AUDIO MODELING AND LOUDNESS ESTIMATION WITH IJDSP MOBILE SIMULATIONS

Girish Kalyanasundaram‡, Mahesh Banavar† and Andreas Spanias‡

‡SenSIP Center, School of ECEE, Arizona State University, Tempe, AZ, USA
†Department of ECE, Clarkson University, Potsdam, NY, USA
gkalyan1@asu.edu, mbanavar@clarkson.edu, spanias@asu.edu

ABSTRACT

Audio signal modeling and simulation is important in several coding, noise removal, and recognition applications. This paper focuses on implementing models for loudness estimation and their use in estimating parameters on iOS mobile devices (iPhones and iPads). We briefly address estimating excitation patterns and loudness through auditory models. These loudness estimation and other algorithms were implemented in the award winning educational iOS app iJDSP for performing DSP simulations on mobile devices. The modules were introduced to graduate students in the general signal processing area, to evaluate their effectiveness as teaching tools. The evaluation process involved giving the students a pre-quiz, guiding them through hands-on activities on the iOS app, and finally, a post-quiz. Assessments results were positive with noticeable improvement of student understanding of topics such as spectrograms and linear predictive coding.

Index Terms**—** Loudness estimation, auditory pattern, iJDSP, mobile education, iOS

1. INTRODUCTION

Digital signal processing techniques are strongly motivated by many popular speech and audio processing applications in which they are used. Hence, the illustration of these DSP techniques shown along with their underlying motivation would strongly benefit students specializing in this field. Advanced signal processing concepts such as time-frequency representation of signals, concepts related to speech coding such as linear predictive coding and line spectral pairs, etc. are widely used in existing DSP systems. Effective illustration of these concepts generally requires the use of examples and visualizations in order to help students better understand them.

In recent years, mobile devices have been identified as powerful platforms for educating students and distance learners. For instance, in a recent study on using the iPad in primary school classrooms, it was found that iPads are effective due to their mobility and that they enhance student engagement [1]. Studies show that mobile tools have several advantages in teaching a broad range of subjects, from the arts, to language and literature, to the sciences [2-5].

Fig. 1. The iJDSP interface, as seen on an iPhone. DSP block diagrams can be created graphically in the main workspace, function blocks can be accessed from the function list, blocks can be configured by changing parameters suitably, and plots can be seen. Tips for use and further information about the app can be accessed from the main workspace.

iJDSP is a mobile educational iOS app for performing DSP simulations through a block diagram based approach on iOS devices such as iPhones and iPads [6]. Students can create simulations on the screen by placing a set of blocks and suitably connecting their inputs and outputs (Fig. 1). Function blocks can be chosen from a list and added to the simulation setup. Parameters that define the input signals and the system (filter coefficients, FFT length, etc.) can be edited using several interfaces (See Fig. 1).

This application has the ability to access the device microphones for recording signals, which are then processed and visualized. Frame-by-frame visualizing capabilities are provided for analyzing the system for each input audio frame. User interfaces have been developed for functions such as the spectrogram, linear predictive coding, line spectrum pair analyses, and viewing masking thresholds and auditory patterns [26,27]. Some of them are described in this paper.

Auditory models are implemented in iJDSP for visualizing auditory patterns and estimating perceptual loud-

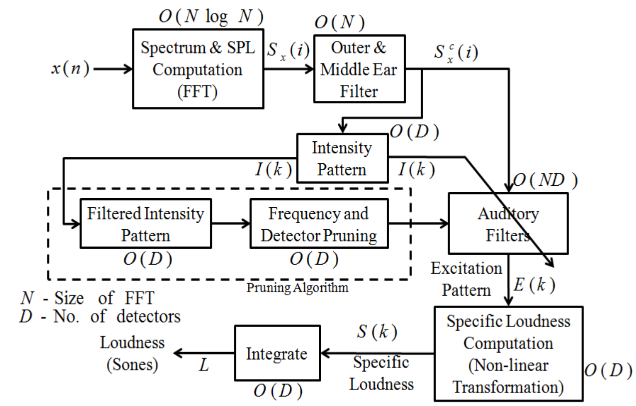


Fig. 2. Block diagram representation of the Moore & Glasberg model along with the pruning scheme, labeled with indicated computational complexities.

