THE EFFECTS OF AUDIO PROCESSING ON THE PERCEIVED EMOTIONAL CHARACTERISTICS OF MUSICAL INSTRUMENT SOUNDS

by

RONALD KYLE MO

A Thesis Submitted to
The Hong Kong University of Science and Technology
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
in Computer Science and Engineering

August 2017, Hong Kong

Authorization

I hereby declare that I am the sole author of the thesis.

I authorize the Hong Kong University of Science and Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Hong Kong University of Science and Technology to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

RONALD KYLE MO

THE EFFECTS OF AUDIO PROCESSING ON THE PERCEIVED EMOTIONAL CHARACTERISTICS OF MUSICAL INSTRUMENT SOUNDS

by

RONALD KYLE MO

This is to certify that I have examined the above Ph.D. thesis

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by

the thesis examination committee have been made.

PROF. ANDREW HORNER, THESIS SUPERVISOR

PROF. HUAMIN QU, ACTING HEAD OF DEPARTMENT

Department of Computer Science and Engineering

3 August 2017

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my thesis supervisor, Professor Andrew Horner, for his kindly guidance and advice in my long journey of postgraduate studies. He always encourages me to express myself freely and tackle those tough research questions with confidence. He has also taught me how to conduct my research scientifically and systematically, as well as how to present my works clearly. Thanks for Professor Horner's supervision and guidance, my writing and presentation skills have been improved quite a lot throughout my PhD studies.

I would like to sincerely thank members of my PhD thesis examination committee, Professor Raymond Wong, Professor Xiaojuan Ma, and Professor Richard So, who also served as the committee members of my PhD qualifying examination and thesis proposal examination. I would also like to express my gratitude to Professor Ricky Kwok for accepting to be my external committee member, and Professor Ajay Joneja for being the chairperson of my PhD thesis examination. I appreciate their valuable time and comment on my PhD thesis which makes my PhD thesis more complete.

In addition, I would like to thank the people in our research group: Chuck-jee Chau, Bin Wu, Hong Yu, Jennifer Chan, Bing Yen Chang, along with all the subjects of my listening tests, for the help of my research in these years. It would not be successful without your help and support. Many thanks to my colleagues Nicko Shek, Cecia Chan, Alex Lam, Kevin Wang, Wallace Mak, Katherine Fung, and Luke Lo for their support in these years.

Last but not least, I owe a debt of gratitude to my mother. Her endless support and love are what support me all the time. A special thanks goes to my fiancée, Yoki Ho, for her unselfish love, care, and support. She is always by my side through the ups and downs. Without your love, it is impossible for me to pursue my dreams. 1437.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THE EFFECTS OF AUDIO PROCESSING ON THE PERCEIVED EMOTIONAL CHARACTERISTICS OF MUSICAL INSTRUMENT SOUNDS

by

RONALD KYLE MO

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

ABSTRACT

Musical instrument sounds have been shown to have distinct timbral and emotional characteristics, and when audio processes are applied to them, their timbral and emotional characteristics are changed. In this thesis, we investigated into how audio processes change the perceived emotional characteristics of musical instrument sounds.

We first investigated the effects of MP3 compression on the emotional characteristics of instrument sounds, which has not been explored previously. Our results showed that MP3 compression strengthened neutral and negative emotional characteristics such as Scary and Sad, and weakened positive emotional characteristics such as Happy and Romantic. Interestingly, Angry was relatively unaffected by MP3 compression.

For artificial reverberation, since our previous research has shown that the distinctive emotional characteristics in musical instruments can be significantly changed with parametric reverberation, we would like to see whether the parametric reverberation results can be applied to real concert hall reverberation, namely convolution reverberation, as well. We would like to know whether these changes in character are relatively uniform or instrument-dependent as well.

Our finding shows that convolution reverberation had more pronounced effects on the emotional characteristics compared to parametric reverberation, yet there was a strong agreement in the results of parametric and convolution reverberations. For investigating into the underlying instrument space with reverberation, our results indicate that the underlying instrument space did not change much with both parametric and convolution reverberations, in terms of emotional characteristics. It means that reverberation time has a remarkably consistent effect on the emotional characteristics no matter whether parametric or convolution reverberation was used. It is also a reflection of their deep underlying functional similarities despite their fundamentally different implementations.

In terms of applications, our MP3 compression study will give listeners and music streaming service providers some preliminary benchmarks for understanding the emotional effects of MP3 compression on music. For the artificial reverberation studies, the relatively consistent rankings of emotional characteristics between the instruments certainly helps each instrument retain its identity in different halls. Moreover, the instrument-independent behavior of concert halls is perhaps what helps distinguish a good music venue from a poor one. This can be an interesting avenue for future work.

CHAPTER 1

INTRODUCTION

1.1 Motivation

When someone thinks of a piece of music, they usually think of its melody. They might also consider other striking features related to tempo, dynamics, pitch range, mode, or harmony. The instrument playing the melody, and the spaciousness of the place where it is played also shape the emotional characteristics of the music, and are the topic of this thesis.

We will give an analogy to try to make the idea easier to visualize. Claude Monet created a series of impressionist paintings of Rouen Cathedral, all from the same viewpoint, where each had a very different character depending on the time of day and weather. He used the same subject, viewpoint, and style, but with different colors and visibility to produce distinctly different moods ranging from majestic to mysterious to foreboding to tranquil.

Translating the idea to our domain, the equivalent of Monet's cathedral with its different colors are tones from different musical instruments with their distinctive sound colors. And the equivalent to the visibility around Monet's cathedral is reverberation of these instrument tones. Just as the moisture in the air smears the details of the cathedral's textures and colors, reverberation adds spaciousness to the sound that smears its temporal and spectral envelopes. For the layman, this smearing is perhaps most obvious in karaoke systems, where reverberation smooths imperfections in the voice (e.g., its waverings in tone color), just as an air-brush smooths facial features in a photo. And just as Monet used different colors and visibility to bring out different moods, we want to investigate more fully how reverberation changes the emotional characteristics of musical instruments.

It is obvious even to a non-musician that an instrument can express a very wide range of emotional characteristics depending on the musical context. But, at the

Figure 1.1: A series of paintings of Rouen Cathedral by Monet. Even in black and white, each has a different character.

same time each instrument has its own distinct sound color that brings out particular emotional characteristics. For example, a trumpet sound can be high or low, loud or soft, muted or open, and played on trumpets in different keys (e.g., Bb, C, or Eb). They are all trumpet sounds, and experienced listeners recognize them all as trumpet sounds. At the same time, each of these trumpet sounds has its particular sound color or timbre, influenced by the temporal and spectral envelopes of the sound. Whenever the timbre changes even a little, at least some of the emotional characteristics of the sound change with it. For example, if a trumpet sound is higher, faster, louder, or brighter, it will generally become higher arousal in character and be perceived as more heroic, joyful, or angry depending on the particular musical context. It is also obvious that audio processing such as reverberation would add or alter the color of an instrument sound, therefore change the emotional characteristics one instrument brought out.

Researchers have considered various relationships between timbre (sound color) and music emotion [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16], and in particular, have found that different instruments have different timbral and emotional characteristics [17, 18, 19, 20, 21, 22, 23, 24]. By changing the pitch and dynamics, the timbre and emotional characteristics also change [25, 26, 27, 28, 29, 30, 31, 32, 33]. These characteristics are further modified by the performance environment —by the amount and length of reverberation in the space [34, 35, 36, 37], which smears the temporal and spectral envelopes and changes the emotional character of the sound. The same idea holds when artificial reverberation is added as a post-process. The main goal motivating our initial study on reverberation was to understand how emotional characteristics vary with different reverberation parameters [38]. In light of this, one may wonder how audio processes effect the emotional characteristics of musical instrument sounds perceived by the audience.

1.2 MP3 Compression

Audio post-processes that have nothing to do with the music, instruments, or recording environment can also change the sound, its color, and emotional characteristics [39]. MP3 compression is such an example, and is often used to speed up downloads and streaming by discarding less audible parts of the sound due to simultaneous and temporal masking [40, 41]. But due to the lossy nature of MP3 compression, the sound is altered. The artifacts of MP3 compression are audible when high compression rates (i.e., low bit rates) are used [42, 43, 44, 45, 46, 47, 48, 49]. These artifacts change the timbre of the sound more and more with higher compression rates [50]. In particular, MP3 compression adds quantization jitter to the amplitude envelopes, making the spectrum more dynamic, and increasing spectral incoherence. Some instruments such as the saxophone are much more effected than the others such as the horn [50].

One of the topics of this thesis is to determine how these MP3 artifacts change the emotional characteristics of the sound. Even though many previous studies have considered the relationship between music emotion and timbre, the relationship between music emotion and MP3 compression is still unexplored. In light of this, one might wonder how much MP3 compression affects the emotional characteristics of musical instruments. In particular, do all emotional characteristics decrease about equally with more compression, or do some increase and others decrease? Are any emotional characteristics relatively unaffected by compression? Which instruments change the most or least with more compression?

Major music streaming service providers such as Spotify use MP3 compression, and typically allow users to select the quality for streaming and downloads. The quality ranges from lower to higher quality bit rates for MP3 compression, and includes an automated option for selecting the bit rate adaptively depending on the speed of the connection. Lower quality bit rates download faster, especially when the connection is poor, and are therefore popular and often automatically chosen. But, a lower quality bit rate means the music quality has been somewhat compromised.

One way to study the effects of MP3 compression is to compare a number of short pieces of music with varying amounts of MP3 compression. A disadvantage with this approach is that the emotional effects of the artifacts might be somewhat obscured by the activity of the music with its different notes, instruments, and textures. For this preliminary study, we have chosen to focus on single instrument tones where MP3 artifacts will be most exposed and obvious. This is actually a rather useful approach since many pieces of music feature a solo instrument prominently. For example, any piece by John Coltrane will automatically feature a prominent saxophone. It would be useful to know how the emotional characteristics of the saxophone in particular are affected by any MP3 compression of his music. This study will give listeners and music streaming service providers some preliminary benchmarks for understanding the emotional effects of MP3 compression on music. It will help quantify how much the emotional characteristics of particular musical instruments such as the saxophone have been changed by MP3 compression, and will give an indication of whether these changes are acceptable or not for particular bit rates and instruments. In light of this, one might wonder how much the emotional characteristics of the music have been changed by different quality bit rate. In particular, it would be interesting to know how MP3 compression affects the emotional characteristics of musical instruments. We will address the following questions in this thesis: Generally, what are the emotional effects of MP3 compression? (Do all emotional characteristics decrease about equally with more compression?) Which emotional characteristics increase or decrease with more compression? Which emotional characteristics are unaffected by more compression? Which instruments change the most and least with more compression? This will give listeners and music streaming service providers a useful benchmark for understanding the emotional effects of MP3 compression on music.

1.3 Parametric and Convolution Reverberation

There are two most popular types of artificial reverberations, namely parametric and convolution reverberation. For parametric reverberations, there are two main parameters, which are reverberation length and amount respectively. Since they are easy to control and manipulate, parametric reverberation is a natural starting place for our initial investigations. Our previous work has shown that the emotional characteristics of instruments are significantly changed with parametric reverberation. For example, parametric reverberation can bring out Mysterious or Heroic from the original recording, or the recording engineer and musicians might use a dry sound to emphasize its Comic character [38, 51]. Of course, the musical context (i.e., the melody) also has its own emotional characteristics that are colored by both the instrument and reverberation.

Convolution reverberation, on the other hand, is a bit more complex or computationally costly in the sense that it depends on particular impulse responses measured in real halls at discrete points. In other words, with convolution reverberation you are working with impulse responses at a few discrete points in the real hall, while with parametric reverberation you have a continuum of parameter values for the reverberation amount and reverberation time. In light of this, one may wonder whether the results we found for parametric reverberation can be applied to convolution reverberation as well. This is a critical next step since convolution reverberation is relatively popular and probably even more frequently used than simple parametric reverberation, and it is probably a better indicator of how instruments sound in real concert hall environments. Therefore, we would like to know whether the results we found for parametric reverberation can be applied to convolution reverberation too, or in other words, the results for parametric and convolution reverberation are consistent.

Concretely, we seek to understand how the emotional characteristics of musical instruments vary with reverberation time in convolution reverberation. In our previous study on parametric reverberation, reverberation time had strongly significant effects on the emotional characteristics Romantic and Mysterious, and medium effects on Sad, Scary, and Heroic. Anechoic tones were judged most Comic. We are particularly interested to compare and contrast these results to those for convolution reverberation, since convolution reverberation is usually regarded as smoother, warmer, and more natural than parametric reverberation, so the emotional characteristics could be basically similar, further enhanced, or completely different.

The answers to this question will give audio engineers and musicians an interesting perspective on reverberation since many recordings are done in studios where the type and quantity of artificial reverberation added is decided by recording engineers and performers. This also has applications in music designed for virtual environments, computer games, film soundtracks, and karaoke systems by adjusting reverberation to emphasize desired emotional characteristics.

1.4 Instrument Space

In the previous section, we discussed the idea of how convolution reverberation changes the emotional characteristics of instrument tones as the way parametric reverberation did. No matter what the results will be, this leads us to another direction: while reverberation can strengthen or deemphasize particular emotional characteristics, does it also change the underlying instrument space? In other words, when reverberation changes the emotional characteristics of the instruments, does it change them uniformly or some instruments more than others? If we compare the instruments in terms of the emotional characteristic Heroic for example and rank them, is the ranking about the same for different amounts and lengths of reverberation? Or, does a bright instrument such as the trumpet increase more in its Heroic character with more reverberation compared to darker instruments such as the horn? These questions have not been investigated previously to our knowledge, even though they have some implications in timbre and music emotion research.

Our second objective for this thesis is to answer the question: does reverberation change the emotional characteristics of instruments uniformly in about the same way, or is the result instrument-dependent? The answer to this question is interesting in itself from the standpoint of music emotion and timbre. Certainly each instrument has a distinct timbre in the sense that a clarinet is identifiable to musically-trained listeners in an anechoic chamber, a practice room, a recital hall, and a large concert hall. The spectral and temporal envelopes of the clarinet are different depending on the room reverberation, but the instrument identity remains unchanged. Similarly, are there distinctive emotional characteristics for each instrument? In other words, for each emotional characteristic, is there a relatively consistent ranking between the instruments that holds up under different types of reverberation? Is there a footprint of emotional characteristics for each instrument? If not, then in each performance environment the instruments will assume different characters, which helps explain their rich versatility. On the other hand, if there is a unique footprint for each instrument, it helps explain why performers can practice in small rehearsal rooms and reasonably predict the emotional blends and balances between the instruments even when the final performance is in a large concert hall (perhaps with some minor adjustments). In either case, the results will be interesting.

To address these issues, we conducted two listening tests to compare instrument sound over various emotional characteristics with different reverberation settings for both parametric and convolution reverberations. In each test, we compared the instruments pairwise and establish a ranking based on statistical methods for each reverberation type and emotional characteristic. We then correlated the rankings to determine their similarity. We also computed statistically significant differences between the instruments using paired t-tests. Finally, we correlated the results we found in both reverberation environments to see whether the results were consistent. This allows us to judge changes to the instruments in the underlying space of emotional characteristics with different reverberation settings for reverberation.

Answering these questions will also introduce some possible music emotion research of single musical instrument tones. For parametric reverberation, most of the sample libraries contain tones with light reverberation (e.g., The McGill University Master Samples Collection [52], Prosonus Sound Library [53], RWC Music Database [54]), and there are only a limited number of anechoic samples available (e.g., University of Iowa Musical Instrument Samples [55]). Most timbre and music emotion studies of single instrument tones do not explicitly state whether the tones are anechoic or with light reverberation, and assume that it does not matter too much. It would be useful to know whether this is a safe assumption. If reverberation changes the emotional characteristics of instruments uniformly in about the same way, we can use the numerous samples that have light reverberation to compare instruments in terms of their emotional characteristics and expect about the same relative characteristics if they had been recorded in an anechoic chamber or a hall with different reverberation. On the other hand, if the change of emotional characteristics is instrument-dependent with reverberation, it would indicate a strong dependence on the type of reverberation, and suggests the limited applicability of studies of single instrument tones only to tones with similar types of reverberation. In this case, it would also suggest the need for more anechoic sample libraries.

For convolution reverberation, the question whether reverberation is uniform or instrument-dependent sheds light on other aspects of music emotion and timbre. Certainly each instrument has a distinct timbre in the sense that a clarinet is identifiable to musically-trained listeners in an anechoic chamber, a practice room, a recital hall, and a large concert hall. The spectral and temporal envelopes of the clarinet are different depending on the room reverberation due to smearing, but the instrument identity remains unchanged. Similarly, are there distinctive emotional characteristics that identify each instrument? In other words, for each emotional characteristic, is there a relatively consistent ranking between the instruments that holds under different types of reverberation? Is there a footprint of emotional characteristics for each instrument? If so, it helps explain why performers can practice in small rehearsal rooms and reasonably predict the emotional blends and balances between the instruments even when the final performance is in a large concert hall. If not, then in each performance environment the instruments will assume different characters, which helps explain their rich versatility. Either way, the results will deepen our understanding of these issues.

1.5 Summary

For the sake of simplicity, the aforementioned studies are numbered as follows:

1. The Effects of MP3 Compression on Perceived Emotional Characteristics in Musical Instruments

2. The effects of convolution reverberation on the emotional characteristics of musical instrument sounds

3. An investigation into how parametric reverberation effects the space of instrument emotional characteristics

4. An investigation into how convolution reverberation effects the space of instrument emotional characteristics

We will discuss the above experiments in detail in the following chapters. Chapter 2 describes the details of Experiment 1. Chapter 3 describes the details of Experiment 2. Chapter 4 describes the details of Experiment 4. Chapter 5 describes the details of Experiment 4. Chapter 6 introduces some possible future work based on the current study and summarizes our work.

CHAPTER 2

EXPERIMENT 1: THE EFFECTS OF MP3 COMPRESSION ON PERCEIVED EMOTIONAL CHARACTERISTICS IN MUSICAL INSTRUMENTS

2.1 Overview

In order to investigate into how MP3 compression effects the perceived emotional characteristics of musical instruments, we conducted a listening test to compare pairs of original and MP3 compressed instrument sounds over different emotional categories. This research follows a similar basic methodology as the research by Wu et al. [21, 20, 22, 19], Chau et al. [23, 24, 32], and Mo et al. [38], but using MP3 compressed stimuli. Paired comparisons were chosen for simplicity. This section gives further details about the listening test.

2.2 Instrument Sounds

We used eight sustained instrument sounds: bassoon (bs), clarinet (cl), flute (fl), horn (hn), oboe (ob), saxophone (sx), trumpet (tp), and violin (vn). The sustained instruments are nearly harmonic, and the chosen sounds had fundamental frequencies close to Eb4 (311.1 Hz). All eight instrument sounds were also used by a number of other timbre studies [56, 57, 58, 59, 60, 61, 49, 62, 50]. Using the same samples makes it easier to compare results. Most of the instrument sounds came from the McGill University instrument sound collection.

Compressed sounds were encoded and decoded using the LAME MP3 encoder [63]. Instrument sounds were compressed with three different bit rates. As a preliminary step, we listened to the sounds compressed at different bit rates, and judged that 112 Kbps sounds were the lowest bit rate that sounded nearly indistinguishable from the original sounds. Sounds at 32 Kbps had obvious atrifacts. We selected 56 Kbps as an intermediate bit rate, representing medium quality between indistinguishable and obvious artifacts (i.e., some artifacts). Though these particular bit rates may not be as common in practice, they are representative of the basic levels of compression. These three bit rates also gave near-perfect (for 32 Kbps), intermediate (for 56 Kbps), and near-random (for 112 Kbps) discrimination results in a previous discrimination study of these MP3-compressed musical instrument sounds [49], meaning listeners in that study could nearly always notice a difference between the original sound and a 32 Kbps compression, sometime notice a difference for 56 Kbps, and were unable to reliably notice a difference for 112 Kbps.

2.3 Emotional Categories

The subjects compared the stimuli in terms of ten emotional categories: Happy, Heroic, Romantic, Comic, Calm, Mysterious, Shy, Angry, Scary, and Sad. Some choices of emotional characteristics are fairly universal and occur in many previous studies (e.g., Happy, Sad, Scary/Fear/Angry, Tender/Calm/Romantic) roughly corresponding to the four quadrants of the Valence-Arousal plane, but there are lots of variations beyond that [64]. For this study, we used the same categories we have used in our previous research on musical instruments [21, 20, 22, 19, 23, 24, 32, 38]. The ratings of the emotional categories according to the Affective Norms for English Words [65] are shown in Figure 2.1 using the Valence-Arousal model. Valence shows the positiveness of an emotional category; Arousal shows the energy level of an emotional category. Romantic, Happy, Comic, and Heroic form one cluster, and Scary and Angry another. Though Scary and Angry are similar in terms of Valence and Arousal, they have distinctly different meanings. Likewise with Romantic, Happy, Comic, and Heroic.

2.4 Listening Test

There were 20 undergraduate students recruited for the listening test. All subjects were fluent in English and were undergraduate students at the Hong Kong University of Science and Technology, where the medium of all instruction is English. None of the subjects reported any hearing problems. Subjects were not musically-trained subjects

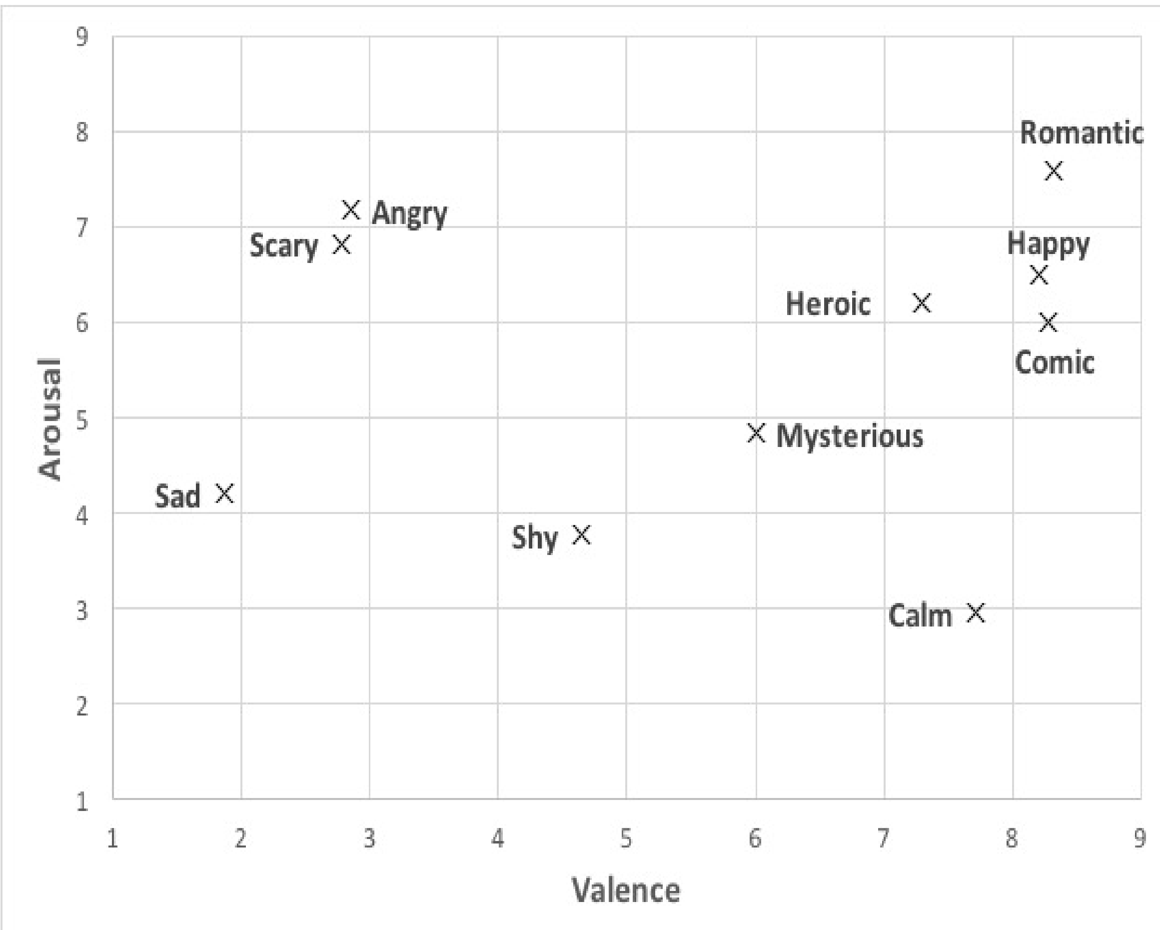

Figure 2.1: Distribution of the emotional characteristics in the dimensions Valence and Arousal. The Valence and Arousal values are given in the 9-point rating in ANEW [65]. Valence shows the positiveness of an emotional category; Arousal shows the energy level of an emotional category.

(e.g., recording engineers, professional musicians, or music conservatory students) but average attentive listeners.

The subjects were provided with an instruction sheet containing definitions of the ten emotional categories from the Cambridge Academic Content Dictionary [66]. The dictionary definitions we used in this experiment are shown in Table 2.1. Every subject made paired comparisons between the sounds (see user interface in Figure 2.2). The test asked listeners to compare four types of compressed sounds for each instrument over ten emotional categories. During each trial, subjects heard a pair of sounds from the same instrument with different types of compression (no compression, 112 Kbps, 56 Kbps, and 32 Kbps) and were prompted to choose which sounded stronger for given emotional characteristics. This method was chosen for simplicity of comparison, since subjects only needed to remember two sounds for each comparison and make a binary decision. This required minimal memory from the subjects, and allowed them to give more instantaneous responses [57, 23]. Each combination of two different compressions was presented for each instrument and emotional category, and the listening test totaled $P_2^4 \times 8 \times 10 = 960$ trials. For each instrument, the overall trial presentation order was randomized (i.e., all combinations of compressed bassoon sounds were in a random order, then all the clarinet comparisons, etc.). However, the emotional categories were presented in order to avoid confusing and fatiguing the subjects. The listening test took about 2 hours, with a short break of 5 minutes after every 30 minutes to help minimize listener fatigue and maintain consistency.

Emotional Category	Definition
Happy	Glad, pleased
Heroic	Exhibiting or marked by courage and daring
Romantic	Relating to love or loving relationship
Comic	Causing laughter or amusement
Calm	A quiet and peaceful state or condition
Mysterious	Strange or unknown
Shy	Disposed to avoid a person or thing
Angry	Having a strong feeling of being upset or annoyed
Scary	Causing fright
Sad	Affected with or expressive of grief or unhappiness

Table 2.1: The dictionary definitions of the emotional categories used in this experiment.

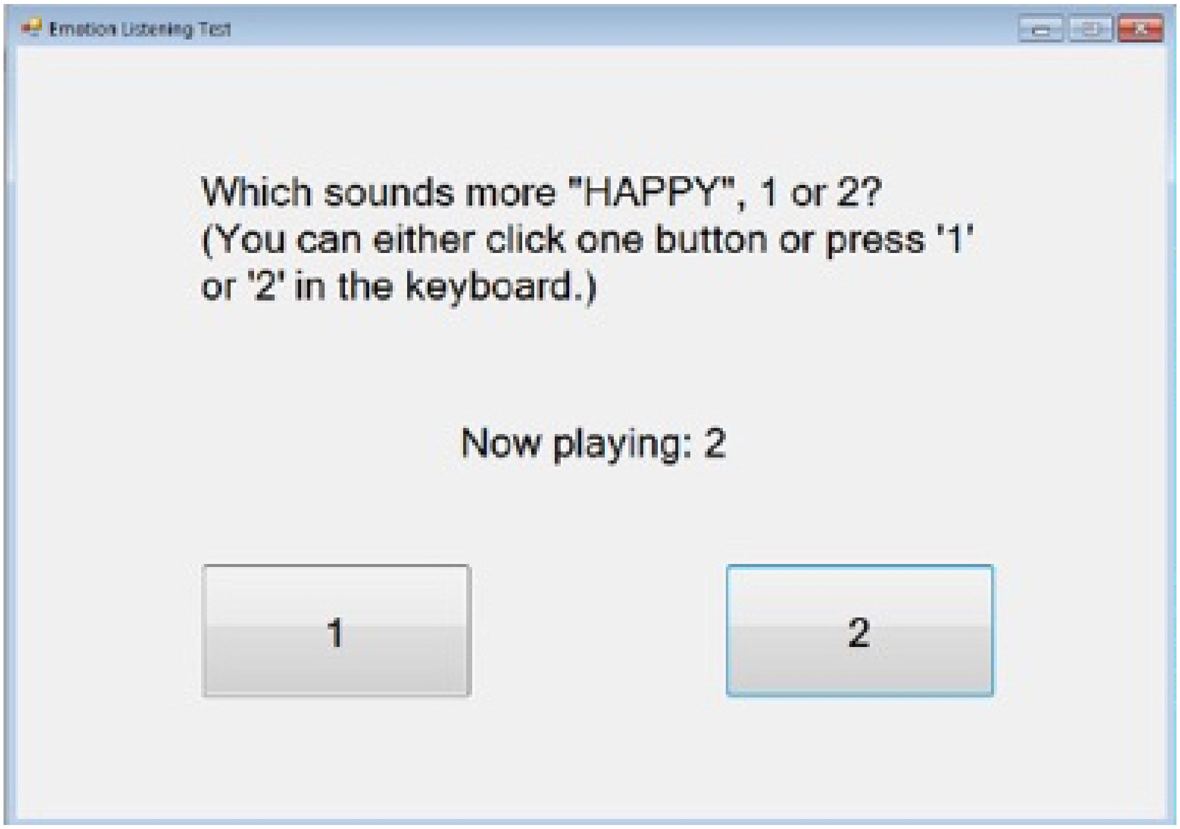

Figure 2.2: Paired comparison listening test interface

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. 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.

