

Collaborative Online Activities for Acoustics Education and Psychoacoustic Data Collection

Youngmoo E. Kim, *Member, IEEE*, Travis M. Doll, *Student Member, IEEE*,
and Raymond Migneco, *Student Member, IEEE*

Abstract—Online collaborative game-based activities may offer a compelling tool for mathematics and science education, particularly for younger students in grades K-12. We have created two prototype activities that allow students to explore aspects of different sound and acoustics concepts: the “cocktail party problem” (sound source identification within mixtures) and the physics of musical instruments. These activities are also inspired by recent work using games to collect labeling data for difficult computational problems from players through a fun and engaging activity. Thus, in addition to their objectives as learning activities, our games facilitate the collection of data on the perception of audio and music, with a range of parameter variation that is difficult to achieve for large subject populations using traditional methods. Our activities have been incorporated into a pilot study with a middle school classroom to demonstrate the potential benefits of this platform.

Index Terms—Computers and education, computer-assisted instruction, games, sound and music computing.

1 INTRODUCTION

RECENTLY, a number of web-based collaborative games have been developed for the large-scale collection of “ground truth” data for difficult computational problems (e.g., tagging of objects in images and features of music) through a fun and engaging activity [1]. In the same spirit, we have developed online activities designed to collect psychoacoustic data from a large number of users, but with the additional aim of educating students about the mathematical and scientific concepts underlying physical acoustics. These web-based interfaces are designed as game activities with minimal complexity so they can be easily used by students without previous training. We present two such collaborative games that explore the perception of audio through the manipulation of speech and music sounds. The first focuses on the “cocktail party problem” [2], i.e., our ability to focus on and understand one voice in a crowded room (a problem that still eludes computational solutions). The second game investigates musical instrument timbre (the sound qualities that differentiate instrument sounds).

To maintain widespread accessibility, our activities require only internet access through a web browser and run independently of external applications. This paper will describe the implementation of these games and their usage in a classroom of eighth grade students (ages 13-14). We will demonstrate the potential of this platform to enhance student learning and to offer fine-grained data on student progress and assessment.

2 BACKGROUND

Since the introduction of personal computers, educators have sought ways of integrating digital games into K-12 curricula to augment traditional teaching methods. Some studies incorporating video games have reported improved spelling and reading skills [3] and increased student motivation [4] at elementary grade levels (K-2). Effective integration of computer games at all levels of education, however, has proven to be difficult because of the inherent challenges in designing a learning activity that is fun and engaging for an increasingly demanding audience, and some so-called “edutainment” software is neither entertaining nor educational [5]. But recent work by Shaffer and colleagues suggests that video games and computer simulations offer an ideal medium for introducing new collections of skills and knowledge (the theory of epistemic frames), valuing innovation over rote learning, into education curricula [6].

There are three common approaches in which computer games are integrated with educational curricula [7]. The first engages the students to design and build their own games under the premise that they will learn desired concepts as they develop the game. But few K-12 teachers have the background and experience to teach programming and game development. The second approach uses games designed by professional designers, who can incorporate the sophisticated graphics and controls found in modern games to engage players. These games require significant development time, and commercial game studios are hesitant to embark on such projects due to the limited success of previous educational games (and there is no guarantee that the resulting games will be fun for students). The third approach integrates commercial-off-the-shelf (COTS) games that facilitate selected learning objectives into the classroom

• Y. Kim, T. Doll, and R. Migneco are with the Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA, 19104. E-mail: {ykim,tdoll,rmigneco}@drexel.edu

Manuscript received October 20, 2008.

[8]. These games are readily available, designed for entertainment, and have credibility with students outside of the school environment, but are limited in what they can teach and may be expensive to obtain for a large number of students.

Our web-based collaborative games offer a hybrid approach that addresses many of these issues. The activities incorporate an inherent design phase, which allows students to participate in the development of portions of the game played by other students. By working in partnership with schools and teachers through existing outreach programs, such as the NSF GK-12 program [9], our games are collaboratively designed and tested by teachers and graduate students. This facilitates rapid prototyping using the university's laboratory resources, pilot testing through partner schools, and eventual integration into existing K-12 curricula.