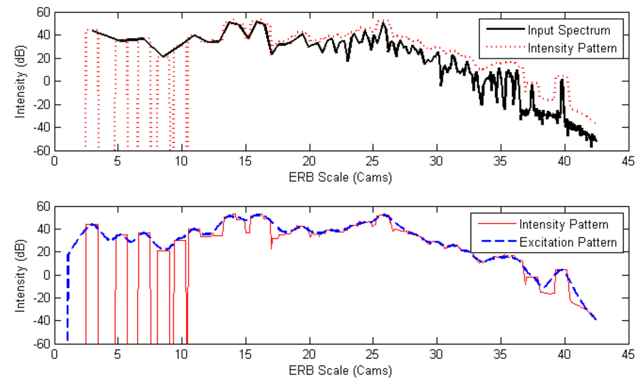


Fig. 3. For a frame of music sampled at 44.1 kHz the intensity pattern along with the spectrum in the ERB scale is shown (top), and the intensity pattern along with the excitation pattern (bottom) are shown.

-ness. Loudness is the measure of perceived intensity of sound. The human auditory system, upon reception of a stimulus, transmits neural impulses to the brain, where the perception of loudness is inferred. Loudness is measured in units called sones [7-9, 10, 11]. Based on the notion of critical bandwidths [9], more sophisticated methods model the human ear as a bank of many overlapping and highly selective bandpass filters [13-15,25]. The pattern of energies of the signal within the filter bands (called excitation pattern) is used to compute the neural excitation (called the auditory pattern or loudness pattern) and the total loudness. This process is explained in Section 2. The Moore and Glasberg model is shown to perform well with a variety of auditory inputs, giving accurate loudness measures [16]. Several applications such as sinusoidal selection, speech enhancement, bandwidth extension, and rate determination make use of auditory patterns.

This article focuses on the implementation of audio and speech processing methods and auditory models. Loudness estimation on iOS devices on the award winning iJDSP application are described along with their use in education.

2. MOORE & GLASBERG MODEL AND LOUDNESS ESTIMATION

This section briefly describes the Moore and Glasberg loudness estimation (see Fig. 2) [16]. The spectrum $S_x(i)$ of a signal $x(n)$ is filtered by the outer-middle ear and the effective spectrum $S_x^c(i)$ reaches the inner ear. The inner ear is modeled as a filter bank with rounded exponential responses. Frequencies of on an *auditory scale* are measured by the Equivalent Rectangular Bandwidth (ERB) at each frequency [17-19]. The frequency f in Hz is mapped to an “ERB number” and the ERB scale, whose unit is denoted ‘Cam’ or ‘ERB units’ is represented by:

$$d(\text{in Cam}) = 21.4\log_{10}\left(4.37f/1000+1\right). \qquad (1)$$

Let D represent the number of auditory filters, and let $L_r=\{d_k \mid |d_k\text{-}d_{k-1}|=0.1,\ k=1,2\ldots D\}$. The frequency response of the auditory filter at detector location d_k is given by:

$$W(k,i) = \left(1+p_{k,i}g_{k,i}\right)\exp\left(-p_{k,i}g_{k,i}\right),\ k = 1\ldots D,\ i = 1\ldots N, \quad (2)$$

where $p_{k,i}$ is the slope of the auditory filter corresponding to the detector d_k at frequency f_i and $g_{k,i}=|(f_i - f_{c_k})/f_{c_k}|$ is the normalized deviation of f_i from the center frequency f_{c_k} of the detector [19]. The auditory filter slope $p_{k,i}$ is dependent on the intensity pattern, $I(k)$, which is the intensity of the spectrum within one ERB at detector d_k.

$$I(k)=\sum_{i\in A_k} S_x^c(i),\ A_k=\{\,i \mid d_k\text{-}0.5< f_i^{erb}\le d_k+0.5, i=1 \text{to } N\}. \quad (3)$$

The excitation pattern is then evaluated from the following expression, requiring $O(ND)$ operations.

$$E(k) = \sum_{i=1}^{D} W(k,i).S_x^c(\omega_i),\ k = 1,\ldots D \text{ and } i = 1,\ldots N. \quad (4)$$

