The Effects of Reverberation on the Emotional Characteristics of Musical Instruments

RONALD MO, BIN WU, AND ANDREW HORNER, *AES Member*

(ronmo@cse.ust.hk)   (bwuaa@cse.ust.hk)   (horner@cse.ust.hk)

Department of Computer Science and Engineering Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong

Though previous research has shown the effects of reverberation on clarity, spaciousness, and other perceptual aspects of music, it is still largely unknown to what extent reverberation influences the emotional characteristics of musical instrument sounds. This paper investigates the effect of simple parametric reverberation on music emotion, in particular, the effect of reverberation length and amount. We conducted a listening test to compare the effect of reverberation on the emotional characteristics of eight instrument sounds representing the wind and bowed string families. We compared these sounds over eight emotional categories. We found that reverberation length and amount had a strongly significant effect on the emotional characteristics Romantic and Mysterious and a medium effect on Sad, Scary, and Heroic for the samples we tested. Interestingly, for Comic, reverberation length and amount had the opposite effect; that is, anechoic tones were judged most Comic. Reverb had a mild effect on Happy and relatively little effect on Shy. These results give audio engineers and musicians an interesting perspective on simple parametric artificial reverberation.

0 INTRODUCTION

Previous research has shown that musical instrument sounds have strong and distinctive emotional characteristics [1–5]. For example, that the trumpet is happier in character than the horn, even in isolated sounds apart from musical context.

In light of this, one might wonder what effect reverberation has on the character of music emotion. This leads to a host of follow-up questions: Do all emotional characteristics become stronger with more reverberation? Or, are some emotional characteristics affected more and others less (e.g., positive emotional characteristics more, negative less)? In particular, what are the effects of reverberation time and amount? What are the effects of hall size and listener position? Which instruments sound emotionally stronger to listeners in the front or back of small and large halls? Are dry sounds without reverberation emotionally dry as well, or, do they have distinctive emotional characteristics?

We cannot address all of the above questions definitively in this paper with only a simple parametric reverberator and a few parameter settings, but we can make a good start. This work will give audio engineers and musicians an interesting perspective on simple parametric artificial reverberation. More studies with different reverberation models and parameters should be carried out to get more definitive answers. Understanding how listeners perceive emotional characteristics in reverberation can help us engineer potentially even more expressive recordings and opens new possibilities for interactive music systems and applications.

1 BACKGROUND

1.1 Music Emotion and Timbre

Previous work has investigated emotion recognition in music, especially addressing melody [6], harmony [7, 8], rhythm [9, 10], lyrics [11], and localization cues [12]. Similarly, researchers have found timbre to be useful in a number of applications such as automatic music genre classification [13], automatic song segmentation [14], and song similarity computation [14].

Researchers have considered music emotion and timbre together in a number of studies. Hevner's early work [15] pioneered the use of adjective scales in music and emotion research. She divided 66 adjectives into 8 groups where adjectives in the same group were related and compatible. The results of their listening test were affective values for the major and minor scales, different types of rhythms, dissonant and consonant harmonies, and rising and falling melodic lines.

Scherer and Oshinsky [16] used a 3D dimensional model to study the relationship between emotional attributes and synthetic sounds by manipulating different acoustic parameters such as amplitude, pitch, envelope, and filter cutoff.

Subjects rated sounds on a 10-point scale for the three dimensions Pleasantness, Activity, and Potency. They also allowed users to label sounds with emotional labels such as Anger, Fear, Boredom, Surprise, Happiness, Sadness, and Disgust. They found that timbre was a salient factor in the rating of synthetic sounds.

Peretz et al. [17] asked listeners to rate musical excerpts on a 10-point scale along the dimension Happy-Sad. They found that listeners could discriminate between Happy and Sad musical excerpts lasting only 0.25 s, sounds so short that factors other than timbre could not have come into play.

Ellermeier et al. [18] investigated whether auditory Unpleasantness was judged consistently across a wide range of acoustic stimuli. They used paired comparisons of all possible combinations of 10 environmental sounds. They used a BTL model to statistically rank the sounds. They found that a linear combination of the psychoacoustic parameters Roughness and Sharpness accounted for more than 94% of the variance in perceived Unpleasantness.

Bigand et al. [19] conducted experiments to study emotion similarities between one-second musical excerpts. They grouped excerpts that conveyed a similar emotional meaning. They then transformed the groupings into an emotional dissimilarity matrix, which was analyzed with multidimensional scaling. A 3D space provided a good fit with Arousal and Valence as the primary dimensions. The average duration of the excerpts was 30 s. They confirmed the consistency of this 3D space using excerpts of only 1 s duration (a result similar to that of Peretz [17]).

Zentner et al. [20] conducted a series of experiments to compile a list of musically-relevant emotional terms (e.g., Enchanted and Amused) and to study the frequency of both felt and perceived emotion across groups of listeners with different musical preferences. They found that responses varied greatly according to musical genre and depending on whether it was a felt or perceived response. They also examined the structure of music-induced emotions using a factor analysis of the emotion ratings.

Hailstone et al. [21] studied the relationship between sound identity and music emotion. They asked participants to select which one of four emotional categories (Happiness, Sadness, Fear, or Anger) was represented in 40 novel melodies that were recorded in different versions using electronic synthesizer, piano, violin, and trumpet, controlling for melody, tempo, and loudness between instruments. They found a significant interaction between instrument and emotion. In a second experiment, they asked participants to identify the emotions represented by the same melodies with four novel synthetic timbres designed to include timbral cues to particular emotions. Their results showed that timbre independently affected perceived emotion in music after controlling for other acoustic, cognitive, and performance factors.

Yang et al. [22] developed a music emotion recognition system to predict the Valence and Arousal values for music excerpts using the representation proposed by Russell [23]. They formulated music emotion recognition as a regression problem to predict the Valence and Arousal values of each music sample directly. Each music sample was a point in the Valence-Arousal plane, so that listeners could specify a desired point and efficiently retrieve matching music.

Krumhansl [24] found that 0.4 s musical excerpts were long enough to allow listeners to identify both the artist and title of popular songs from 1960 to 2010 more than 25% of the time. Even when not correctly identified, listeners were able to gather information about emotional content, style, and the decade of release.

Similarly, Filipic et al. [25] found that 0.5 s musical excerpts were long enough to allow feelings of familiarity to be triggered. They also found that 0.25 s excerpts were long enough to allow distinctions between emotionally-moving and neutral responses.

Eerola and Vuoskoski [26] compared categorical and dimensional models for perceived emotion using 110 film music excerpts. Subjects rated the excerpts based on the emotional categories Happy, Sad, Tender, Fearful, and Angry using a nine-point scale. Separately, they also rated the music excerpts based on another nine-point scale for the dimensions Valence, Energy, and Tension. They observed a high correspondence between the categorial and dimensional results. That is, the results for either model could be predicted from the other with a high degree of accuracy. They also found that the three dimensions Valence, Energy, and Tension could be reduced to the two dimensions Valence and Arousal without significantly reducing the goodness of fit.

Vuoskoski and Eerola [27] further compared the same categorical and dimensional models with Zentner's [20] model (described above) for perceived emotion in 16 film music excerpts. Subjects were most consistent in the dimensional model. Principal component analysis revealed that almost 90% of the variance in the mean ratings for perceived emotion in all three models was accounted for by two principal components that could be labeled as Valence and Arousal.

Eerola et al. [1] studied the correlation of perceived emotion with temporal and spectral sound features. They asked listeners to rate the perceived affect qualities of 1 s instrument tones using five dimensions: Valence, Energy, Tension, Preference, and Intensity. They correlated the ratings with acoustic features such as attack time and brightness. They found strong correlations between these acoustic features and the emotion dimensions Valence and Arousal.

Asutay et al. [28] studied Valence and Arousal along with loudness and familiarity in subjects' responses to environmental and processed sounds. Subjects were asked to rate each sound on nine-point scales for Valence and Arousal. Subjects were also asked to rate how Annoying the sound was. They found that the processed sounds were emotionally neutral. They also found that even though most of the processed sounds decreased in measured loudness compared to the original sounds, neither perceived loudness nor auditory-induced emotion changed accordingly. This result suggested the importance of factors other than physical sound characteristics in sound design.

Liebetrau et al. [29] compared different methods for measuring music emotion including paired comparisons and free-choice profiling (FCP). They tested paired

comparisons of Valence and Arousal on musical phrases using a relatively small number of listening subjects (10). They found subjects were able to efficiently assess each paired comparison, especially compared to free-choice profiling where subjects had to first define their own attributes. However, they suggested FCP could obtain more interpretable results in situations when only a relatively small number of subjects was available.

Baume [30] evaluated the usefulness of acoustic and musical features for classifying about 2400 music tracks into four mood categories: Terror, Peace, Joy, and Excitement. He evaluated how well each feature performed as part of an SVM for classifying music using these four mood categories. He found that spectral and harmonic features performed better than rhythm, temporal, and energy features.

