

INTERACTIVE JAVA MODULES FOR THE MPEG-1 PSYCHOACOUSTIC MODEL

Yu Song, Andreas Spanias, Venkatraman Atti, and Visar Berisha¹

MIDL Lab, Department of Electrical Engineering
Arizona State University, Tempe, AZ 85287-5706, USA
[yu.song, spanias, atti, and visar]@asu.edu

ABSTRACT

This paper presents a collection of interactive Java modules for the purpose of introducing undergraduate DSP students to perceptual audio coding principles. This effort is part of a combined research and curriculum program funded by NSF that aims towards exposing undergraduate students to advanced concepts and research in signal processing. A computer laboratory with several supporting exercises and Java functions has been developed for use in our undergraduate DSP course. This exercise along with the accompanying Java software was assigned and assessed in the Summer of 2004 and will be reassessed in the Fall of 2004. Results of this assessment along with student comments are presented at the end of the paper.

1. INTRODUCTION

Psychoacoustic analysis is an integral part of several audio coding standards (see Fig. 1), e.g., the MPEG-1 Layer-III (MP3) [1], the Dolby² AC-3 [2] [4], and the DTS [3]. Most current audio coders achieve compression by incorporating into the coder several strategies that make use of psychoacoustic principles and the human auditory psychophysics [4]-[6]. Although audio codec designers and practitioners are well aware of the fundamental ideas used in perceptual audio coding algorithms, undergraduate students typically do not get an opportunity in courses to study these concepts. During the last five years, there have been several notable efforts to introduce advanced multimedia technologies in the undergraduate education [7]-[10]. Introducing advanced multimedia concepts at the undergraduate level requires extensive reference to applications that appeal to undergraduates, e.g., cellular telephony, MP3 players, and surround sound systems. Results from our earlier assessment studies [11] revealed an elevated student interest and enhanced learning when DSP concepts were bundled with exciting applications and presented through interactive Java-based experiments. In this paper, we describe innovative software-modules and computer experiments for exposing students to – *i*) the perceptual masking properties of the human ear; *ii*) the perceptual entropy notions; and *iii*) the psychoacoustic models employed in the MP3 players. We also present assessment results and student comments collected from

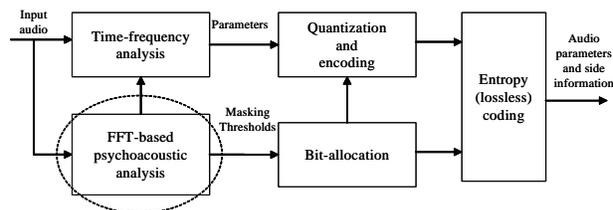


Fig. 1. A general perceptual audio encoder block diagram

actual laboratory use. The presented work is part of an NSF Combined Research and Curriculum Development and Educational Innovation (CRCD-EI) program that aims at developing and injecting advanced signal processing modules in undergraduate DSP-related courses.

The Java modules presented, enable students and distance learners to perform on-line psychoacoustic simulations and visualize web-based interactive demos from remote locations. These Java modules have been integrated in the Arizona State University's award-winning on-line simulation software called Java-DSP (J-DSP) [12]-[14]. J-DSP is an object-oriented programming environment that enables students to establish and run DSP simulations on the internet. The J-DSP simulation environment is intuitive and the visual programming allows students to learn to establish and run simulations in minutes. The Java software is accompanied by laboratory exercises that complement classroom and textbook content. The laboratories cover several concepts including, the absolute threshold of hearing, critical band analysis, the Bark scale, the spread of masking, the simultaneous and temporal masking effects, and the perceptual entropy. In addition to the on-line laboratories, a fast Fourier transform (FFT)-based computer project has been designed to provide students with hands-on experiences to the psychoacoustic model employed in the ISO/IEC MPEG-1 audio standard [1]. Web-based assessment instruments have been developed to assess whether our "interactive modules" enhanced learning in an undergraduate DSP class. The rest of the paper is organized as follows. Section 2 describes the Java modules. Section 3 addresses the on-line laboratories. In section 4, we present the graphical design of the MPEG-1 psychoacoustic model. Section 5 presents the assessment results and conclusions.

