating conditions, but also present data that is easily interpretable. Often it is desirable to image over large fields-of-view in real-time such as for screening or surgical guidance. Given the high-resolution capabilities of OCT, this necessitates an extremely high data acquisition rate which provides a challenge for real-time interpretation of OCT data [4].

An effective method to enhance the interpretation of diagnostic measurements may be the rendering of the data as audio signals. The process of converting data into non-speech audio signals or waveforms for the purpose of conveying information about the data is known as sonification. Sonification is a relatively new field that has drawn considerable interest in recent years due to the challenges of visualizing and interpreting the increasingly large amounts of scientific data being generated by the availability of inexpensive hardware

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

resources and computational power. The complementary nature of sound data, the superior temporal resolution of the human auditory system, and the ability to monitor multiple parallel audio streams have enabled sonification to be used as warning alarms, for process monitoring, and data exploration [5]. In general, auditory information has been utilized in a number of different instruments or devices, such as in Geiger counters, electrophysiological recordings, warning alarms, human-computer interfaces, and representation of multi-dimensional and multi-modal data [6,7]. In biomedical applications, sonification has been used for providing audio feedback for manual positioning of surgical instruments [8], surgical navigational systems, analysis of EEG signals from the brain [9], heart rate variability [10] and interpretation of image data [11] and texture [12]. Audio output has also been utilized in Doppler ultrasound [13] and Doppler OCT [14].

In some situations, auditory representation of OCT data may be more beneficial than the conventional visual display. This is especially true in the highly demanding operating room environment where the surgeon has to simultaneously keep track of a number of parameters while attending to the surgical field and following the surgical plan. The addition of an audio channel can free the visual sense for other tasks. It is known that human auditory perception is very sensitive to slight changes in the temporal characteristics of sound and can detect even small changes in the frequency of a signal [15]. These properties can be exploited to provide a faster method of tissue classification and identification of morphological landmarks in time-sensitive image-guided surgical procedures such as screening, tumor resection, or needle biopsy, and may complement the visual representation of OCT data. Sonification may also find applications where non-image data is collected such as optical needle biopsy procedures which use forward sensing devices to collect and analyze A-scan data [16,17].

Various methods of auditory representation of data have been reported in literature [18]. The simplest form is audification which is an audible playback of the data samples where each data value directly corresponds to a sound pressure value. Re-sampling, shifts in frequencies, choice of an appropriate time compression, and a level scaling factor may be required prior to audible playback of the sound [19]. Earcons [20] and auditory icons [21] have also been used, which are based on associating a unique sound to a specific message (or signal). Although simple, these methods only work on data than can be categorized. Another approach is model-based sonification where the data parameters control a parameterized sound model to generate non-speech audio [18]. The most widely used techniques are based on mapping data parameters onto sound attributes such as pitch, loudness, tempo, and duration. However, these interfaces are more difficult to design and interpretation may require extensive user training.

In this paper, we have applied sonification to OCT data and images of human breast adipose and tumor tissue with the aim of distinguishing these tissue types based on the rendered audio signals. To the best of our knowledge, this is the first work describing sonification for OCT data. In section 2, we describe the psycho-acoustic properties and sonification principles that would form the basis of any sonification system design. Section 3 shows results of sonification of OCT data. The limitations and future work are discussed in section 4 followed by the conclusions in section 5.

2. Methodology

A well designed sonification system must be fast. The sound must make intuitive sense, and the listener must be able to effectively extract important diagnostic features from the OCT data. A simple method could be the use of earcons or auditory icons. This requires the use of classification algorithms to first distinguish between tissue types. Based on the classification, each tissue type can then be assigned a unique sound. Real-time image interpretation is thus performed by the classification algorithm, which is subject to its own sensitivity and specificity limitations [22]. Another method could be the audification of spatial-domain intensity data which has the potential advantage of high speed due to low processing

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

requirements, but may be noisy due to the presence of speckle. In this paper we have used the method of parameter-mapped sonification which is based on extracting features from the data and mapping them into sound attributes. In the absence of a reliable classification scheme, this method utilizes the sensitivity of the human auditory sensory system as a tissue classification mechanism, and hence does not require prior classification of the data.