Our games are also designed to facilitate data collection for research in auditory perception and machine listening. As in other collaborative games, the design of our activities uses players' responses to verify each others' results. Scoring is based upon validation by other players, which provides a strong incentive for producing high-quality responses for listening tasks. We have used this collaborative framework to design games around the identification of sounds (a particular voice in a crowded room and musical instruments).

2.1 The cocktail party problem and speaker identity

The task of focusing on a single sound within a mixture of sounds is known as the *cocktail party problem* [2], referring to the common situation in which we are able to carry on an individual conversation in a crowded room amidst a number of other conversations and background noises. This is a task that we as humans perform quite naturally, but that has thus far eluded attempts at a general computational solution [10]. Furthermore, the auditory system has the ability to quickly identify sound sources with relatively little data, and again even in the presence of other interfering sounds. This allows us to recognize the sound of a person's voice calling out our name in a crowded room, or in music where we can follow a solo instrument as being distinct from complementary background instruments [11].

2.2 The perception of musical instrument timbre

Timbre refers to the character of a sound, more specifically those qualities that distinguish two sounds of equal pitch and loudness (i.e., what makes a violin sound different than a clarinet). Although the precise dimensions of timbre remain an open area of research, the relative strengths of the different frequency components produced by an instrument, as originally proposed by Helmholtz [12], have a significant impact on the perception of a sound. Other work has shown that the distribution of energy over time (the amplitude envelope) also has a large effect on the perceived timbre [13].

Relatively little research has been conducted on human performance in the identification of musical instruments after modifications along these dimensions [14], [15], although listening experiments have demonstrated that humans can identify an instrument's family with greater accuracy than the instrument itself [16].

3 DESCRIPTION OF DEVELOPED ACTIVITIES

Both of the activities described below are available online at <http://schubert.ece.drexel.edu/research/webapps>.

3.1 Hide & Speak

Hide & Speak is a web-based activity designed to illustrate how source locations and acoustic spaces affect identification of a speaker and the intelligibility of speech. This game consists of two components: room creation and listening room simulators. The room creation component introduces the concepts of the cocktail party problem and illustrates the effects of reverberation and multiple interfering sounds. The listening room component evaluates the ability of a listener to detect a known person's voice within a room mixture.

The game interface is designed to be simple and minimal so that players can easily experiment with sound source locations and reverberation intensities and immediately listen to the results. The graphical layout of the game represents actual acoustic space so that players can visually correlate speaker positions with the resulting sound. The activity requires no specific training in engineering or music, which broadens its appeal.

3.1.1 The Room Creation activity

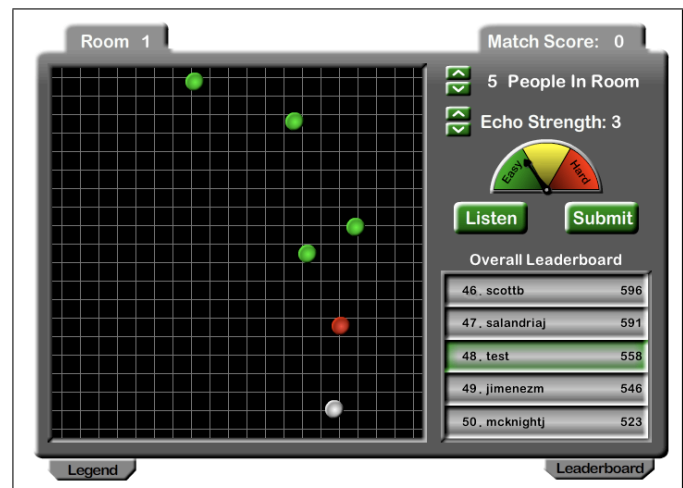


Fig. 1. The Room Creation interface.