¹ V. Berisha is supported by a NSF Graduate Research Fellowship. This work is sponsored by NSF CRCD-EI award 0417604.

² "Dolby," "Dolby AC-3," are trademarks of Dolby Laboratories. "DTS" is the trademark of Digital Theatre Systems.

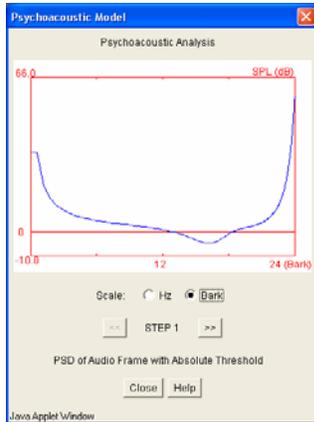


Fig. 2. The absolute threshold of hearing

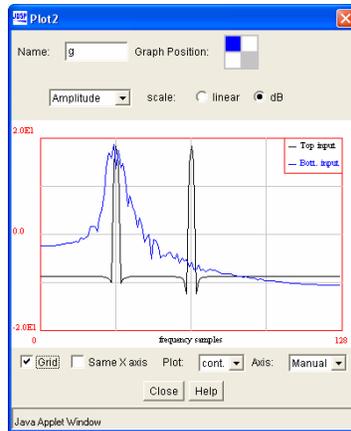


Fig. 3. The GUI for tone and noise experiments

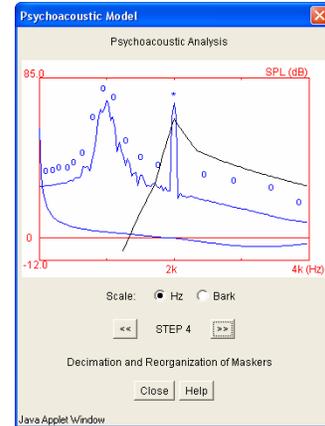


Fig. 4. The GUI for masking experiments

2. ORGANIZATION OF THE JAVA MODULES

We describe below some of the Java modules developed to assist in teaching the principles of psychoacoustics. Introductory information is given to students before they engage themselves with computer simulations. This includes theoretical details associated with each of the Java modules, and a step-by-step procedure to help students become acquainted with the graphical user interface (GUI).

2.1. Absolute threshold of hearing

Figure 2 shows a sample simulation that teaches students how the absolute threshold of hearing characterizes the amount of energy needed in a tone such that it can be detected by a listener in a quiet environment. This module also enables students to learn the frequency dependence of the threshold of hearing. In addition, this module helps students visualize the absolute threshold of hearing on a Bark scale. This gives students a brief exposure to the frequency selectivity properties of human hearing and auditory filterbanks before they proceed to the critical band frequency analysis module.

2.2. Critical band frequency analysis

This module was designed primarily to teach students (i) the concepts of frequency-to-place transformation in the cochlea (inner ear) along the basilar membrane; (ii) how this transformation can be interpreted from the signal processing perspective as a bank of bandpass filters; (iii) how these non-uniform, overlapping bandpass filters are quantified using *critical bandwidths* as a function of frequency; and (iv) the significance of the Bark scale in critical band frequency analysis.

2.3. Simultaneous masking

This module enables two types of masking experiments, namely, tone-masking-noise (TMN) and noise-masking-tone (NMT). As an example of such an experiment, figure 3 represents a GUI in which two tones located at 1kHz and 2kHz, along with narrowband noise of 160Hz bandwidth are present. From Figure 4, note that the narrowband noise masks the pure

tone at 1kHz, whereas the 2 kHz tone is still audible. Students can learn through graphics the asymmetry of masking power. In particular, experiments that demonstrate the strong masking power of narrowband noise can be performed. On-line laboratories that highlight TMN and NMT are given in Section 3.2.