The specific loudness is computed from the excitation pattern as per the following expression.

$$S(k) = c\left(\left(E(k)+A(k)\right)^{\propto}\text{-}A^{\propto}(k)\right) \text{ for } k = 1,\ldots D. \quad (5)$$

$A(k)$ is a frequency dependent constant [16]. The specific loudness $S(k)$ is integrated to obtain the total loudness, which requires $O(D)$ operations. Evaluation of excitation pattern (see Fig. 2) is computationally expensive. As an illustration, the intensity pattern and excitation pattern of a frame of music signal are shown in Fig. 3. In iJDSP, the implementation is simplified by ignoring the dependency of the auditory filter slopes on the intensity pattern, which does not significantly affect the excitation pattern values for most ranges of frequencies [19].

3. IJDSP IMPLEMENTATIONS

In addition to the auditory models, iJDSP [21] also has implementations of functions such as the spectrogram and LPC. These functions are described in this section.

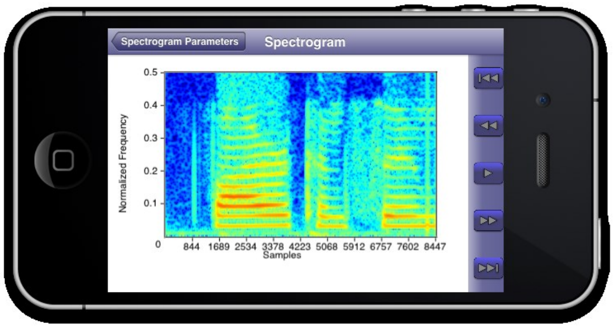


Fig. 4. The spectrogram function in iJDSP for visualizing speech spectra varying with time.

3.1. Spectrogram

The *Spectrogram* function has been developed in iJDSP to compute and display the spectrogram of a signal (Fig. 4). Varying the window length can configure the spectral and temporal resolution of the spectrogram. Other parameters such as the shape of the windowing function, the overlap between successive windows, and the length of the FFT can also be modified.

3.2. Loudness Estimation and Masking Thresholds

The *Psychoacoustic Model* function in iJDSP, displays the masking thresholds, the loudness pattern of the signal, and the total loudness of the signal in sones.

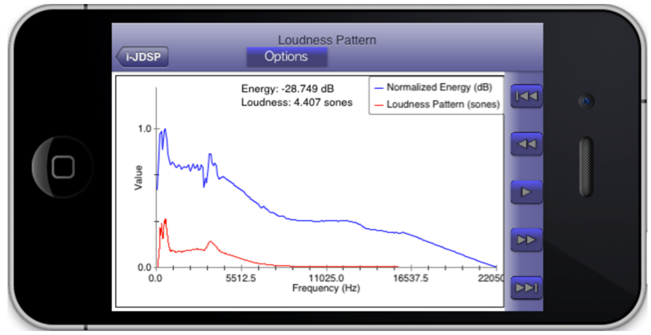
Sound can be recorded using the *Sound Recorder* block and connected to the *Psychoacoustic Model* block. Fig. 5(a) shows the loudness pattern of a signal input, and the total loudness of the input signal frame. The masking threshold of the signal is shown in Fig. 5(b). The signal after truncating the masked frequencies is also available for visualization from the options provided in the interface.

3.3. LPC in iJDSP – An Exercise

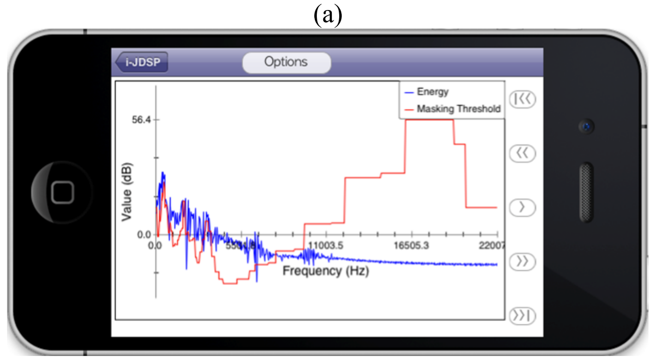
The speech of any signal can be resynthesized from its LPC coefficients and the excitation residual signal [22,23]. Hence, the resynthesized signal can be observed by filtering the residual signal with the LPC filter using a *Filter* block. The quality of this signal can be evaluated by calculating the signal-to-noise ratio (SNR) between the original signal and the signal resynthesized from the LPC coefficients [24]. This can be used to demonstrate the concepts of filtering and the effects of parameter quantization to students, with the help of hands-on activities and exercises. One such exercise is described below.