In this component of the game, the player simulates a cocktail party situation by positioning multiple talkers, including the target person of interest, in a reverberant room. The goal is to create a situation where the sound from the person of interest is obscured, but still identifiable, and more points potentially will be awarded to the

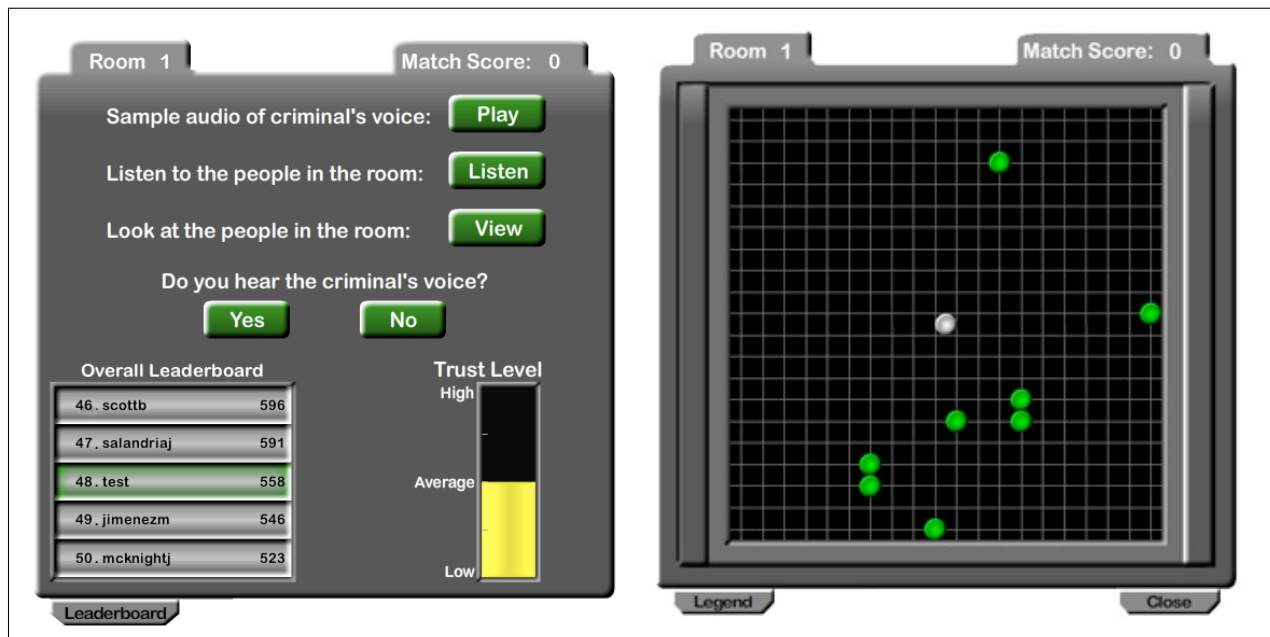


Fig. 2. The Listening Room interface (left) revised with the “surveillance” scenario described in Section 4; expanded to view room configuration (right).

player based on the “degree of difficulty” of the designed room. Initially, the game displays a room (20' x 20') containing two people: the listener and the person of interest, represented as white and red circles, respectively. The player has the option to add or remove people from the room, change the strength of the room reverberation, and alter the position of the people (Figure 1). Audio for each of the speakers in the room is randomly drawn from the TIMIT speech database [17].

A room’s potential score is based on the resulting Signal to Interferers plus Noise Ratio (SINR), treating the target voice as the signal and the others as interferers. These points are only added to the room creator’s score if a player in the Room Listening phase correctly determines whether or not the speaker is present.

3.1.2 The Listening Room activity

In the listening component of the game, the goal is simply to determine if a target person’s voice is present within the mixture of voices in the room. The player is provided with an isolated sample of the target’s voice as well as the mixed room audio, which is generated in exactly the same manner as in the room creation component. Each of the audio samples may be listened to as many times as desired, and the player also has the option to graphically view the physical configuration of the persons in the room (of course, without the target identified as depicted on the right in Figure 2). To complete each round, the player must decide whether or not the target person’s voice is present, after which the player is informed of the correct answer. Scoring is based on the difficulty of the room as calculated by the SINR and points are awarded only when the player chooses the correct answer. After the information is submitted,

the player continues on to the next round until all rounds are completed for the current match.