Wu et al. [2, 4, 31, 32] and Chau et al. [3, 5] compared the emotional characteristics of sustaining and non-sustaining instruments. Like Ellermeier [18], they used a BTL model to rank paired comparisons of eight sounds. Wu compared sounds from eight wind and bowed string instruments, while Chau compared eight non-sustaining sounds such as the piano, plucked violin, and marimba. Eight emotional categories for expressed emotion were tested including Happy, Sad, Heroic, Scary, Comic, Shy, Joyful, and Depressed. The results showed distinctive emotional characteristics for each instrument. Wu found the timbral features spectral centroid and even/odd harmonic ratio were significantly correlated with emotional characteristics for sustaining instruments. Chau found decay slope and density of significant harmonics were significantly correlated for non-sustaining instruments.

Table 1 summarizes the above literature, showing the model, emotional categories/dimensions, whether perceived, induced, felt, or expressed emotion, and the stimuli type and evaluation.

1.2 Reverberation

1.2.1 Artificial Reverberation Models

Various models have been suggested for reverberation using different methods to simulate the build-up and decay of reflections in a hall as the sound is absorbed by surfaces of objects in the space. They include simple reverberation algorithms using several feedback delays to create a decaying series of echoes, such as Schroeder reverb [33]. More sophisticated reverb algorithms simulate the time and frequency response of a hall, using its dimensions, absorption, and other properties [34–38]. There are also models that convolve the impulse response of the space being modeled with the audio signal to be reverberated [39, 40].

These models use different parameters, but in all of them it is possible to characterize the reverberation by characteristics such as reverberation time and early decay time. Reverberation time (RT_{60}) is one of the most important characteristics of reverberation, and measures the time reverberation takes to decay by 60 dB SPL from an initial impulse [41]. Jordan [42] suggested an alternative measurement called Early Decay Time (EDT), which is defined as either: (1) six times the time interval that it takes for

an impulse response to decay from 0 dB to –10 dB, or (2) by the straight line that best fits an impulse response as it decays from 0 dB to –10 dB.

1.2.2 Subjective Evaluation of Reverberation

Some previous research has considered the subjective evaluation of reverberation. In a preliminary study, Kaczmarek et al. [43] subjectively evaluated reverberation amount using individual anechoic instrument tones. They ran two experiments. In the first, listeners rated tones with 0%, 30%, and 60% reverb based on sound characteristics such as Bright, Dark, Natural, Rumbling, and Sharp. However, the reported results were brief and inconclusive. In their second experiment, they used A-B-A comparisons of various levels of reverb in terms of naturalness, which decreased with more reverberation.

1.2.3 Reverberation and Music Emotion

Though various research has shown the effects of reverberation and room geometry on clarity, spaciousness, and other perceptual aspects of speech and music (e.g., Cremer and Müller [44]), only a few studies have considered the emotional effect of reverberation. Västfjäll et al. [45] studied how reverberation time influences emotion in musical excerpts. They used a dimensional model to measure the effects on Valence and Arousal. They found that long reverberation times were perceived as most unpleasant. More recently, Tajadura-Jiménez et al. [46] studied the correlation between emotion and room size for four natural and four artificial sounds. They also used a dimensional model with measurements for Valence, Arousal, and perceived Safeness. Their results suggested that smaller rooms were considered more pleasant, calmer, and safer than big rooms, although these differences seemed to disappear for threatening sound sources. Even with these studies, it is still largely unknown to what extent reverberation influences the emotional characteristics of musical instrument sounds.

2 METHODOLOGY

2.1 Overview

For this investigation, we used a relatively simple parametric reverberation model to measure the emotional effects of two of the most important reverb parameters: reverberation length and amount. Future experiments with other reverberation parameters and models will further deepen our understanding, but reverberation length and amount provide an obvious starting place for understanding reverberation's effect on music emotion.

Through a listening test with paired comparisons and statistical analysis we will investigate the effects of simple parametric reverberation on the emotional characteristics of musical instruments. In particular, we will address the following questions:

- Do all emotional characteristics become stronger with more reverberation, or are some affected more

Table 1. Summary of the literature on music emotion and timbre.

Year	Author(s)	Emotion Model	Emotional Categories/ Dimensions	Emotion Type	Stimuli Type	Stimuli Evaluation
1936	Hevner	Categorical	8 groups of adjectives	Expressed	Musical Excerpts	Rating
1977	Scherer and Oshinsky	Categorical and Dimensional	Dimension: Pleasantness, Activity, and Potency Category: Anger, Fear, Boredom, Surprise, Happiness, Sadness, and Disgust	Perceived	Instrument Tones	Rating
1998	Peretz, Gagnon, and Bouchard	Dimensional	Happy/Sad	Perceived	Musical Excerpts	Rating
2004	Ellermeier, Mader, and Daniel	Dimensional	Unpleasantness	Felt	Environmental Sounds	Paired Comparison
2005	Bigand, Vieillard, Madurell, Marozeau, and Dacquet.	Dimensional	Valence, Arousal, and a third dimension expressing the influence of body gestures	Induced	Musical Excerpts	Paired Comparison
2008	Zentner, Grandjean, and Scherer	—	—	Felt and Perceived	—	Rating
2009	Hailstone, Omar, Henley, Forst, Kenward, and Warren	Categorical	Happiness, Sadness, Fear, and Anger	Perceived	Novel Melodies	Rating
2009	Yang, Lin, Su, and Chen	Dimensional	Valence and Arousal	Induced	Musical Excerpts	Rating
2010	Krumhansl	Categorical	Happiness, Sadness, Anger, Fear, and Tenderness	Perceived	Musical Excerpts	Rating
2010	Filipic, Tillmann, and Bigand	Dimensional	Degree of Emotionally Touching	Felt	Musical Excerpts	Rating
2011	Eerola and Vuoskoski	Categorical and Dimensional	Category: Happy, Sad, Tender, Fearful, Angry Dimension: Valence, Energy, Tension	Perceived	Film Music Excerpts	Rating
2011	Vuoskoski and Eerola	Categorical, Dimensional, and Geneva Emotional Music Scale	Geneva Emotional Music Scale Category: Sadness, Happiness, Tenderness, Fear, and Anger Dimension: Valence, Arousal, and Tension	Induced	Film Music Excerpts	Rating
2012	Eerola, Ferrer, and Alluri	Dimensional	Valence, Energy, Tension. Preference, and Intensity	Perceived	Instrument Tones	Rating
2012	Asutay, Västfjäll, Tajadura-Jiménez, Genell, Bergman, and Kleiner	Dimensional	Valence, Arousal, Loudness, Familiarity, and Annoyingness	Felt	Environmental Sounds	Rating
2013	Liebetrau, Nowak, Sporer, Krause, Rekitt, and Schneider	Dimensional	Valence and Arousal	Induced	Music Excerpts	Paired Comparison
2015	Baume	Categorical	Terror, Joy, Peace, and Excitement	Induced	Music Tracks	–
2014–15	Wu, Horner, and Lee	Categorical	Happy, Sad, Heroic, Scary, Comic, Shy, Joyful, and Depressed	Perceived	Instrument Tones	Paired Comparison
2014–15	Chau, Wu, and Horner	Categorical	Happy, Sad, Heroic, Scary, Comic, Shy, Joyful, and Depressed	Perceived	Instrument Tones	Paired Comparison

and others less (e.g., positive characteristics more, negative less)?

- What are the effects of reverberation time (i.e., what are the effects of hall size)?
- What are the effects of reverberation amount (i.e., what are the effects of listener position relative to the front or back of the hall)?
- Which instruments sound emotionally stronger to listeners in the front or back of small and large halls?
- Are dry sounds without reverberation emotionally neutral, or, do they have distinctive emotional characteristics (e.g., strong negative emotional characteristics)?

To begin to address these questions, we conducted a listening test to compare the effect of reverberation on the emotional characteristics of individual instrument sounds. We tested eight sustained musical instruments representing the wind and bowed string families. We compared anechoic recordings of these sounds and sounds where artificial reverberation had been added in varying amounts.

We compared these sounds over eight emotional categories that are commonly expressed by composers in tempo and expression marks (Happy, Sad, Heroic, Scary, Comic, Shy, Romantic, and Mysterious). The following sections describe the details of the listening test and the statistical analysis used to address the questions raised above.

2.2 Listening Test

Our test had listeners compare five types of reverberation over eight emotional categories for each instrument. The basic stimuli consisted of eight sustained wind and bowed string instrument sounds without reverberation: bassoon (bs), clarinet (cl), flute (fl), horn (hn), oboe (ob), saxophone (sx), trumpet (tp), and violin (vn). They were obtained from the *University of Iowa Musical Instrument Samples* [47]. These sounds were all recorded in an anechoic chamber and were thus free from reverberation.

The sustained instruments are nearly harmonic and the chosen sounds had fundamental frequencies close to Eb4 (311.1 Hz). They were analyzed using a phase-vocoder algorithm where bin frequencies were aligned with the signal's harmonics [48]. Attacks, sustains, and decays were equalized by interpolation to 0.05 s, 0.8 s, and 0.15 s respectively, for a total duration of 1.0 s. The sounds were resynthesized by additive sinewave synthesis at exactly 311.1 Hz. Since loudness is a potential factor in emotional characteristics, the sounds were equalized by loudness by manual adjustment.