2.4. The spread of masking

While the previous module teaches simultaneous masking effects within a critical band, this module gives insight to the spread of masking across several critical bands. From Figure 5, a masking tone generates an excitation (along the basilar membrane) that has been modeled by a corresponding minimum masking threshold. From the minimum masking threshold, students can compute graphically, the noise-to-mask-ratio (NMR) and signal-to-mask-ratio (SMR). In coding applications, the spread of masking is typically modeled using an approximately triangular masking function as shown in Figure 5.

2.5. Non-simultaneous or temporal masking

This module enables students to perform temporal masking experiments. For example, for a masker of finite duration, temporal masking occurs both prior to masking onset as well as after masker removal. This masking phenomena result in an increase of audibility thresholds for masked sounds. A sound player has been implemented for the students to listen to the post-masking sounds. Typically, pre-masking tends to last for 2ms, while post-masking may extend for more than 200ms. Non-simultaneous masking experiments that involve maskers with varying strengths and durations can also be performed.

3. ON-LINE LABORATORIES FOR PSYCHOACOUSTIC EXPERIMENTS

A total of ten computer exercises have been designed and grouped into three on-line laboratories. These include, the critical band analysis lab, the masking experiments lab, and the perceptual entropy computation lab. These labs would also serve as a set of preparatory experiments (see section 4.1) for the students before they perform the ISO/IEC MPEG-1 psychoacoustic model-1 simulations.

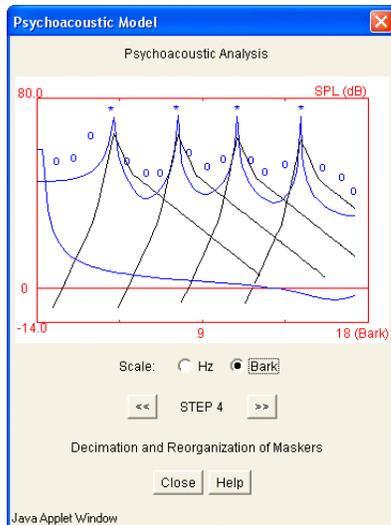


Fig. 5. An example simulation depicting the spread of masking

3.1. Exercise-1: Critical band analysis

The first exercise involves presenting the concept of critical bands to students. An introductory experiment has the students model a filterbank that mimics the critical band structure of the human auditory filterbank. The second experiment involves performing a Bark scale transformation using the bilinear transform. The third experiment involves a simple computer simulation to generate a signal with two pure tones within a critical bandwidth (e.g., at 650Hz and 700Hz) and to differentiate them for varying tone amplitudes (e.g., 0.2 to 1). A similar experiment performed for tones present in two different critical bands gives the students additional practice with critical bands.

3.2. Exercise-2: Masking experiments

The second exercise deals with the concept of masking. The first experiment in this exercise has the students simulate the two important masking scenarios. In the first case (TMN), a pure tone occurring in the center of a critical band masks noise of any sub-critical bandwidth or shape. In the second case (NMT), a narrow-band noise signal of bandwidth 1 Bark masks a tone in the same critical band. The experiment requires that the students find the threshold for each scenario and determine the SMR. This experiment gives the students insight to the asymmetry of masking power between noise and tone. A more advanced experiment presents how the noise masking power depends on frequency. The students repeat the NMT experiment in each Bark and plot the threshold as a function of the center frequency of the noise masker.