3.1.3 Educational objectives

Hide & Speak is designed to educate and inform students about several sound and acoustics concepts through guided exploration, including:

- Acoustic changes due to listener and source position
- Energy conservation of sound
- Human auditory perception

In the game, altering the listener and source locations directly affects the direction, loudness, and reverberation intensity of each source in the sound mixture received by the listener. As can be expected, increasing the distance from a speaker results in a decrease in loudness of the received sound. The activity also establishes a clear relationship between the distance from a source and the relative intensity of the reverberation (i.e., the closer the speaker, the less the amount of reverberation).

Acoustic energy conservation is incorporated as students audibly experience the declining sound energy of reflections from room surfaces. As students modify the echo strength level (reflection coefficient) of room surfaces, the amount of sound energy that continues propagating throughout the room after a reflection changes. This illustrates how energy within sound is transferred and why the duration of a sound is finite.

Hide & Speak also introduces students to characteristics of human hearing, particularly the ability to isolate one source within a mixture of sounds. This concept appears in both components of Hide & Speak, as students experiment with the limits of speech intelligibility in the Room Creation component and as they assess whether

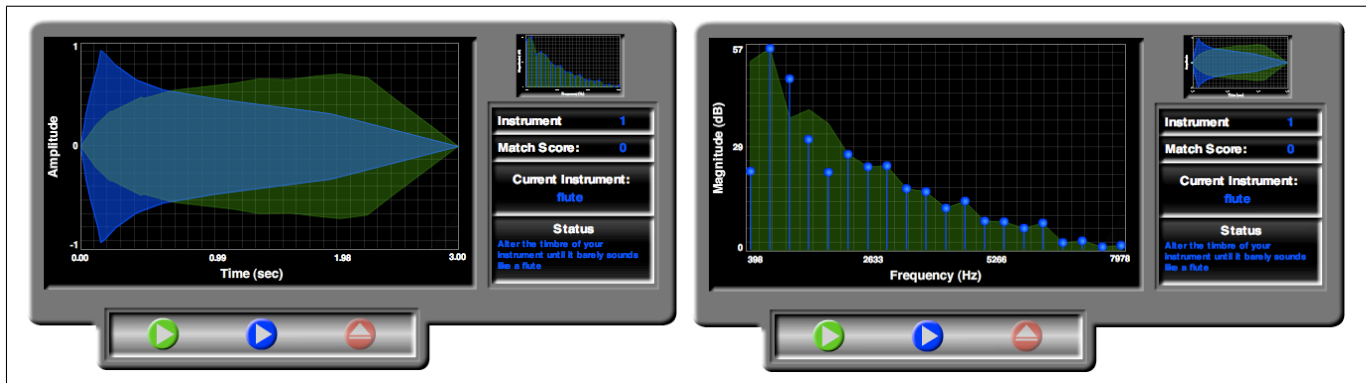


Fig. 3. The tone modification interface, with time-domain (left) and frequency-domain (right) views expanded.

a target voice is contained within the mixture of voices in the Listening Room component. The concept of the cocktail party effect is easily transferable to students' classroom environment as they focus on a teacher's voice while filtering out other interfering sounds.

3.2 Tone Bender

Tone Bender is an online activity to explore the acoustic components of musical instrument sounds. As in *Hide & Speak*, *Tone Bender* has two separate user interfaces (Create and Listen), which allow players to manipulate the timbre of musical sounds and identify sounds subjected to timbral modifications, respectively. A player submits a modified sound using the Create interface, and other players using the Listen interface attempt to determine the original instrument source. Points are awarded to both players (modifier and listener) if the listening player enters the sound source's identity correctly.

3.2.1 Create: The tone modification interface

The tone modification phase of the activity involves time and frequency analysis of a single instrument sound and provides a visual interface that allows a player to modify timbre. The sounds available for analysis are recordings of real musical instruments, each producing a single note with a duration of less than five seconds.