In addition to the resynthesized anechoic sounds, we compared sounds with reverberation lengths of 1 s and 2 s, which according to Hidaka and Beranek [49] and Beranek [50] typically correspond to small and large concert halls. We used the reverberation generator provided by *Cool Edit* [51]. Its "Concert Hall Light" preset is a reasonably natural sounding reverberation. This preset uses 80% for the amount of reverberation corresponding to the back of the hall, and we approximated the front of the hall with

20%. Thus, in addition to the dry sounds, there were four reverberated sounds for each instrument:

Hall Type and Position	Reverb Length	Reverb Amount	RT_{60}
Small Hall Front	1 s	20%	0.95
Small Hall Back	1 s	80%	1.28
Large Hall Front	2 s	20%	1.78
Large Hall Back	2 s	80%	2.37

Figs. 1 to 4 show the impulse responses and RT_{60} values for the different types of reverberation we used. The Early Decay Time (EDTs) were near-zero for all four reverberation types.

We hired 34 subjects without hearing problems to take the listening test. All subjects were fluent in English. They were all undergraduate students at the Hong Kong University of Science and Technology where all courses are taught in English.

The subjects compared the stimuli in paired comparisons for eight emotional categories: Happy, Sad, Heroic, Scary, Comic, Shy, Romantic, and Mysterious. Some choices of emotional characteristics are fairly universal and occur in many previous studies as shown in Table 1 (e.g., Happy, Sad, Scary/Fear/Calm, Tender/Calm/Romantic) roughly corresponding to the four quadrants of the Valence-Arousal plane, but there are lots of variations beyond that [52]. We carefully picked the emotional categories based on terms we felt composers were likely to write as expression marks for performers (e.g., mysteriously, shyly, etc.) and at the same time would be readily understood by lay people. Simple English emotional categories were chosen as they would be familiar and self-apparent to subjects rather than Italian music expression marks traditionally used by classical composers to specify the character of the music. The emotional categories we chose and the related Italian expression marks [53–56] are listed in Table 2.

We tried to include a well-balanced group of emotional categories, and these eight categories roughly correspond to the eight adjective groups of Hevner [15]. Other researchers have also used some of these (or related) emotional categories [16, 20, 21]. Our previous research showed the

Table 2. Our emotional categories and related music expression marks commonly used by classical composers.

| Emotional Category	Commonly-used Italian musical expression marks
Happy	*allegro, gustoso, gioioso, giocoso, contento, gaudioso*
Sad	*dolore, lacrimoso, lagrimoso, mesto, triste, mesto, freddo*
Heroic	*eroico, grandioso, epico*
Scary	*sinistro, terribile, allarmante, feroce, furioso*
Comic	*capriccio, ridicolosamente, spiritoso, comico, buffo*
Shy	*timido, riservato, timoroso*
Romantic	*romantico, appasionato, afectuoso*
Mysterious	*misterioso*
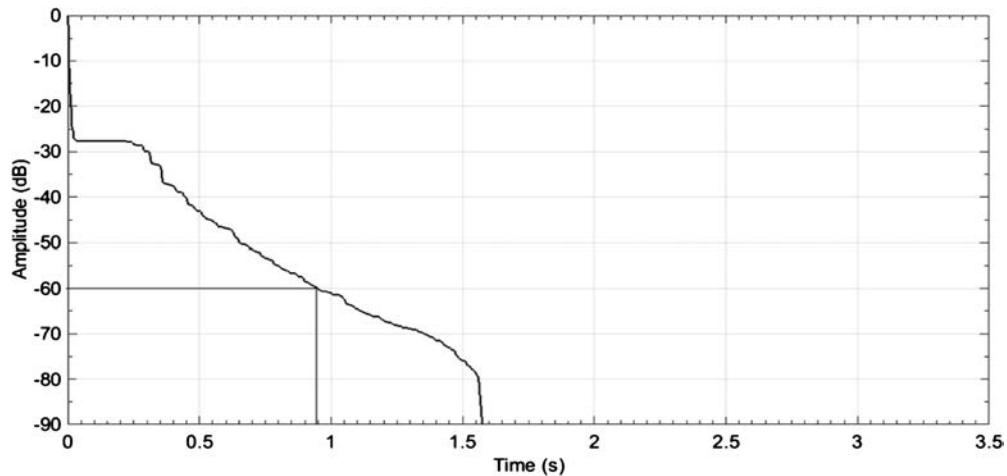
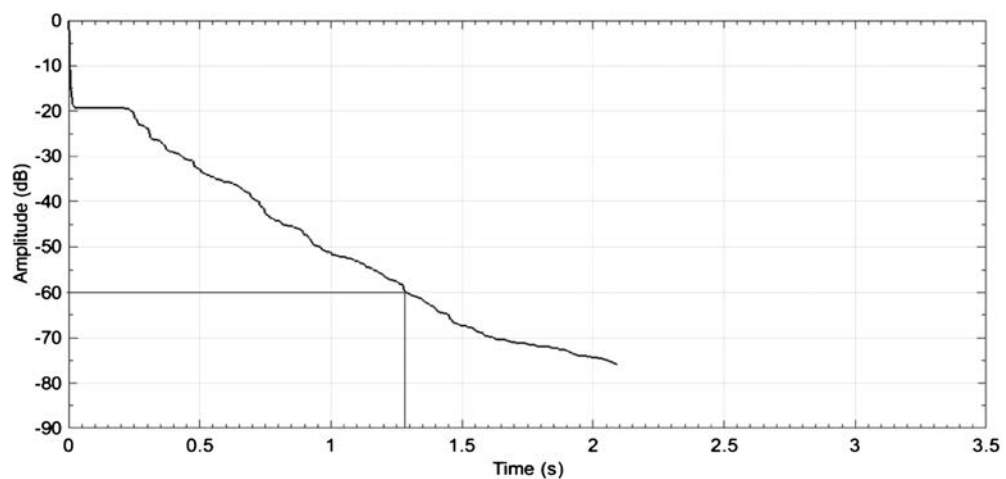Fig. 1.   Impulse response and RT_{60} for Small Hall Front.

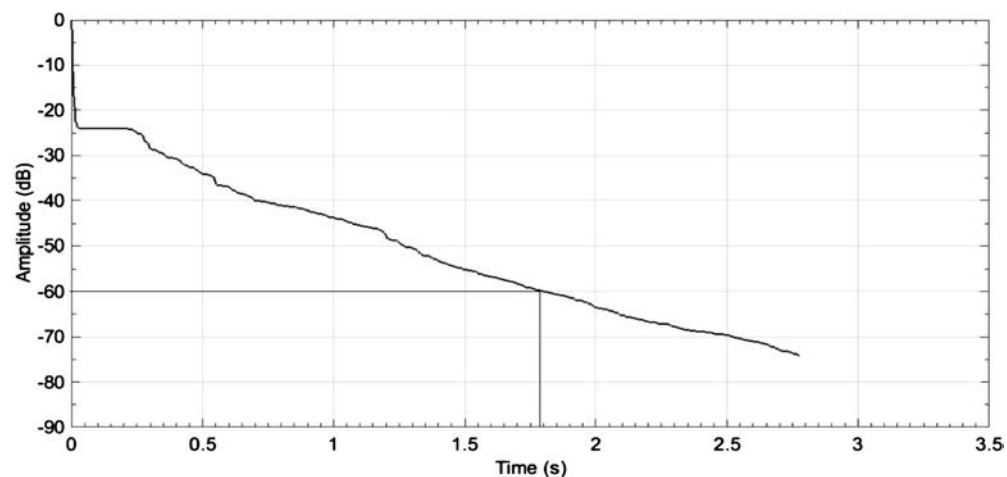Fig. 2.   Impulse response and RT_{60} for Small Hall Back.

Fig. 3.   Impulse response and RT_{60} for Large Hall Front.

statistical significance of the correlation of these terms for single instrument tones [2–5, 31, 32].