2.5 Results

For the listening test, subjects compared pairs of original and compressed instrument sounds for each of the ten emotional categories. The subjects' responses were checked for consistency. Consistency was defined based on the two comparisons of a pair of sounds A and B for a particular instrument and emotional category as follows:

$$consistency_{A,B} = \frac{max(v_A, v_B)}{2} \tag{2.1}$$

where v_A and v_B are the number of votes a subject gave to each of the two sounds. A consistency of 1 represents perfect consistency, whereas 0.5 represents approximately random guessing. The mean average consistency of the 20 subjects was 0.795. 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 statistics. It was calculated at 0.22, indicating a fair agreement among subjects [67].

We ranked the compressed sounds 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 [68, 69]. For each instrument-category pair, the BTL scale values for the original and three compressed sounds sum to 1. The BTL value for each sound is the probability that listeners will choose that compression rate when considering a certain instrument and emotional category. For example, if all four sounds (the original and three compressed sounds) are judged equally happy, the BTL scale values would be 1/4=0.25. We also derived the corresponding 95% confidences intervals for the compressed sounds using the method proposed by Bradley [68].

Figures 2.3 to 2.12 show the BTL values and corresponding 95% confidence intervals for each emotional category. Though there is some variation, the trend generally decreased with more compression for positive-Valence categories such as Happy, and increased for negative-Valence categories such as Sad. In order to examine the significance of the results, Paired t-tests were conducted on the voting data. Table 2.2 shows the number of instruments that were significantly different from the original sound for each compression rate and emotional category. The table shows that there were almost no significant differences for 112 Kbps, some for 56 Kbps, and about half for 32 Kbps. This agrees with the results of Lee et al. [49], which found low, medium, and high discrimination rates for the same bit rates.

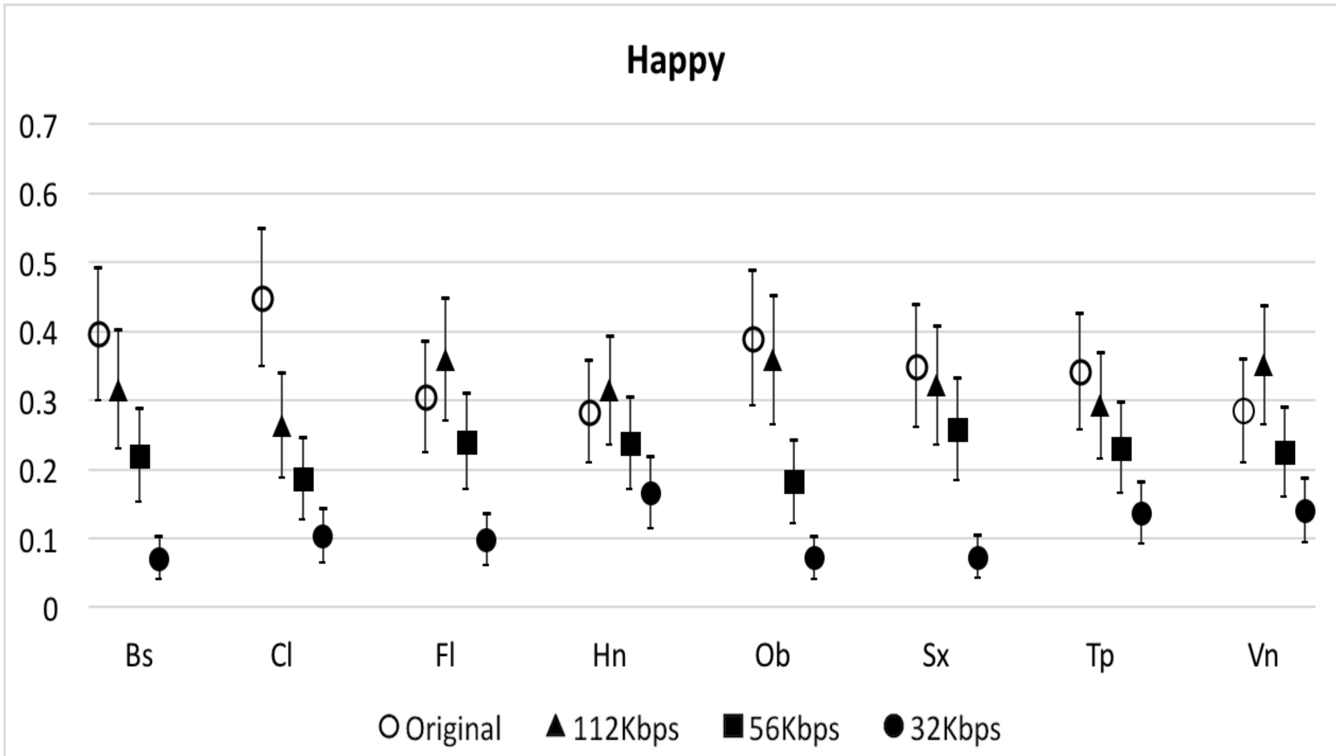

Figure 2.3: BTL scale values and the corresponding 95% confidence intervals for the emotional category Happy.

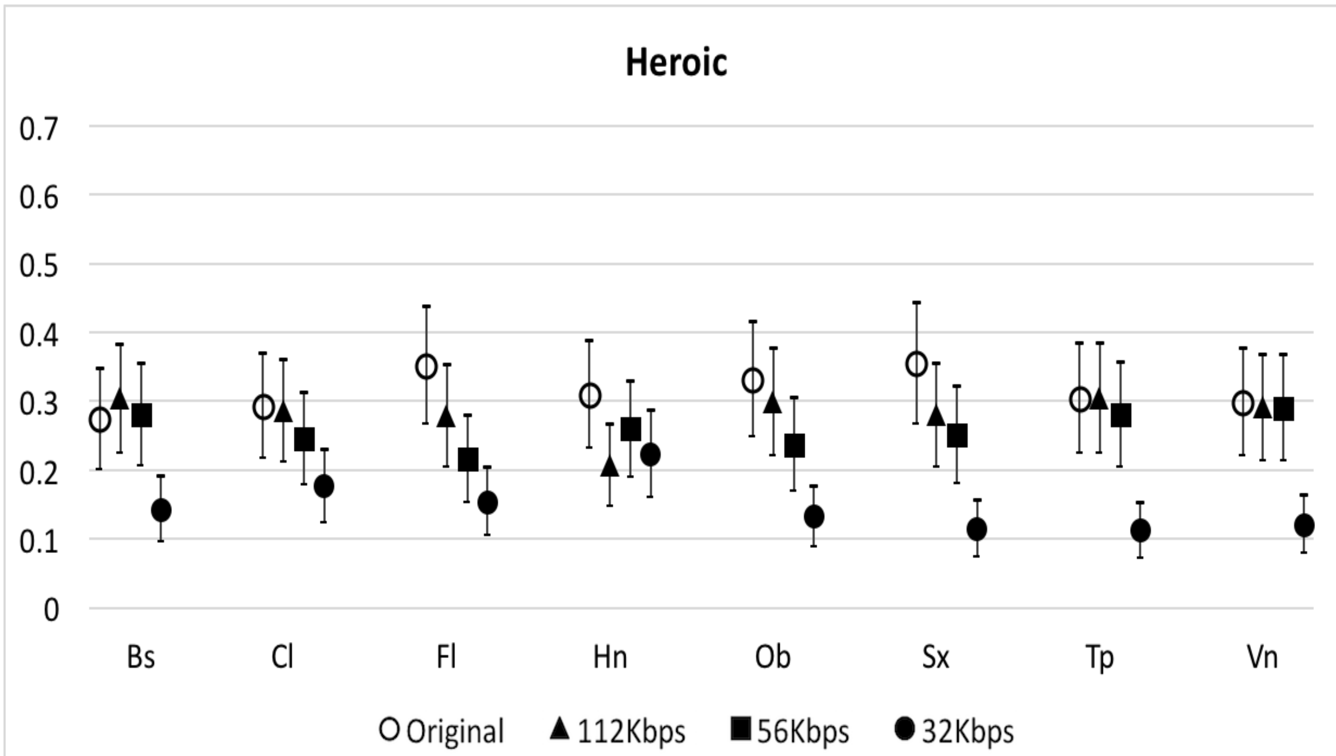

Figure 2.4: BTL scale values and the corresponding 95% confidence intervals for Heroic.

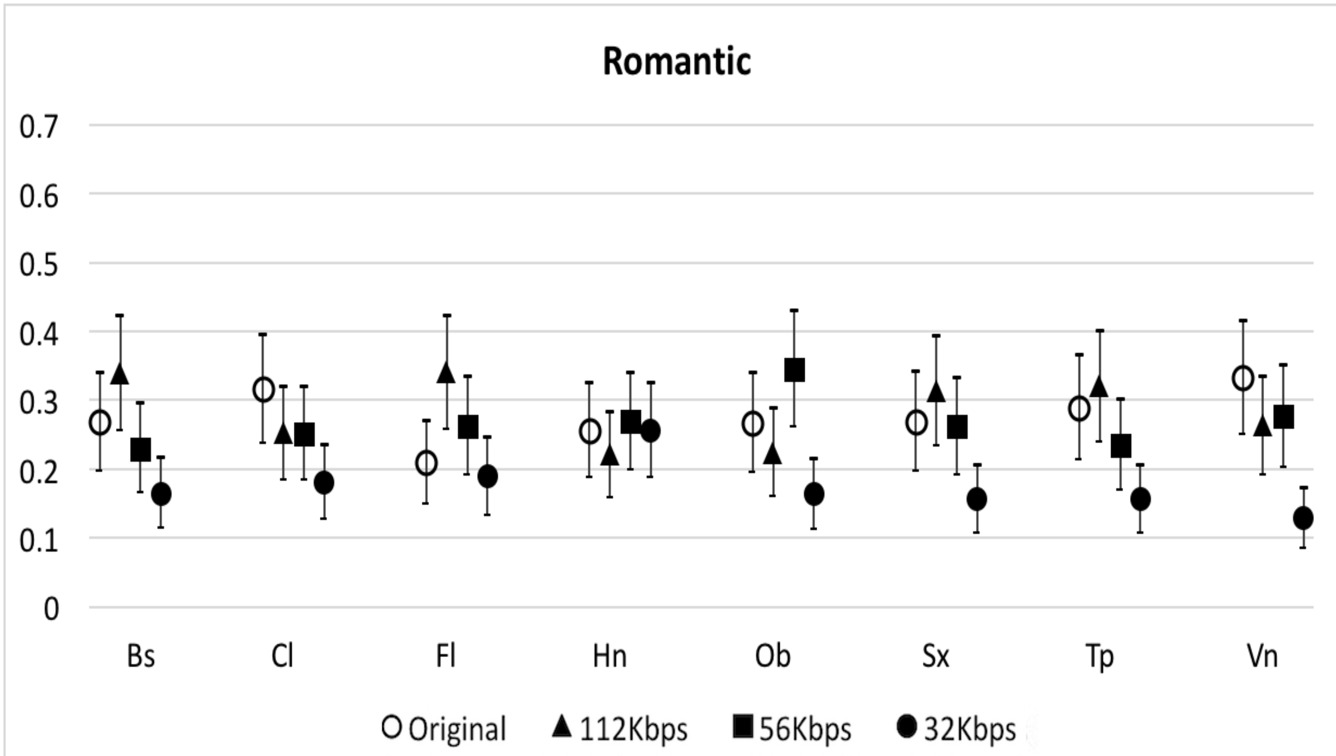

Figure 2.5: BTL scale values and the corresponding 95% confidence intervals for Romantic.

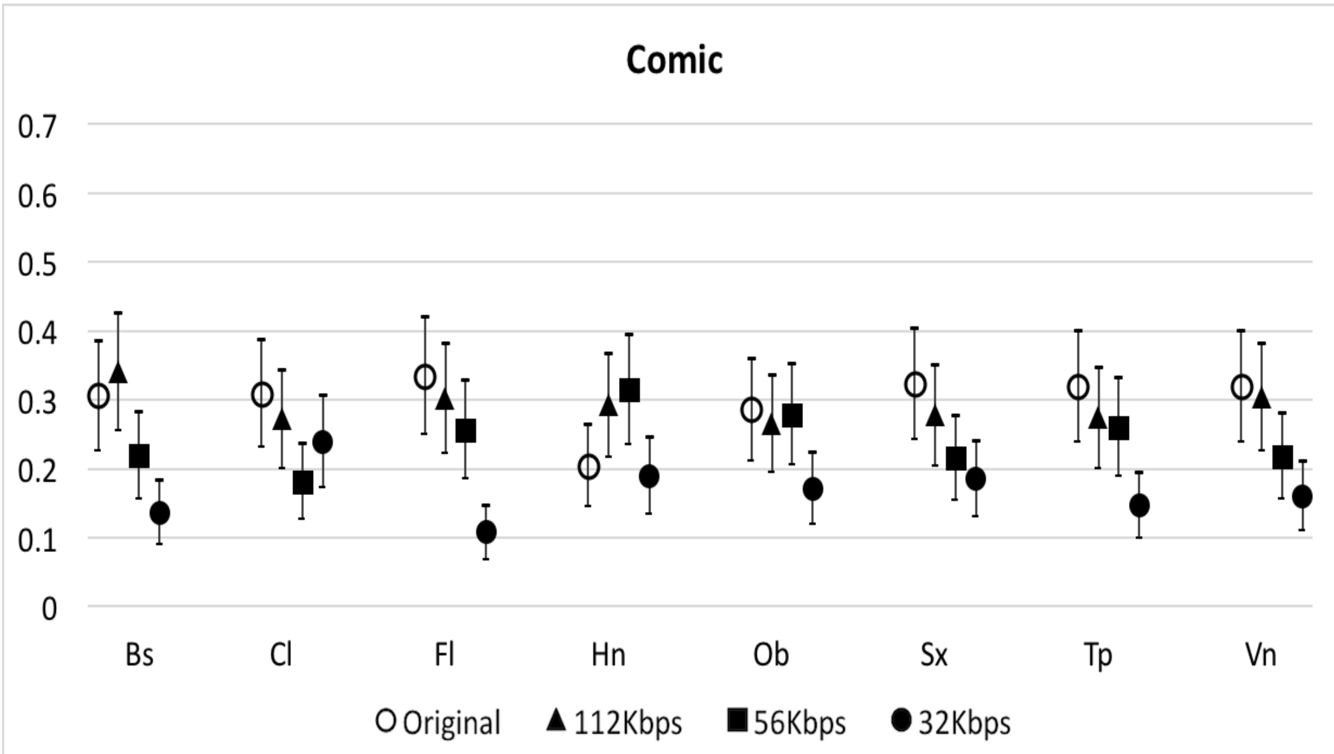

Figure 2.6: BTL scale values and the corresponding 95% confidence intervals for Comic.

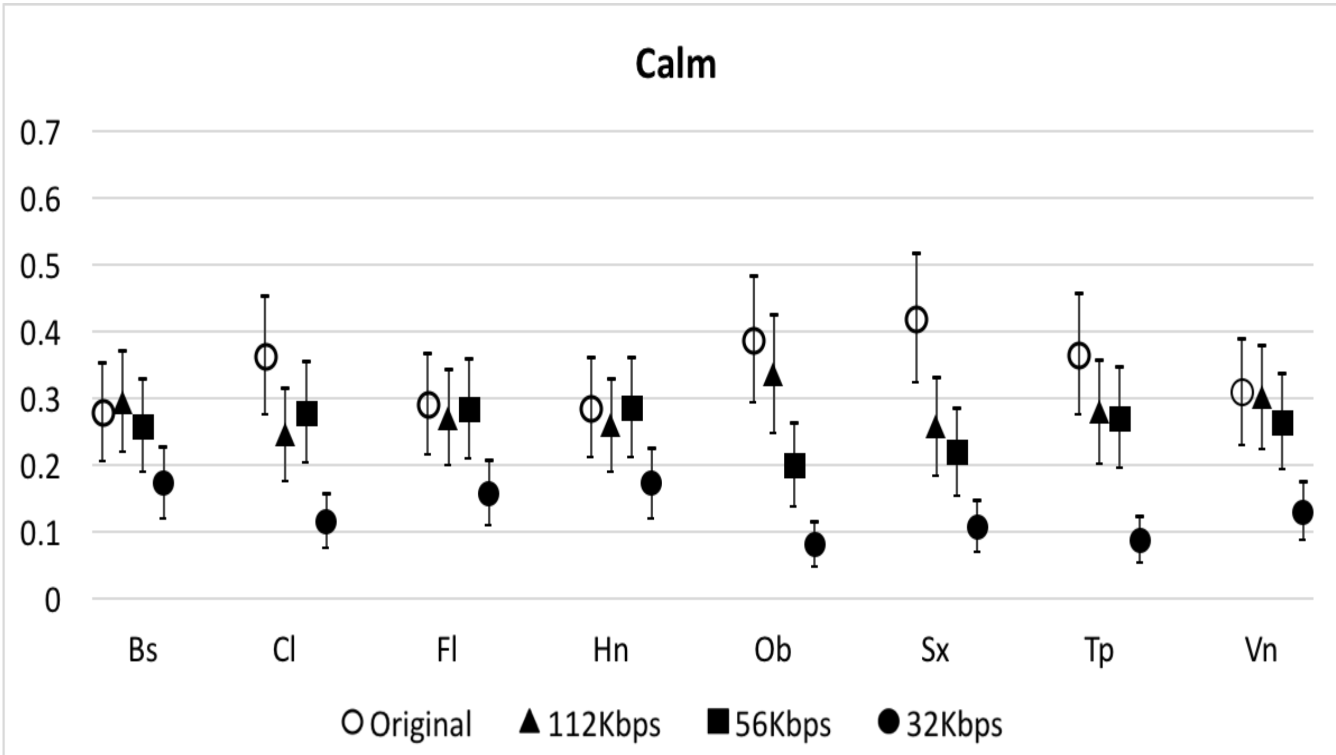

Figure 2.7: BTL scale values and the corresponding 95% confidence intervals for Calm.

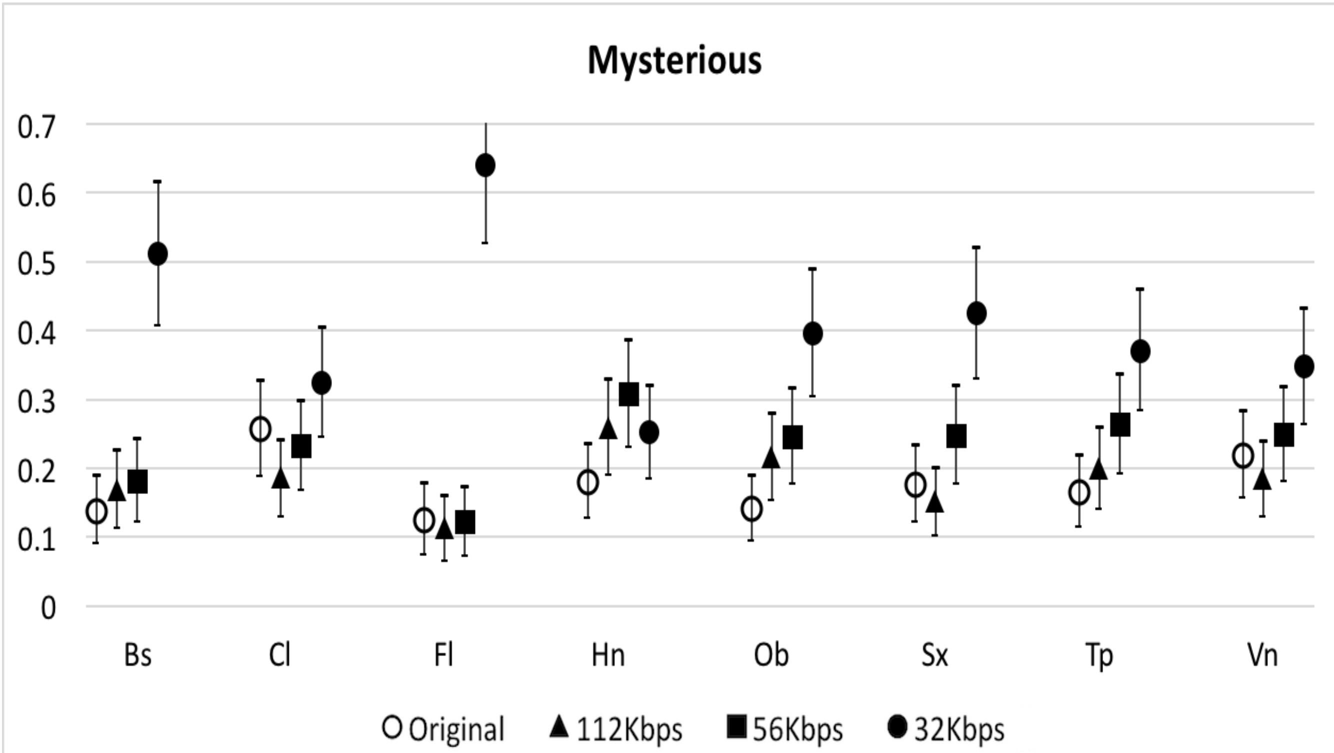

Figure 2.8: BTL scale values and the corresponding 95% confidence intervals for Mysterious.

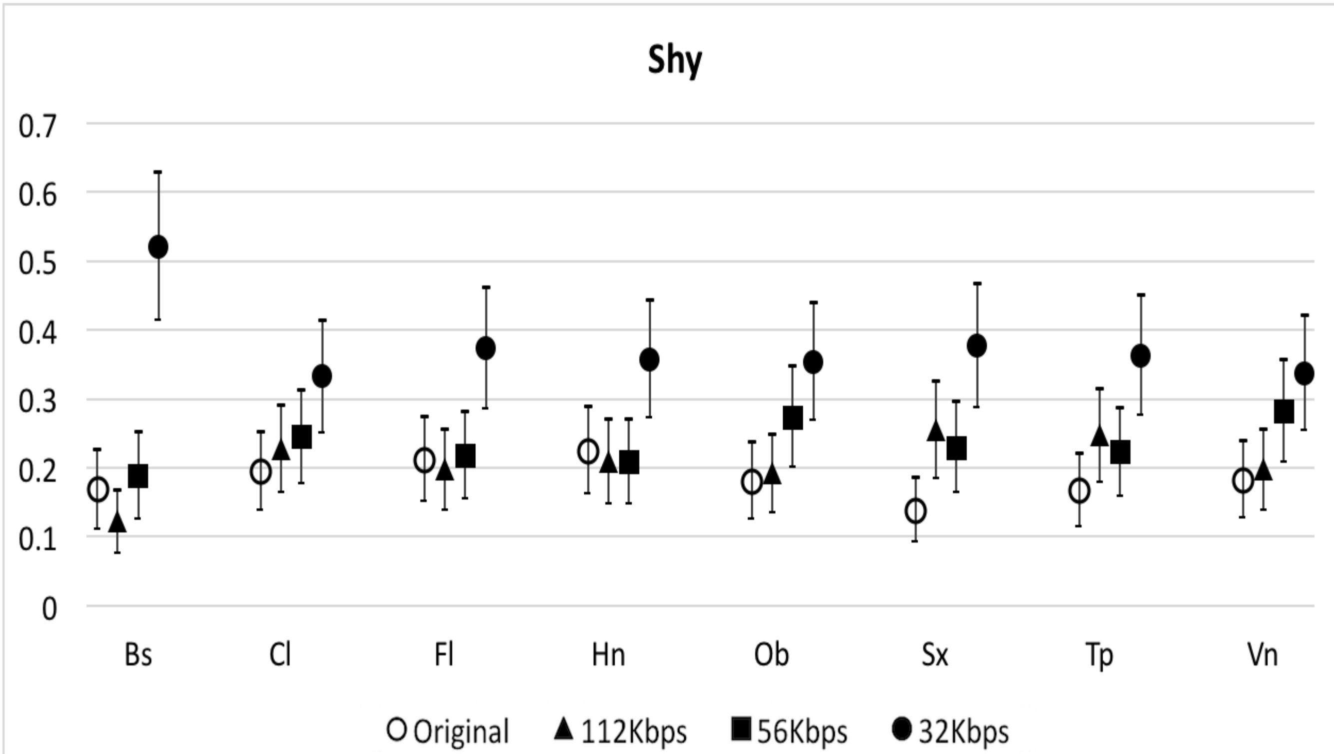

Figure 2.9: BTL scale values and the corresponding 95% confidence intervals for Shy.

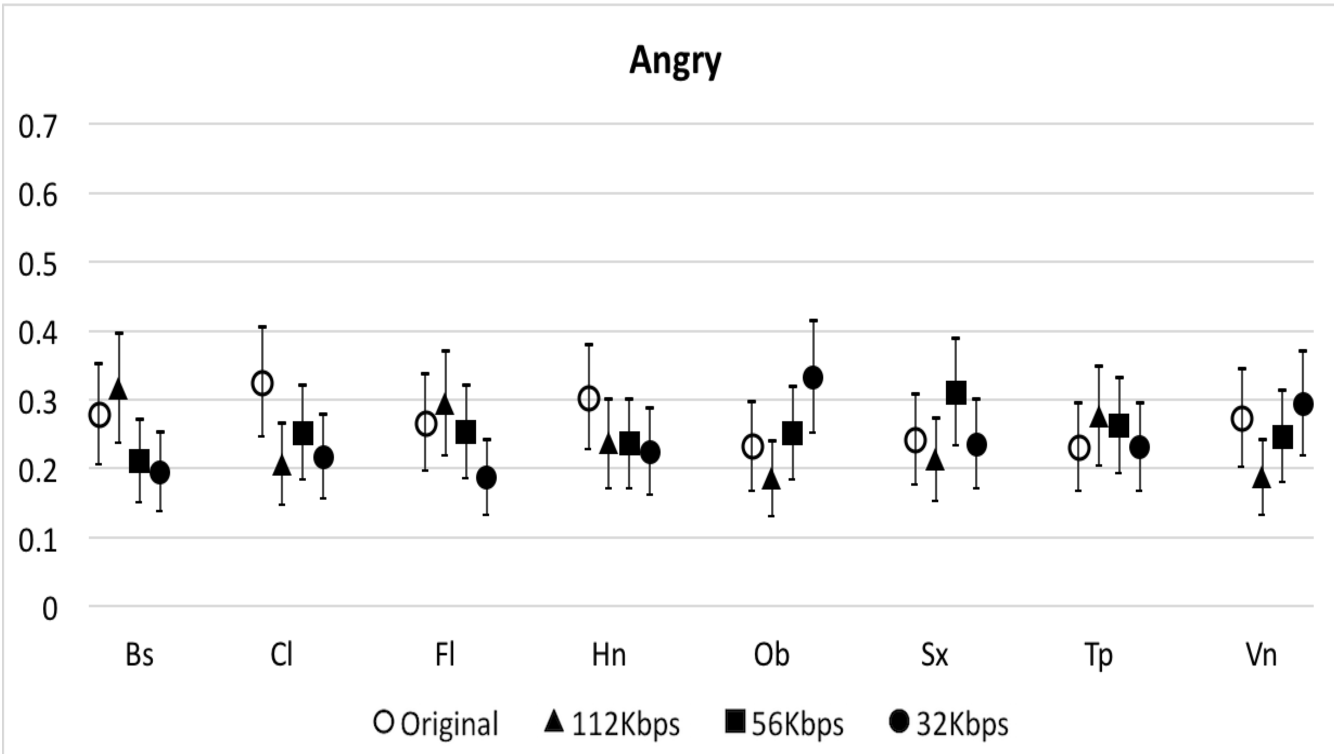

Figure 2.10: BTL scale values and the corresponding 95% confidence intervals for Angry.

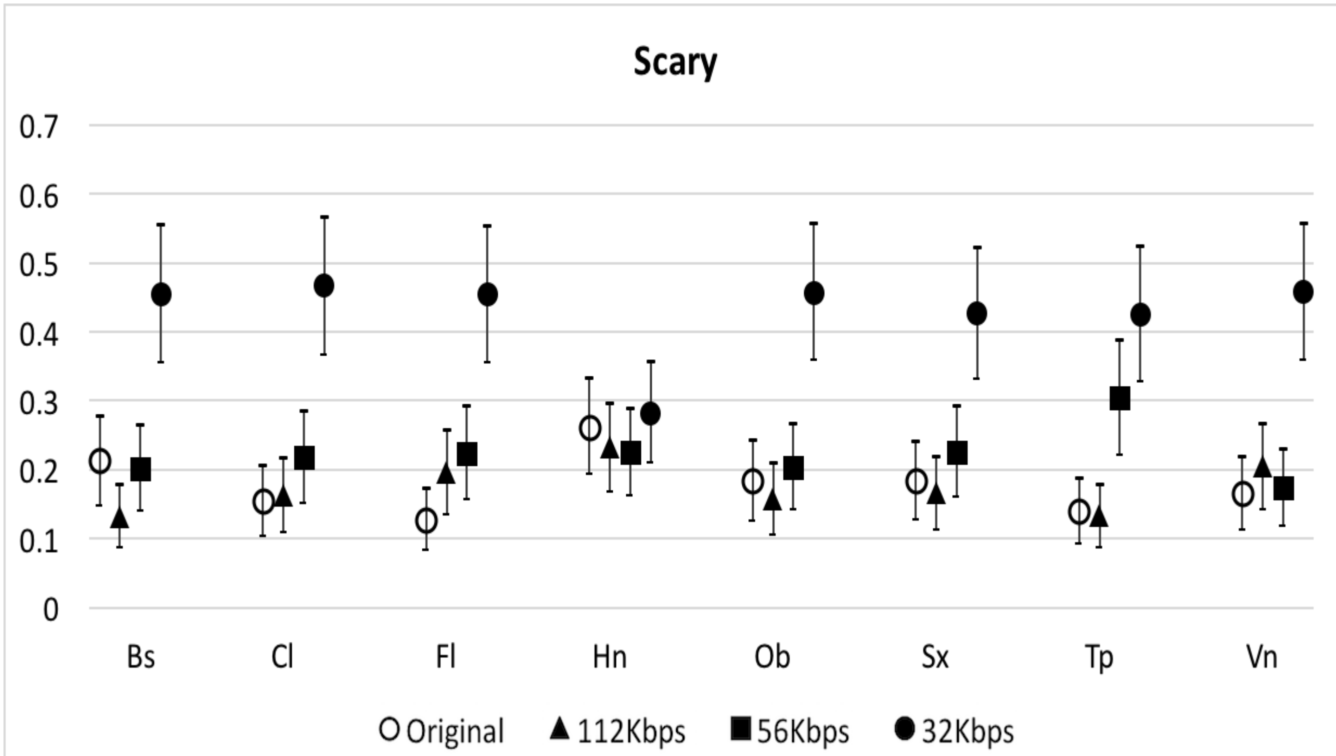

Figure 2.11: BTL scale values and the corresponding 95% confidence intervals for Scary.