The visual representation of the sound's timbre is displayed to the player in two separate 'XY' plots in the user interface as shown in Figure 3. In the time-domain plot, the sound wave is shown with the extracted amplitude control points. The player is allowed to manipulate the shape of the amplitude envelope by directly drawing a new shape in the plot window. In the frequency-domain window, the estimated spectral envelope of the instrument is shown with the spectral control points, similar to a "graphic equalizer", which allows component frequencies to be attenuated or accentuated. The player is allowed to move control points vertically to adjust the harmonic weights of the sound without affecting pitch. After modifying the spectral envelope, the sound wave is re-synthesized using sinusoidal additive synthesis [18].

After each modification, the player is presented with a potential score based on the difference between the original sound and the altered sound, calculated using the Signal-to-Noise Ratio (SNR):

$$SNR = 10 \log_{10} \left| \sum_{n=0}^N \frac{s[n]^2}{(s[n] - \hat{s}[n])^2} \right| \quad (1)$$

This score is intended to reflect the potential difficulty in correctly identifying the original instrument from the modified sound where $s[n]$ and $\hat{s}[n]$ are the original and modified sounds, respectively. The resulting difficulty score ranges from 1-25, where 1 corresponds to a high SNR (little change) and 25 represents a low SNR (significant change). Since a player will be awarded points based on the difficulty score of their modifications only if a listener can correctly determine the original instrument, the goal is to modify the timbre of the sound as greatly as possible while still maintaining the instrument's identity. This encourages the player to be creative with their timbre adjustments yet still produce recognizable musical sounds.

3.2.2 Listen: The instrument identification interface

In the listening phase, a player is presented with an interface that allows them to assess sounds created (modified) by other players, where the objective is to correctly name the family and identity of the instrument from the sound presented. The modified sounds are randomly selected from the database and the sounds are re-synthesized using the amplitude and spectral envelope data. The player is allowed to listen to each sample as many times as needed prior to submitting a response.

The listening player is first asked to classify the sound as belonging to one of three instrument families: strings, wind, and brass, which populates another list consisting of individual instruments. The player will receive points only if they name either the instrument family and/or the specific instrument. If the player correctly identifies the instrument, they receive a score proportional to the difficulty rating. If only the instrument family is correct, half of the potential point value is awarded.

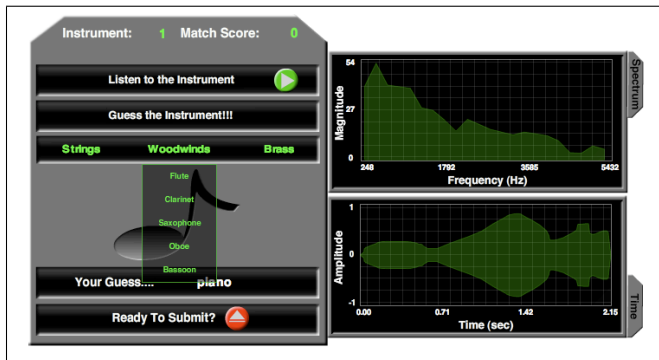


Fig. 4. The listening/instrument identification interface (expanded to show the time- and frequency-domain envelopes of the sound).

3.2.3 Educational objectives

The Tone Bender activity was developed to educate students regarding some of the acoustic properties affecting the timbre of sound, particularly:

- The distribution of energy (loudness) over time
- The distribution of energy over frequency

In one phase of the activity, players learn how the *time amplitude envelope* is related to the loudness over a sound's duration. By manipulating this envelope, students can explore how the shape relates to the identity of the source instrument (e.g. the loudness of a plucked instrument is initially strong and decays quickly, while a wind instrument features sustained loudness from applied breath energy).

Tone Bender also allows students to explore the spectral characteristics of musical sounds by altering the energy distribution across frequency. This capability is provided by an adjustable *spectral envelope* in the Tone Bender activity, which allows students to individually manipulate the relative strengths of a signal's overtones and listen to the change in the resulting instrument sound. This demonstrates how spectral energy distribution is directly associated with the character of musical sounds (e.g. the spectral contour distinguishes different instruments, and excessive high frequency energy leads to unnatural sounds).