In picking these categories, we particularly had dramatic musical genres such as opera and musicals in mind, where there are typically heroes, villains, and comic-relief characters with music specifically representing each. The emo-

tional characteristics in these genres are generally more obvious and less abstract than in pure orchestral music. Their ratings according to the *Affective Norms for English Words* [57] are shown in Fig. 5 using the Valence-Arousal model. Happy, Comic, Heroic, and Romantic form a cluster, but they represent distinctly different emotional categories.
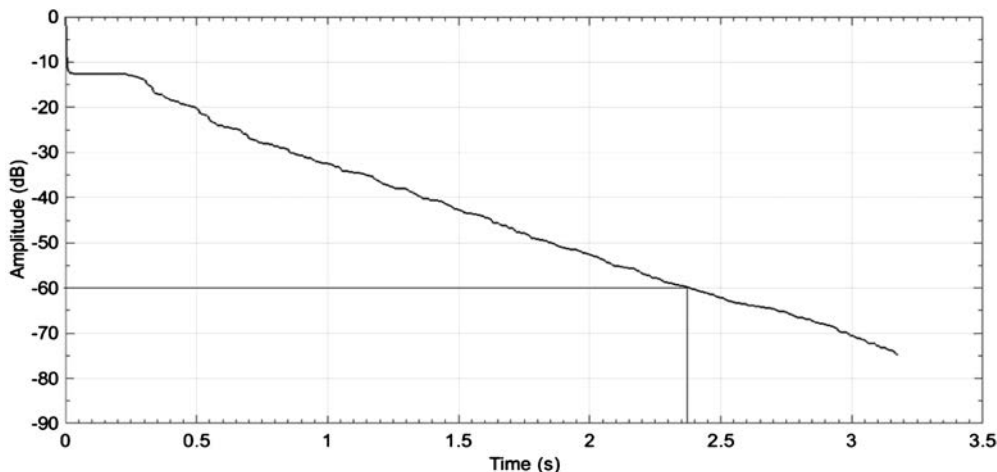
Fig. 4.   Impulse response and RT_{60} for Large Hall Back.
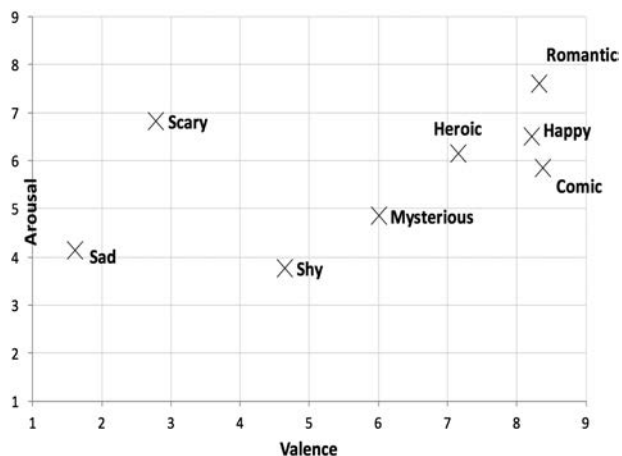

Fig. 5. Distribution of the emotional characteristics in the dimensions Valence and Arousal. The Valence and Arousal values are given in the nine-point rating in ANEW [57]. Valence shows the positiveness of an emotional category; Arousal shows the energy level of an emotional category.

In the listening test, every subject heard paired comparisons of all five types of reverberation for each instrument and emotional category. During each trial, subjects heard a pair of sounds from the same instrument with different types of reverberation and were prompted to choose which more strongly aroused a given emotional category. There was not a training period for this listening test because each trial was a single paired comparison requiring minimal memory from the subjects. In other words, subjects did not need to remember all of the tones, just the two in each comparison. Fig. 6 shows a screenshot of the paired comparison listening test interface. One big advantage of using paired comparisons of emotional categories is that it allows faster decision-making by the subjects. Paired comparison is also a simple decision and is easier than absolute rating.

Each permutation of two different reverberation types were presented for each of the eight instruments and eight emotional categories, and the listening test totaled $P_2^5 \times 8 \times 8 = 800$ trials. For each instrument, the overall trial presentation order was randomized (i.e., all the bassoon comparisons were first in a random order, then all the clarinet comparisons second, etc.).

Before the first trial, subjects read online definitions of the emotional categories from the *Cambridge Academic Content Dictionary* [58]. The dictionary definitions we used in our experiment are shown in Table 3. Subjects were not
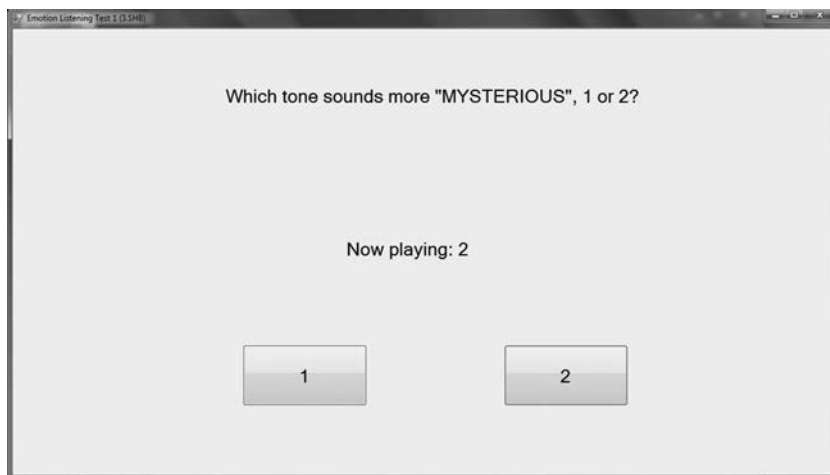

Fig. 6.   Paired comparison listening test interface.

Table 3. The dictionary definitions of the emotional categories
used in our experiment.

Emotional Category	Definition
Happy	Glad, pleased
Sad	Affected with or expressive of grief or unhappiness
Heroic	Exhibiting or marked by courage and daring
Scary	Causing fright
Comic	Causing laughter or amusement
Shy	Disposed to avoid a person or thing
Romantic	Relating to love or loving relationship
Mysterious	Strange or unknown

"golden ear" subjects (e.g., recording engineers, professional musicians, or music conservatory students) but average attentive listeners. The listening test took about 2 hours, with breaks every 30 minutes.

The subjects were seated in a "quiet room" with 39 dB SPL background noise level (mostly due to computers and air conditioning). The noise level was reduced further with headphones. Sound signals were converted to analog by a Sound Blaster X-Fi Xtreme Audio sound card and then presented through Sony MDR-7506 headphones at a level of approximately 78 dB SPL, as measured with a sound-level meter. The Sound Blaster DAC utilizes 24 bits with a maximum sampling rate of 96 kHz and a 108 dB S/N ratio. We felt that basic-level professional headphones were adequate in representing the simple reverberated sounds for this test as the lengths and amounts of reverberation were quite different and readily distinguishable. A big advantage of the Sony MDR-7506 headphones is their relative comfort in a relatively long listening test such as this one, especially for subjects not used to tight-fitting studio headphones.

3 RANKING RESULTS FOR THE EMOTIONAL CHARACTERISTICS WITH DIFFERENT TYPES OF REVERBERATION

The subjects' responses were first checked for inconsistencies. Consistency was defined based on the two comparisons of a pair of tones A and B for a particular instrument and emotional category as follows:

$$consistency_{A,B} = \frac{max(v_A, v_B)}{2} \quad (1)$$

where v_A and v_B are the number of votes a subject gave to each of the two tones. A consistency of 1 represents perfect consistency, whereas 0.5 represents approximately random guessing. The mean average consistency of all subjects was 0.78. Subjects were fairly consistent in their responses. That is, subjects voted for the same tone in both comparisons (AB and BA) about 80% of the time. We measured the level of agreement among the subjects with an overall Fleiss' Kappa statistic. It was calculated at 0.026, indicating a statistically significant agreement among subjects [59].

We ranked the tones by the number of positive votes they received for each instrument and emotional category and derived scale values using the Bradley-Terry-Luce (BTL) statistical model [60, 61].

For each graph, the BTL scale values for the five tones sum up to 1. The BTL value for each tone is the probability that listeners will choose that reverberation type when considering a certain instrument and emotional category. For example, if all five reverb types (Anechoic, Small Hall Front, Small Hall Back, Large Hall Front, Large Hall Back) were judged equally happy, the BTL scale values would be $1/5 = 0.2$.

Figs. 7 to 14 show BTL scale values and the corresponding 95% confidence intervals for each emotional category and instrument. Based on Figs. 7 – 14, Table 4 shows the number of times each reverb type was significantly greater than the other four reverb types (i.e., where the
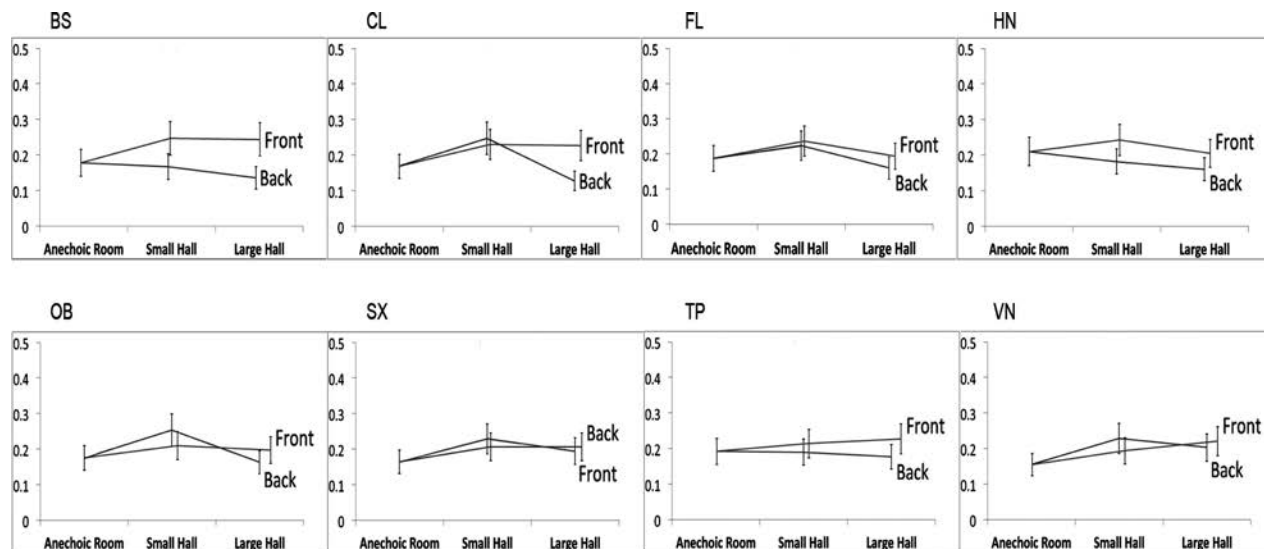