3.3. Exercise-3: Perceptual entropy experiments

The third exercise illustrates the importance of perceptual entropy (PE) in audio coding. The first experiment of this exercise requires that the students determine the PE histogram for different types of audio. This will give students a visual

representation of PE and it will also give them a sense of what types of audio require more bits for artifact-free representation. In addition, numerical examples are also formulated in order to make the students familiar with the PE formula and to reinforce the graphical results obtained in the previous experiment. A more complex experiment has the students perform bit allocation on a simplified coding scheme for three cases. In the first case, the overall bit rate is less than the PE. In the second case, the overall bit rate is the PE. Finally, in the third case, the overall bit rate is much larger than the PE. The students will subjectively assess the three audio files and compare them with the original PCM encoded file.

4. SINUSOIDAL SYNTHESIS BASED ON THE ISO/IEC MPEG-1 PSYCHOACOUSTIC MODEL-1

Below we outline a number of term-projects for use in a DSP course. In order to prepare the students for longer, more involved projects, the basics of psychoacoustics must first be introduced. The exercises outlined in Section 3 can serve as stand alone lab experiments; however they can also serve as preparatory examples for longer projects. These exercises will provide the students with the basics of psychoacoustics in a visual manner. In addition to these experiments, depending on the project, content specific preparatory examples not covered in the experiments are also introduced. As an example, in an audio synthesis project, in addition to the psychoacoustic concepts, the students should also be introduced to the concept of peak-picking. This can be done with an exemplary experiment in which they perform peak picking by using the simple least-squares method.

4.1. Part – A: Preparatory exercise

We first assign an FFT-based least squares peak-picking method to select a sufficient number of FFT components such that a signal is synthesized from a constrained DFT basis. The challenge is now to use the global masking thresholds, and in particular, the JND curve, to select the perceptually relevant FFT components. All the FFT components below the JND curve are assigned a minimal value (for example, -50 dB SPL), such that these perceptually irrelevant FFT components receive a minimum number of bits or no bits at all. The two methods are compared using both a signal-to-noise-ratio (SNR) and subjective evaluations.

4.2. Part – B (I): Masking asymmetry

The purpose of this portion of the term-project is to understand the asymmetry in masking power between tonal and noise maskers and how this concept is used in the psychoacoustic model. For each temporal analysis frame, the student selects the tonal and noise components in the frequency domain and applies the appropriate masking relationships in a manner similar to the MPEG 1 model. Finally, the global masking threshold will be obtained. Although most of the code for the psychoacoustic model will be provided, the student will be responsible for minor changes in the algorithm, such as looking at the effects of slight modifications to the rules for detection of tonal components. In addition, the student is also asked to look at the effects of tonal and noise components on the global masking threshold.

4.3. Part – B (II): Audio synthesis

In this section of the project, the students are required to make use of the MPEG 1 psychoacoustic model in a simple audio coding algorithm. The students are required to use the global masking threshold in order to perform peak-picking in such a manner that the synthesized signal is perceptually-identical to the original. In addition to subjective evaluations, the student is also asked to find objective measures (SNR, per frame SNR) of the output audio compared to the original. Figure 6 gives an example of the graphical programming required in J-DSP to perform the audio synthesis using the psychoacoustic model.