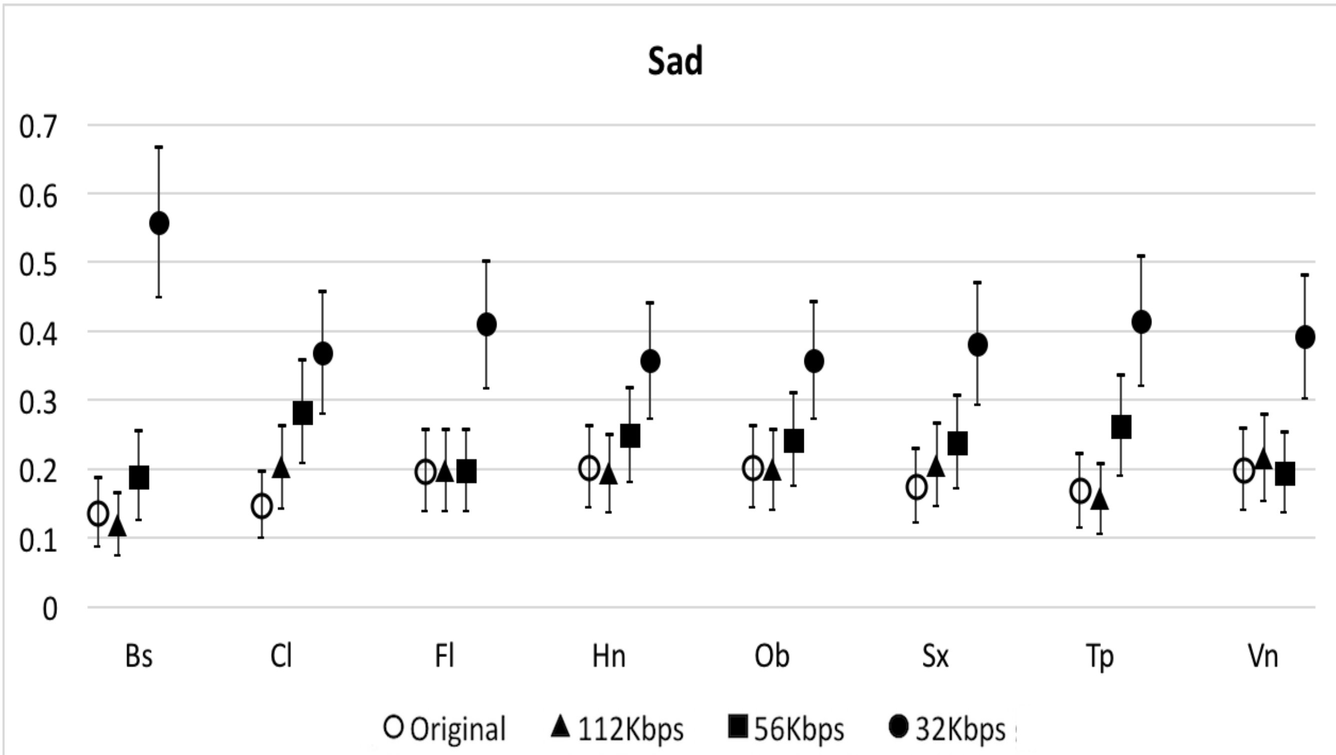

Figure 2.12: BTL scale values and the corresponding 95% confidence intervals for Sad.

Emotional Category	112 Kbps	56 Kbps	32 Kbps
Happy	1	3	6
Heroic	1	1	4
Romantic	1	0	2
Comic	0	1	3
Calm	1	2	5
Mysterious	0	3	5
Shy	1	0	2
Angry	0	0	0
Scary	1	2	6
Sad	0	1	3
Average over all emotional categories	0.6	1.3	3.6

Table 2.2: The number of instruments that were significantly different ($p < 0.05$) from the original sound for each compression rate and emotional category. The maximum for each entry is 8, since there were 8 instruments.

To understand which instruments and emotional categories were most and least affected by MP3 compression, Table 2.3 shows the number of compressed sounds that were significantly different from the original sound for each instrument and emotional category. Based on Table 2.3, the trumpet was the most effected instrument, while the horn was by far the least effected instrument. Lee et al. [49] also found the MP3-compressed horn the most difficult to discriminate compared to the other instruments. Among the emotional categories in Table 2.3, Happy was the most affected (though several other categories were close behind), and Angry was by far the least affected with no significant differences.

Emotional Category \ Instrument	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn	Sum of significant differences over all instruments
Happy	2	3	1	0	2	1	1	0	10
Heroic	0	0	1	1	1	1	1	1	6
Romantic	0	0	1	0	0	0	1	1	3
Comic	1	1	1	0	0	0	1	0	4
Calm	0	1	0	0	2	3	1	1	8
Mysterious	1	0	1	1	2	1	2	0	8
Shy	1	0	0	0	0	2	0	0	3
Angry	0	0	0	0	0	0	0	0	0
Scary	2	1	2	0	1	0	2	1	9
Sad	1	1	0	0	0	0	2	0	4
Sum of significant differences for all emotional categories	8	7	7	2	8	8	11	4	

Table 2.3: The number of compressed sounds that were significantly different from the original sound for each instrument and emotional category. The maximum for each entry is 3, corresponding to the 3 compression rates under consideration.

To see which emotional categories were strengthen or weakened by MP3 compression, Figure 2.13 shows how often the original instruments sounds were statistically significantly greater than the three compressed sounds. The values in Figure 2.13 are different from the sum in the final column of Table 2.3 which counts any significant difference as +1 for both those significantly greater and those significantly less. In Figure 2.13, when a compressed sound is significantly greater than the original sound it is counted as +1, and when a compression sound is significantly less then the original sound it is counted as -1. So, sometimes these cancel. Therefore, a positive value indicates an increase in an emotional characteristic, and a negative value a decrease. Again, Happy was the most affected emotional category, and Angry the least. Emotional categories with larger Valence (e.g., Happy, Heroic, Romantic, Comic, Calm) tended to decrease with more MP3 compression, while emotional categories with smaller Valence (e.g., Mysterious, Shy, Scary, and Sad) tended to increase with more MP3 compression. As an exception, Angry was relatively unaffected by MP3 compression for the compression rates we tested.

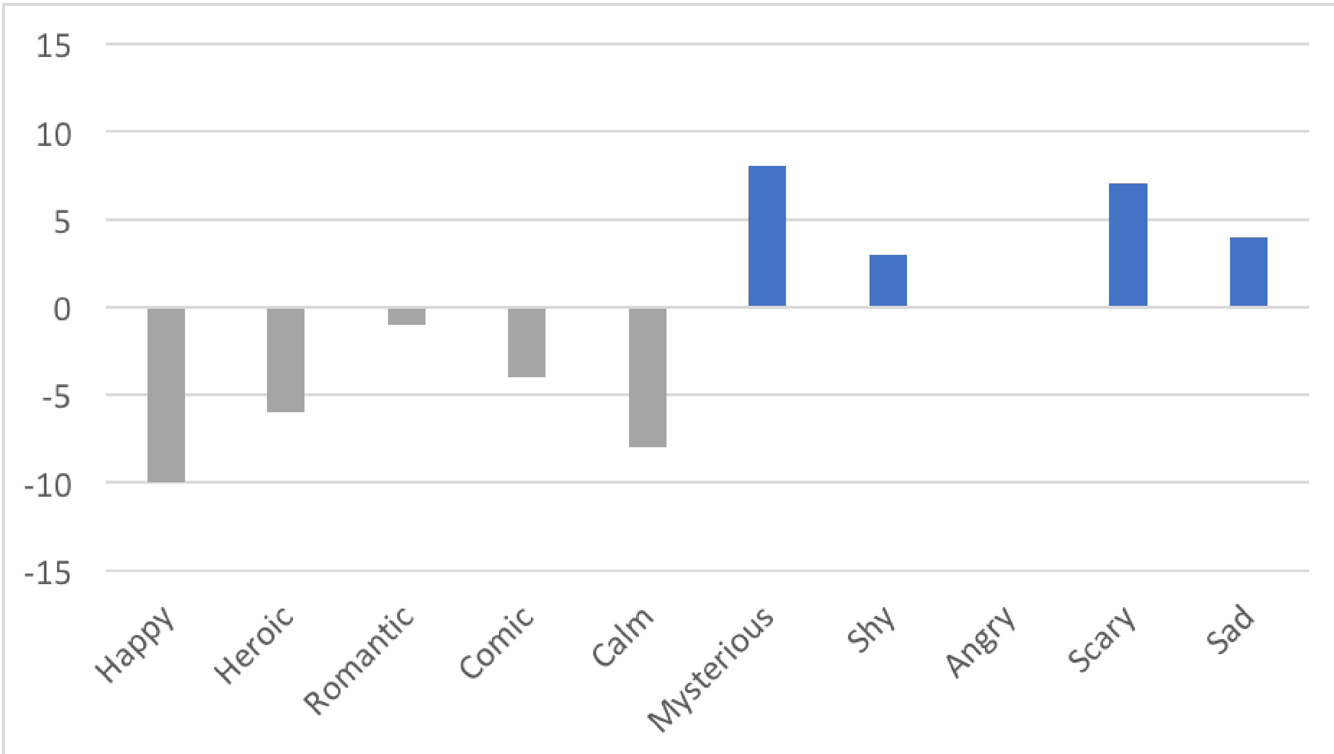

Figure 2.13: The number of significant differences between the original and compressed sounds, where strengthened emotional categories are positive, and weakened emotional categories are negative.

2.6 Discussion

The goal of our work was to understand how emotional characteristics of instruments vary with MP3 compression. Based on Table 2.3 and Figure 2.13, our main findings are as follows:

1. Neutral and negative emotional characteristics (Mysterious, Shy, Scary, and Sad) increased with more MP3 compression in the samples we tested (see Figure 2.13).
2. Positive emotional characteristics (Happy, Heroic, Romantic, Comic, and Calm) decreased with more MP3 compression in the samples we tested (see Figure 2.13).
3. Angry was relatively unaffected by MP3 compression for the rates we tested (see Figure 2.13).
4. MP3 compression effected some instruments more and others less. The trumpet was the most effected, and the horn by far the least (see Table 2.3).

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.

As a possible explanation for these results, perhaps quantization jitter introduced into the amplitude envelopes by MP3 compression decreased positive emotional characteristics such as Happy, Heroic, Romantic, Comic, and Calm while increasing others such as Mysterious and Scary by changing the quality of the sounds to be somewhat different and unnatural. Lee et al. [50] previously noted an increase in spectral incoherence in MP3-compressed sounds, and attributed it to quantization jitter in the amplitude envelopes. More jitter makes the spectrum more dynamic, thus increasing spectral incoherence. Indeed, the artifacts introduced by MP3 compression in these instruments, especially at 32 Kbps, added an audible background “growl” to the sounds, so it is easy to imagine why listeners perceived them as more Mysterious or Scary.

Among instruments, the horn was by far the least effected. This suggests that the horn is less sensitive to deterioration caused by MP3 compression. This makes sense since the original horn had much less spectral incoherence than the other instruments

[49], so quantization jitter had less impact on its already smooth amplitude envelopes. Conversely, the trumpet was the most effected instrument, and it had about the highest level of spectral incoherence. The dynamic spectra of the sound seems to have accentuated the added variations of quantization jitter.

However, we should once again emphasize our results are for basic-level professional headphones. We informally compared the compressed sounds on higher-quality professional headphones (Sennheiser HD25-1 and AKG K240 MKII) where the level of detail in the artifacts and instrument sounds were more than in the Sony's. We also compared the sounds on standard iPhone earphones, where the level of detail in the artifacts was similar to the Sony's though not as good for listening to the sounds in isolation.

It was interesting that though Scary and Angry are very close to each one another in terms of Valence and Arousal (see Figure 2.1), yet Scary significantly increased with more compression while Angry was relatively unaffected. The results indicate that they were interpreted as distinctively different emotional characteristics by listeners.

CHAPTER 3

EXPERIMENT 2: THE EFFECTS OF CONVOLUTION REVERBERATION ON THE EMOTIONAL CHARACTERISTICS OF MUSICAL INSTRUMENT SOUNDS

3.1 Overview

In this study, we seek to understand how the emotional characteristics of musical instruments vary with reverberation time in convolution reverberation. In our previous study on parametric reverberation, we found strong effects for Romantic and Mysterious, medium effects for Sad, Scary, and Heroic, mild effects for Happy, little effect on Shy, and the opposite effect for Comic. We are interested to determine whether this pattern basically holds for convolution reverberation as well, or if another pattern emerges.

To easily compare and contrast the convolution and parametric reverberation results, we conducted a listening test for convolution reverberation in the same way that we did for parametric reverberation. Listeners compared the convolution reverberations pairwise for each instrument and emotional characteristic. Below are some of the main points, especially the differences from the parametric reverberation tests.

3.2 Listening Test

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* [55]. 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 [70]. Attacks, sustains, and decays were equalized by time-compression/expansion of the amplitude envelopes to 0.05s, 0.8s, and 0.15s respectively, for a total duration of 1.0s. 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 in loudness by manual adjustment.

In addition to the anechoic sounds, we compared sounds with reverberation lengths of approximately 1s and 2s, which according to Hidaka and Beranek [71] and Beranek [72] typically correspond to small and large concert halls. To do this, we selected several representative hall convolution reverberations based on the impulse responses in Altiverb [73]. We measured their reverberation lengths based on their reverberation time RT_{60}, and picked those that most closely matched the reverberation times we tested in our previous study of parametric reverberation. Table 3.1 shows the parametric and convolution reverberation RT_{60} values and other parameters. We also included the cathedral impulse response of King's College Chapel with a 5.44 second reverberation time to determine the effects for a more extreme case. Figures 3.5 to 3.9 show the energy decay curve for the different halls we tested, along with their RT_{60} and EDT values. The convolution reverberation decays were basically linear from the very beginning, while the parametric reverberation decays had an immediate drop-off and shelf at -10 to -30 dB that lasted about 0.25 seconds before the linear portion of the decay commenced (see Figures 3.1 - 3.4). This reflects the presence of finely detailed and smoothly decaying early reflections in the convolution reverberation responses compared to the parametric reverberation responses.

Parametric				
Hall Type and Position	Reverb Length	Reverb Amount	RT_{60}	EDT
Small Hall Front	1s	20%	0.95	0
Small Hall Back	1s	80%	1.28	0
Large Hall Front	2s	20%	1.78	0
Large Hall Back	2s	80%	2.37	0
Convolution				
Hall Impulse Response		Distance	RT_{60}	EDT
Royal National Theatre		8.9m	0.94	0.96
Empire Hall		10.1m	1.31	1.56
Disney Hall		14.1m	1.80	1.62
Concertgebouw Hall		14m	2.32	2.58
King's College Chapel		20.2m	5.44	5.4

Table 3.1: Parametric and convolution reverberation parameters

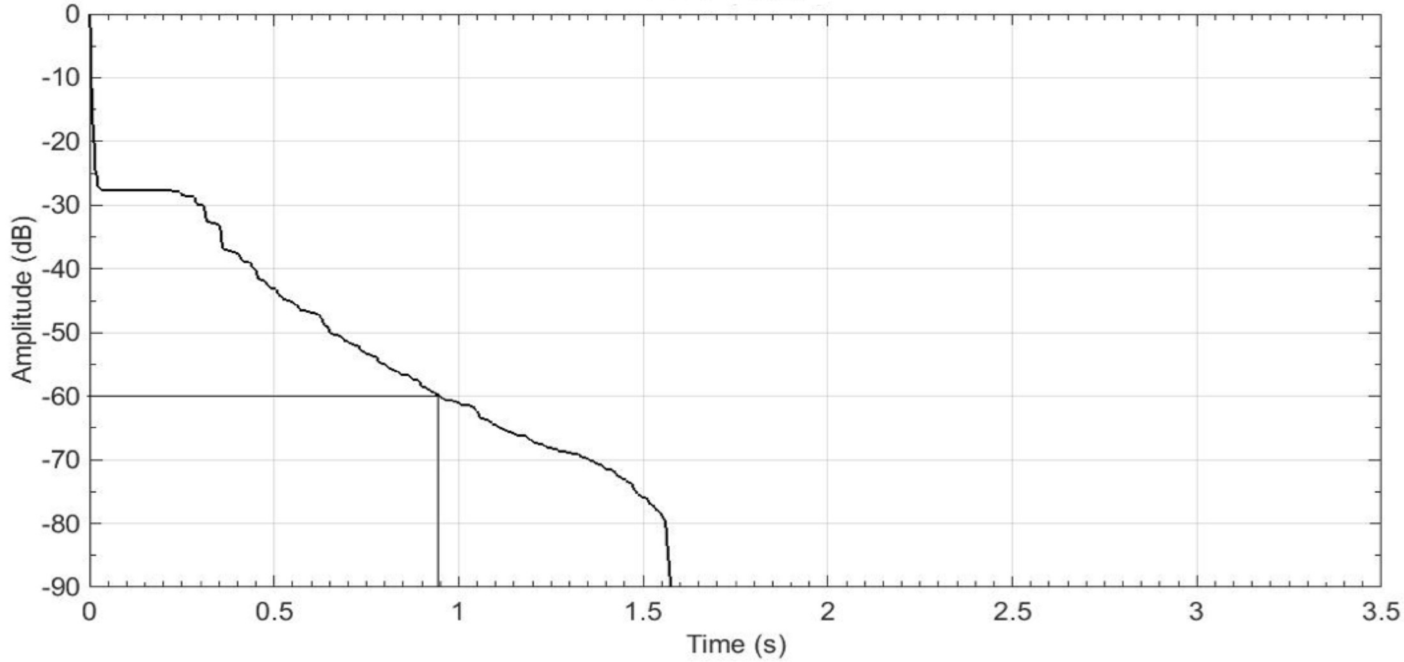

Figure 3.1: Energy decay curve and RT_{60} for Small Hall Front.

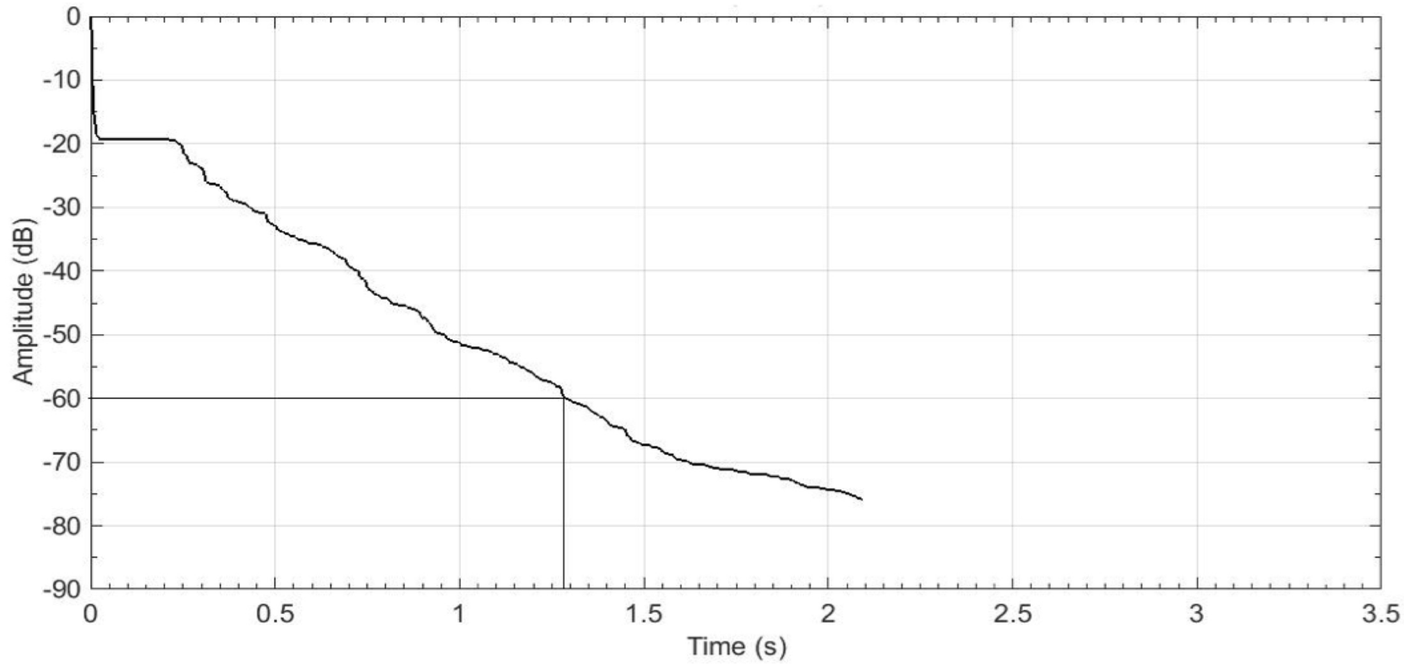

Figure 3.2: Energy decay curve and RT_{60} for Small Hall Back.

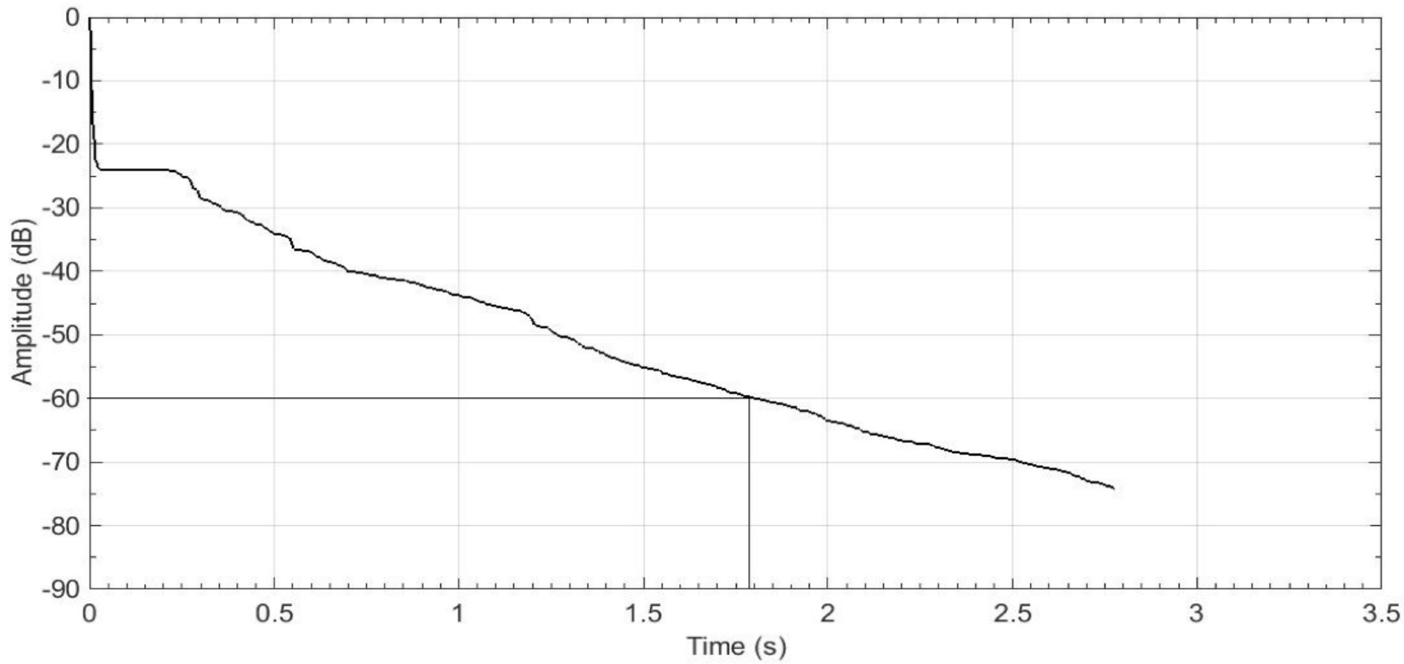

Figure 3.3: Energy decay curve and RT_{60} for Large Hall Front.

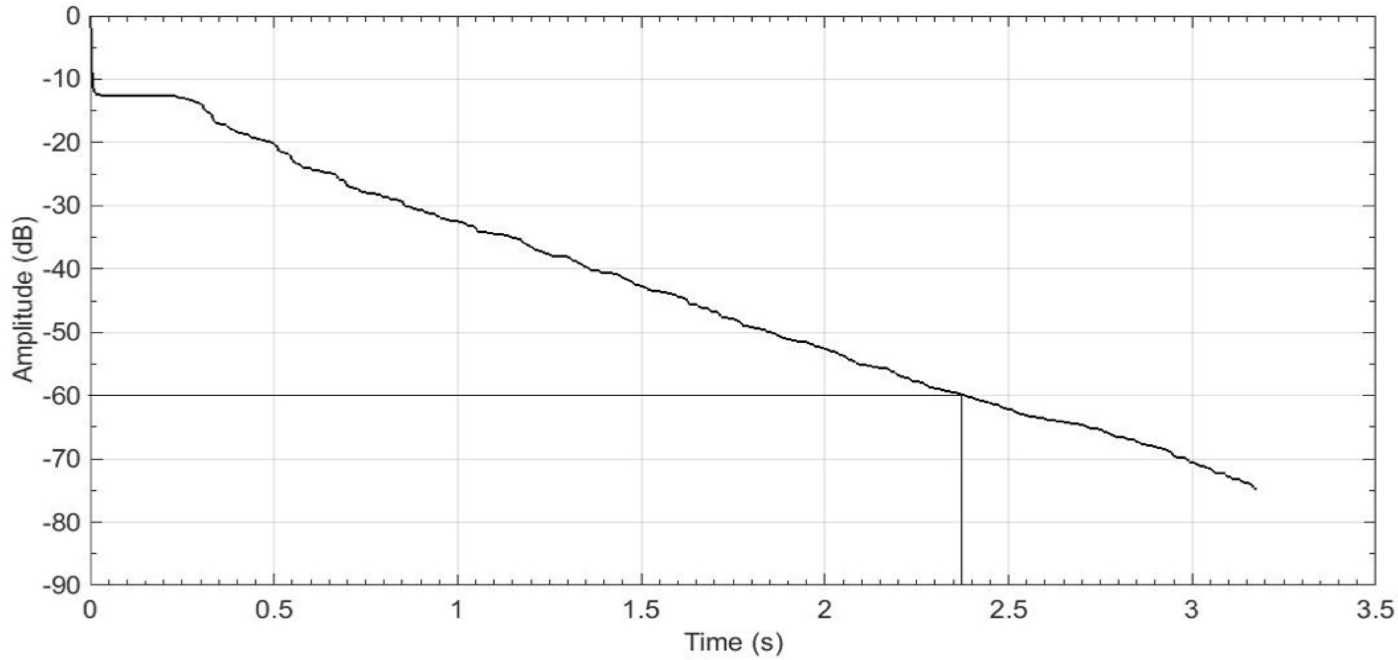

Figure 3.4: Energy decay curve and RT_{60} for Large Hall Back.

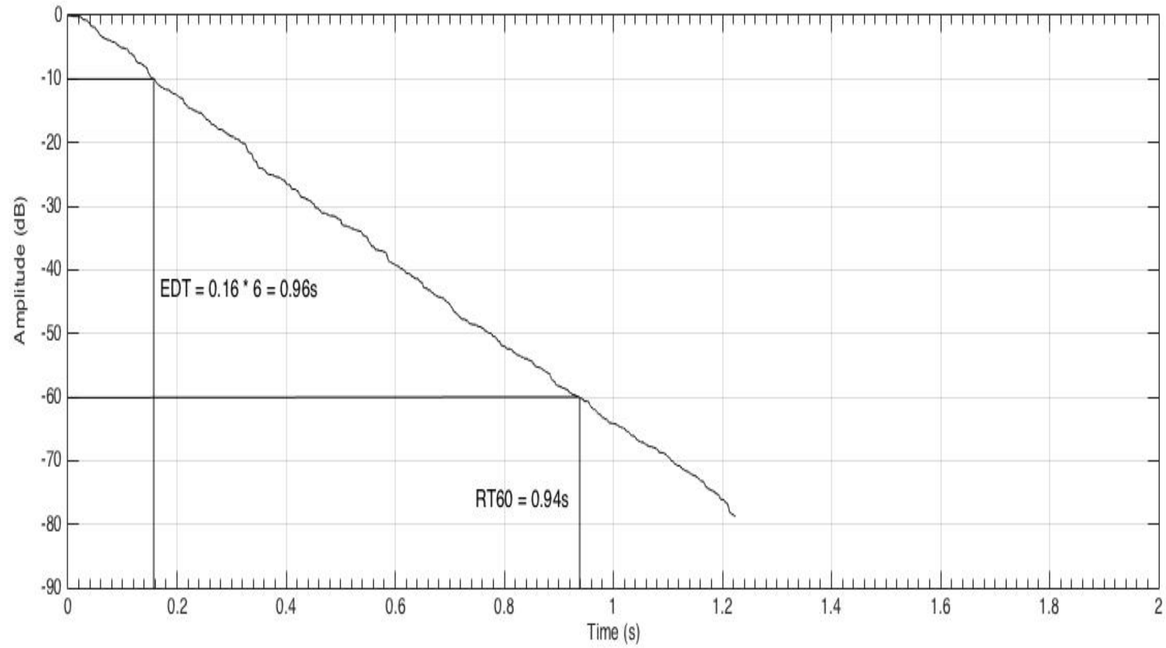

Figure 3.5: Energy decay curve and RT_{60} for Royal National Theatre.