Fig. 7.   BTL scale values and the corresponding 95% confidence intervals for the emotional category Happy.
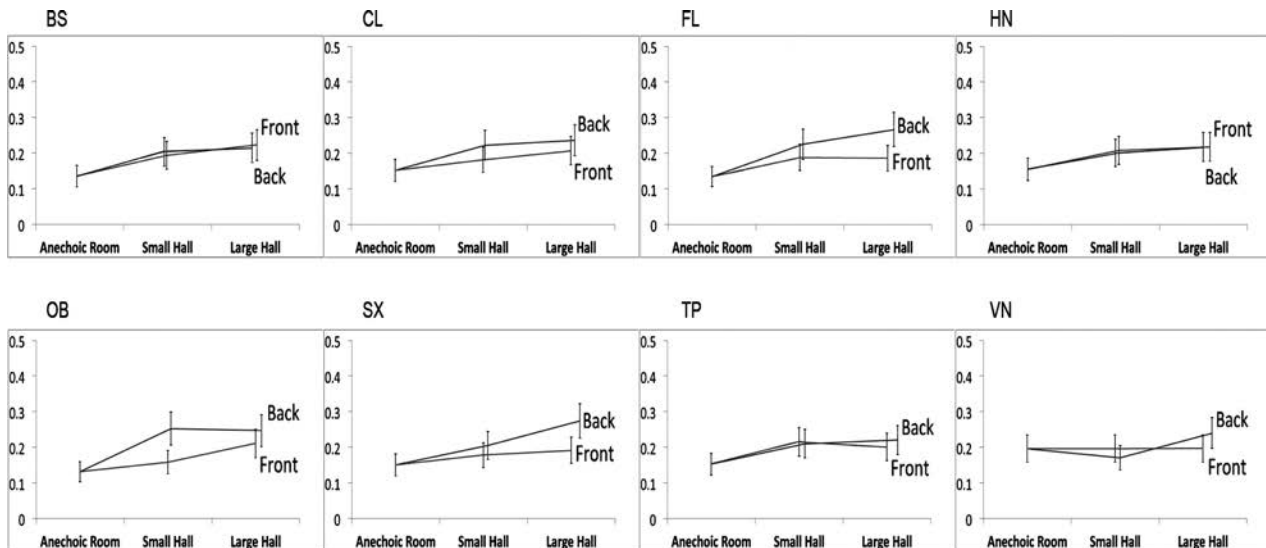
Fig. 8. BTL scale values and the corresponding 95% confidence intervals for Heroic.
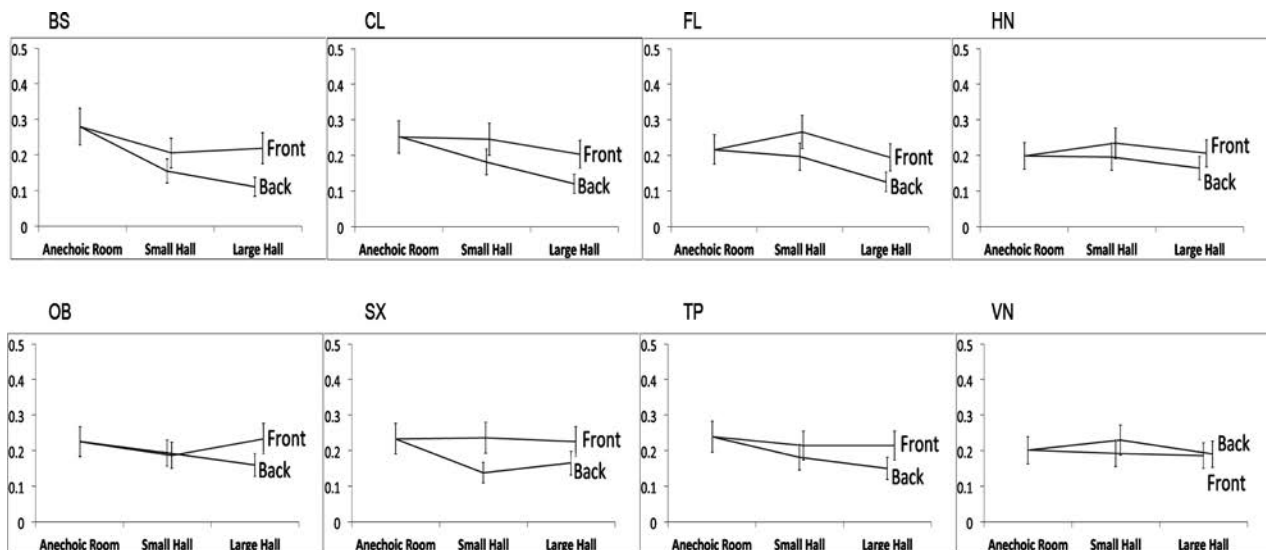
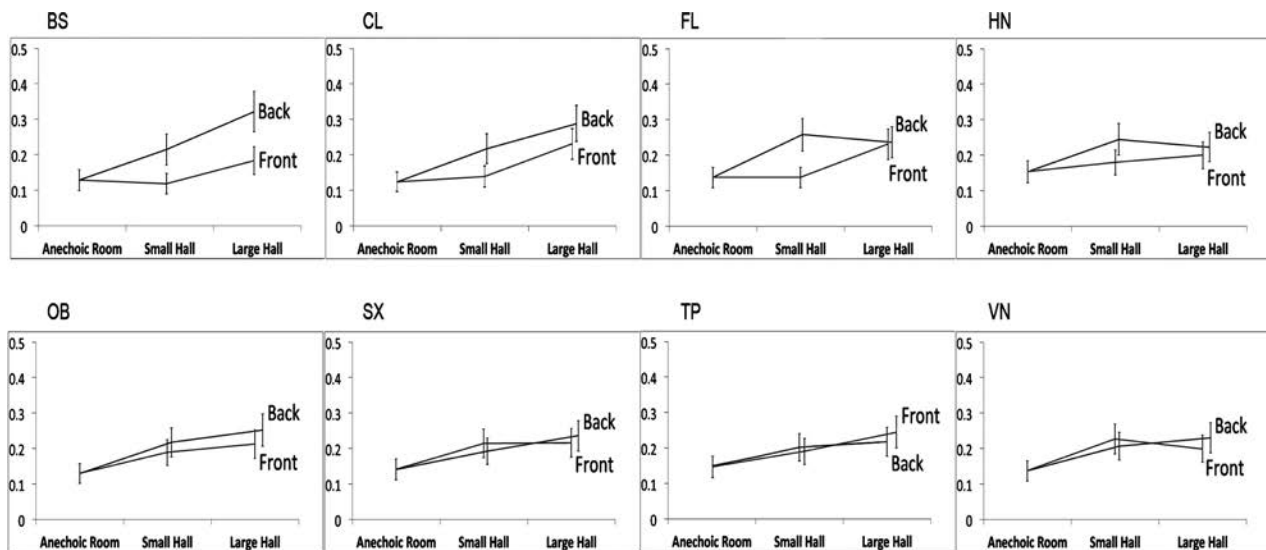Fig. 9. BTL scale values and the corresponding 95% confidence intervals for Comic.