The method of parameter-mapped sonification is illustrated in Fig. 1. Characteristics of A-scans or image parameters that can be used to classify tissue types are chosen, and the extracted parameters are then mapped to a set of sound attributes for sonification. The mapping is done while considering the psycho-acoustic response of the human auditory system to different physical sound attributes. Below we discuss the key parameters used for the classification of breast tissues, the psycho-acoustic principles relevant to the synthesis of sounds, and the chosen method of sonification.

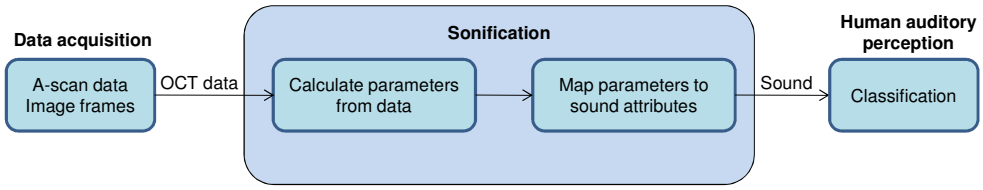

Fig. 1. Parameter-mapped sonification for OCT.

2.1 Parameter extraction from OCT data

A range of different parameters have been reported for tissue classification in OCT. These parameters depend on the nature of the data and may typically be the slope, intensity variations, spatial frequencies, periodicity of A-scans, and textural features of OCT images [17, 22–24]. The extracted parameters must have good discriminating power and must be computationally simple in order to meet the real-time requirements of a sonification system. Previous work on OCT-based classification of breast tissue has found that the slope and Fourier spectrum possess good discriminating power [17, 22]. Therefore in our sonification, we have used these parameters to distinguish between adipose and tumor tissue.

OCT images of human breast adipose and tumor tissue, and the corresponding A-scans, are shown in Fig. 2. Adipose tissue has a more regular scattering structure and lower attenuation compared to tumor tissue, which is more highly scattering and exhibits greater signal attenuation. The different attenuation causes a difference in the slopes of the (logarithmically mapped) A-scans with tumor having a higher slope compared to adipose tissue [17]. The slopes were calculated by selecting values in the A-scans that were above a certain threshold (~30% of the maximum value) and then fitting a linear function to these values as shown in Figs. 2(c) and 2(d).

Previous work has shown that the unique Fourier signatures from tissue types can be used for classification [22]. The Fourier transforms of A-scans are shown in Fig. 2(e). Adipose, due to its regular structure, has more energy concentrated in the low frequency regions while tumor has more high frequency content. The Fourier transform was normalized to a unit area and truncated to half of the maximum spatial frequency (reciprocal of the system axial resolution) as the latter half mostly consisted of noise and did not show any differences between the tissue types. The truncated Fourier space was divided into three non-overlapping regions representing the low (from 0 to 6% of the truncated spatial frequencies), middle (22-44%), and high frequencies (66-100%) as shown in Fig. 2(e). The squares of the areas under the curve corresponding to each of these regions (labeled as region I, II and III respectively) were selected as the three spectral parameters in our sonification.

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

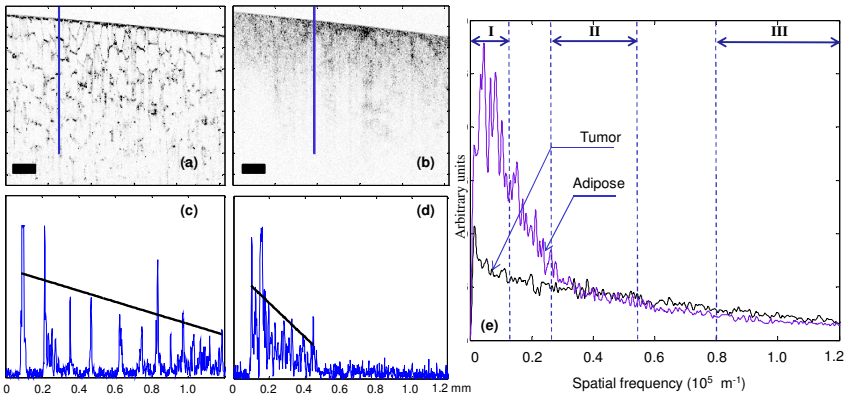

Fig. 2. OCT image and data analysis from human breast (a) adipose tissue and (b) tumor tissue. (c) A-scan corresponding to the highlighted line from the adipose tissue. (d) A-scan corresponding to the highlighted line from the tumor tissue. (e) Normalized Fourier transforms (average of 150 A-scans). Roman numerals indicate the regions corresponding to the three spectral parameters. Scale bars represent 200 μm. The superimposed lines in (c) and (d) represent the 1st order fit to the A-scan for calculation of slope.