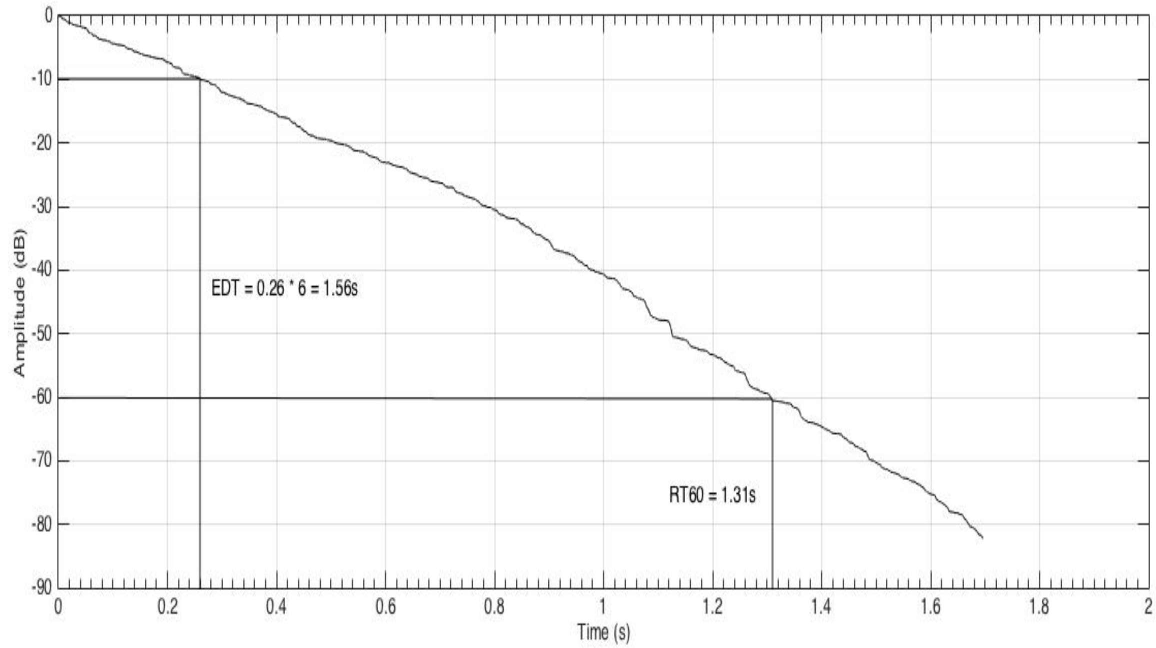

Figure 3.6: Energy decay curve and RT_{60} for Empire Hall.

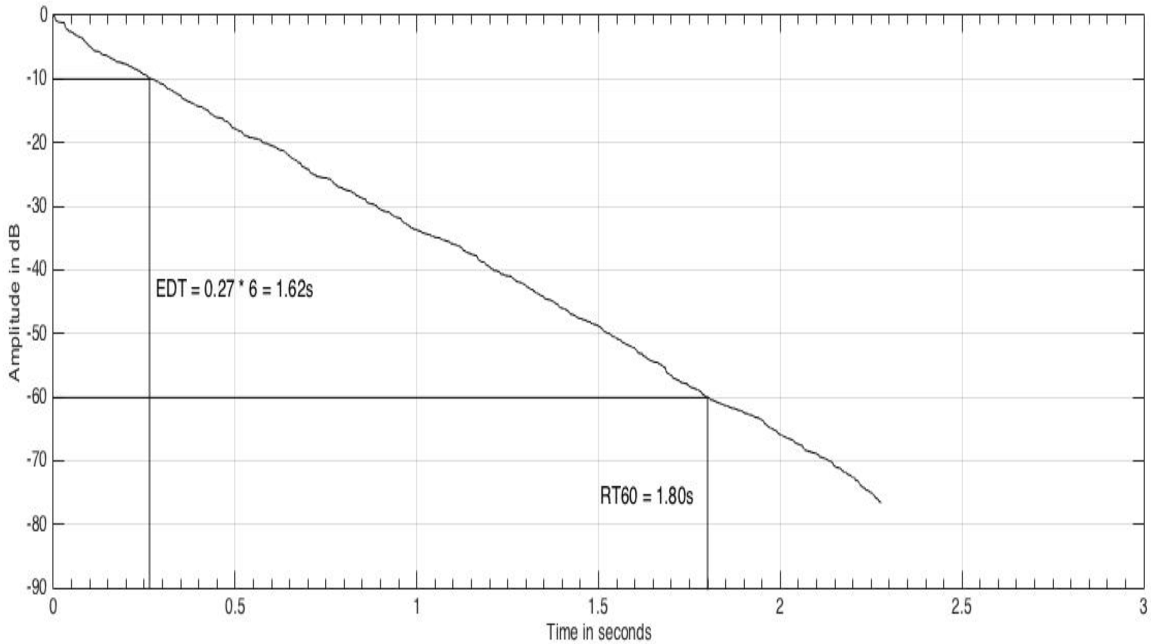

Figure 3.7: Energy decay curve and RT_{60} for Disney Hall.

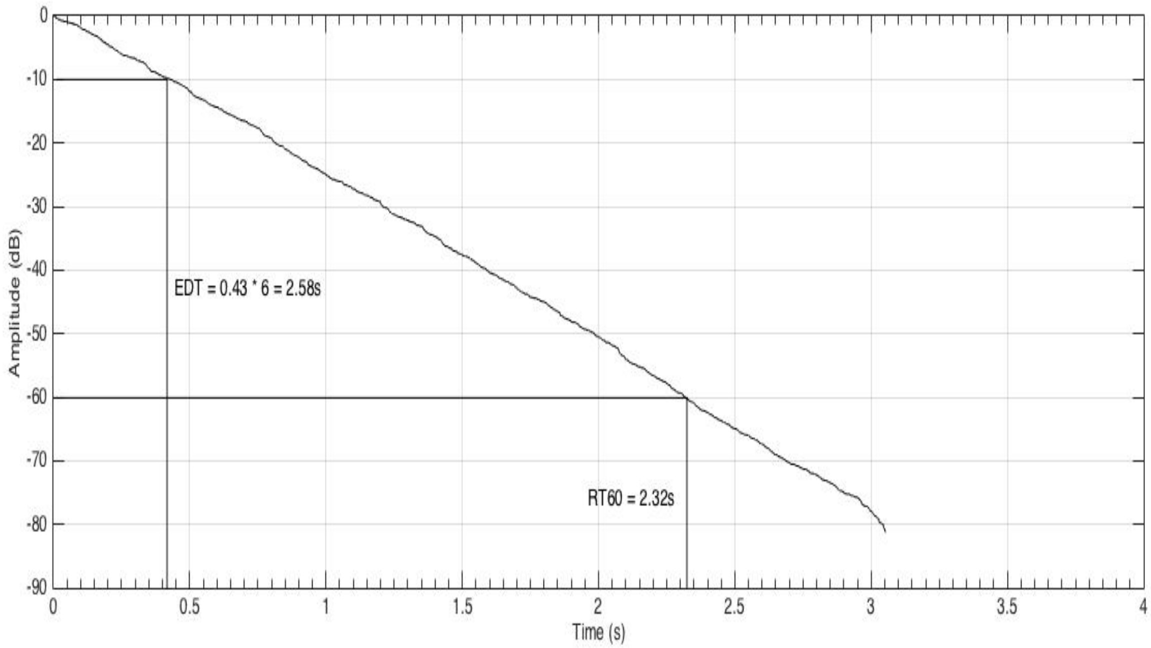

Figure 3.8: Energy decay curve and RT_{60} for Concertgebouw Hall.

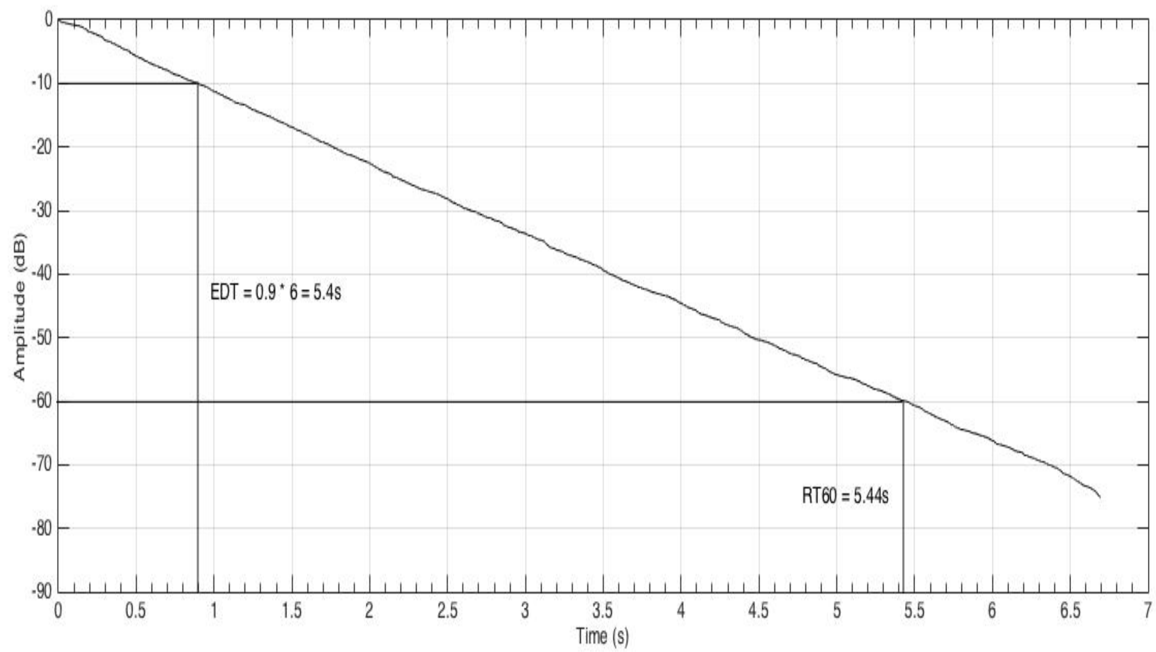

Figure 3.9: Energy decay curve and RT_{60} for King's College Chapel.

In addition to the eight emotional categories we tested (e.g., Comic, Happy, Heroic, Mysterious, Romantic, Sad, Scary, and Shy) in the parametric reverberation test, for this study we picked two extra emotional categories, namely Angry and Calm. These categories were actually used in our previous research on musical instruments and parametric reverberation [21, 20, 22, 19, 23, 24, 32, 74, 51, 75, 76, 33, 38, 77, 78, 79]. Readers may refer to Figure 2.1 for the ratings of the emotional categories according to the Affective Norms for English Words [65]. Just before the listening test, subjects read online definitions of the emotional categories used in this experiment, which were taken from the Cambridge Academic Content Dictionary [66], and were shown in Table 2.1 in Section 2.4.

To limit the length of the test and to minimize listener fatigue, we divided listeners into two groups, where each group heard a different set of five emotional categories. We recruited 72 subjects to take the listening test, and since they each heard half the emotional categories, 36 subjects compared each emotional category. 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. Subjects were not musical experts (e.g., recording engineers, professional musicians, or music conservatory students) but average attentive listeners. Among the 72 subjects, there were 51 males and 21 females. The subjects ranged in age from 18 to 24. In terms of musical experience, 40 subjects had some experience playing an instrument (an average of 6.3 years), and 32 subjects did not have experience playing an instrument. In recruiting the subjects, all indicated they had no known hearing problems.

In this listening tests, subjects heard paired comparisons between the hall impulse responses for the each instrument and emotional category. During each trial, subjects heard a pair of instrument sounds from the same type of reverberation and were prompted to choose which more strongly aroused a given emotional category. Since each trial was a single paired comparison requiring minimal memory from the subjects, subjects did not need to remember all of the tones, just the two in each comparison. Figure 3.10 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.

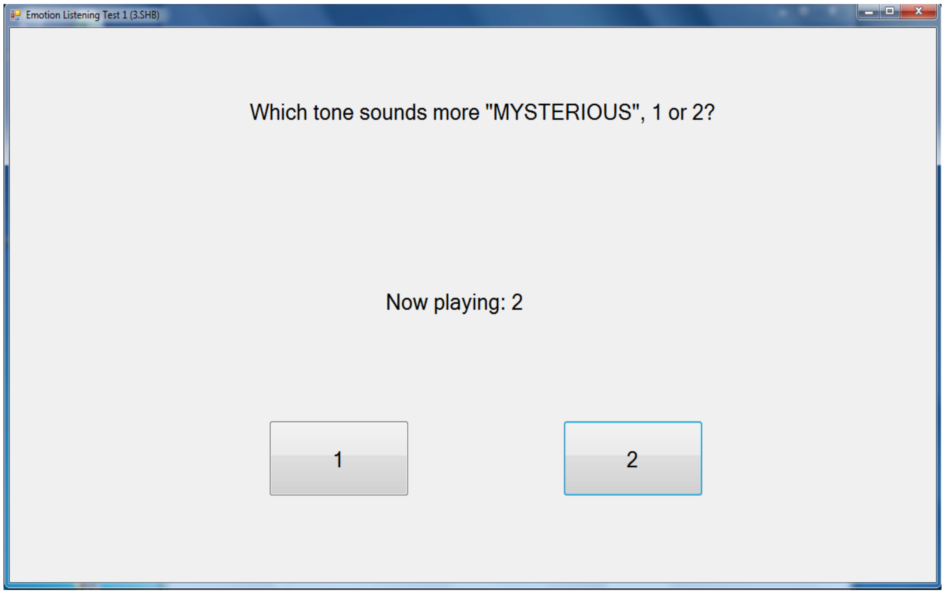

Figure 3.10: Paired comparison listening test interface.

Each combination of two different hall impulse responses were presented for each of the eight instruments and ten emotional categories, and the complete listening test totaled $C_2^6 \times 8 \times 10 = 1200$ trials (600 trials per listener since we divided the task into two groups). For each listener, the overall trial presentation order was randomized to average out effects due to learning or fatigue. For the two sounds A and B, they heard AB where the order of A and B was random for each comparison (but if they heard AB, they did not hear BA later). The listening test took about 75 minutes, with forced short breaks every 25 minutes. The subjects were seated in the same "quiet room" as we specified in Section 2.4.

3.3 Results

For our listening test, subjects compared each pair of hall impulse responses for each instrument and emotional category. Originally, there were 36 subjects for each group of five emotional categories. We screened their responses, and found 7 subjects from group A and 6 subjects from group B were obviously spamming the same key responses toward the end of the test, so we excluded all of their data. We scanned the remaining subjects' data, especially at the end of the test, and based on the consistency of their responses, felt that they were giving sincere and attentive responses to the questions, so we did not exclude any further subjects.

Based on the filtered listening test data (29 subjects for group A, and 30 for group B), we ranked the hall impulse responses 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 [68, 69]. Figures 3.11 to 3.20 show BTL scale values and the corresponding 95% confidence intervals for each emotional category and instrument. For each instrument in each graph, the BTL scale values for the six hall impulse responses (Anechoic, Royal National Theatre, Empire Hall, Disney Hall, Concertgebouw Hall, and King's College Chapel respectively) sum to 1. The BTL value for each hall impulse response is the probability that listeners will choose that hall impulse response when considering a certain instrument and emotional category. For example, if all six hall impulse responses were judged equally Happy, the BTL scale values would all be $1/6 \approx 0.167$.

Though there are certainly individual differences, for each emotional category the trend from anechoic to cathedral usually follows the same direction for the different instruments. For example, Angry trends down, Calm trends up, and Shy arches up to the Disney Hall and then down. The trumpet for Mysterious was the most strongly effected among all the instruments and emotional categories with a BTL value of more than 0.5 for King's College Chapel.

We also wanted to determine the number of times each hall impulse response was significantly greater than the other five hall impulse responses over the eight instruments for each emotional characteristic. As a preliminary step, the normality of the data was calculated for each hall impulse response, instrument, and emotional characteristic. Since most of them were not normally distributed (see Tables A.1 - A.6 in Appendix A), both parametric and nonparametric statistical tests (parametric: Paired t-tests, Pearson correlation ; nonparametric: Wilcoxon signed-rank tests, Spearman correlation) were used to analyze the voting data (i.e., the number of positive votes received by each hall impulse response for each instrument and emotional category). The results from the two tests showed some minor differences, but basically they were in agreement. Table 3.2 shows the paired t-test results and Table B.1 in Appendix B shows the Wilcoxon signed-rank test results. For each hall impulse response, the maximum possible value is 40 and the minimum possible value is 0. For example, with the emotional characteristic Mysterious, the value of the King's College Chapel is 34 in Table 3.2 since it was statistically significantly greater than all the other hall impulse responses for Mysterious in Figure 3.16 except for the Concertgebouw Hall in 6 instances. The maximum value for each emotional characteristic is shown in bold and shaded in both tables, which are in agreement.

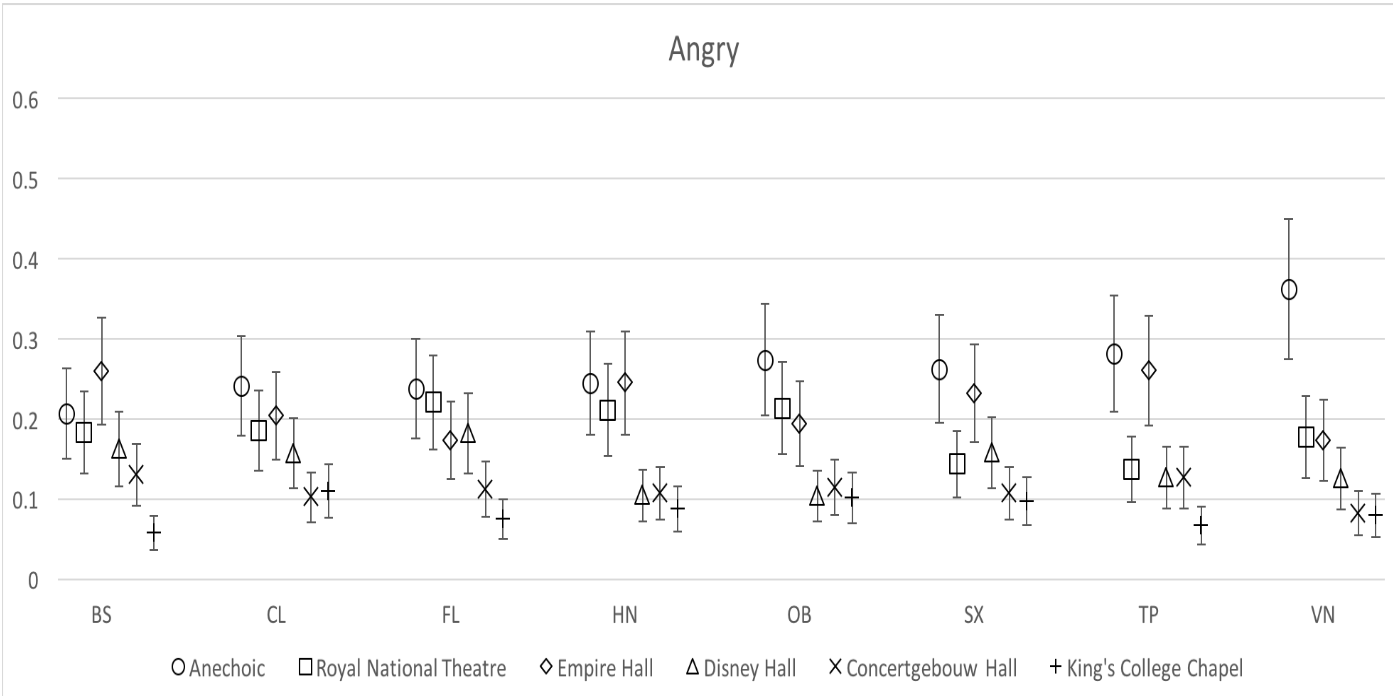

Figure 3.11: BTL scale values and the corresponding 95% confidence intervals for the emotional category Angry.

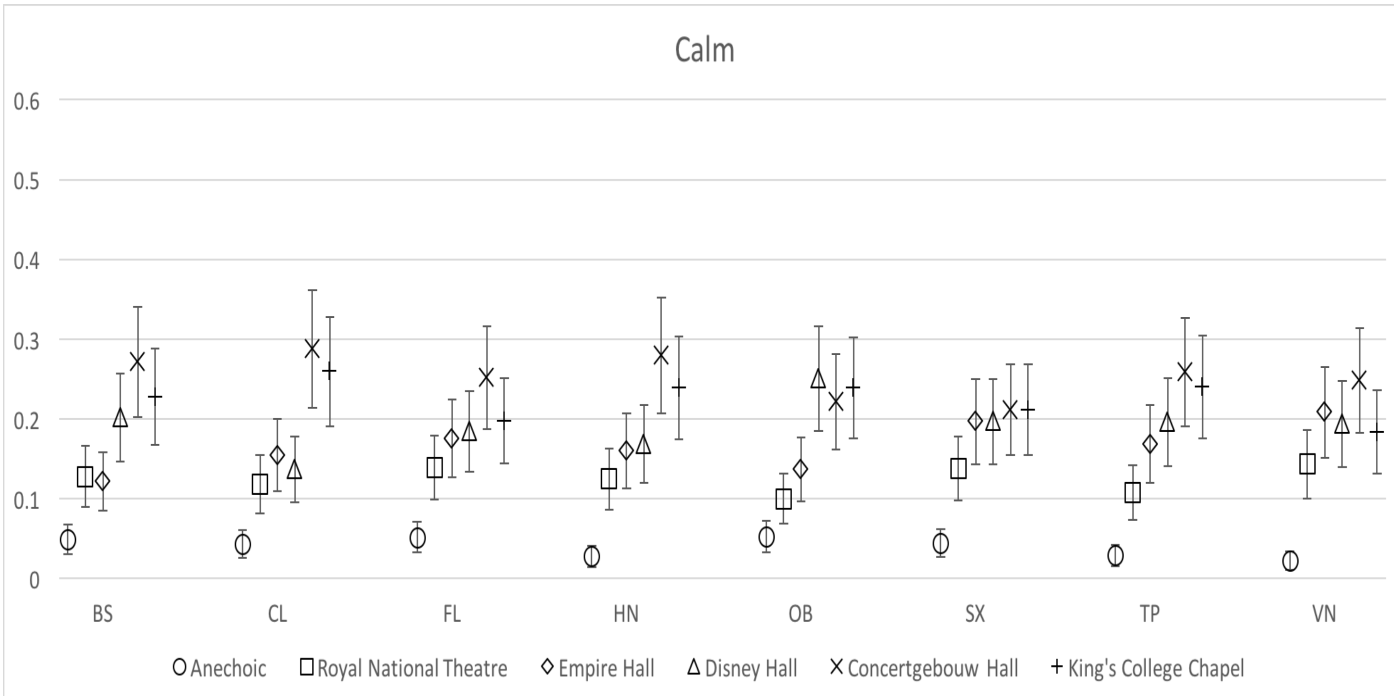

Figure 3.12: BTL scale values and the corresponding 95% confidence intervals for the emotional category Calm.

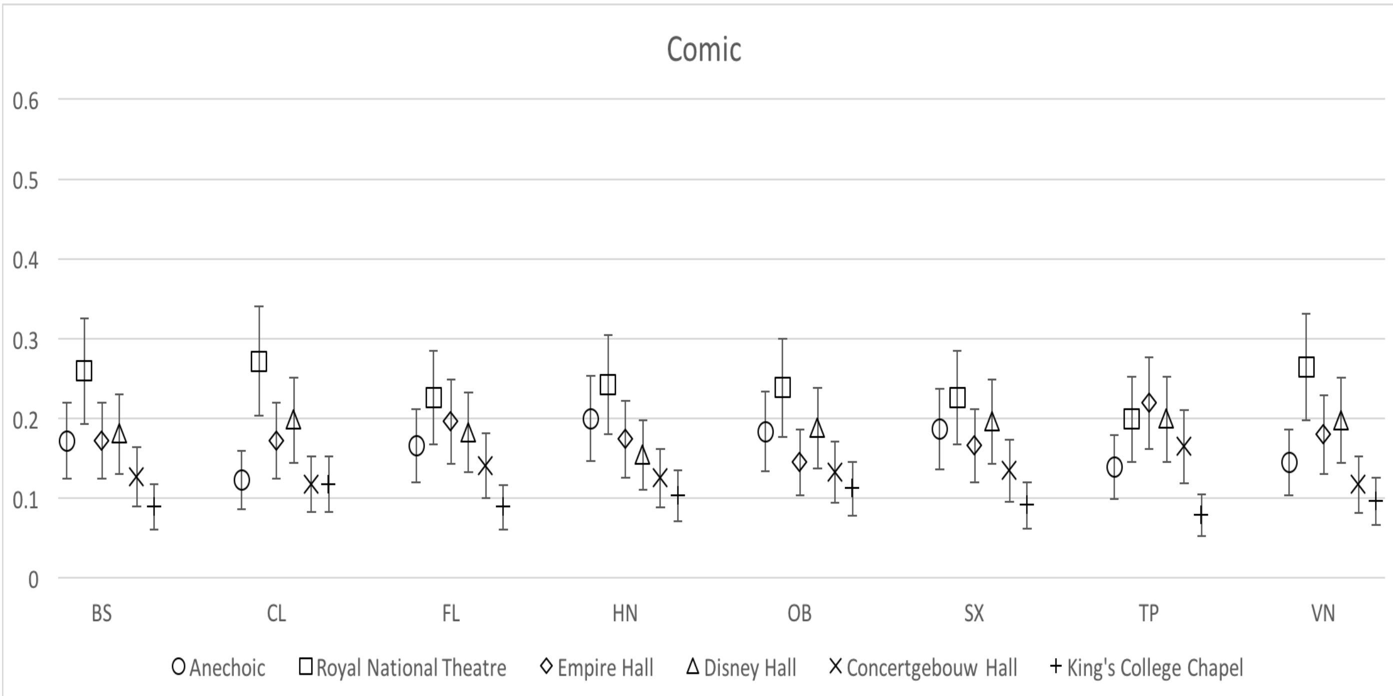

Figure 3.13: BTL scale values and the corresponding 95% confidence intervals for the emotional category Comic.

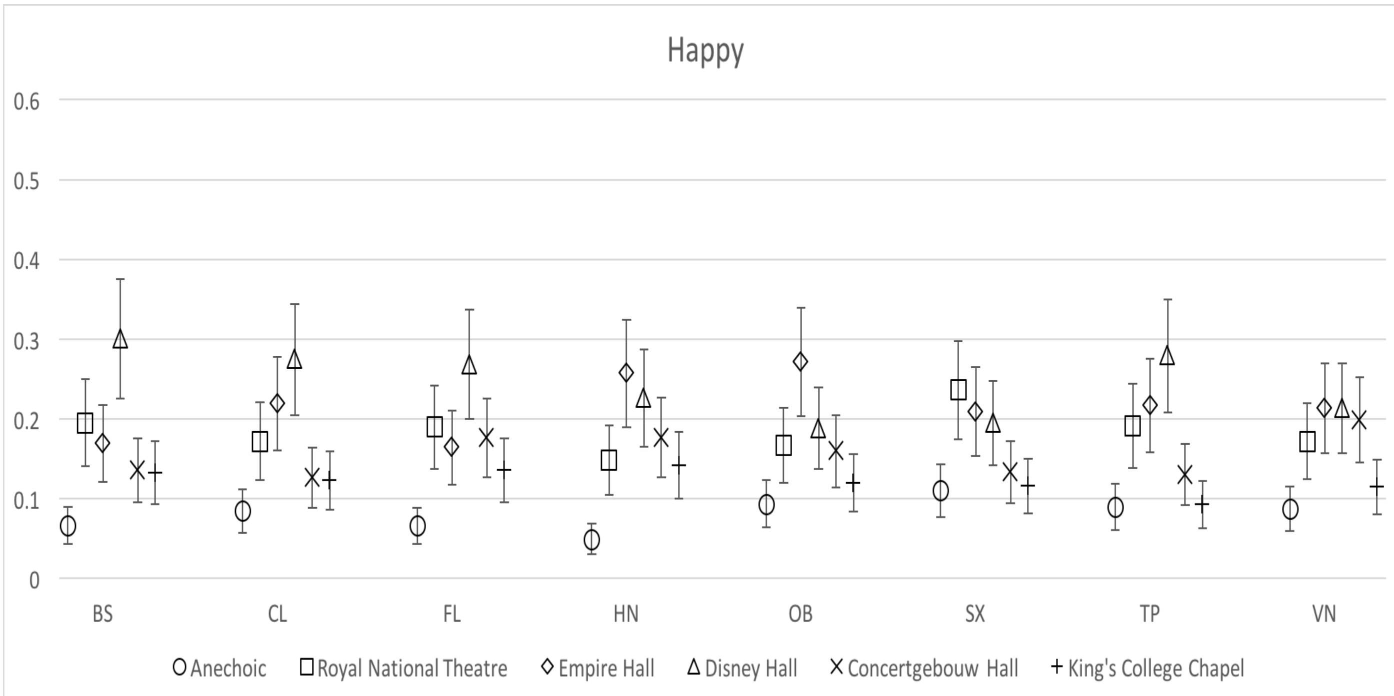

Figure 3.14: BTL scale values and the corresponding 95% confidence intervals for the emotional category Happy.