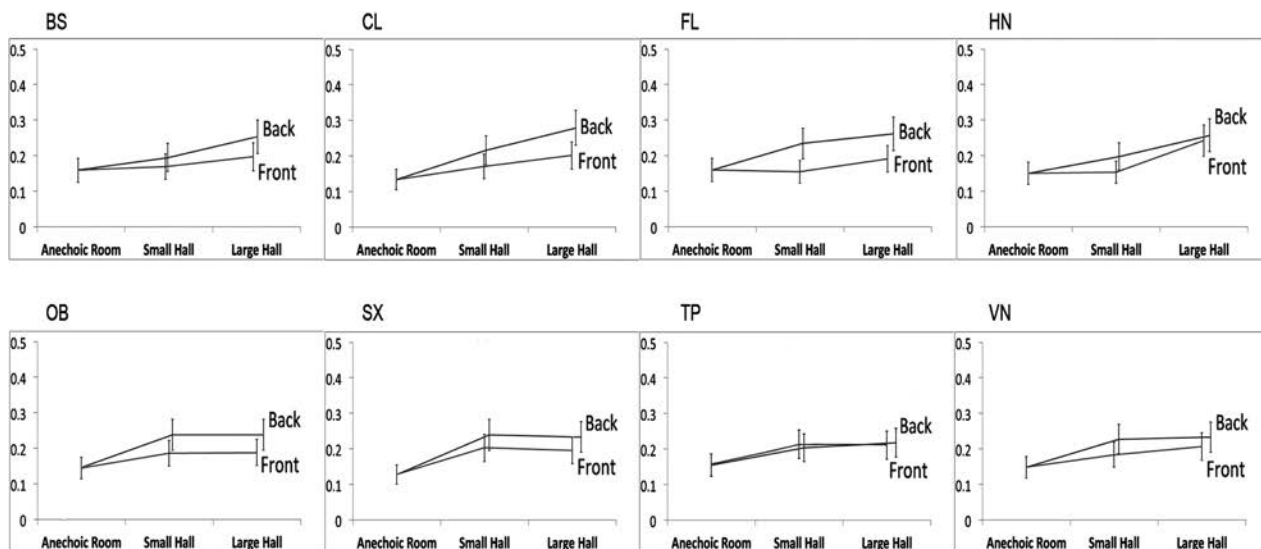Fig. 10. BTL scale values and the corresponding 95% confidence intervals for Sad.

Fig. 11.   BTL scale values and the corresponding 95% confidence intervals for Scary.
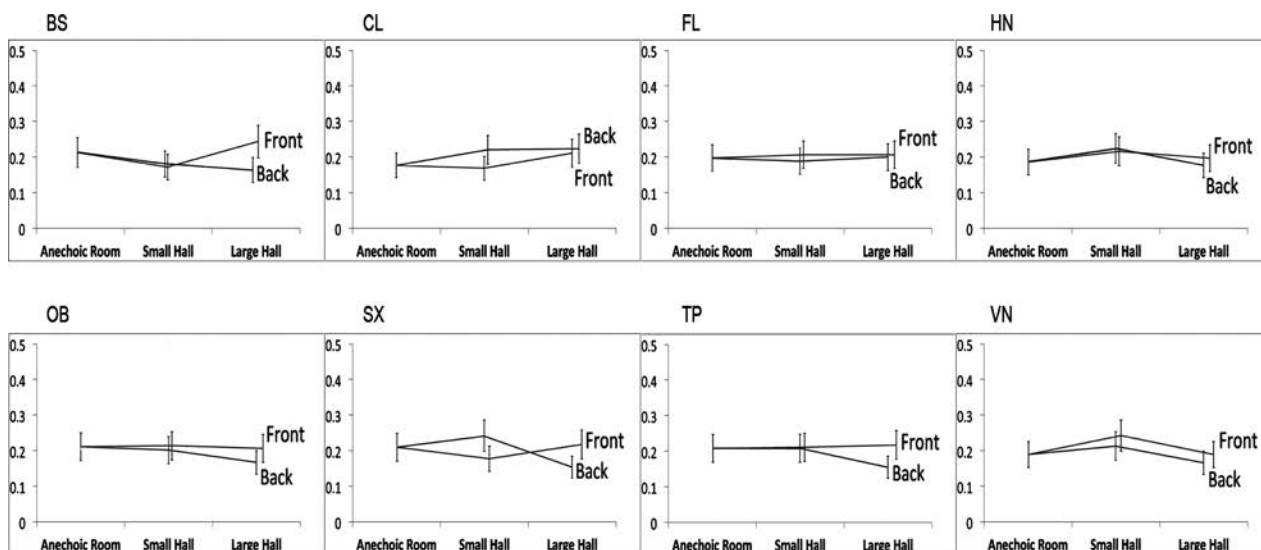

Fig. 12.   BTL scale values and the corresponding 95% confidence intervals for Shy.
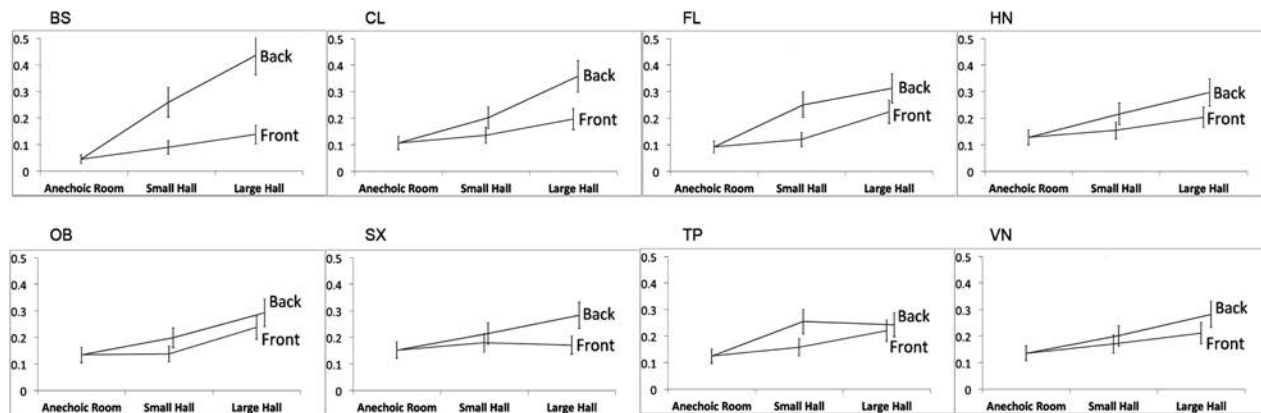

Fig. 13.   BTL scale values and the corresponding 95% confidence intervals for Romantic.

bottom of its 95% confidence interval was greater than the top of their 95% confidence interval) over the eight instruments. The maximum possible value is 32 and the minimum possible value is 0. Table 4 shows the maximum value for each emotional category in bold in a shaded box (except for Shy since all its values are zero or near-zero).

Table 4 shows that for the emotional category Happy, Small Hall Front and Small Hall Back together had most of the significant rankings. This result agrees with that
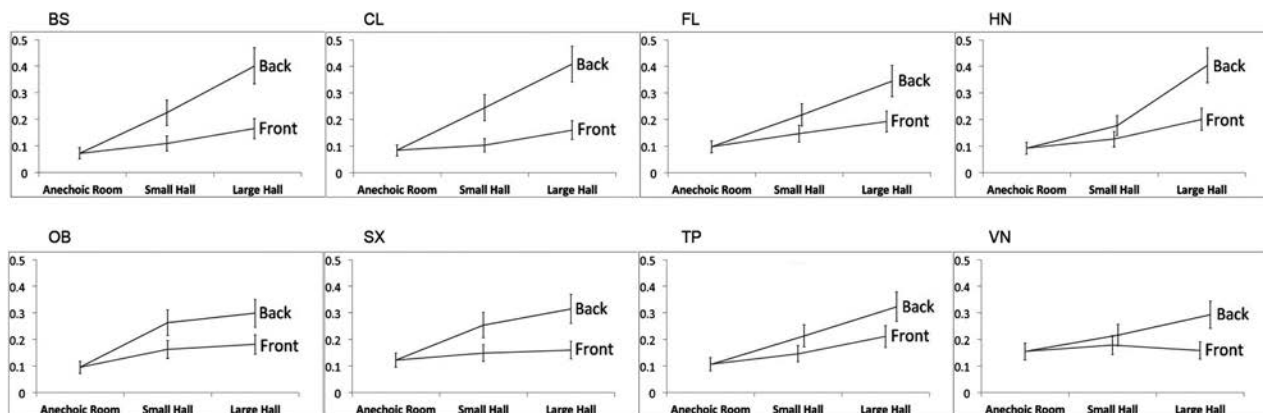
Fig. 14. BTL scale values and the corresponding 95% confidence intervals for Mysterious. The clarinet was moderately well-described ($0.01 < p < 0.05$) by the BTL model while all the other instruments were well-described. (Note that other instruments and emotional categories were well-described by the BTL model.)

Table 4. How often each reverb type was statistically significantly greater than the others over the eight instruments. The maximum possible value is 32 and the minimum possible value is 0. The maximum for each emotional category is shown in bold (except for Shy since all its values are zero or near-zero).

Emotion Category	Reverb Type	Anechoic	Small Hall Front	Small Hall Back	Large Hall Front	Large Hall Back
Happy		0	**4**	3	2	0
Heroic		0	0	3	2	**7**
Comic		**6**	4	1	4	0
Sad		0	2	9	7	**11**
Scary		0	1	5	4	**11**
Shy		0	0	1	0	0
Romantic		0	1	9	9	**23**
Mysterious		0	1	12	7	**29**

found by Tajadura-Jiménez [46], who found that smaller rooms were most pleasant (Fig. 5 indicates that Happy is high-Valence or very pleasant). The result also agrees with Västfjäll [45], who found that larger reverberation times were more unpleasant than shorter ones.