2.2 Parameter mapping to sound attributes

The psycho-acoustic properties of human hearing have been widely studied and are well understood. Psychological perception of sound can be described in terms of loudness, pitch, timbre, and time. These sound attributes in turn depend on the physical properties of intensity, frequency, waveform, and duration of the sound waves. The relationship between the physical properties and the psycho-acoustic perception of sound must be considered for sonification design. Moreover, the interactions amongst the sound attributes must also be taken into account, especially when more than one of these variables is to be manipulated simultaneously in the same sonification.

2.2.1Psycho-acoustic principles

The psycho-acoustic concepts of *critical band* and *just noticeable difference (jnd)* have the strongest implication for a good sonification system design. Critical band is the frequency-dependent bandwidth at which sound intensities interact with each other. It can be approximated by the relationship $\Delta f = 25 + 75(1 + 1.4 f^2)^{0.69}$ where f is the center frequency in kHz [25]. The concept of *just noticeable difference (jnd)* characterizes the ability to differentiate between two nearly equal stimuli. In general, people are much better at making relative judgments of the sound attributes than at making an absolute judgment.

Loudness L is related to the sound intensity I by the relationship $L = kI^{0.3}$ where k is a constant which depends upon the units used and individual perception. However, the same intensity may produce a different sensation of loudness depending on the frequency of the waves, as described by the Fletcher-Munson curves [26]. Based on these curves it can be seen that variation in perceived loudness may be minimized by using a frequency range of approximately 800-2000 Hz. Moreover, these curves show that the sensitivity of the human ear to loudness is greatest between 3000 and 4000 Hz. The *jnd* in intensity between two sounds is about 1 dB (12% change in intensity). However, in real life scenarios, a change of 3 dB (50% change in intensity) is easily detectable by humans [25]. Loudness is also dependent on duration. With an increasing duration of a sound, the perception of loudness stabilizes after about 100 ms. Using loudness as a sound attribute for sonification is challenging as the human auditory system adjusts to the loudness level of sound. Loudness is also affected by the distance from the source, and the loudness level of a sound may be masked by other sounds in the environment [25,27].

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

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.

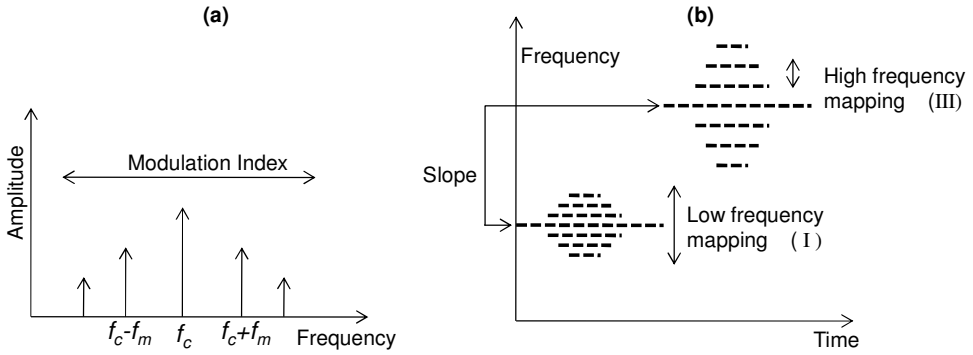

Fig. 3. Frequency Modulation (FM) synthesis. (a) Spectral components in FM synthesis. (b) Mapping of parameters for sonification via FM synthesis.

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

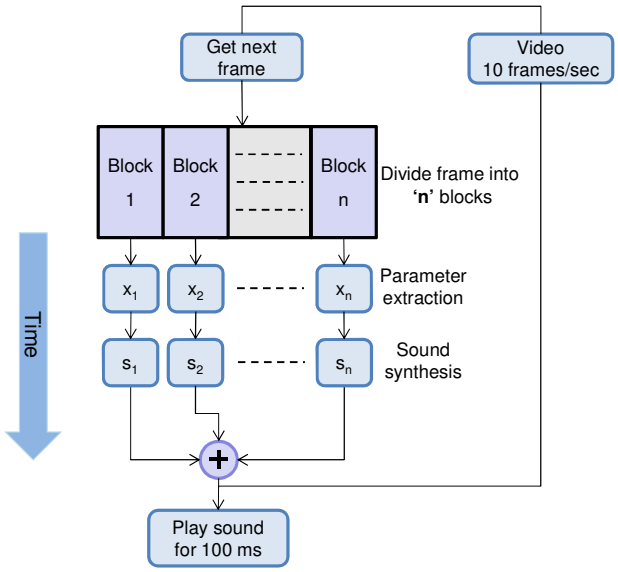

Fig. 4. Block diagram for image-mode sonification.