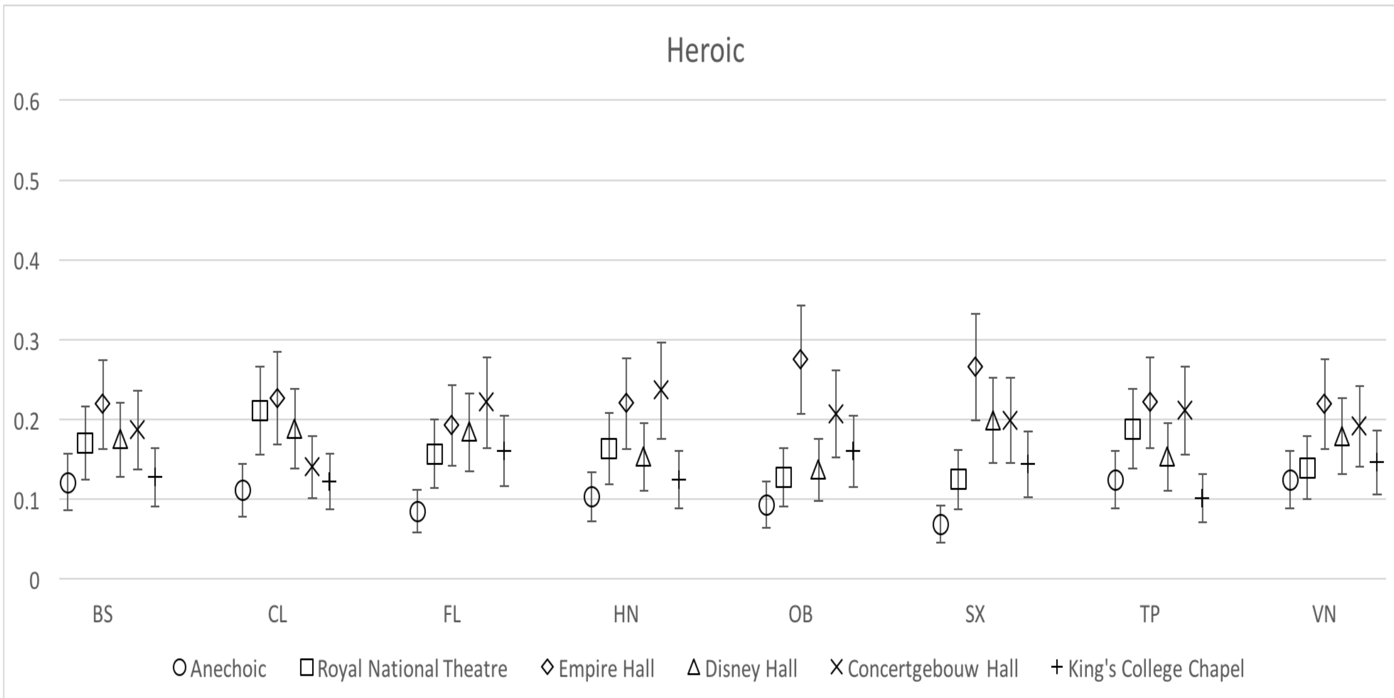

Figure 3.15: BTL scale values and the corresponding 95% confidence intervals for the emotional category Heroic.

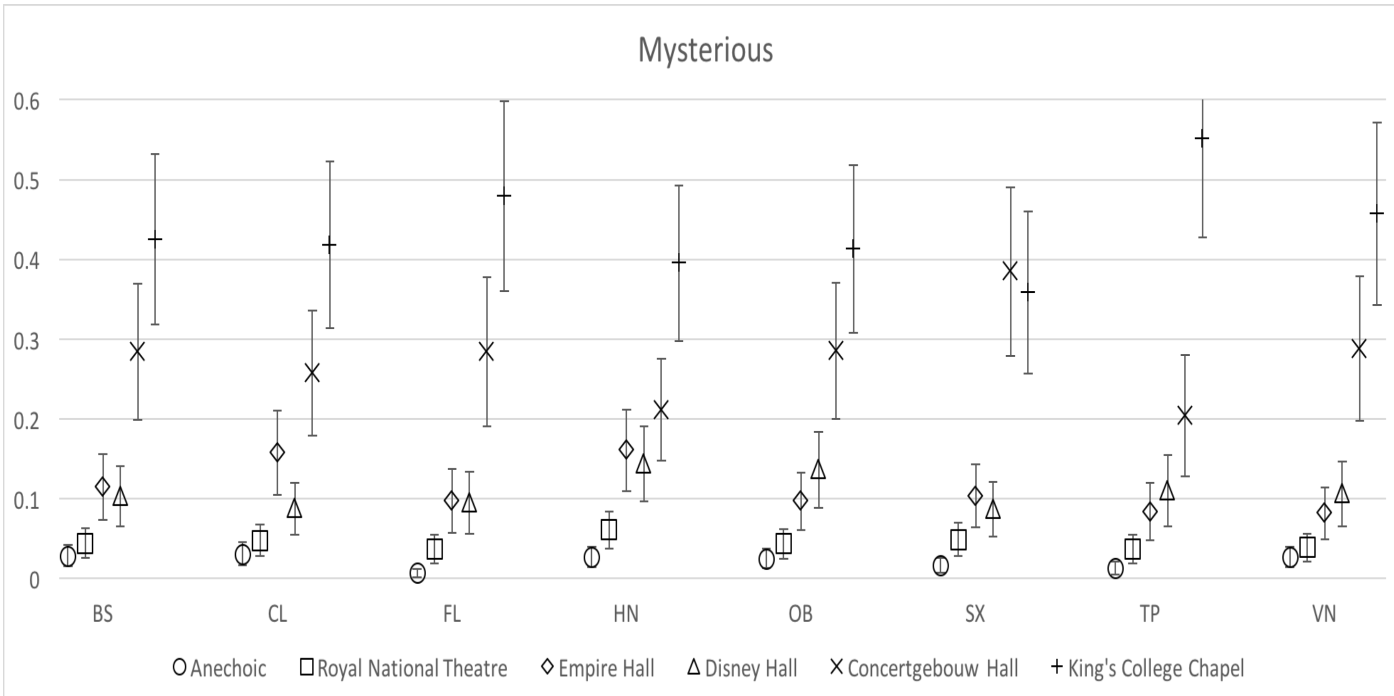

Figure 3.16: BTL scale values and the corresponding 95% confidence intervals for the emotional category Mysterious.

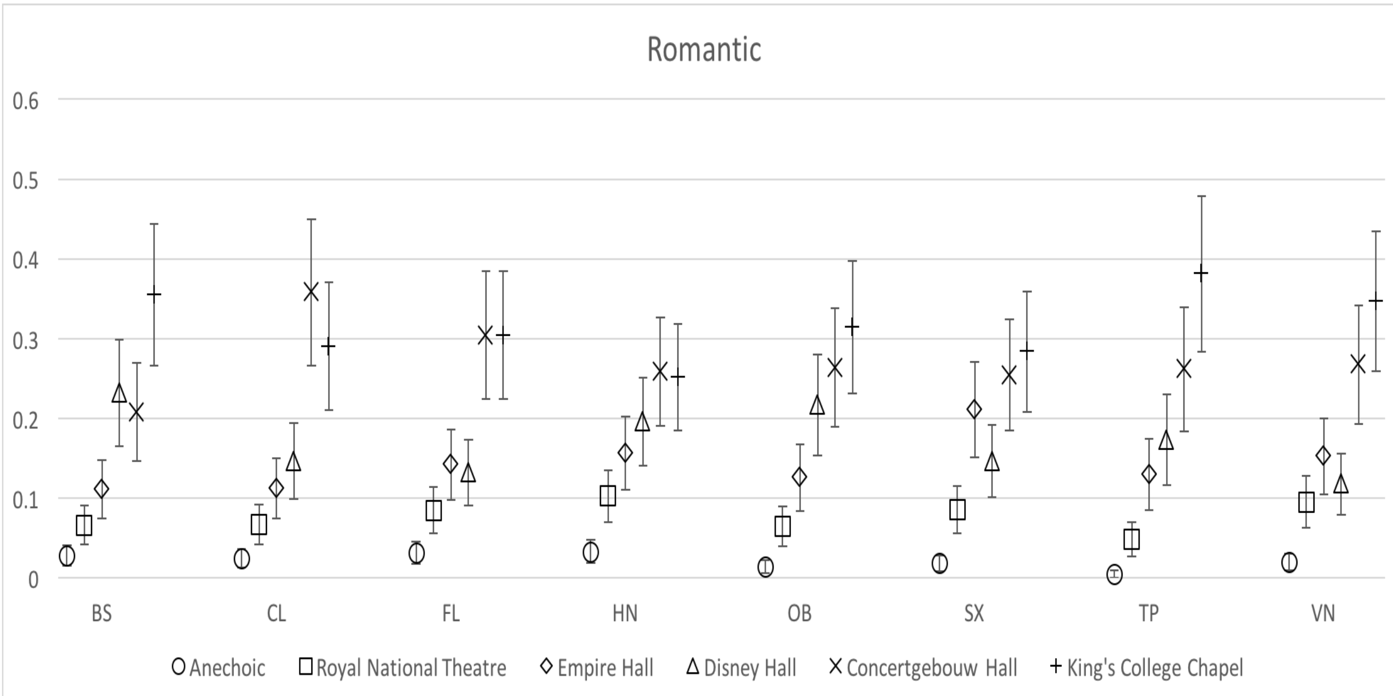

Figure 3.17: BTL scale values and the corresponding 95% confidence intervals for the emotional category Romantic.

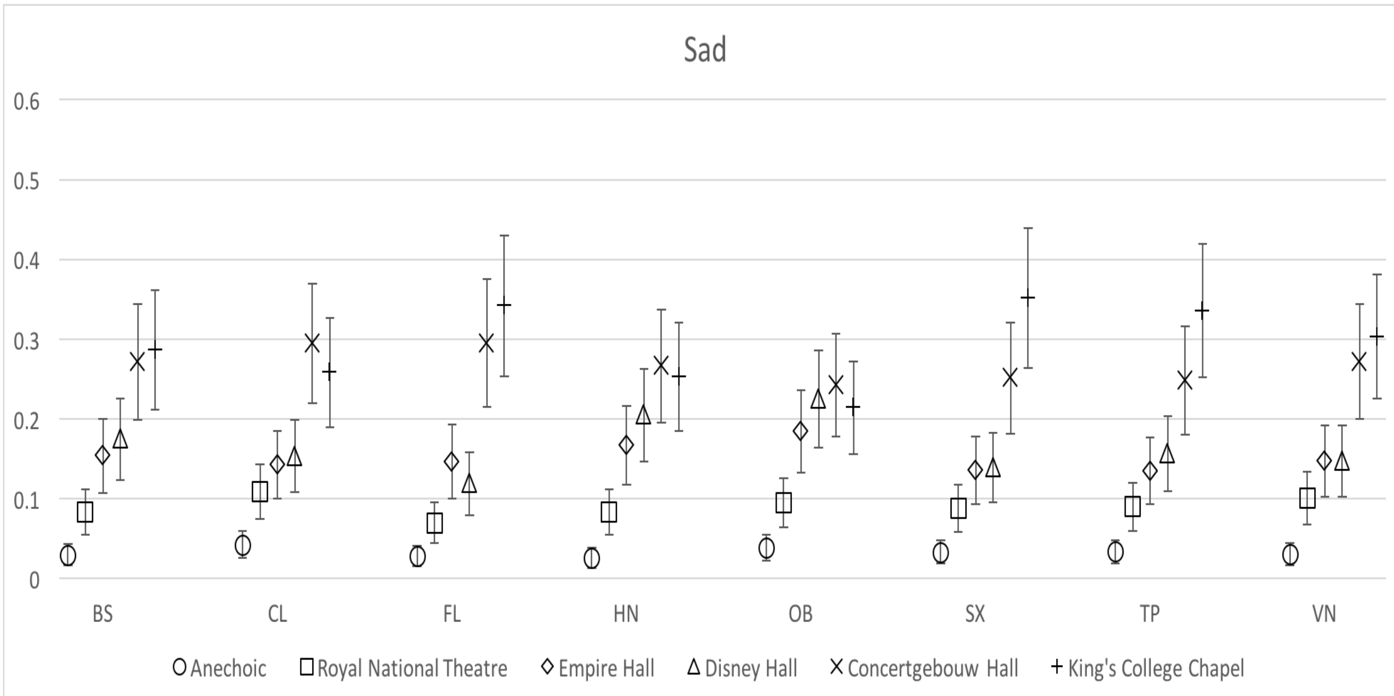

Figure 3.18: BTL scale values and the corresponding 95% confidence intervals for the emotional category Sad.

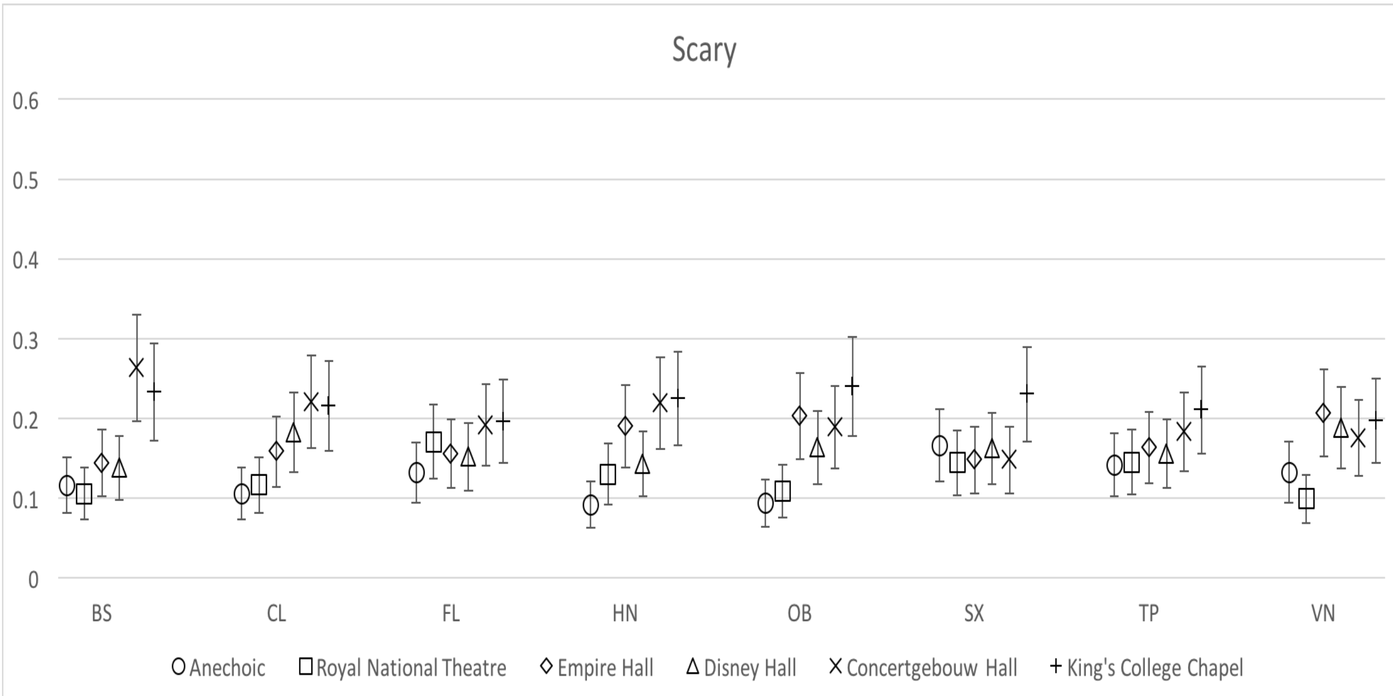

Figure 3.19: BTL scale values and the corresponding 95% confidence intervals for the emotional category Scary.

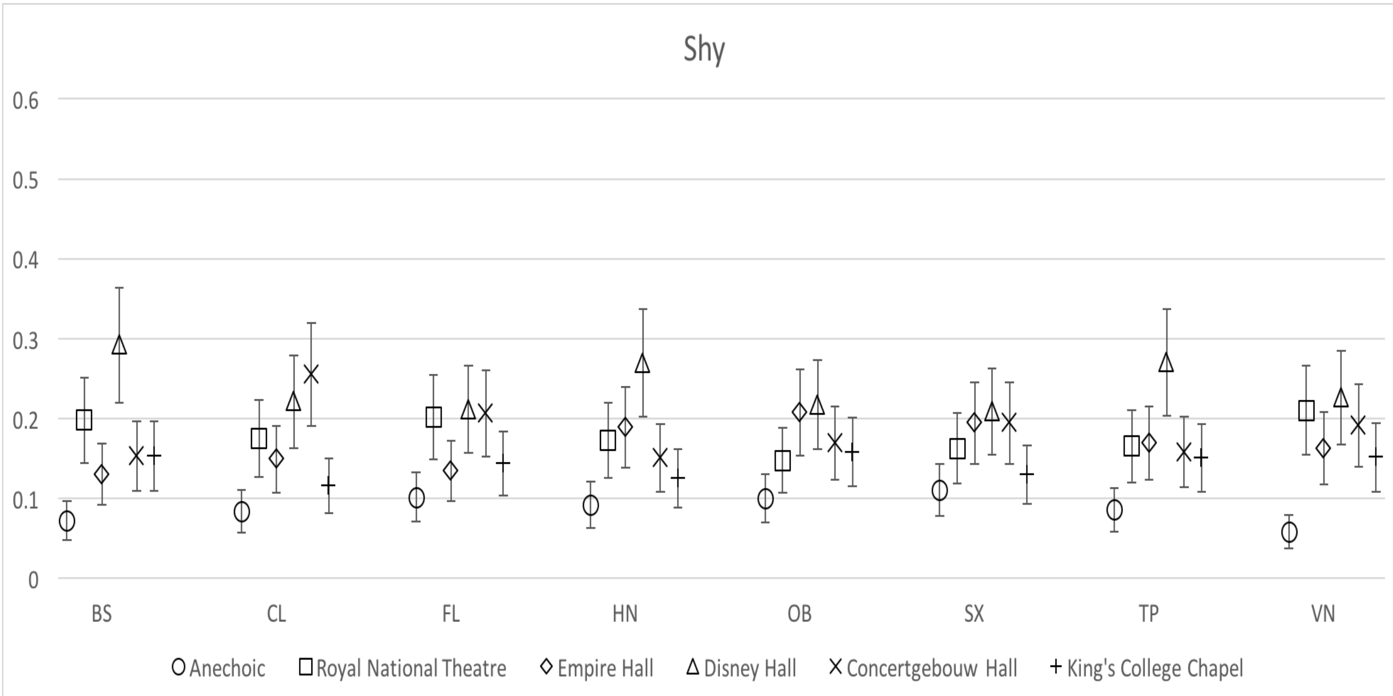

Figure 3.20: BTL scale values and the corresponding 95% confidence intervals for the emotional category Shy.

	Anechoic Chamber RT_{60}=0	Royal National Theatre RT_{60}=0.94	Empire Hall RT_{60}=1.31	Disney Hall RT_{60}=1.80	Concertgebouw Hall RT_{60}=2.32	King's College Chapel RT_{60}=5.44
Angry	**21**	13	20	4	2	0
Calm	0	8	8	13	**19**	15
Comic	0	**17**	5	9	2	0
Happy	0	10	17	**18**	5	2
Heroic	0	3	**15**	3	6	0
Mysterious	0	4	17	16	29	**34**
Romantic	0	8	15	17	26	**28**
Sad	0	8	12	13	**28**	26
Scary	0	0	4	1	**8**	7
Shy	0	6	5	**18**	4	1

Table 3.2: Based on paired t-tests, how often each instrument was statistically significantly greater (for $p < 0.05$) than the others for each hall impulse response and emotional characteristic. The maximum possible value is 40 and the minimum possible value is 0. The maximum for each emotional characteristic is shown in bold and shaded.

Based on Table 3.2, we see that, halls with shorter reverberation times were the strongest for the emotional categories Angry and Comic. Halls with medium reverberation times were the strongest for Happy, Heroic, and Shy. Halls with long reverberation times were the strongest for Calm, Mysterious, Romantic, Sad, and Scary. We compared this characterization with our results from parametric reverberation [38] and found agreement in 7 of the 8 emotional categories that we tested in that study (Heroic was the only one that was different).

As a further comparison, we correlated the BTL data from the hall impulse responses with the BTL data from our previous study of parametric reverberation [38], and found a correlation of 0.74 over all emotional categories, indicating a rather remarkable level of agreement. Table 3.3 also shows the correlations between the individual emotional categories. Seven of the emotional categories were significant, but Shy was not. The seven significant correlations were fairly strong ranging from about 0.47 to 0.87 for the individual categories. The emotional categories with the strongest correlations were Mysterious, Romantic, and Sad which also had the largest number of significant differences in Table 3.2.

Emotional Category	Pearson Correlation	Spearman Correlation
Comic	**0.495	**0.475
Happy	**0.500	**0.500
Heroic	**0.607	**0.521
Mysterious	**0.863	**0.876
Romantic	**0.785	**0.827
Sad	**0.791	**0.765
Scary	**0.709	**0.656
Shy	*0.277	*0.275
Overall	**0.742	**0.684

Table 3.3: Pearson and Spearman correlation between the BTL values for the convolution reverberation and parametric reverberation. **: $p < 0.05$; *:$0.05 < p < 0.1$.

3.4 Discussion

The main goal of this experiment was to see how the emotional characteristics of musical instruments changed with reverberation time in convolution reverberation. Based on Table 3.2, our main findings are the following:

1. Halls with shorter reverberation times tended to emphasize the emotional characteristics Angry and Comic.
2. Halls with medium reverberation times tended to emphasize the emotional characteristics Happy, Heroic, and Shy.
3. Halls with longer reverberation times tended to emphasize the emotional characteristics Calm, Mysterious, Romantic, Sad, and Scary.

We were curious to see how the results for convolution reverberation would compare to the results from our previous study on parametric reverberation [38]. The biggest difference between them was that the convolution reverberation emotional characteristics were more pronounced. There were 62% more significant differences for the eight categories and five reverberation times that we tested in both experiments (see Table 3.4). This difference shows up most clearly in the emotional category Shy where there was only one significant difference for parametric reverberation and 27 for convolution reverberation.

Emotional Category	Parametric	Convolution
Comic	15	8
Happy	9	36
Heroic	12	16
Mysterious	49	65
Romantic	42	64
Sad	29	59
Scary	21	14
Shy	1	27
Total	178	289

Table 3.4: Number of significant differences for eight emotional categories (Angry and Calm were not included in our parametric reverberation experiment) and five reverberation times (Cathedral reverb time was not tested in our parametric reverberation experiment).

While the effects on the emotional characteristics were more pronounced in convolution reverberation compared to parametric reverberation, there was also a striking basic agreement in the results. The BTL rankings for convolution and parametric reverberations were significantly and strongly correlated with a correlation coefficient of 0.74 over all emotional categories. Seven out of eight individual emotional were also significantly correlated with correlation coefficients ranging from 0.47 to 0.87. Shy was the only category not significantly correlated since its BTL rankings were very flat for parametric reverberation. These strong correlations indicate that reverberation time has a remarkably consistent effect on the emotional characteristics regardless of whether using convolution or parametric reverberation. It makes sense that these two different processes would have some deep underlying similarities since they are both reverberation techniques. It also makes sense that convolution reverberation would have a more pronounced effect on the emotional characteristics, since convolution reverberation is usually regarded as warmer, more natural, and smoother than parametric reverberation, which is usually regarded as more bland in comparison.

It was really surprising that the emotional characteristics Angry and Scary, which have near-identical Valence and Arousal values (see Figure 2.1), were so completely opposite in their results (see Table 3.2). Angry was the strongest for halls with short reverberation times, and Scary was the strongest for halls with long reverberation times. The Scary results agree with our previous parametric reverb results [38], as well as the results by Västfjäll et al. [35] and Tajadura-Jiménez et al. [36], who found that larger reverberation times and larger rooms were more unpleasant.

Similarly, the emotional characteristics Comic, Happy, Heroic, and Romantic all have similar Valence and arousal values (see Figure 2.1), yet they showed distinctly different results in Table 3.2. Comic, Happy, and Heroic were the strongest in halls with short and medium reverberation times, though their patterns were distinctly different. Romantic was in a different class altogether, and was very strong for long reverberation times. The results of Västfjäll [35] and Tajadura-Jiménez [36] suggested all four of these characteristics would be stronger in smaller rooms, but Table 3.2 shows the differences between these emotional characteristics. For example, Table 3.2 shows that Happy was the strongest in the medium-sized Disney Hall.

In a sense the columns of Table 3.2 represent the footprints of emotional characteristics of the individual halls relative to one another. The Anechoic Chamber was singularly Angry in its response. The crisp sound of the Royal National Theatre was strong for Angry and Comic, appropriately given that it is a venue for Shakespeare comedies and dramas. The hard bright surfaces of the Empire Hall brought out emotional characteristics such as Angry, Happy, and Heroic, which are good for the witty and stately music of Haydn which was composed for this venue. The warm sound of the Disney Hall was especially unique in bringing out Shy and Happy. The sophisticated elegance of the Concertgebouw Hall was apparent in the categories Calm, Mysterious, Romantic, and Sad with a touch of Scary. Finally, the spacious King's College Chapel also brought out the characteristics Mysterious, Romantic, and Sad, the first two even a bit more than the Concertgebouw Hall with its 5-second reverberation time.

CHAPTER 4

EXPERIMENT 3: INVESTIGATING INTO HOW PARAMETRIC REVERBERATION EFFECTS THE SPACE OF INSTRUMENT EMOTIONAL CHARACTERISTICS

4.1 Overview

In order to see how parametric reverberation changes the underlying instrument space as we mentioned in Section 1.4, we used a relatively simple parametric reverberation model to measure the emotional characteristics of instruments for two of the most important reverberation parameters: reverberation length and amount. Through a listening test with paired comparisons and statistical analysis similar to Experiment 1, we will investigate whether simple parametric reverberation changes the emotional characteristics of instruments uniformly or in an instrument-dependent way.

To address this question, we conducted a listening test to compare instruments in order to determine how the ranking of the instruments varied with different types of reverberation and different emotional characteristics. We tested eight sustained musical instruments representing the wind and bowed string families. We compared these sounds over eight emotional categories used in previous studies [21, 20, 22, 19, 23, 24, 32, 38] and that are commonly expressed by composers in tempo and expression marks (Happy, Sad, Heroic, Scary, Comic, Shy, Romantic, and Mysterious). The following section describes the details of the listening test.

4.2 Listening Test

Our test had listeners compare eight instrument tones over eight emotional categories for each type of reverberation. The anechoic stimuli we used in this experiment were exactly the same as we used in our previous parametric reverberation study [38] and Experiment 1 in this thesis (see Section 3.2 for more information). To recap, these anechoic tones were obtained from the University of Iowa Musical Instrument Samples [55]. These instrument sounds have fundamental frequencies close to Eb4 (311.1 Hz), and were further analyzed and normalized.

In addition to the anechoic sounds, we compared sounds with reverberation lengths of 1s and 2s, which typically correspond to small and large concert halls. We used the reverberation generator provided by *Cool Edit* [80] to generate reverberated tones. 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 types of reverberation. Note that these reverberation types were also used in our previous reverberation study [38]:

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

Readers may refer to Figures 3.1 to 3.4 in Section 3.2 for the energy decay curves for the different types of reverberation we used. The Early Decay Times (EDTs) were near-zero for all four reverberation types.

We recruited 36 subjects 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. Among the 36 subjects, there were 24 males and 12 females. The subjects ranged in age from 19 to 27. In terms of musical experience, 17 subjects had some experience playing an instrument (an average of 4.8 years), and 19 subjects did not have experience playing an instrument. In recruiting the subjects, all 36 indicated they had no known hearing problems.

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 (e.g., Happy, Sad, Scary/Fear/Angry, Tender/Calm/Romantic) roughly corresponding to the four quadrants of the Valence-Arousal plane, but there are lots of variations beyond that [64]. For this study, we used the same categories we have used in our previous research on musical instruments [21, 20, 22, 19, 23, 24, 32, 38]. The ratings of the emotional categories can be found in Figure 2.1.

In the listening test, every subject heard paired comparisons of all eight instruments for each type of reverberation and emotional category. During each trial, subjects heard a pair of instrument sounds from the same type of reverberation and were prompted to choose which more strongly aroused a given emotional category. Since each trial was a single paired comparison requiring minimal memory from the subjects, subjects did not need to remember all of the tones, just the two in each comparison. Readers may refer to Figure 3.10 in Section 3.2 for the paired comparison listening test interface.

Each combination of two different instruments tones were presented for each of the five reverberation types and eight emotional categories, and the listening test totaled $C_2^8 \times 5 \times 8 = 1120$ trials. For each instrument, the overall trial presentation order was randomized. For the two sounds A and B, they heard AB where the order of A and B was random for each comparison (but if they heard AB, they did not hear BA later).

Before the first trial, subjects read online definitions of the emotional categories from the *Cambridge Academic Content Dictionary* [66]. The dictionary definitions we used in this experiment can be found in Table 2.1 in Section 2.4. Subjects were not musical experts (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 the same "quiet room" as we specified in Section 2.4.