For Heroic, Large Hall Back was ranked significantly greater more often than all the other options combined. This result is in contrast to that found by Västfjäll [45] and Tajadura-Jiménez [46] since Heroic, like Happy, is also high-Valence, and they would have predicted that Heroic would have had a similar result as Happy.

Table 4 also shows that Anechoic (and to a lesser extent Small Hall Front and Large Hall Front) was the most Comic, while Large Hall Back was the least Comic. This basically agrees with Västfjäll [45] and Tajadura-Jiménez [46].

Large Hall Back was the most Sad in Table 4 (though Small Hall Back and Large Hall Front were not far behind). Large Hall Back was more decisively on top for Scary. Since Sad and Scary are both low-Valence (see Fig. 5), these results agree with Västfjäll [45] and Tajadura-Jiménez [46] who found that larger reverberation times and larger rooms were more unpleasant.

Reverb had very little effect on Shy in Table 4. There were almost no significant differences between the reverb types and instruments.

The Romantic rankings in Fig. 13 were more widely spaced than the other categories, and Table 4 indicates that

Large Hall Back was significantly more Romantic than most other reverb types. Like Heroic, this result is in contrast to the results of Västfjäll [45] and Tajadura-Jiménez [46] since Romantic is high-Valence. The bassoon for Romantic was the most strongly affected among all instruments and emotional categories.

Similar to Romantic, the Mysterious rankings in Fig. 14 were also widely spaced. Table 4 indicates Large Hall Back was significantly more Mysterious than nearly all other reverb types across the eight instruments. Also, Small Hall Back was significantly more Mysterious than Large Hall Front for about half the instruments.

In summary, our results show distinctive differences between the high-Valence emotional categories Happy, Heroic, Comic, and Romantic. In this respect our results contrast with the results of Västfjäll [45] and Tajadura-Jiménez [46].

4 DISCUSSION

The main goal motivating our work is to understand how emotional characteristics vary with reverberation length and amount in simple parametric reverberation. In other words, roughly how emotional characteristics vary with hall size and listener position relative to the front or back of the hall. Based on Table 4 our main findings are the following:

1. Simple parametric reverberation had a strongly significant effect on Mysterious and Romantic for Large Hall Back.
2. Simple parametric reverberation had a medium effect on Sad, Scary, and Heroic for Large Hall Back.
3. Simple parametric reverberation had a mild effect on Happy for Small Hall Front.
4. Simple parametric reverberation had relatively little effect on Shy.
5. Simple parametric reverberation had an opposite effect on Comic, with listeners judging anechoic sounds most Comic.

We should emphasize that these results apply to basic-level professional headphones and that higher-quality professional headphones could perhaps show even more pronounced differentiation.

The above results demonstrate how the categorical emotional model can give added emotional nuance and detail than a 2D model with only Valence and Arousal. Table 4 shows very different results for the high-Valence emotional categories Happy, Heroic, Comic, and Romantic. The results of Västfjäll [45] and Tajadura-Jiménez [46] suggested that all four of these emotional characteristics would be stronger in smaller rooms. Only Happy and Comic were stronger for Small Hall or Anechoic, while Heroic and Romantic were stronger for Large Hall.

The above results give audio engineers and musicians an interesting perspective on simple parametric artificial reverberation since many recordings are done in studios where the type and quantity of artificial reverberation added is decided by the recording engineer and performers.

One possible area for future research would be to investigate the effects of even longer reverberation times (such as 4 seconds, representing cathedral-like spaces) on the emotional characteristics of musical instruments. Also, it would be interesting to investigate the change in emotional characteristics for other reverberation models such as plate reverberation.

5 ACKNOWLEDGMENTS

REFERENCES

[1] T. Eerola, R. Ferrer, and V. Allure "Timbre and Affect Dimensions: Evidence from Affect and Similarity Ratings and Acoustic Correlates of Isolated Instrument Sounds," *Music Perception: An Interdisciplinary J.,* vol. 30, no. 1, pp. 49–70 (2012), doi: http://dx.doi.org/10.1525/mp.2012.30.1.49.

[2] B. Wu, A. Horner, and C. Lee "Musical Timbre and Emotion: The Identification of Salient Timbral Features in Sustained Musical Instrument Tones Equalized in Attack Time and Spectral Centroid," *International Computer Music Conference (ICMC), Athens, Greece*, pp. 928–934 (14–20 Sept 2014).

[3] C. Chau, B. Wu, and A. Horner "Timbre Features and Music Emotion in Plucked String, Mallet Percussion, and Keyboard Tones," International Computer Music Conference (ICMC), Athens, Greece, pp. 982–989 (14-20 Sept 2014).

[4] B. Wu, C. Lee, and A. Horner "The Correspondence of Music Emotion and Timbre in Sustained Musical Instrument Tones," *J. Audio Eng. Soc.,* vol. 62, pp. 663–675 (2014 Oct.), doi: http://dx.doi.org/10.17743/jaes.2014.0037.

[5] C. Chau, B. Wu, and A. Horner "The Emotional Characteristics and Timbre of Nonsustaining Instrument Sounds," *J. Audio Eng. Soc.,* vol. 63, pp. 228–244 (2015 Apr.), doi: http://dx.doi.org/10.17743/jaes.2015.0016.

[6] L.-L. Balkwill and W. F. Thompson "A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues," *Music Perception*, pp. 43–64 (1999). doi: http://dx.doi.org/10.2307/40285811.

[7] J. Liebetrau, S. Schneider, and R. Jezierski "Application of Free Choice Profiling for the Evaluation of Emotions Elicited by Music," *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012): Music and Emotions*, pp. 78–93 (2012).

[8] I. Lahdelma and T. Eerola "Single Chords Convey Distinct Emotional Qualities to both Naíve and Expert Listeners," *Psychology of Music*, p. 0305735614552006 (2014).

[9] J. Skowronek, M. McKinney, and S. Van De Par "A Demonstrator for Automatic Music Mood Estimation," *Proceedings of the International Conference on Music Information Retrieval* (2007).

[10] M. Plewa and B. Kostek "A Study on Correlation between Tempo and Mood of Music," presented at the *133rd Convention of the Audio Engineering Society* (2012 Oct.), convention paper 8800.

[11] Y. Hu, X. Chen, and D. Yang "Lyric-Based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method," *Proceedings of ISMIR* (2009).

[12] I. Ekman and R. Kajastila "Localization Cues Affect Emotional Judgments—Results from a User Study on Scary Sound," presented at the *AES 35th International Conference: Audio for Games* (2009 Feb.), conference paper 23.

[13] G. Tzanetakis and P. Cook "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing,* vol. 10, no. 5, pp. 293–302 (2002), doi: http://dx.doi.org/10.1109/TSA.2002.800560.

[14] J-J. Aucouturier, F. Pachet, and M. Sandler "'The Way it Sounds": Timbre Models for Analysis and Retrieval of Music Signals," *Multimedia, IEEE Transactions on,* vol. 7, no. 6, pp. 1028–1035 (2005). doi: http://dx.doi.org/10.1109/TMM.2005.858380.

[15] K. Hevner "Experimental Studies of the Elements of Expression in Music," *Amer. J. Psych.*, pp. 246–268 (1936), doi: http://dx.doi.org/10.2307/1415746.

[16] K. R. Scherer and J. S. Oshinsky "Cue Utilization in Emotion Attribution from Auditory Stimuli," *Motivation and Emotion,* vol. 1, no. 4, pp. 331–346 (1977), doi: http://dx.doi.org/10.1007/BF00992539.

[17] I. Peretz, L. Gagnon, and B. Bouchard "Music and Emotion: Perceptual Determinants, Immediacy, and Isolation after Brain Damage," *Cognition,* vol. 68, pp. 111–141 (1998), doi: http://dx.doi.org/10.1016/S0010-0277(98)00043-2.

[18] W. Ellermeier, M. Mader, and P. Daniel, "Scaling the Unpleasantness of Sounds According to the BTL Model: Ratio-Scale Representation and Psychoacoustical Analysis," *Acta Acustica United with Acustica,* vol. 90, no. 1, pp. 101–107 (2004).

[19] E. Bigand et al., "Multidimensional Scaling of Emotional Responses to Music: The Effect of Musical Expertise and of the Duration of the Excerpts," *Cognition & Emotion,* vol. 19, no. 8, pp. 1113–1139 (2005), doi: http://dx.doi.org/10.1080/02699930500204250.

[20] M. Zentner, D. Grandjean, and K. R Scherer "Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement," *Emotion,* vol. 8, p. 494 (2008). doi: http://dx.doi.org/10.1037/1528-3542.8.4.494.

[21] J. C Hailstone et al., "It's Not What You Play, It's How You Play It: Timbre Affects Perception of Emotion in Music," *Quarterly J. Exper. Psych.,* vol. 62, no. 11, pp. 2141–2155 (2009), doi: http://dx.doi.org/10.1080/17470210902765957.

[22] Y.-H. Yang et al., "A Regression Approach to Music Emotion Recognition," *IEEE TASLP,* vol. 16, no. 2, pp. 448–457 (May 15, 2009), doi: http://dx.doi.org/10.1109/tasl.2007.911513.