In addition to the educational objectives related to timbre, Tone Bender also reinforces basic graph reading skills since the time and spectral interfaces are presented using 2-D plots. When a student modifies the amplitude and spectral envelope of an instrument, the resulting audio reflects the changes they impart on the signal, enabling the student to correlate the visual representations with the audio.

3.3 Technical implementation

Both Hide & Speak and Tone Bender employ a client-server architecture that allows distributed online game play. The user interface for each of the activities is implemented in Adobe Flash, which has rather limited

audio processing and computational capabilities. Therefore, it was necessary to use a Java helper applet for audio playback flexibility and its concurrent computation capabilities using threading. In Hide & Speak, the room impulse response for each source speaker location is determined based on the listener's position in the room using a Java implementation of the well-known room image model [19]. For Tone Bender, 10th-order linear prediction [20] is implemented in Java to establish a threshold for harmonic detection in the frequency domain. The resulting filter magnitudes from each frame are averaged to form a spectral envelope that approximates the time-averaged spectrum. This prediction curve is then uniformly lowered so that the first twenty most prominent harmonics can be extracted using a heuristic peak-picking algorithm that finds local maxima. The software architecture is described in greater detail in [21].

4 ACTIVITY EVALUATIONS

Evaluations of both activities were performed on a population of 56 eighth grade students attending a public magnet school specializing in music performance. The students were divided into groups of approximately 10 students for each of six sessions lasting 40 minutes per day. They were switched from the creative component of the game to the listening component midway through the session to have an opportunity to both create and objectively listen to the sounds. An initial analysis of the data collected from these sessions is available in [21].

During a second visit to the school, Hide & Speak was introduced differently than the first visit by presenting the game in a real-world context. For the Room Listening component, the students were told they were working for the CIA to find criminals on the most wanted list. For each round, the CIA planted a microphone in a random room and provided the student with a sample of a criminal's voice. For the Room Creation component, the students were told to pretend that the CIA is searching room by room for criminals on the most wanted list and it is their job to hide the target in a room. In each round, the student must hide, but still be able to communicate with the target.

4.1 Observations

As expected, the room creation component of Hide & Speak raised more questions and requests for clarification from the students than the listening component, due to the greater understanding required to complete the task. The lack of information provided to room creators regarding the performance of their rooms when listened to by other players was also frustrating and reduced one of the motivating competitive aspects of the game. Overall, the room creation component may need further simplification in order for middle school students to better understand the objectives of the activity.

The revised Hide & Speak game scenario increased student interest in the game as evidenced by the audible

reactions of the students and the overall number of rounds played. During this visit approximately 5-10 more minutes were given to play the game than in the first visit, but the students tripled the number of rounds played in the second visit. Also, during game play, the students were in constant competition by sharing the success or failure of the last room played with their friends sitting near them.

With the Tone Bender activity, the listening interface had the broadest appeal among the students. Again, this was due to the simple objective of the activity, which only required players to listen to sounds and respond. Additionally, the instant feedback and scoring added a distinct competitive aspect that encouraged the students to keep playing and comparing scores with each other. In the case of the tone modification component, some students appeared intimidated by or uninterested in the visual representations of instrument timbre, evinced by repetitive questions to demonstrators or by simply not participating. The students who were more engaged with the activity, however, attempted to modify many different sounds without assistance.

5 FUTURE WORK AND CONCLUSIONS

One enhancement to the activities being pursued is to allow players to record their own sounds for immediate use in the games: an instrumentalist could record their own instrument to be used in Tone Bender, and players could record their own voices for use in Hide & Speak. This feature would enable continuous expansion of the sound databases, keeping the game fresh for frequent users. A straightforward extension of Hide & Speak would be to extend the source sounds beyond voices (e.g., musical instruments). We are also investigating the utility of time limits for different phases of the game, in order to keep the activity moving and to increase competition. We continue to pursue a detailed analysis of acquired performance data for cases that deviate from anticipated "difficulty" in terms of S(I)NR, and we plan to investigate other acoustic factors affecting listening performance as well as other metrics that may be better correlated to perceptual task performance than S(I)NR.