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.

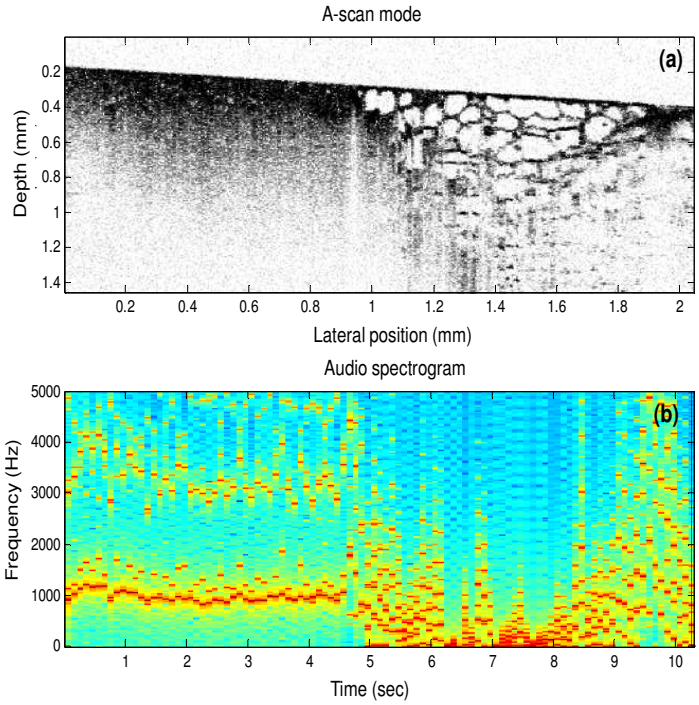

Fig. 5. Sonification using the A-scan mode (Media 1 – both video and audio). (a) Human breast tissue containing a tumor margin with tumor (left side of the image) and adipose (right side of the image). (b) Audio spectrogram of the output sound, where each column in the spectrogram corresponds to 10 A-scans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

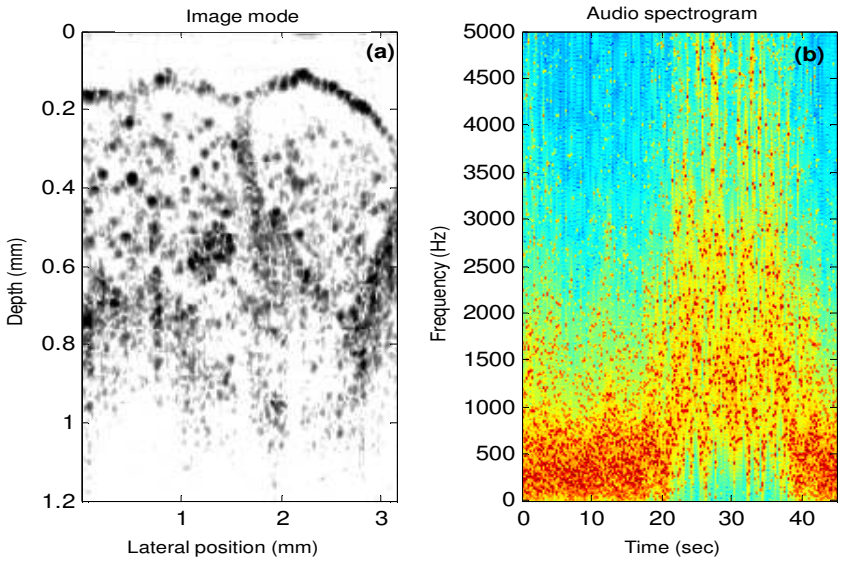

Fig. 6. Sonification using the image mode (Media 2 (13 MB) – both video and audio), (Media 3 (4 MB) – low display resolution video and audio). (a) A single frame from a three-dimensional volumetric data set, which consists of 450 frames played at 10 frames per second. (b) Audio spectrogram of the output 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