4.3 Results

For our listening test, listeners compared each pair of instruments for each emotional category and each reverberation type. Originally, we had 36 subjects since the test was rather long at 2 hours. We screened the responses, and found 3 subjects were obviously spanning the same key responses toward the end of the test, so we excluded all of their data. We scanned the remaining subjects' data, especially at the end of the test, and based on the consistency of their responses, felt that they were giving sincere and attentive responses to the questions, so we did not exclude any further subjects.

Based on the filtered listening test data of 33 subjects, we derived scale values using the Bradley-Terry-Luce (BTL) statistical model [68, 69]. Figures 4.1 to 4.5 show the BTL scale values and the corresponding 95% confidence intervals for each reverberation type (anechoic, small hall front, small hall back, large hall front, and large hall back respectively). For each graph, the BTL scale values for the eight instruments sum up to 1. The BTL value for each instrument is the probability that listeners will choose that instrument when considering a certain reverberation type and emotion category. For example, if all eight instruments (Bs, Cl, Fl, Hn, Ob, Tp, Sx, and Vn) were judged equally Happy, the BTL scale values would be $1/8 = 0.125$.

Though there are certainly differences between Figures 4.1 - 4.5, overall they are remarkably similar to one another. For example, the trumpet was consistently ranked highest for Heroic with all reverberation types, while the clarinet was ranked highest for Sad. Going further, the trumpet ranked the highest for all five reverberation types for Happy, Heroic, and Comic, while the clarinet was highest for Sad, Shy, Romantic, and Mysterious (except a close second for Small Hall Front), and the flute and violin shared top-rankings for Scary. Heroic consistently had the widest range among all reverberation types, and Scary the narrowest.

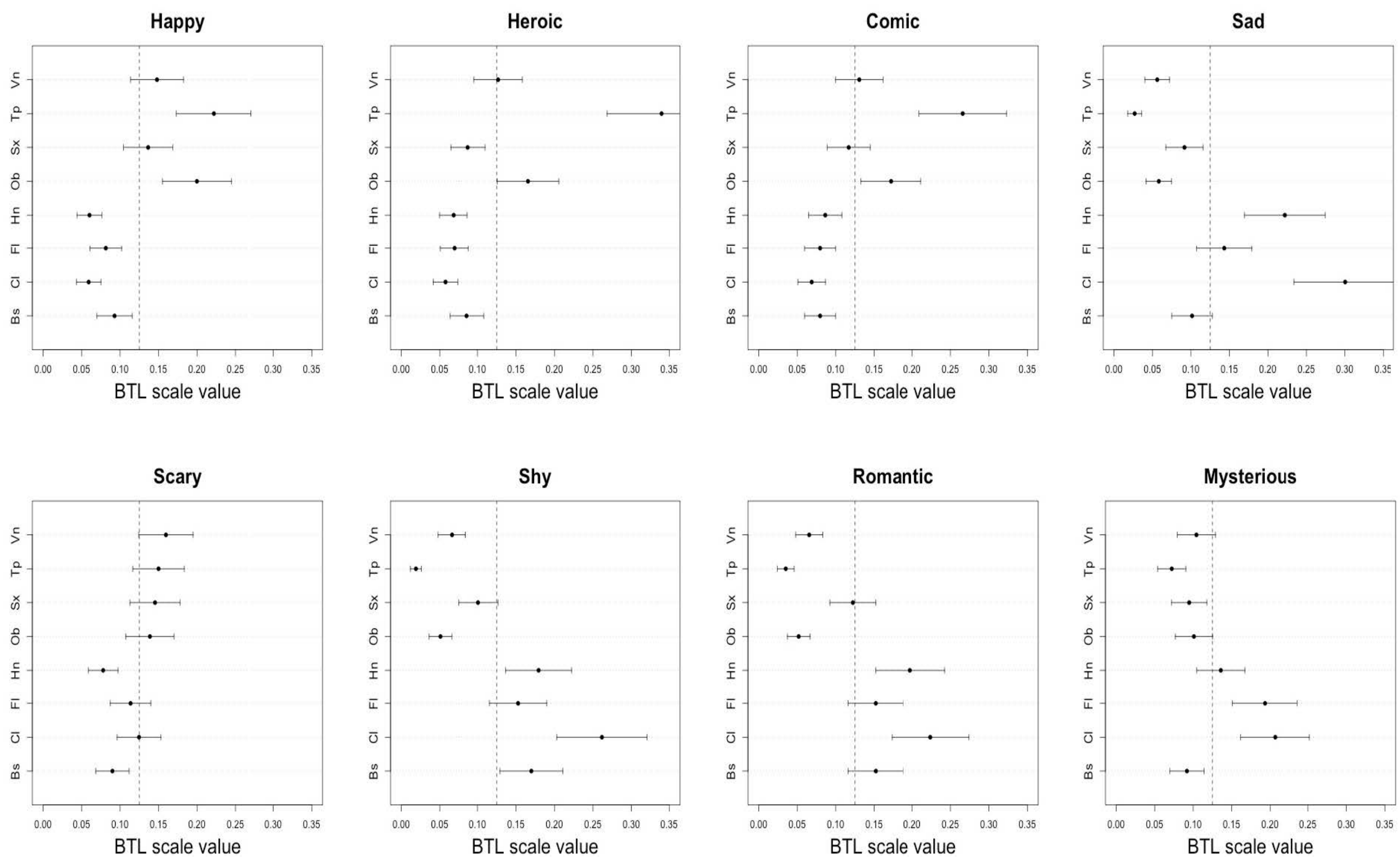

Figure 4.1: BTL scale values and the corresponding 95% confidence intervals of the original anechoic instrument sounds for each emotional characteristic. The dotted line represents the average, with instrument sounds to the right more Happy for example, and instrument sounds to the left less.

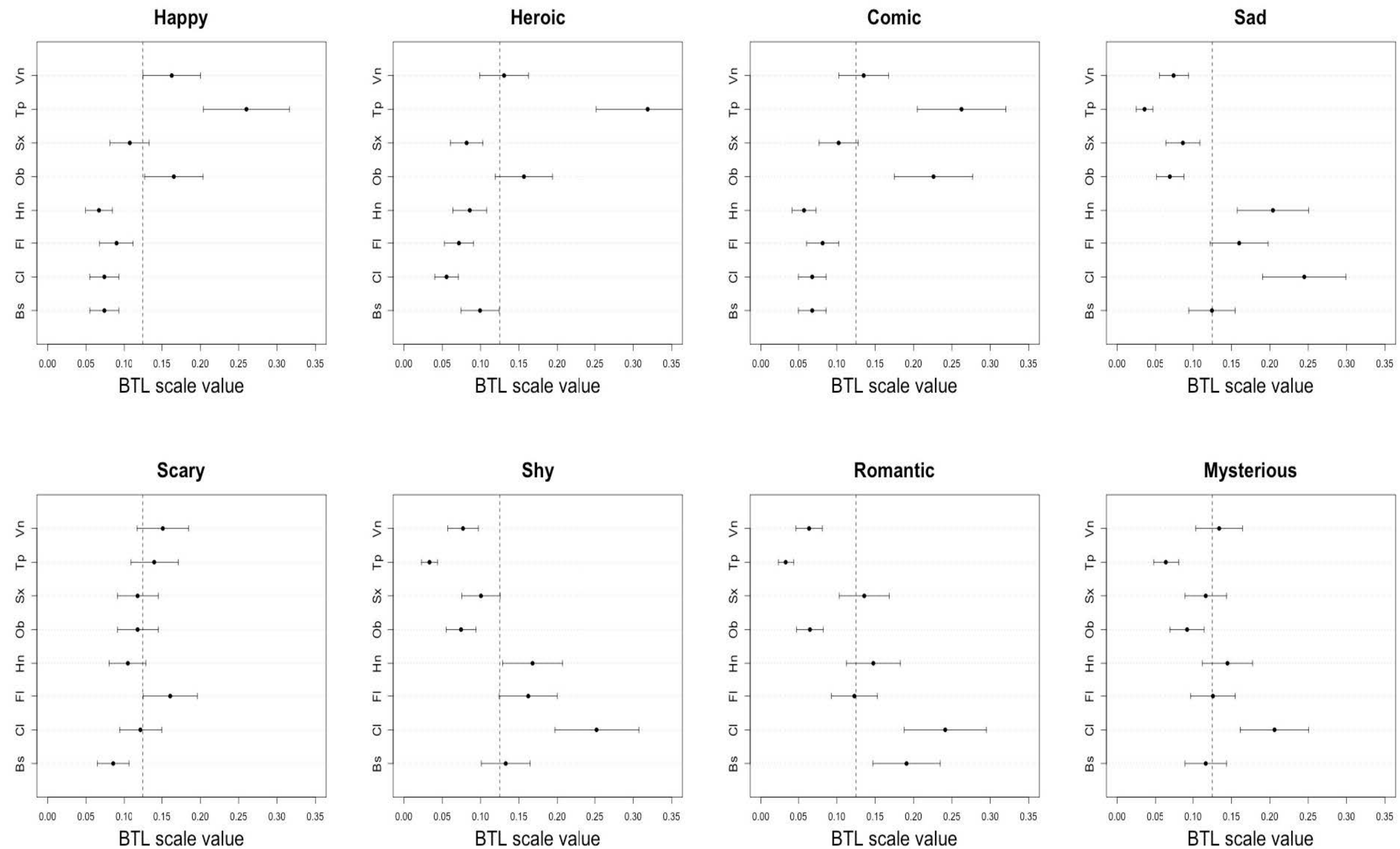

Figure 4.2: BTL scale values and the corresponding 95% confidence intervals of the instrument sounds with Small Hall Front reverberation for each emotional characteristic.

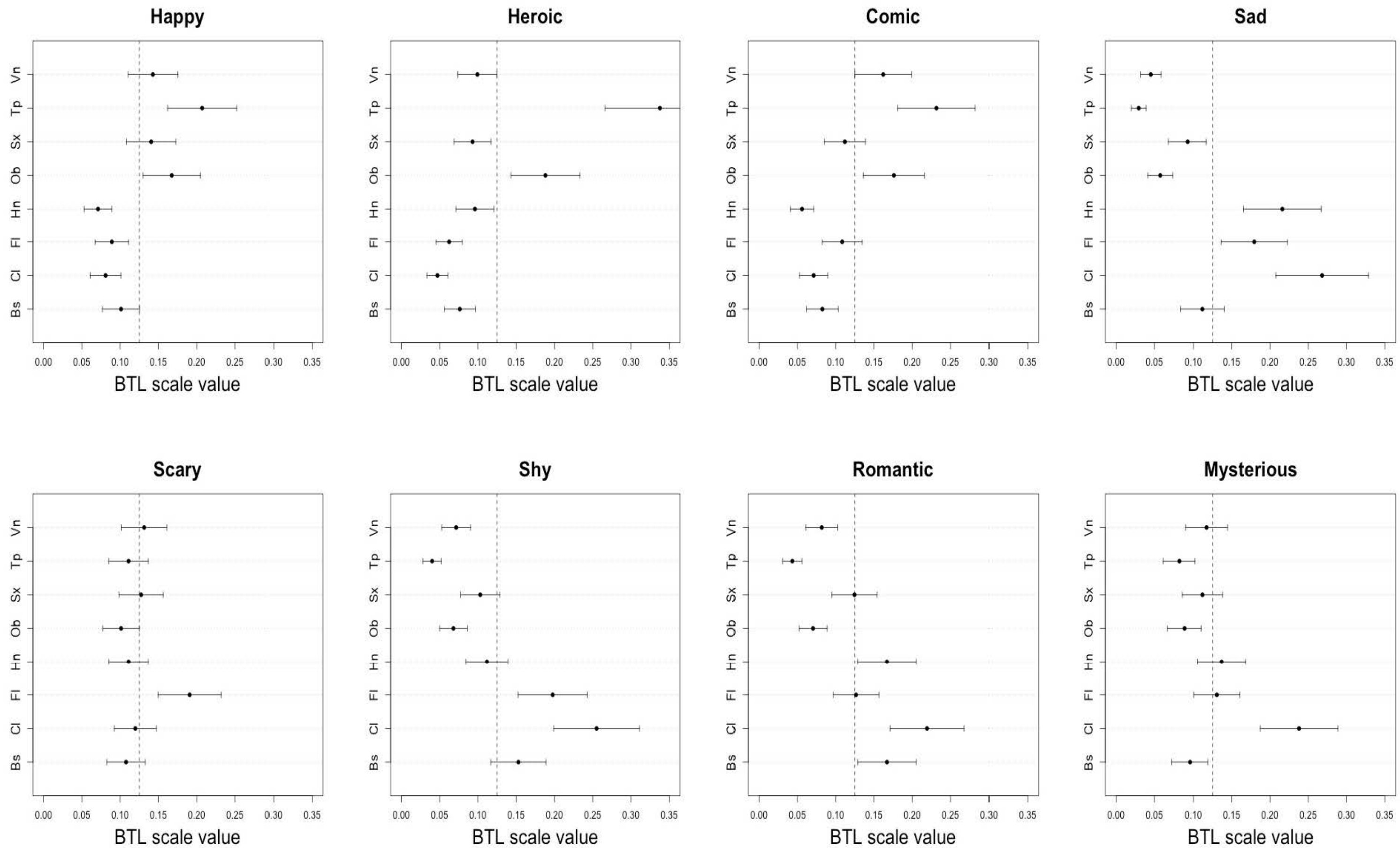

Figure 4.3: BTL scale values and the corresponding 95% confidence intervals of the instrument sounds with Small Hall Back reverberation for each emotional characteristic.

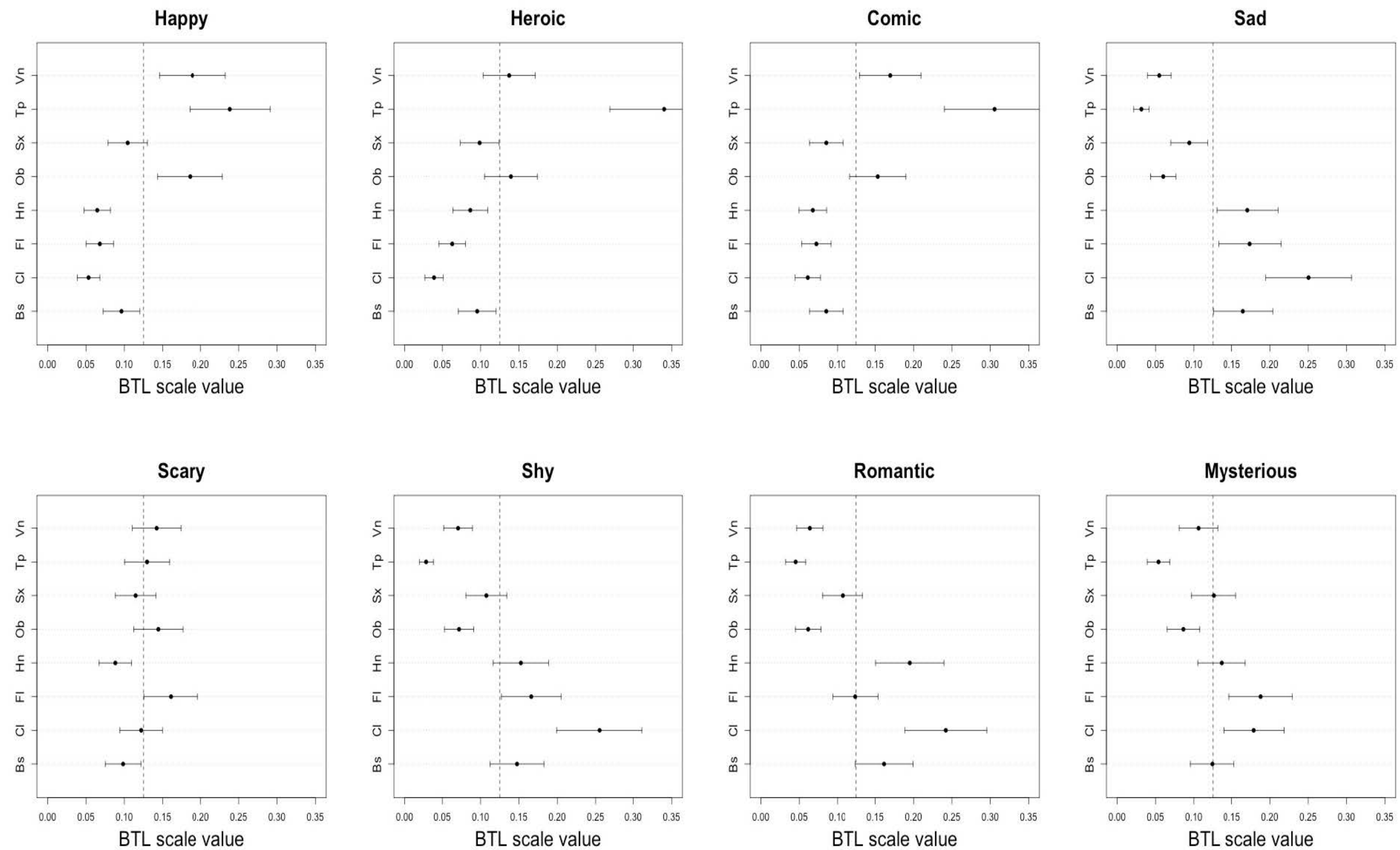

Figure 4.4: BTL scale values and the corresponding 95% confidence intervals of the instrument sounds with Large Hall Front reverberation for each emotional characteristic.

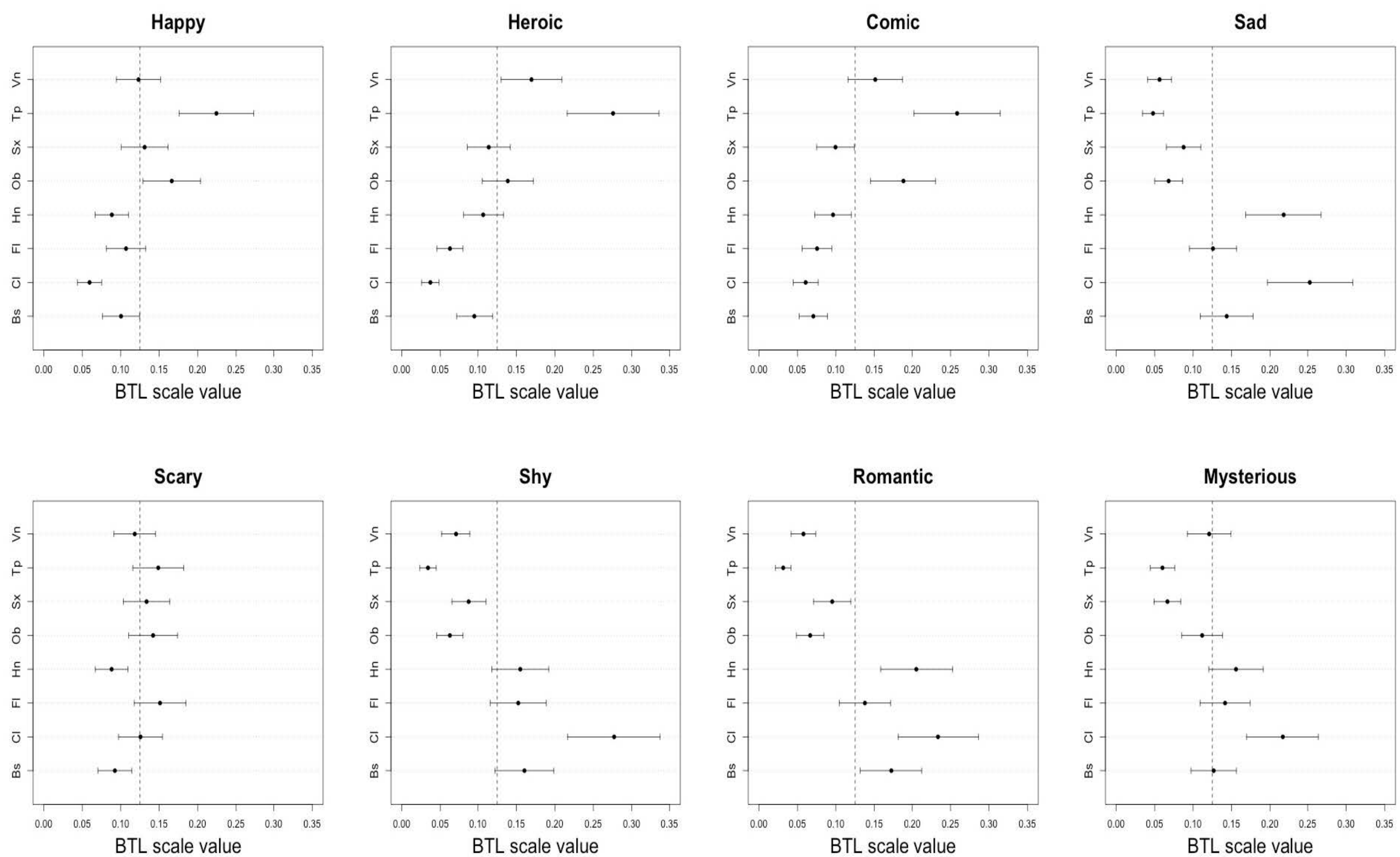

Figure 4.5: BTL scale values and the corresponding 95% confidence intervals of the instrument sounds with Large Hall Back reverberation for each emotional characteristic.

We wanted to determine the number of times each instrument was significantly greater than the other seven instruments for each reverberation type and emotional characteristic. As a preliminary step, the normality of the data was calculated for each instrument, emotional characteristic, and reverberation type. Since most, though not all, were normally distributed (see Tables C.1 - C.5 in Appendix C), both parametric and nonparametric statistical tests (parametric: Paired t-tests, Pearson correlation ; nonparametric: Wilcoxon signed-rank tests, Spearman correlation) were used to analyze the voting data (i.e., the number of positive votes received by each instrument for each emotional category and reverberation type). The results from the two tests showed some minor differences, but basically they were in agreement. Table 4.1 shows the paired t-test results and Table D.1 in Appendix D shows the Wilcoxon signed-rank test results. For each instrument, the maximum possible value is 7 and the minimum possible value is 0. For example, with the original anechoic sounds and the emotional characteristic Heroic, the value of the trumpet is 7 since it was statistically significantly greater than all seven of the other instruments for the Heroic subgraph in Figure 4.1. The maximum value for each reverberation type and emotional characteristic is shown in bold and shaded for both tables.

Table 4.2 sums the sub-tables in Table 4.1 and shows the number of times each instrument was significantly greater than the other seven instruments over all five reverberation types for each emotional characteristic. The maximum possible value is 35 and the minimum possible value is 0. For example, for Heroic the trumpet was statistically significantly greater than all the other seven instruments for four reverberation types and six for Large Hall Back, so its value is 34. The maximum value for each emotional characteristic is shown in bold and shaded. Table 4.2 makes it obvious that the trumpet was ranked the highest for Happy, Heroic, and Comic, the clarinet for Sad, Shy, Romantic, and Mysterious, and the flute for Scary.

Anechoic	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	2	0	0	0	**4**	3	**4**	3
Heroic	0	0	0	0	5	0	**7**	3
Comic	0	0	0	0	4	2	**7**	2
Sad	3	**6**	3	**6**	1	2	0	1
Scary	0	1	0	0	1	**2**	1	**2**
Shy	4	**6**	4	4	1	3	0	1
Romantic	3	**5**	3	3	0	3	0	1
Mysterious	0	**6**	5	1	0	0	0	0

Small Hall Front	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	0	0	0	0	5	1	**7**	5
Heroic	1	0	0	0	5	0	**7**	3
Comic	0	0	0	0	**6**	3	**6**	4
Sad	3	**6**	4	5	1	1	0	1
Scary	0	0	**2**	0	0	0	0	1
Shy	2	**7**	4	4	1	1	0	1
Romantic	4	**5**	3	3	1	3	0	1
Mysterious	1	**5**	1	1	0	1	0	1

Small Hall Back	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	0	0	0	0	**4**	3	**4**	3
Heroic	1	0	0	1	6	2	**7**	1
Comic	1	0	2	0	**5**	2	**5**	3
Sad	3	**5**	**5**	**5**	1	2	0	1
Scary	0	0	**4**	0	0	0	0	0
Shy	3	**6**	5	1	1	1	0	1
Romantic	3	**5**	3	3	1	3	0	1
Mysterious	0	**7**	0	1	0	0	0	0

Large Hall Front	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	2	0	0	0	**5**	2	**5**	**5**
Heroic	2	0	1	1	3	1	**7**	2
Comic	0	0	0	0	5	0	**7**	5
Sad	**4**	**4**	**4**	**4**	1	3	0	1
Scary	0	0	**2**	0	0	0	0	0
Shy	3	**7**	4	3	1	3	0	1
Romantic	4	**5**	3	**5**	0	3	0	0
Mysterious	1	2	**4**	1	1	1	0	1

Large Hall Back	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	1	0	1	0	3	1	**6**	1
Heroic	2	0	1	2	2	2	**6**	3
Comic	0	0	0	1	5	1	**6**	4
Sad	4	**6**	4	**6**	0	2	0	0
Scary	0	0	**2**	0	0	0	0	0
Shy	4	**7**	4	4	1	1	0	1
Romantic	4	**5**	3	4	1	2	0	1
Mysterious	2	**6**	2	2	2	0	0	2

Table 4.1: Based on paired t-tests, how often each instrument was statistically significantly greater (for $p < 0.05$) than the others for each reverberation type and emotional characteristic. The maximum possible value is 7 and the minimum possible value is 0. The maximum for each reverberation type and emotional characteristic is shown in bold and shaded.

	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	5	0	1	0	21	10	**26**	17
Heroic	6	0	2	4	21	5	**34**	12
Comic	1	0	2	1	25	8	**31**	18
Sad	17	**27**	20	26	4	10	0	4
Scary	0	1	**10**	0	1	2	1	3
Shy	16	**33**	21	16	5	9	0	5
Romantic	18	**25**	15	18	3	14	0	4
Mysterious	4	**26**	12	6	3	2	0	4

Table 4.2: How often each instrument was statistically significantly greater than the others over the five reverberation types. The maximum possible value is 35 and the minimum possible value is 0. The maximum for each emotional characteristic is shown in bold and shaded. This table is simply the sum of the individual sub-tables in Table 4.1.

We wanted to determine how similar were the sub-tables in Table 4.1 and the BTL data in Figures 4.1 - 4.5 for the different reverberation types. Therefore, we ran correlations for both of these as well as for the voting data (i.e., the number of positive votes received by each instrument for emotional category and reverberation type). In all cases, the correlations were statistically significant (at the $p < 0.0001$ level) and very strong, ranging from 90 to 95%, indicating a near-linear relationship and a very high level of agreement. In particular, Table 4.3 shows Pearson and Spearman correlation between the different reverberation types based on the voting data, since it is the most precise and direct measure of correlation in the sense that it is correlation of the original data and not correlation of statistics based on the original data (e.g., the BTL data in Figures 4.1 - 4.5 and paired t-test data in Table 4.1).

Let's take another look at the question of the consistency of the listeners during this long 2-hour listening test. As further evidence that the 33 subjects were giving sincere and attentive responses, if they had been giving random responses at the end of the test due to fatigue, it would have decreased the number of significant differences in Table 4.1, making the footprints less clear and less consistent. As it turned out, they were very consistent, suggesting listeners remained reasonably attentive. We don't claim that they were perfect, but the 90 - 95% correlation in Table 4.3 indicates that listeners were amazingly consistent.

Reverberation Types	Pearson Correlation	Spearman Correlation
Anechoic & Small Hall Front	0.944	0.930
Anechoic & Small Hall Back	0.924	0.922
Anechoic & Large Hall Front	0.946	0.941
Anechoic & Large Hall Back	0.934	0.932
Small Hall Front & Small Hall Back	0.946	0.944
Small Hall Front & Large Hall Front	0.950	0.944
Small Hall Front & Large Hall Back	0.927	0.917
Small Hall Back & Large Hall Front	0.930	0.922
Small Hall Back & Large Hall Back	0.902	0.897
Large Hall Front & Large Hall Back	0.935	0.922

Table 4.3: Pearson and Spearman correlation between the different reverberation types based on the listener voting data. All correlations were statistically significantly (for $p < 0.05$).

4.4 Discussion

Previous work has shown that different musical instruments have distinct emotional characteristics [17, 18, 19, 20, 21, 22, 23, 24, 32], and that reverberation can greatly change these characteristics [35, 36, 38]. And while these emotional characteristics can be greatly changed with reverberation, the results in this experiment have shown that they are changed uniformly in about the same way for different instruments. In other words, the underlying instrument space does not change much with reverberation in terms of emotional characteristics. For example, added reverberation might bring out characteristics such as Mysterious or Heroic, but in a uniform way for the instruments, and not some more than others. There seems to be a relatively consistent ranking of emotional characteristics between the instruments that holds with different reverberation amounts and lengths, at least for simple parametric reverberation.

We should also emphasize that our results are for basic-level professional headphones. Higher-quality professional headphones could perhaps show even more pronounced differentiation between the emotional characteristics though we expect it would also be in a uniform way for the instruments.

This uniformity is contrasting to our previous study [38], where distinct and significant changes occurred in every instrument and emotional characteristic with different types of reverberation. The strong distinct changes found in our first study led us to expect some instrument-dependencies in this study, which used exactly the same tones. But, the two studies are from contrasting perspectives. In our first study, tones with different types of reverberation were compared for each instrument and emotional characteristic, allowing us to identify which reverberation types heightened each emotional characteristic for each instrument. In this study, tones from different instruments were compared for each reverberation type and emotional characteristic, allowing us to rank the instruments for each reverberation type and emotional characteristic. There is no contradiction in their results: reverberation distinctly changes the character of the sound, but does so in a uniform way across the instruments. It makes sense that reverberation changes the character uniformly across the instruments: if it were not uniform, then performers in orchestras and chamber groups would not be able to practice in small rehearsal rooms in a reliable way if reverberation affected the character in an instrument-dependent way. Musicians would need to carefully rehearse in the performance venue, not just to get used to the hall, but to adjust their blends and balances differently for each different venue.