[23] J. A. Russell "A Circumplex Model of Affect," *J. Personality and Social Psych.,* vol. 39, no. 6, p. 1161 (1980), doi: http://dx.doi.org/10.1037/h0077714.

[24] C. L. Krumvansl, "Plink: 'Thin Slices' of Music," *Music Perception: An Interdisciplinary J.,* vol. 27, no. 5 (2010), doi: http://dx.doi.org/10.1525/mp.2010.27.5.337.

[25] S. Filipic, B. Tillmann, and E. Bigand "Judging Familiarity and Emotion from Very Brief Musical Excerpts," *Psychonomic Bulletin & Rev.,* vol. 17, pp. 335–341 (2010), doi: http://dx.doi.org/10.3758/pbr.17.3.335.

[26] T. Eerola and J. K. Vuoskoski "A Comparison of the Discrete and Dimensional Models of Emotion in Music," *Psychology of Music,* vol. 39, no. 1, pp. 18–49 (2011), doi: http://dx.doi.org/10.1177/0305735610362821.

[27] J. K. Vuoskoski and T. Eerola "Measuring Music-Induced Emotion: A Comparison of Emotion Models, Personality Biases, and Intensity of Experiences," *Musicae Sciential,* vol. 15, no. 2, pp. 159–173 (2011), doi: http://dx.doi.org/10.1177/1029864911403367.

[28] E. Asutay et al., "Emoacoustics: A Study of the Psychoacoustical and Psychological Dimensions of Emotional Sound Design," *J. Audio Eng. Soc.,* vol. 60, pp. 21–28 (2012 Jan./Feb.).

[29] J. Liebetrau et al., "Paired Comparison as a Method for Measuring Emotions," presented at the *135th Convention of the Audio Engineering Society* (2013 Oct.), convention paper 9016.

[30] C. Baume "Evaluation of Acoustic Features for Music Emotion Recognition," presented at the *134th Convention of the Audio Engineering Society* (2013 May), convention paper 8811.

[31] B. Wu et al., "Investigating Correlation between Musical Timbres and Emotions," *International Society for Music Information Retrieval Conference (ISMIR), Curitiba, Brazil* (2013), pp. 415–420.

[32] B. Wu, A. Horner, and C. Lee "Emotional Predisposition of Musical Instrument Timbres with Static Spectra," *International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan,* vol. 253–258 (Nov. 2014).

[33] M. R. Schroeder "Natural Sounding Artificial Reverberation," *J. Audio Eng. Soc.,* vol. 10, pp. 219–223 (1962 July).

[34] M. R. Schroeder "Digital Simulation of Sound Transmission in Reverberant Spaces," *J. Acous. Soc. Amer.,* vol. 47, no. 2A, pp. 424–431 (1970), doi: http://dx.doi.org/10.1121/1.1971383.

[35] J. A. Moorer "About this Reverberation Business," *Computer Music J.,* vol. 3, no. 2, pp. 13–28 (1979 June). doi: http://dx.doi.org/10.2307/3680280.

[36] J.-M. Jot and A. Change "Digital Delay Networks for Designing Artificial Reverberators," presented at the *90th Convention of the Audio Engineering Society* (1991 Feb.), convention paper 3030.

[37] W. G. Gardner "A Realtime Multichannel Room Simulator," *J. Acoust. Soc. Amer.,* vol. 92, no. 4, p. 2395 (1992). doi: http://dx.doi.org/10.1121/1.404752.

[38] W. G. Gardner "*The Virtual Acoustic Room," Ph.D. thesis,* Citeseer (1992).

[39] A. Reilly and D. McGrath "Convolution Processing for Realistic Reverberation," presented at the *98th Convention of the Audio Engineering Society* (1995 Feb.), convention paper 3977.

[40] A. Farina "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," presented at the *108th Convention of the Audio Engineering Society* (2000 Feb.), convention paper 5093.

[41] W. C. Sabine and M. David Egan "Collected Papers on Acoustics," *J. Acous. Soc. Amer.,* vol. 95, no. 6, pp. 3679–3680 (1994), doi: http://dx.doi.org/10.1121/1.409944.

[42] V. L. Jordan "Acoustical Criteria for Auditoriums and Their Relation to Model Techniques," *J. Acous. Soc. Amer.,* vol. 47, no. 2A, pp. 408–412 (1970), doi: http://dx.doi.org/10.1121/1.1911535.

[43] M. Kaczmarek, C. Szmal, and R. Tomczyk "Influence of the Sound Effects on the Sound Quality," presented at the *106th Convention of the Audio Engineering Society* (1999 May), convention paper 4902.

[44] L. Cremer, H. .A Müller, and T. J. Schaultz *Principles and Applications of Room Acoustics,* Vol. 1 (Applied Science, NY, 1982).

[45] D. Västfjäll, P. Larsson, and M. Kleiner "Emotion and Auditory Virtual Environments: Affect-Based Judgments of Music Reproduced with Virtual Reverberation Times," *CyberPsychology & Be-*

havior, vol. 5, no. 1, pp. 19–32 (2002), doi: http://dx.doi.org/10.1089/109493102753685854.

[46] A. Tajadura-Jiménez et al., "When Room Size Matters: Acoustic Influences on Emotional Responses to Sounds," *Emotion,* vol. 10, no. 3, pp. 416–422 (2010), doi: http://dx.doi.org/10.1037/a0018423.

[47] "University of Iowa Musical Instrument Samples," University of Iowa (2004). http://theremin.music.uiowa.edu/MIS.html.

[48] J. W. Beauchamp "Analysis and Synthesis of Musical Instrument Sounds," in *Analysis, Synthesis, and Perception of Musical Sounds* (Springer, 2007), pp. 1–89, doi: http://dx.doi.org/10.1007/978-0-387-32576-7_1.

[49] T. Hidaka and L. L. Beranek "Objective and Subjective Evaluations of Twenty-Three Opera Houses in Europe, Japan, and the Americas," *J. Acous. Soc. Amer.,* vol. 107, no. 1, pp. 368–383 (2000), doi: http://dx.doi.org/10.1121/1.428309.

[50] L. Beranek *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* (Springer Science & Business Media, 2004), doi: http://dx.doi.org/10.1007/978-0-387-21636-2.

[51] "Cool Edit," Adobe Systems (2000). https://creative.adobe.com/products/audition.

[52] P. N. Juslin and J. Slobodan *Handbook of Music and Emotion: Theory, Research, Applications* (Oxford University Press, 1993), doi: http://dx.doi.org/10.1093/acprof:oso/9780199230143.001.0001.

[53] M. Kennedy and K. J. Bourne *The Oxford Dictionary of Music* (Oxford University Press, 2012),
doi: http://dx.doi.org/10.1093/acref/9780199578108.001.0001.

[54] Connect for Education Inc. OnMusic Dictionary. url: http://dictionary.onmusic.org/ (visited on 12/29/2014).

[55] Classical.dj. Classical Musical Terms. url: http://www.classical.dj/musical_terms.html (visited on 12/29/2014).

[56] Dolmetsch Organisation. Dolmetsch Online - Music Dictionary. url: http://www.dolmetsch.com/musictheorydefs.htm (visited on 12/29/2014).

[57] M. M. Bradley and P. J. Lang, *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings,* Tech. rep. (Citeseer, 1999).

[58] *Cambridge Academic Content Dictionary.* url: http://dictionary.cambridge.org/dictionary/american-english.

[59] F. L. Joseph "Measuring Nominal Scale Agreement among Many Raters," *Psychological Bulletin,* vol. 76, no. 5, pp. 378–382 (1971), doi: http://dx.doi.org/10.1037/h0031619.

[60] R. A. Bradley, "14 Paired Comparisons: Some Basic Procedures and Examples," Nonparametric Methods,vol. 4, pp. 299–326 (1984), doi: http://dx.doi.org/10.1016/s0169-7161(84)04016-5.

[61] F. Wickelmaier and C. Schmid "A Matlab Function to Estimate Choice Model Parameters from Paired-comparison Data," *Behavior Research Methods, Instruments, and Computers,* vol. 36, no. 1, pp. 29–40 (2004), doi: http://dx.doi.org/10.3758/bf03195547.

THE AUTHORS

Ronald Mo   Bin Wu   Andrew Horner

Ronald Mo is currently a Ph.D. student in the department of computer science and engineering at the Hong Kong University of Science and Technology. His research interests include timbre of musical instruments and music emotion recognition. He received his B.Eng. of computer science and M.Phil. of computer science and engineering from the Hong Kong University of Science and Technology in 2007 and 2015 respectively.

Bin Wu is currently a senior research engineer at Baidu. He obtained his Ph.D. in computer science and engineering

from the Hong Kong University of Science and Technology in 2015. His research interests include music emotion recognition, data mining, and musical timbre.

Andrew Horner is a professor in the department of computer science and engineering at the Hong Kong University of Science and Technology. His research interests include music analysis and synthesis, timbre of musical instruments, and music emotion. He received his Ph.D. in computer science from the University of Illinois at Urbana-Champaign.