In iJDSP, a simulation setup is provided for illustrating the variation of the SNR of the resynthesized speech signal with respect to the original speech signal. The simulation setup can be chosen from the "LPC Quantization Setup" option, accessed through the "+" button in the toolbar at the bottom of the screen. The block diagram is automatically generated, and appears as shown in Fig. 6(a).

The *Quantizer* block quantizes any input according to the

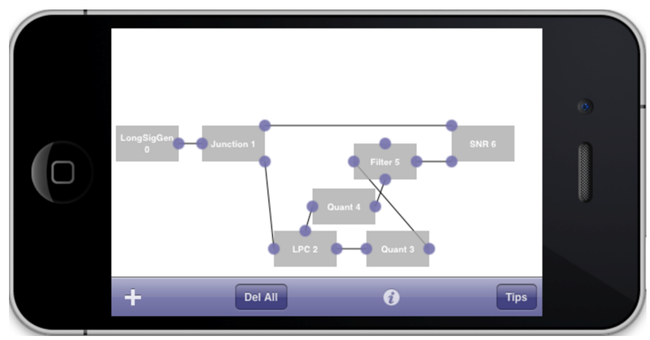


(a)

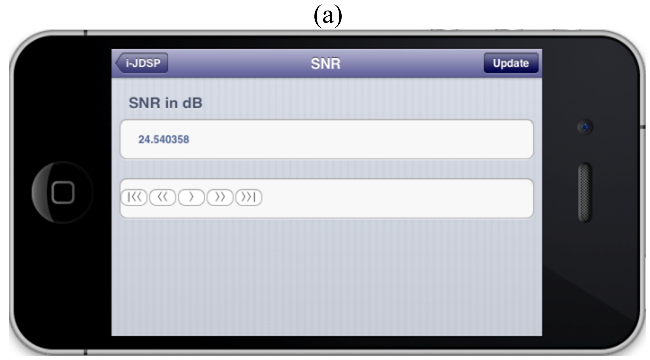


(b)

Fig. 5. (a) The loudness pattern is shown in the psychoacoustic model function user interface. (b) The masking threshold of the signal.



(a)



(b)

Fig. 6. (a) LPC Quantization and analysis-synthesis setup. (b) User interface of the SNR block. SNR is displayed in decibels. The playback buttons allow traversal through the input frames to view the resulting SNR for each input frame

bit depth that is specified by the user. By setting a particular bit depth in a quantizer, the output SNR for the signal can be calculated using the *SNR* block. The interface of the SNR

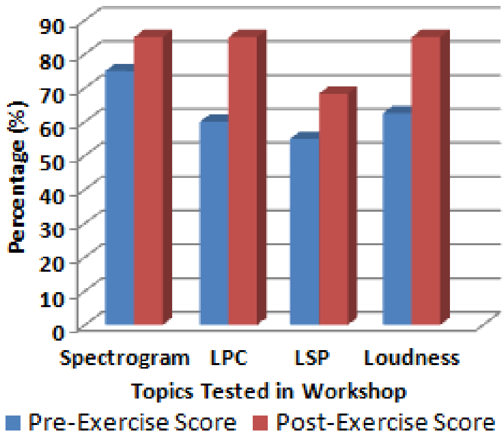


Fig. 7. Test scores averaged over all students, before and after performing iJDSP exercises. Improvement is shown in all areas.