The uniform effects of reverberation on instruments is in contrast to another post-process that we studied in Experiment 1, MP3 compression, where the results were instrument-dependent. There, the trumpet was much more effected than other instruments with more compression, and the horn much less effected. But, for the tones we tested in our study of MP3 compression, the artifacts of excessive compression were obvious. Readers may refer to Section 2.5 for more information. If we had tested tones where the compression rate was lower, and the tones sounded the same as the original, we feel pretty confident that the emotional characteristics would have been the same as the original, and the instruments would have shown a trivially uniform response.

Admittedly, MP3 compression and reverberation are different. MP3 compression is a lossy process, and reverberation is in a sense an additive one - so it may be the results are simply different for the two processes. On the other hand, perhaps they are similar. Perhaps with concert hall levels reverberation the results are uniform, and with very large amounts or lengths of reverberation instrument-dependencies emerge. Why? It is not difficult to imagine that with excessive smearing of the temporal and spectral envelopes (e.g., a 5-second cathedral reverberation), that instruments with strong spectral variations in either the temporal or spectral envelopes (e.g., the clarinet with its strong odd harmonics) would be changed more than other instruments with smoother temporal or spectral envelopes (e.g., the horn). It is likely that the distinctive emotional characteristics of instruments such as the clarinet would erode in Tables 4.1 and 4.2 with very large amounts or lengths of reverberation. So, it may be that we did not happen to test a wide enough range of reverberations to be able to see the onset of these effects. Further work will be needed to confirm this. But it is remarkable how uniform the instruments were within the concert hall range of reverberation that we did test.

More broadly, perhaps the relatively consistent ranking of emotional characteristics between the instruments is what allows each instrument to identify each instrument regardless of room reverberation, or at least helps. Perhaps each instrument has a characteristic footprint, that varies with pitch and dynamic level, which makes it identifiable.

So, where do these footprints appear in our data? The columns of each sub-table in Table 4.1 represent the unites of the emotional characteristics for each instrument and reverberation type. In general, the footprints for each instrument were very similar for the different reverberation types (e.g., the trumpet had large values for Happy, Heroic, and Comic and small values for the others across all reverberation types).

The columns of Table 4.2 represent the overall footprints of the emotional characteristics for each instrument (for our Eb4-*forte* tones). The instruments clustered into two fairly distinct groups: those where the positive energetic emotional characteristics were strong (e.g., oboe, trumpet, violin), and those where the low-arousal characteristics were strong (e.g., bassoon, clarinet, flute, horn). The saxophone was an outlier, and was uniquely somewhat strong for most emotional characteristics. Looking in more detail, the oboe, trumpet, and violin had similar footprints, but the trumpet's footprints were deeper for Happy, Heroic, and Comic than the other two instruments. In the same way, the clarinet and horn had similar footprints, though the clarinet was deeper especially for Shy and Mysterious. The flute also had a similar footprint to the clarinet and horn, but was deeper for Scary. The bassoon was similar to the horn except deeper for Happy, less for Sad. The saxophone had the most even distribution, with medium values for most emotional categories.

As a disclaimer, probably the footprint for each instrument varies depending on its pitch and dynamics as well as other factors of each particular tone. What is useful to note here is that the footprints of each instrument for different types of reverberation were very similar, as we can see by comparing the respective columns for each sub-table in Table 4.1.

CHAPTER 5

EXPERIMENT 4: INVESTIGATING INTO HOW CONVOLUTION REVERBERATION EFFECTS THE SPACE OF INSTRUMENT EMOTIONAL CHARACTERISTICS

5.1 Overview

For this investigation, we conducted a listening test to investigate whether convolution reverberation changes the emotional characteristics of instruments uniformly or in an instrument-dependent way. In Experiment 3 (see Section 4.3) where we investigated into how parametric reverberation changes the underlying instrument space, they were changed uniformly.

To easily compare and contrast the convolution and parametric reverberation results, we conducted the listening test for convolution reverberation in the same way as we did for parametric reverberation. Listeners compared the instruments pairwise for each hall and emotional characteristic. Below are some the main points, especially the differences from the parametric reverberation test.

5.2 Listening Test

We tested sustained musical instruments representing the wind and bowed string families obtained from the University of Iowa Musical Instrument Samples [55], identical to the stimuli we used in Experiments 1 and 2. They included the bassoon (bs), clarinet (cl), flute (fl), horn (hn), oboe (ob), saxophone (sx), trumpet (tp), and violin (vn). These sounds were all recorded in an anechoic chamber, and were thus free from reverberation. These instruments were nearly harmonic and had fundamental frequencies close to Eb4 (311.1 Hz).

In addition to the anechoic sounds, we compared convolution reverberated sounds with reverberation lengths of approximately 1s and 2s, which typically correspond to small and large concert halls. To do this, we selected several representative hall convolution reverberations (see Section 3.2). We measured their reverberation lengths based on their reverberation time RT_{60}, and picked those that most closely matched the reverberation times we tested in Experiment 3. Readers may refer to Table 3.1, as well as Figures 3.5 to 3.9 in Section 3.2, for the details of the hall impulse responses we used in this experiment.

For this study we tested the ten emotional categories: Angry, Calm, Comic, Happy, Heroic, Mysterious, Romantic, Sad, Scary, and Shy. Some choices of emotional characteristics are fairly universal and occur in many previous studies roughly corresponding to the four quadrants of the Valence-Arousal plane (e.g., Happy, Sad, Scary/Fear/Angry, Tender/Calm/Romantic), but there are lots of variations beyond that [64]. For this study, we used the same categories that we used in our previous research on musical instruments and reverberation [21, 20, 22, 19, 23, 24, 32, 75, 33, 74, 76, 38, 77, 78, 79, 51]. Just before the listening test, subjects read online definitions of the emotional categories used in this experiment which were taken from the Cambridge Academic Content Dictionary. Readers may refer to Table 2.1 for the definitions of these categories.

To limit the length of the test and to minimize listener fatigue, we divided listeners into two groups, where each group heard a different set of five emotional categories. We recruited 74 subjects to take the listening test, and since they each heard half the emotional categories, 37 subjects compared each emotional category. 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. Subjects were not musical experts (e.g., recording engineers, professional musicians, or music conservatory students) but average attentive listeners. Among the 74 subjects, there were 44 males and 30 females. The subjects ranged in age from 18 to 23. In terms of musical experience, 37 subjects had some experience playing an instrument (an average of 6.2 years), and 37 subjects did not have experience playing an instrument. In recruiting the subjects, all indicated they had no known hearing problems.

In the listening tests, subjects heard paired comparisons between the instruments for the each hall and emotional category. During each trial, subjects heard a pair of sounds and were prompted to choose which more strongly aroused a given emotional category (e.g., “Which tone sounds more Mysterious, 1 or 2?”). Since each trial was a single paired comparison requiring minimal memory from the subjects, subjects did not need to remember all of the tones, just the two in each comparison.

Each combination of two different instruments were presented for each of the eight instruments over the six halls and ten emotional categories, and the complete listening test totalled $C_2^8 \times 6 \times 10 = 1680$ trials (840 trials per listener since we divided the task into two groups). For each listener, the overall trial presentation order was randomized to average out effects due to learning or fatigue. For the two sounds A and B, they heard AB where the order of A and B was random for each comparison (but if they heard AB, they did not hear BA later). The listening test took about 95 minutes, with forced short breaks about every 30 minutes. The subjects were seated in the same "quiet room" as we specified in Section 2.4.

5.3 Results

For this listening test, listeners compared each pair of instruments for each hall and emotional characteristic. Originally, we had 37 subjects for each group of five emotional characteristics. We screened the responses, and found 6 subjects from each group were obviously spamming the same key responses toward the end of the test, so we excluded all of their data. We scanned the remaining subjects' data, especially at the end of the test, and based on the consistency of their responses, felt that they were giving sincere and attentive responses to the questions, so we did not exclude any further subjects.

Based on the filtered listening test data of 31 subjects for each group, we derived scale values using the Bradley-Terry-Luce (BTL) statistical model [68, 69]. Figures 5.1 to 5.10 show the BTL scale values and the corresponding 95% confidence intervals for each emotional category. For each graph, the BTL scale values for the eight instruments sum up to 1. The BTL value for each instrument is the probability that listeners will choose that instrument when considering a certain hall and emotion category. For example, if all eight instruments (Bs, Cl, Fl, Hn, Ob, Tp, Sx, and Vn) were judged equally Happy, the BTL scale values would be $1/8 = 0.125$.

Though there are differences between the individual graphs in each of Figures 5.1 - 5.10, there are striking similarities. For example, with Angry in Figure 5.1, the trumpet is usually first, the oboe second, and clarinet last. Extending this idea, we observe that the trumpet or oboe were ranked highest for nearly all halls in the high-arousal categories Angry, Comic, Happy, Heroic, and Scary. The only exception was that the violin was narrowly ranked first for Disney Hall with the categories Happy and Scary. The clarinet and horn were ranked highest for the low-arousal categories Calm, Romantic, Sad, and Shy. Scary consistently had the narrowest range among all halls.

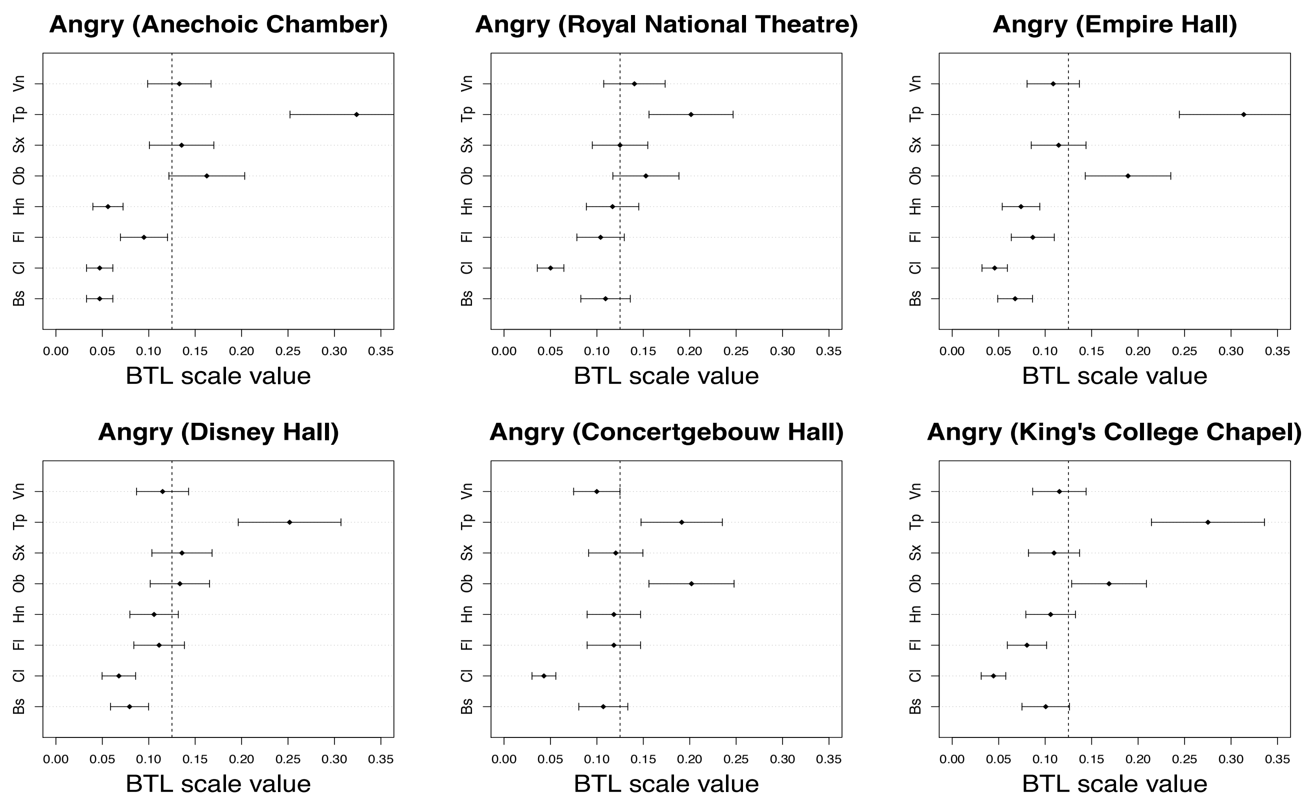

Figure 5.1: BTL scale values and the corresponding 95% confidence intervals with different halls for the category Angry. The dotted line represents the average, with instrument sounds to the right more Angry, and instrument sounds to the left less.

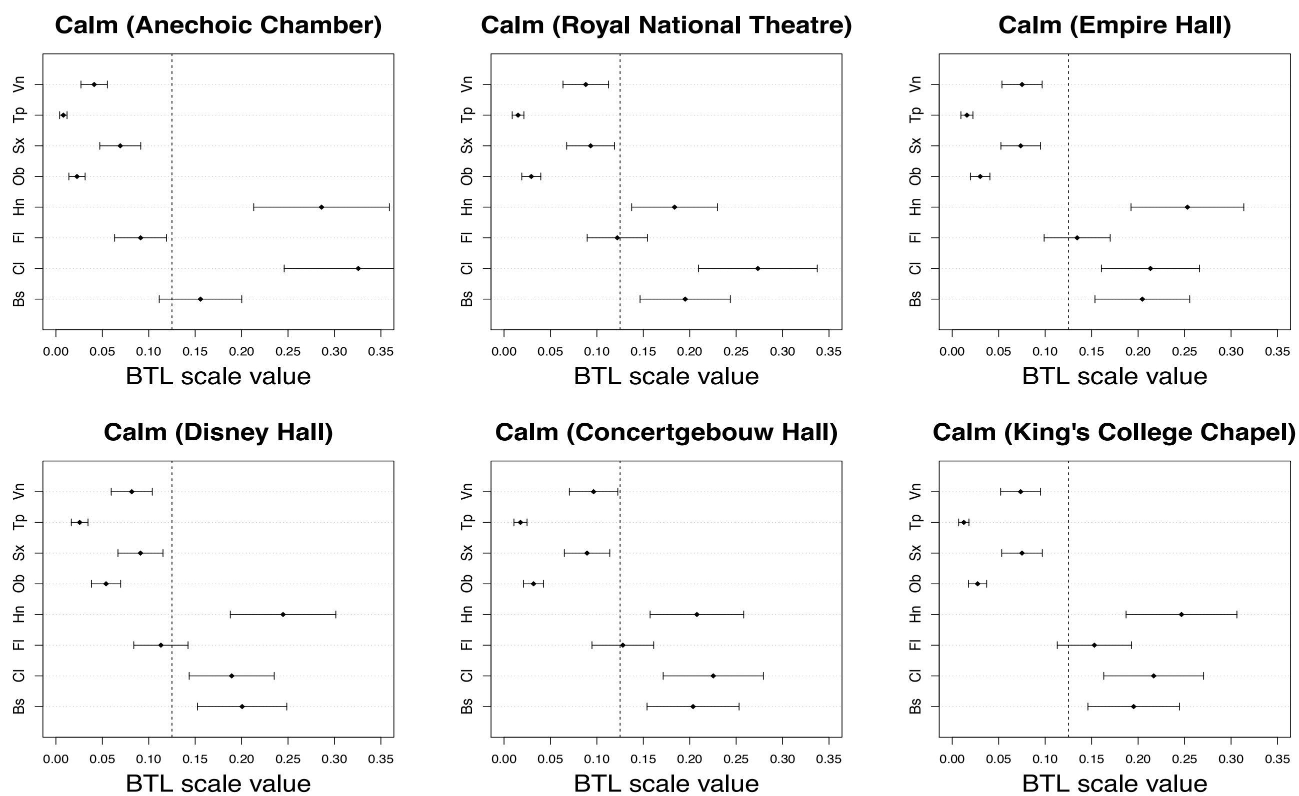

Figure 5.2: BTL scale values and the corresponding 95% confidence intervals with different halls for Calm.

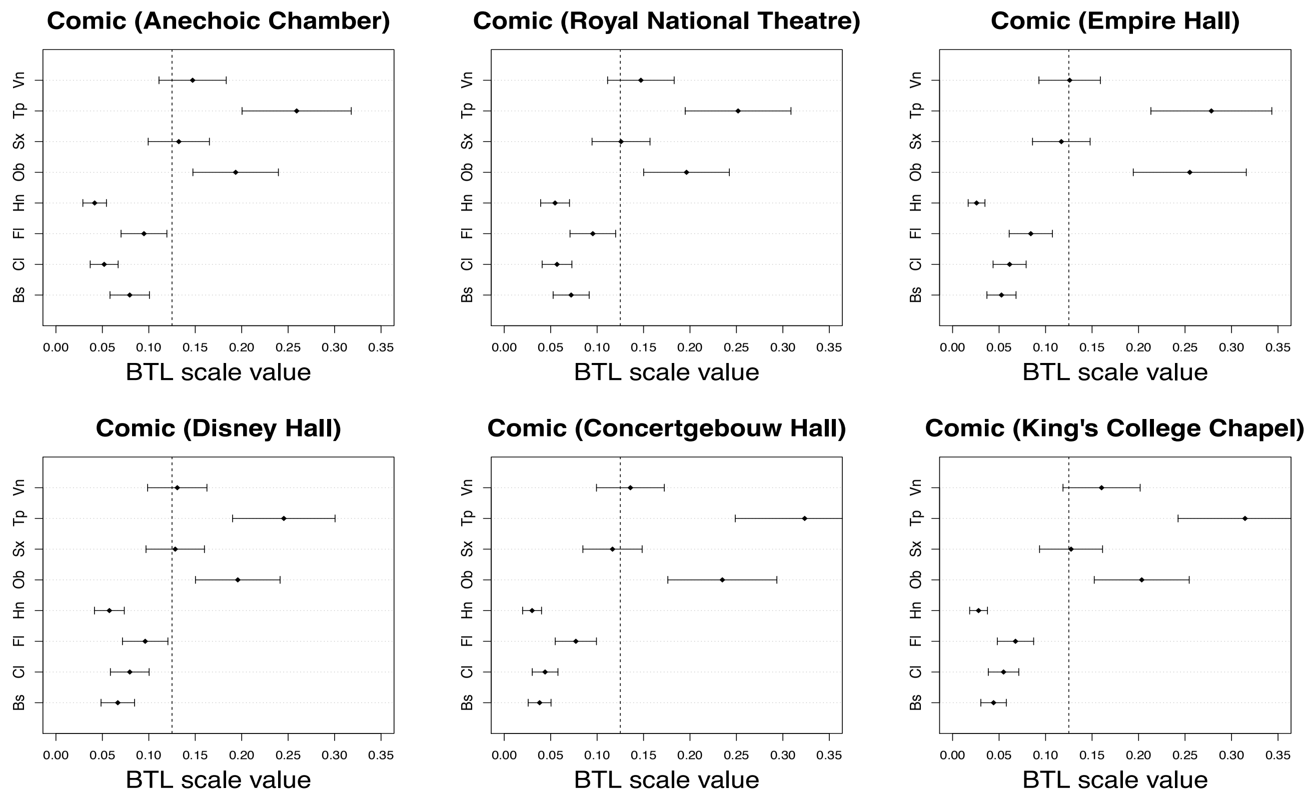

Figure 5.3: BTL scale values and the corresponding 95% confidence intervals with different halls for Comic.

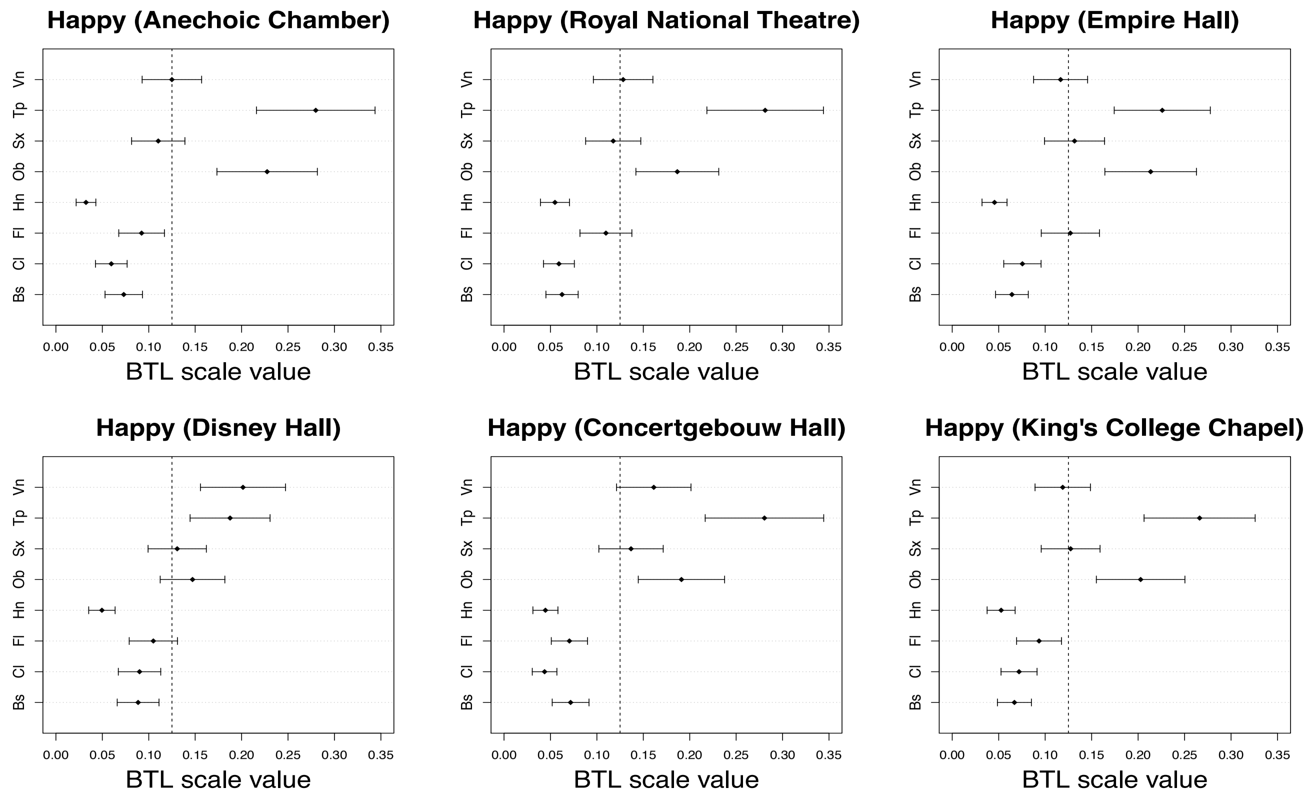

Figure 5.4: BTL scale values and the corresponding 95% confidence intervals with different halls for Happy.

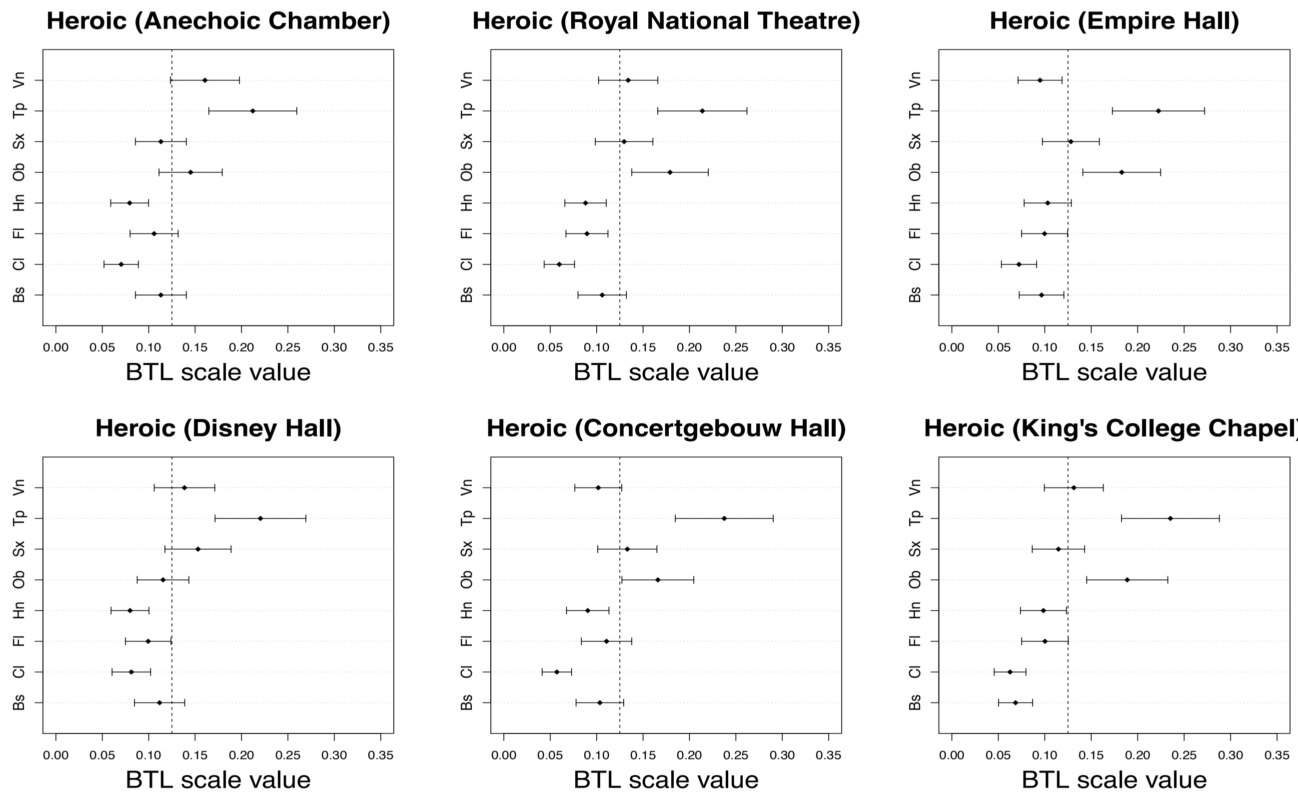

Figure 5.5: BTL scale values and the corresponding 95% confidence intervals with different halls for Heroic.

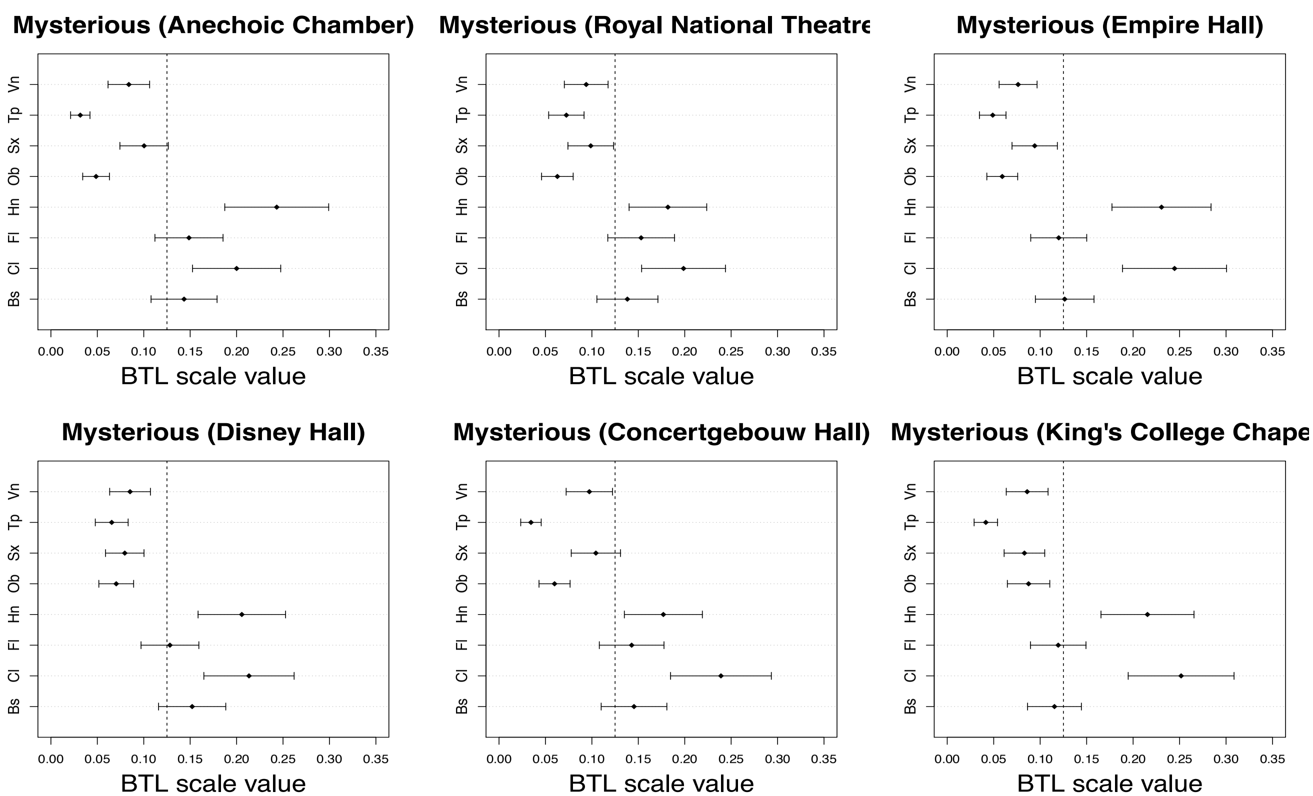

Figure 5.6: BTL scale values and the corresponding 95% confidence intervals with different halls for Mysterious.