Both activities are available through the web for testing, although data is not yet collected from this public interface. We believe these activities have great potential for very large-scale data collection, but we are still refining the implementation to control and scale appropriately for a large general user population. As noted above, placing one of the activities in a real-world, role-playing scenario piqued students interest, thus we are currently exploring different versions of the interface that enhance this aspect.

We have developed two prototype activities that allow students to explore aspects of different sound and acoustics concepts. In addition to their objectives as learning activities, our games facilitate the collection of data on the perception of audio and music. Our pilot study

with a class of middle school students demonstrated the potential benefits of this platform for education, but also revealed many new directions for further exploration and enhancement.

ACKNOWLEDGMENTS

This work is supported by NSF grants IIS-0644151 and DGE-0538476.

REFERENCES

- [1] L. von Ahn, "Games with a purpose," *Computer*, vol. 39, no. 6, pp. 92–94, 2006.
- [2] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, 1953.
- [3] F. S. Din and J. Caleo, "Playing computer games versus better learning," in *Paper presented at the 70th Annual Conference of the Eastern Educational Research Association*, Clearwater, FL, February 16-19 2000.
- [4] R. Rosas, M. Nussbaum, M. Cumsille, V. Marianov, M. Correa, P. Flores, V. Grau, F. Lagos, X. Lopez, V. Lopex, P. Rodriguez, and M. Salinas, "Beyond nintendo: Design and assessment of educational video games for first and second grade students," *Computers & Education*, vol. 40, pp. 71–94, 2003.
- [5] S. Papert, "Does easy do it? children, games, and learning," *Game Developer*, p. 88, June 1998.
- [6] D. W. Shaffer, *How Computer Games Help Children Learn*. Palgrave MacMillan, 2006.
- [7] R. Van Eck, "Digital game-based learning: It's not just the digital natives who are restless," *EDUCAUSE Review*, vol. 41, no. 2, pp. 16–30, 2006.
- [8] K. Squire, "Video games in education," *International Journal of Intelligent Simulations and Gaming*, vol. 2, no. 1, 2003.
- [9] National Science Foundation (NSF), "Graduate Teaching Fellows in K-12 Education (GK-12)." [Online]. Available: http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5472&org=DGE
- [10] S. Haykin and Z. Chen, "The Cocktail Party Problem," *Neural Computation*, vol. 17, no. 9, pp. 1875–1902, Sept. 2005.
- [11] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, 1990.
- [12] H. Helmholtz and A. J. Ellis (trans), *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik (On the Sensations of Tone as a Physiological Basis for the Theory of Music)*, 2nd (English) ed. London: Longmans, Green, 1885.
- [13] M. Clark, D. Luce, R. Abrams, H. Schlossberg, and J. Rome, "Preliminary experiments on the aural significance of parts of tones of orchestral instruments and on choral tones." *Journal of the Audio Engineering Society*, vol. 11, no. 1, pp. 45–54, 1963.
- [14] E. Saldanha and J. Corso, "Timbre cues and the identification of musical instruments," in *Journal of Acoustic Society of America*, 1964, pp. 2021–2026.
- [15] P. Iverson and C. L. Krumhansl, "Isolating the dynamic attributes of musical timbre," in *Journal of Acoustic Society of America*, vol. 94, no. 5, 1993, pp. 2595–2603.
- [16] K. Martin, "Sound-source recognition: A theory and computational model," Ph.D. dissertation, Massachusetts Institute of Technology, 1999.
- [17] V. Zue, S. Seneff, and J. Glass, "Speech database development at MIT: TIMIT and beyond," *Speech Communication*, vol. 9, no. 4, pp. 351–356, August 1990.
- [18] R. J. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, pp. 744–754, 1986.
- [19] J. Allen and D. Berkley, "Image method for efficiently simulating small room acoustics," in *Journal of Acoustic Society of America*, April 1979, pp. 912–915.
- [20] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, pp. 1973–1986, 1975.
- [21] T. M. Doll, R. Migneco, and Y. E. Kim, "Online activities for music information and acoustics education and psychoacoustic data collection," in *Proc. International Conference on Music Information Retrieval*. Philadelphia, PA: ISMIR, September 2008.