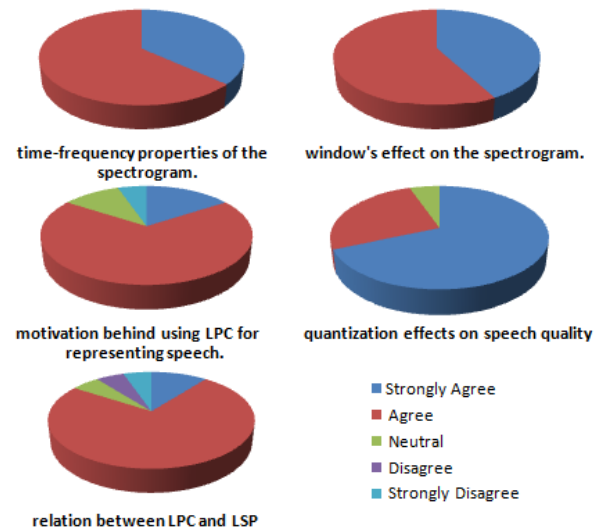


Fig. 8. Response of students indicative of subjective opinions on effectiveness of iJDSP in understanding delivered speech/audio DSP concepts.

block is shown in Fig. 6(b). Playback buttons have been added to the interface to allow the user to view the SNR frame-by-frame for any input speech. The user can view the SNR for the current frame being processed, or the SNR for the entire speech signal by pressing the “>>|” button.

4. SOFTWARE ASSESSMENTS

The effectiveness of reinforcing relevant speech and audio processing concepts through illustrative simulations in iJDSP was evaluated through assessments conducted at ASU in the fall semester in 2013. Graduate students specializing in DSP and Communications were introduced to the speech/audio DSP concepts of spectrograms and their properties; linear predictive coding and the associated motivation and applications; line spectral pairs and their properties; and the concept of perceptual loudness, all with hands-on exercises using iJDSP. The process adopted for the evaluation exercises comprised of (a) a pre-quiz on the concepts involved in the exercise, (b) a brief lecture on the relevant signal processing concepts and simulation exercises using iJDSP, and (c) a post-quiz to test the efficacy of iJDSP in improving student understanding of the concepts.

The pre- and post-quizzes were aimed at understanding the effect of iJDSP in learning the concepts taught in the lecture. Fig. 7 shows the improvement in student understanding using iJDSP. We show the average scores of students for questions asked in each topic. The average scores on spectrograms showed an improvement of about 10% over the pre-assessment quiz. An improvement of 25% was seen with understanding of linear predictive coding and its properties. Similarly, improvements were seen with LSP and loudness related questions.

In the post assessment questionnaire, students provided subjective opinions on whether the speech processing modules in iJDSP were useful in improving understanding in each of the exercises. The responses were used to evaluate the usefulness of the iJDSP functions in improving student understanding of the topics covered in the exercise. The students were asked to respond with one of the following options: Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree. The results of the evaluation are shown in Fig. 8. Almost unanimously, the students responded that iJDSP helped them better understand the speech/audio DSP concepts taught in the assessment session.

In addition to the technical assessments, the app’s quality was also assessed through a set of questions recording the user’s opinion on the qualities of the app such as the fluidity of the interface, aesthetics, and user-friendliness. It was found that more than 90% students familiarized themselves with the iJDSP user environment in less than 15 minutes. They also indicated that the interface is appealing and the simulations are easy to setup. The interface was also reported to be responsive to user inputs. The students also found the speed of the application’s execution to be satisfactory.

5. CONCLUSIONS

Audio processing functions such as auditory models, spectrograms, and linear predictive coding were developed as part of iJDSP to add to the visualization tools in the software package. Studies showed an improvement of student understanding of relevant speech/audio processing concepts when students were made to use the software to perform exercises on specific DSP concepts. The students also provided feedback indicating that the developed functions in the app along with the exercises were effective in improving their understanding of the topics covered. The students also found the software quality and user friendliness appealing.

ACKNOWLEDGEMENTS

The authors from ASU are supported in part by the SenSIP Center and the NSF Phase 3 Grant Award 0817596.

REFERENCES

[1] S. Henderson and J. Yeow, "iPad in Education: A case study of iPad adoption and use in a primary school," in the Proceedings of 45th Hawaii International Conference on System Sciences, 2012.

[2] G. Engel, "Using Mobile Technology to Empower Student Learning," in the Proceedings of 27th Annual Conference on Distance Teaching & Learning, 2011.

[3] G. Engel, "Using Mobile Technology to Empower Student Learning," in the Proceedings of 27th Annual Conference on Distance Teaching & Learning, 2011.