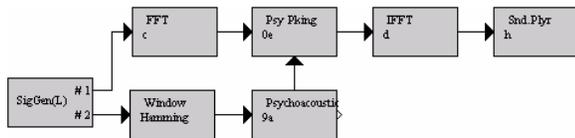


Fig. 6. An example of sinusoidal synthesis using the psychoacoustic model

5. ASSESSMENT RESULTS AND CONCLUSIONS

Concept-specific and general evaluation forms have been developed to obtain an overall assessment on the computer laboratories and to collect a subjective opinion on the Java modules, respectively. We describe here the concept-specific evaluation. Details on general evaluation are given in [12]. The concept-specific forms focus on each exercise by posing questions that determine whether the student has learned a specific psychoacoustic concept. For instance, 75% of the students agreed that they learnt the significance of the absolute threshold of hearing and 50% of the students reported that they understood the scenarios of TMT and NMT. More results are given in Table 1. In order to obtain even more consistent assessment results, we are developing *pre/post-lab* assessment questionnaire. In the *pre/post-lab* evaluation, the questions are technical and are posed to evaluate student's understanding of the key psychoacoustic concepts before and after performing a particular lab assignment. Statistical methods such as the effect size measures [15] are employed to analyze the *pre/post*-assessment results and quantify the degree of student learning attributed specifically to the Java modules and on-line laboratories. The computer laboratories (section 3) and the ISO/IEC MPEG-1 psychoacoustic model-1 project (section 4) were first used in a multimedia class in summer 2004 at Arizona State University (ASU). Some students suggested to extend the documentation for each laboratory and to provide more "hints" for analyzing the results. Most students found the web-based modules highly intuitive.

6. REFERENCES

- [1] ISO/IEC JTC1/SC29/WG11, "Information Technology-Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/sec, IS11172-3: Audio," 1992.
- [2] G. Davidson, "Digital Audio Coding: Dolby AC-3," in *The Digital Signal Processing Handbook*, V. Madisetti and D. Williams, Eds., CRC Press, pp. 41.1-41.21, 1998.

TABLE I
CONCEPT-SPECIFIC ASSESSMENT FROM A MULTIMEDIA CLASS
(SUMMER 2004) AT ARIZONA STATE UNIVERSITY[†]

Evaluation question	Yes (%)	Have minor questions (%)	No (%)
1. I learnt the critical band filterbank properties and the use of Bark scale in psychoacoustic analysis.	50	50	-
2. I understood how a JND curve or the masking threshold is computed	75	-	25
3. I can say whether a tone masks another tone given the JND curve.	50	50	-
4. I understood the effects of omitting some of the FFT components below the JND curve	75	-	25

[†] The assessment results are preliminary. In the final paper, we will submit more comprehensive statistics obtained from the Fall 2004 DSP class.

- [3] The Digital Theater Systems (DTS). web-page: www.dtsonline.com
- [4] T. Painter and A. Spanias, "Perceptual Coding of Digital Audio," *Proc. of IEEE*, vol. 88, no. 4, pp. 451-513, Apr. 2000.
- [5] B.C.J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, Fifth Edition, Jan. 2003.
- [6] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer-Verlag, second edition, Apr. 1999.
- [7] E. A. Lee, "Overview of the Ptolemy Project," *Technical Memorandum UCB/ERL M03/25*, University of California, Berkeley, CA, 94720, USA, July 2, 2003.
- [8] S. C. Douglas, G. C. Orsak, M. A. Yoder, "DSP in high schools: new technologies from the Infinity Project," in *Proc. of IEEE ICASSP*, Vol.4, pp. 4152-4154, May 2002.
- [9] J. H. McClellan, R. W. Schafer, and M. A. Yoder, *DSP First: A Multimedia Approach*, Pearson Education, Dec. 1997.
- [10] J. H. McClellan, *et al*, *Computer-based Exercises for Signal Processing Using MATLAB ver.5*, Pearson Education, Oct. 1997.
- [11] A. Spanias, V. Atti, A. Papandreou-Suppappola, *et al.*, "On-line signal processing using J-DSP," *IEEE Trans. Sig. Proc. Letters*, vol. 11, no. 10, pp. 1-5, Oct. 2004.
- [12] The Java-DSP web-page [on-line]; MIDL LAB, Arizona State University: <http://jdsp.asu.edu>
- [13] A. Spanias, *et al.*, "Assessment of the Java-DSP (J-DSP) on-line laboratory and software," in *Proc. of 33rd IEEE FIE-03*, vol. 1, pp. T2E_10 - T2E_15, Nov. 2003.
- [14] V. Atti and A. Spanias, "Web-based experiments for introducing speech recognition basics in a DSP course," in *Proc. of ICASSP-2004*, May 17-21, 2004, Montreal, Canada.
- [15] J. Cohen, *Statistical power analysis for the behavioral sciences*, Lawrence Earlbaum Associates, second edition, 1988.