[4] N. Ostashewski and D. Reid, "iPod, iPhone, and now iPad: The evolution of multimedia access in a mobile teaching context," in the Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, 2010.

[5] N. Ostashewski, D. Reid, M. Ostashewski, "Mobile Teaching and Learning Technologies: Ukrainian Dance Instruction in Canada," in IADIS Mobile Learning 2009, 2009.

[6] J. Liu, et. al, Interactive DSP laboratories on mobile phones and tablets, in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012.

[7] S.S. Stevens, A scale for the measurement of a psychological magnitude: loudness, *Psychological Review*, vol. 43, no. 5, pp. 405-416, Sep. 1936.

[8] H. Fletcher and W.A. Munson, Loudness, its definition, measurement and calculation, *The Journal of the Acoustical Society of America*, Oct 1933.

[9] American National Standard Specification for Sound Level Meters, ANSI S1.4-1983 (R2006), 2006.

[10] T. Painter, A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451-515, April 2000.

[11] T. Painter, A. Spanias, "Perceptual segmentation and component selection for sinusoidal representations of audio," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 2, pp. 149-162, March 2005.

[12] H. Fletcher, Auditory Patterns, *Reviews of Modern Physics*, vol. 12, no. 1, pp. 47-65, Jan. 1940.

[13] E. Zwicker and B. Scharf, A model of loudness summation, *Psychological Review*, vol. 72, 1965.

[14] B.C.J. Moore and B.R. Glasberg, A revision of Zwicker's loudness model, *Acustica - Acta Acustica*, vol. 82, pp. 335-345, 1996.

[15] T. Dau, D. Püschel, and A. Kohlrausch, A quantitative model of the effective signal processing in the auditory system. i. model structure, *Journal of the Acoustical Society of America*, 2001.

[16] B. C. J. Moore, B. R. Glasberg, and T. Baer, A model for the prediction of thresholds, loudness, and partial loudness, *Journal of Audio Engineering Society*, vol. 45, no. 4, pp. 224-240, Apr. 1997.

[17] H. Krishnamoorthi, V. Berisha and A. Spanias, A low-complexity loudness estimation algorithm, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* , 2008.

[18] H. Krishnamoorthi, A. Spanias, and V. Berisha, A frequency/detector pruning approach for loudness estimation, *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 997-1000, Nov. 2009.

[19] B.R. Glasberg and B.C.J. Moore, Derivation of auditory filter shapes from notched-noise data, *Hearing Research*, 1990.

[20] *Sound Quality Assessment Material Recordings for Test Subjects*. Brussels: EBU Technical Centre, Sep. 2008.

[21] A. Spanias, *Digital Signal Processing: An Interactive Approach*, Lulu Press, 2nd Edition, ISBN 978-1-4675-9892-7, May 2014.

[22] A.S. Spanias, "Speech Coding: A Tutorial Review," *Proceedings of the IEEE*, vol. 82, no. 10, pp. 1441-1582, October 1994.

[23] K. N. Ramamurthy, A. S. Spanias, *MATLAB® Software for the Code Excited Linear Prediction Algorithm: The Federal Standard-1016*, Morgan and Claypool Publishers, ISBN 1608453847, Jan 2010.

[24] A. Spanias, V. Atti, "Interactive On-line Undergraduate Laboratories Using J-DSP," *IEEE Transactions on Education Special Issue on Web-based Instruction*, vol. 48, no. 4, pp. 735-749, Nov. 2005.

[25] J. J. Thiagarajan, A. Spanias, *Analysis of the MPEG-1 Layer III (MP3) Algorithm Using MATLAB*, Morgan and Claypool Publishers, ISBN: 978-1608458011, Nov. 2011.

[26] S. Sandoval, V. Berisha, R. Utianski, J. Liss, A. Spanias, "Automatic assessment of vowel space area," *Journal of the Acoustic Society of America*, vol. 134, pp. El1-El5, Nov. 2013

[27] A. Fink, A. Spanias, P. Cook, "Derivation of a new banded waveguide model topology for sound synthesis," *Journal of the Acoustic Society of America*, vol. 133, no. 2, pp. EL76-EL81, 2013.