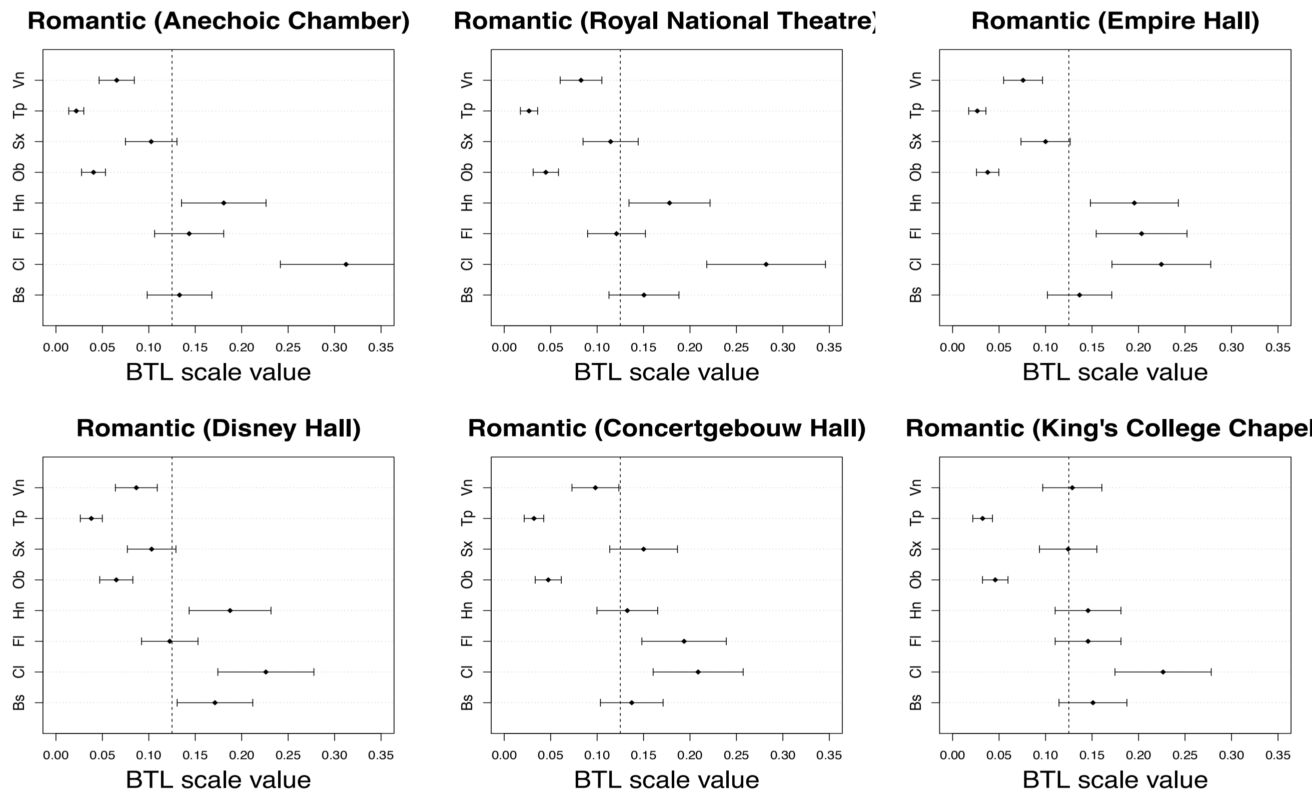

Figure 5.7: BTL scale values and the corresponding 95% confidence intervals with different halls for Romantic.

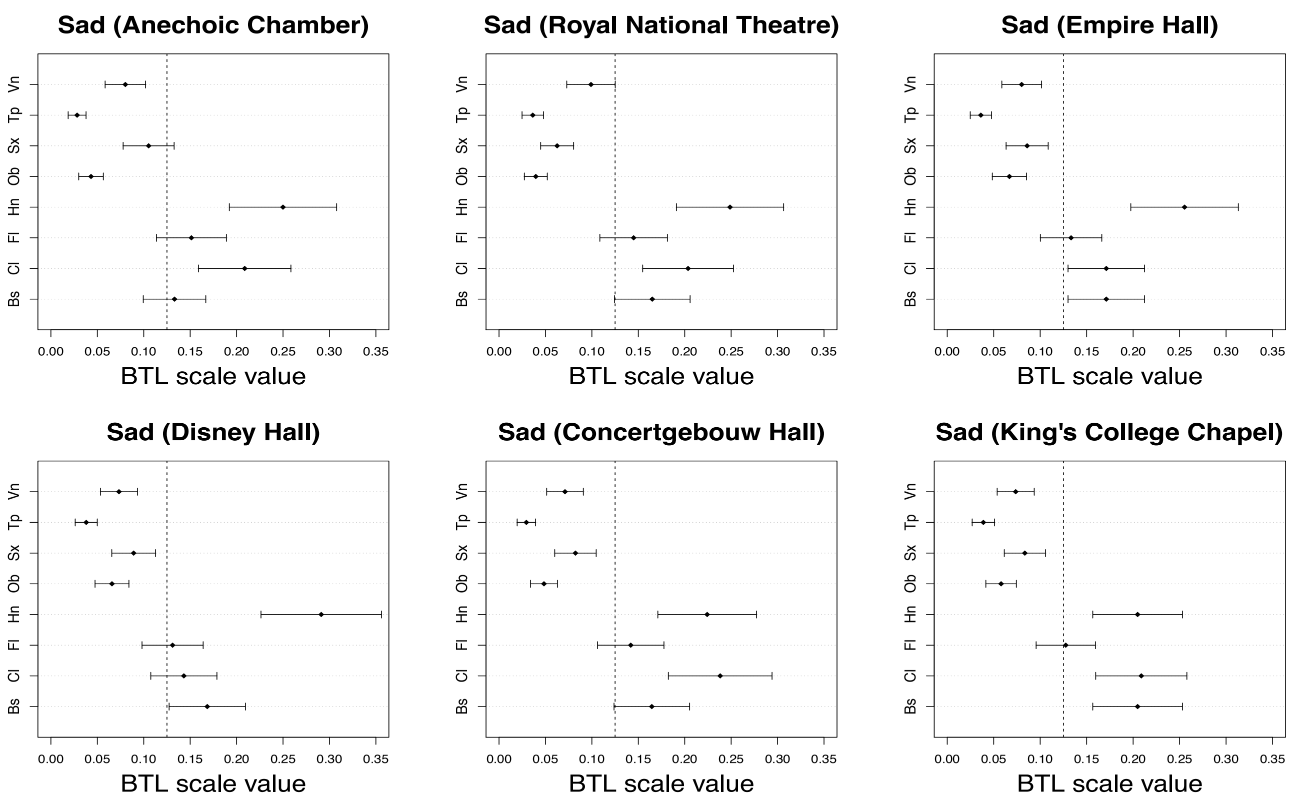

Figure 5.8: BTL scale values and the corresponding 95% confidence intervals with different halls for Sad.

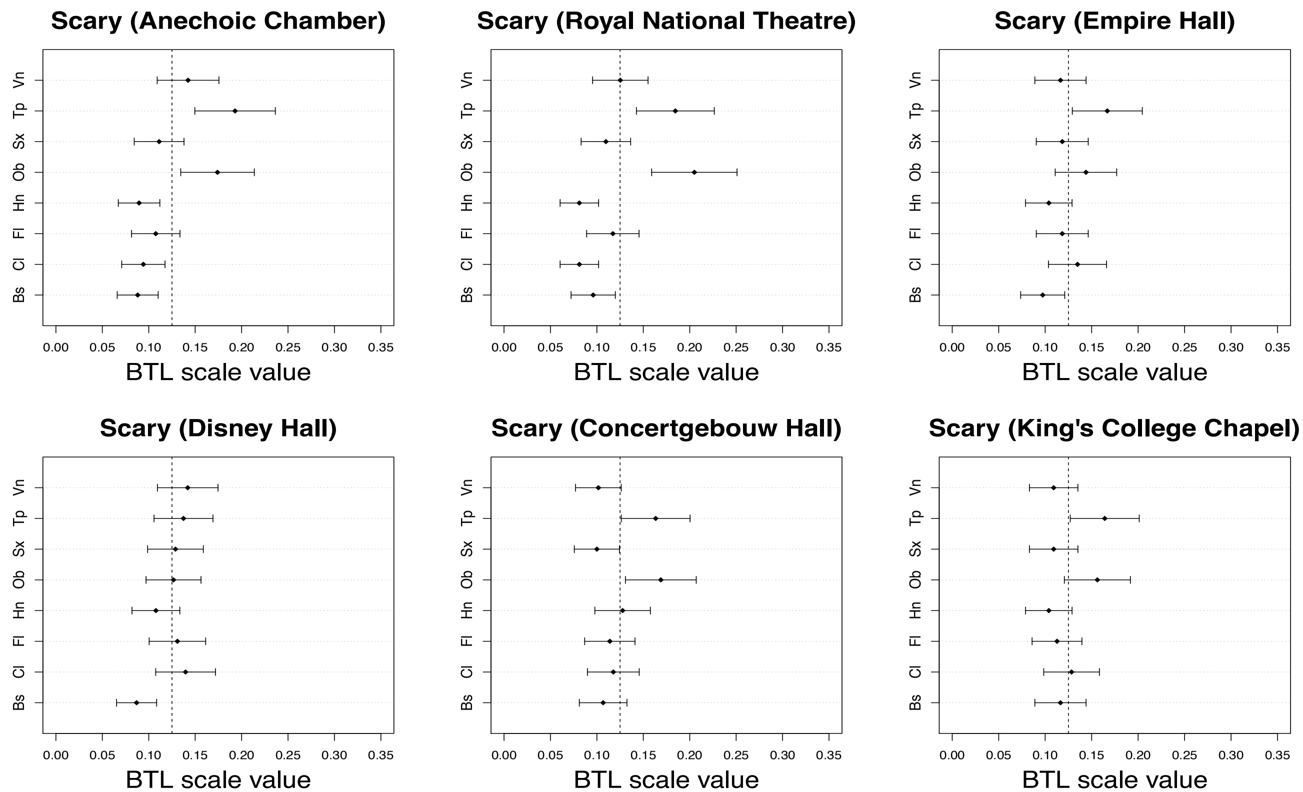

Figure 5.9: BTL scale values and the corresponding 95% confidence intervals with different halls for Scary.

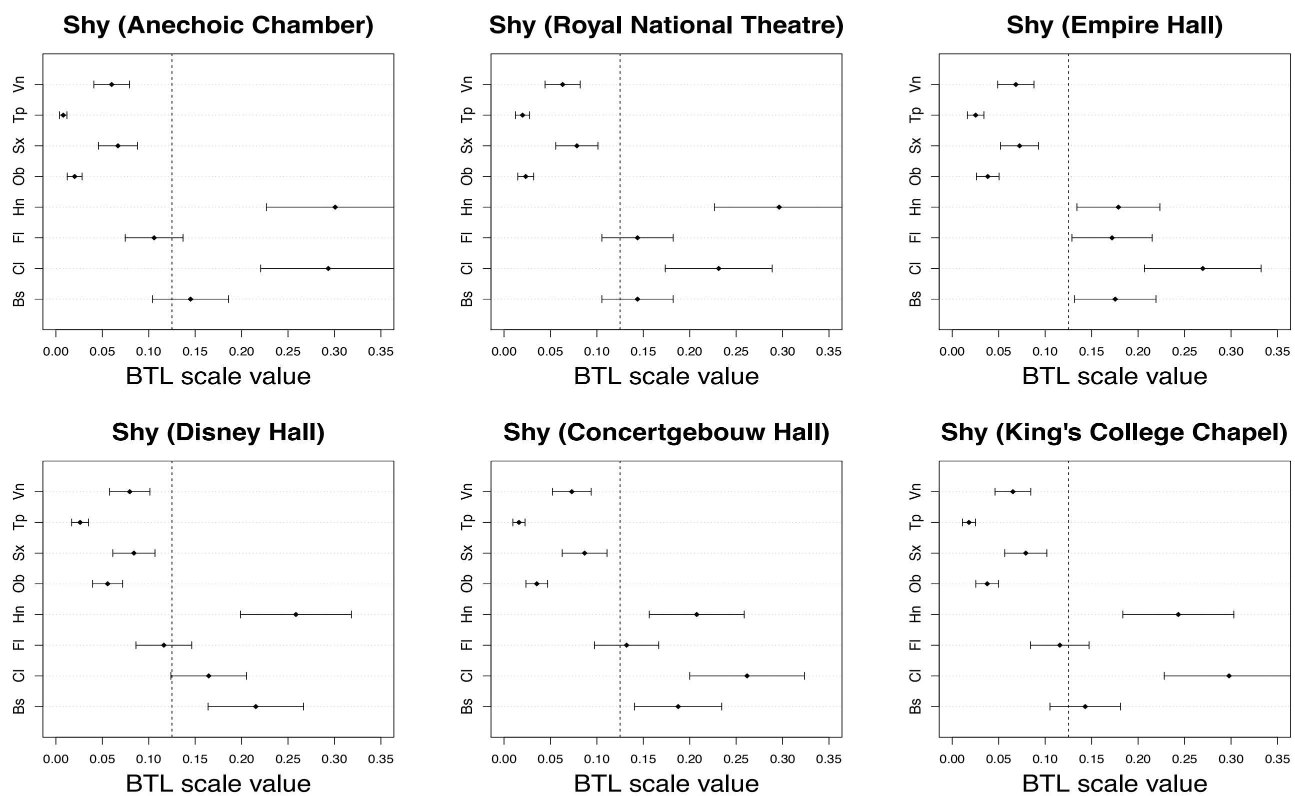

Figure 5.10: BTL scale values and the corresponding 95% confidence intervals with different halls for Shy.

We wanted to determine the number of times each instrument was significantly greater than the other seven instruments for each hall and emotional characteristic. As a preliminary step, the normality of the data was calculated for each instrument, emotional characteristic, and hall. Since most, though not all, were normally distributed (see Tables E.1 - E.6 in Appendix E), both parametric and nonparametric statistical tests (parametric: Paired t-tests, Pearson correlation ; nonparametric: Wilcoxon signed-rank tests, Spearman correlation) were used to analyze the voting data (i.e., the number of positive votes received by each instrument for each emotional category and hall). The results from the two tests showed some minor differences, but basically they were in agreement. Table 5.1 shows the paired t-test results and Table F.1 in Appendix F shows the Wilcoxon signed-rank test results. For each instrument, the maximum possible value is 7 and the minimum possible value is 0. For example, with the original anechoic sounds and the emotional characteristic Angry, the value of the trumpet is 7 since it was statistically significantly greater than all seven of the other instruments for the Anechoic subgraph in Figure 5.1. The maximum value for each hall and emotional characteristic is shown in bold and shaded.

Table 5.2 sums the sub-tables in Table 5.1 and shows the number of times each instrument was significantly greater than the other seven instruments over all six halls for each emotional characteristic. The maximum possible value is 42 and the minimum possible value is 0. For example, for Angry the trumpet was statistically significantly greater than all the other seven instruments most of the time, and its value is 34. Table 5.2 makes it obvious that the trumpet was ranked the highest for Angry, Comic, Happy, and Heroic, while it was a close second to the oboe for Scary. The clarinet was ranked highest for Calm, Mysterious, and Romantic, while the horn was highest for Sad, and they were tied for Shy. In general, the trumpet, oboe, saxophone, and violin were strong for the high-arousal characteristics, and the clarinet, horn, bassoon, and flute were strong for low-arousal characteristics. These two groups represent the high- and low-arousal timbres, though Table 5.2 also indicates that there is some crossover as well.

Anechoic Chamber (RT_{60}=0)								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	1	1	2	0	**6**	2	**6**	3
Heroic	1	0	0	0	1	1	**5**	4
Comic	1	0	2	0	4	3	**6**	4
Sad	3	5	3	**6**	0	2	0	2
Scary	0	0	0	0	3	0	**4**	0
Shy	4	**6**	3	**6**	1	2	0	2
Romantic	3	**6**	3	3	1	3	0	2
Mysterious	3	4	4	**5**	1	2	0	2
Angry	0	0	2	0	4	3	**7**	3
Calm	5	**6**	3	**6**	1	3	0	2

Royal National Theatre (RT_{60}=0.94)								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	0	0	3	0	5	3	**7**	3
Heroic	1	0	1	0	3	2	**5**	2
Comic	0	0	2	0	4	3	**6**	4
Sad	4	4	3	**5**	0	2	0	3
Scary	0	0	2	0	**6**	0	4	0
Shy	4	**6**	4	**6**	0	2	0	2
Romantic	3	**6**	2	3	1	3	0	2
Mysterious	2	**4**	**4**	**4**	0	1	0	1
Angry	1	0	1	1	1	1	**4**	1
Calm	**5**	**5**	2	4	1	2	0	2

Empire Hall (RT_{60}=1.31)								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	0	1	3	0	**5**	3	4	2
Heroic	0	0	0	0	5	2	**6**	0
Comic	1	1	2	0	**6**	3	**6**	3
Sad	4	4	4	**6**	1	1	0	1
Scary	0	**1**	0	0	0	0	0	0
Shy	4	**6**	4	4	1	2	0	2
Romantic	3	**5**	4	**5**	0	2	0	2
Mysterious	3	**6**	3	**6**	0	2	0	0
Angry	0	0	1	0	6	2	**7**	1
Calm	**5**	**5**	4	**5**	1	2	0	2

Disney Hall (RT_{60}=1.80)								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	1	1	1	0	2	1	**4**	**4**
Heroic	2	0	0	0	0	3	**6**	1
Comic	0	0	2	0	**6**	3	**6**	3
Sad	4	4	3	**7**	1	1	0	1
Scary	0	**1**	**1**	0	0	0	0	**1**
Shy	5	4	2	**6**	1	2	0	1
Romantic	4	**5**	2	4	1	2	0	1
Mysterious	4	**5**	3	**5**	0	0	0	0
Angry	0	0	1	0	1	2	**7**	1
Calm	**5**	**5**	2	**5**	1	2	0	1

Concertgebouw Hall (RT_{60}=2.32)								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	2	0	0	0	4	4	**7**	4
Heroic	1	0	1	1	2	1	**6**	1
Comic	0	0	3	0	**6**	4	**6**	4
Sad	4	**5**	4	**5**	1	2	0	1
Scary	0	0	0	0	**2**	0	**2**	0
Shy	4	**5**	4	**5**	1	2	0	2
Romantic	2	3	**4**	2	0	3	0	2
Mysterious	2	**6**	2	4	1	2	0	1
Angry	1	0	1	1	**4**	1	2	1
Calm	**5**	**5**	2	4	1	2	0	2

King's College Chapel (RT_{60}=5.44)								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	0	0	2	0	**6**	3	**6**	2
Heroic	0	0	2	0	4	2	**6**	2
Comic	1	1	1	0	4	4	**7**	4
Sad	**5**	**5**	4	**5**	0	1	0	1
Scary	0	0	0	0	**1**	0	0	0
Shy	4	**6**	4	**6**	1	2	0	2
Romantic	2	**7**	2	2	0	2	0	2
Mysterious	1	**6**	2	**6**	1	1	0	1
Angry	1	0	1	1	4	1	**7**	1
Calm	4	4	4	**5**	1	2	0	2

Table 5.1: Based on paired t-tests, how often each instrument was statistically significantly greater (for $p < 0.05$) than the others for each hall and emotional characteristic. The maximum possible value is 7 and the minimum possible value is 0. The maximum for each hall and emotional characteristic is shown in bold and shaded.

	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	4	3	11	0	28	16	**34**	18
Heroic	5	0	4	1	15	11	**34**	10
Comic	3	2	12	0	30	20	**37**	22
Sad	24	27	21	**34**	3	9	0	9
Scary	0	2	3	0	**12**	0	10	1
Shy	25	**33**	21	**33**	5	12	0	11
Romantic	17	**32**	17	19	3	15	0	11
Mysterious	15	**31**	18	30	3	8	0	5
Angry	3	0	7	3	20	10	**34**	8
Calm	29	**30**	17	29	6	13	0	11

Table 5.2: How often each instrument was statistically significantly greater than the others over the six halls. The maximum possible value is 42 and the minimum possible value is 0. The maximum for each emotional characteristic is shown in bold and shaded. This table is simply the sum of the individual sub-tables in Table 5.1. The circles indicate the high-and low-arousal timbre groupings.

We wanted to determine how similar were the sub-tables in Table 5.1 and the BTL data in Figures 5.1 - 5.10 for the different halls. Therefore, we ran correlations for both of these as well as for the voting data (i.e., the number of positive votes received by each instrument for emotional category and hall). In all cases, the correlations were statistically significant (at the $p < 0.0001$ level) and very strong, ranging from 90 to 96%, indicating a near-linear relationship and a very high level of agreement. In particular, Table 5.3 shows Pearson and Spearman correlation between the different halls based on the voting data, since it is the most precise and direct measure of correlation in the sense that it is correlation of the original data and not correlation of statistics based on the original data (e.g., the BTL data in Figures 5.1 - 5.10 and paired t-test data in Table 5.1).

Let's take another look at the question of the consistency of the listeners during this somehow long 95-minute listening test. As further evidence that the 31 subjects for each group were giving sincere and attentive responses, if they had been giving random responses at the end of the test due to fatigue, it would have decreased the number of significant differences in Table 5.1, making the distinctive patterns less clear and less consistent. As it turned out, they were amazingly consistent, indicating listeners remained vigilant and attentive.

Reverberation Types	Pearson Correlation	Spearman Correlation
Anechoic & Royal National Theatre	0.950	0.952
Anechoic & Empire Hall	0.949	0.944
Anechoic & Disney Hall	0.937	0.935
Anechoic & Concertgebouw Hall	0.929	0.911
Anechoic & King's College Chapel	0.933	0.918
Royal National Theatre & Empire Hall	0.936	0.939
Royal National Theatre & Disney Hall	0.917	0.921
Royal National Theatre & Concertgebouw Hall	0.944	0.933
Royal National Theatre & King's College Chapel	0.944	0.948
Empire Hall & Disney Hall	0.935	0.929
Empire Hall & Concertgebouw Hall	0.948	0.934
Empire Hall & King's College Chapel	0.962	0.947
Disney Hall & Concertgebouw Hall	0.915	0.902
Disney Hall & King's College Chapel	0.919	0.912
Concertgebouw Hall & King's College Chapel	0.962	0.944

Table 5.3: Pearson and Spearman correlation between the different halls based on the listener voting data. All correlations were statistically significant (for $p < 0.05$).

5.4 Discussion

This section discusses the results for this experiment on convolution reverberation. It compares contrasts these results with those found in Experiment 3 in this thesis (see Section 4.3). It also considers the implications of the results in wider applications and research.

5.4.1 Comparison of Results for Parametric and Convolution Reverberation

We were curious to find out to what degree the results for parametric (Experiment 3, see Section 4.3) and convolution reverberation were in agreement. To answer this question, we correlated the BTL rankings for our parametric and convolution reverberation experiments. Eight emotional categories were tested in both experiments since Angry and Calm were not included in the parametric experiment. Also, five reverberation times were tested in both experiments since we did not test a cathedral-length reverberation time in the parametric experiment.

Table 5.4 shows the correlation results for each category and overall. All the correlations were statistically significant (at the $p < 0.0001$ level), and seven out of eight categories were near 80% or more, indicating a very high level of agreement. Scary had the weakest correlation. The overall correlation was also very strong at about 85%.

We also correlated the number of significant differences from both experiments. Table 5.5 shows the number of significant differences from both experiments for easy comparison. There are obvious similarities in emotional categories other than Scary. The correlation results are shown in Table 5.6. Scary was not significantly correlated by this measure, but the other categories were near 90%. This indicates a very strong agreement in the patterns of significant differences, meaning the instrument footprints of emotional characteristics strongly match in parametric and convolution reverberation, except for Scary.

	Pearson Correlation	Spearman Correlation
Comic	0.914	0.874
Happy	0.914	0.892
Heroic	0.898	0.828
Mysterious	0.772	0.801
Romantic	0.896	0.907
Sad	0.849	0.896
Scary	0.378	0.459
Shy	0.862	0.900
Overall	0.837	0.872

Table 5.4: Pearson and Spearman correlation between parametric and convolution reverberation results for eight emotional categories (Angry and Calm were not included in Experiment 3) and five reverberation times (the 5.44s Cathedral reverb time was not tested in Experiment 3) based on the BTL rankings. All correlations were statistically significant (for $p < 0.05$).

	Parametric Reverberation									Convolution Reverberation								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn	Sum	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn	Sum
Comic	1	0	2	1	25	8	31	18	86	2	1	11	0	26	16	30	18	104
Happy	5	0	1	0	21	10	26	17	80	4	3	9	0	22	13	28	16	95
Heroic	6	0	2	4	21	5	34	12	84	5	0	2	1	11	9	28	8	64
Mysterious	4	26	12	6	3	2	0	4	57	14	25	16	24	2	7	0	4	92
Romantic	18	25	15	18	3	14	0	4	97	15	25	15	17	3	13	0	9	97
Sad	17	27	20	26	4	10	0	4	108	19	22	17	29	3	8	0	8	106
Scary	0	1	10	0	1	2	1	3	18	0	2	3	0	11	0	10	1	27
Shy	16	33	21	16	5	9	0	5	105	21	27	17	27	4	10	0	9	115

Table 5.5: Number of significant differences between parametric and convolution reverberation results for eight emotional categories (Angry and Calm were not included in Experiment 3) and five reverberation times (the 5.44s Cathedral reverb time was not tested in Experiment 3) based on the BTL rankings.

	Pearson Correlation	Spearman Correlation
Comic	****0.947**	****0.958**
Happy	****0.958**	****0.970**
Heroic	****0.944**	****0.905**
Mysterious	****0.737**	****0.886**
Romantic	****0.968**	****0.982**
Sad	****0.955**	****0.934**
Scary	-0.046	***0.302**
Shy	****0.867**	****0.891**
Overall	****0.884**	****0.883**

Table 5.6: Pearson and Spearman correlation between parametric and convolution reverberation results based on the significant differences in Table 5.5. **: $p < 0.05$; *:$0.05 < p < 0.1$.

5.4.2 Implications of the Results

I work has shown that different musical instruments have distinct emotional characteristics [17, 18, 19, 20, 21, 22, 23, 24, 32], and that reverberation can greatly change these characteristics [35, 36, 38, 78, 51, 81]. And while these emotional characteristics can be greatly changed with reverberation, the results in this experiment and in Experiment 3 (see Section 4.3) have shown that they are changed uniformly in about the same way for different instruments. In other words, the underlying instrument space does not change much with reverberation in terms of emotional characteristics. For example, added reverberation might bring out characteristics such as Angry or Romantic, but in a uniform way for the instruments, and not some instruments more than others. There seems to be a relatively consistent ranking of emotional characteristics between the instruments that holds with different reverberation lengths..

We should emphasize that our results are for basic-level professional headphones. Higher-quality professional headphones could perhaps show even more pronounced differentiation between the emotional characteristics though we expect it would also be in a uniform way for the instruments.

It makes sense that reverberation changes the character uniformly across the instruments: if it were not uniform, then performers in orchestras and chamber groups would not be able to practice in small rehearsal rooms in a reliable way if reverberation changed the character in an instrument-dependent way. Musicians would need to carefully rehearse in the performance venue, not just to get used to the hall, but to adjust the blends and balances of their emotional characteristics differently for each different venue.

At this point, one may wonder whether there are other halls that we did not test where the results might be instrument-dependent. It is possible. Perhaps in cathedrals with long reverberation lengths of 10 seconds or more the excessive smearing of the temporal and spectral envelopes effects instruments with strong isolated resonances more than other instruments with smoother spectral envelopes. More generally, perhaps there are non-music halls (e.g., stadiums, arenas, lecture theaters) where undesirable acoustic features (e.g., standing waves) result in instrument dependencies in the relative emotional characteristics. Perhaps the instrument-independent behavior of concert halls is what helps distinguish a good music venue from a poor one. Further work will be needed to confirm this, but it was remarkable how uniform the instruments were within the concert halls and anechoic chamber we tested.

In any case, it will be interesting to see whether these same overall results hold for other instruments, pitches, and dynamics, as we only tested Eb4-*forte* tones for eight instruments. It will also be interesting to see whether these results hold for other types of reverberation such as plate reverberation.

More broadly, perhaps the relatively consistent ranking of emotional characteristics between the instruments is what helps us to identify each instrument regardless of room reverberation. It seems each instrument has a characteristic footprint, that varies with pitch and dynamic level and other musical factors, which makes it identifiable. The columns of each sub-table in Table 5.1 represent the footprints of the emotional characteristics for each instrument and hall. In general, the footprints for each instrument were very similar for the different halls (e.g., the trumpet had larger values for Angry, Comic, Happy, and Heroic and small values for the other characteristics across all halls).

The columns of Table 5.2 represent the overall footprints of the emotional characteristics for each instrument for our Eb4-*forte* tones. The instruments clustered into two fairly distinct groups: those where the positive energetic emotional characteristics were strong (e.g., trumpet, oboe, saxophone, violin), and those where the low-arousal characteristics were strong (e.g., clarinet, horn, bassoon, flute). Looking in more detail, the trumpet, oboe, saxophone, and violin had similar deep footprints, but the trumpet's footprints were deepest, the oboe second deepest, and saxophone and violin similar but shallower. In the same way, the clarinet and horn had similar footprints, and the bassoon and flute were shallower. For Scary, the oboe and trumpet were uniquely strong. The saxophone and violin had the most even distribution, with medium values for most emotional categories.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 MP3 Compression

Based on the results we found in Experiment 1 (see Section 2.5), one might wonder whether the change in emotional characteristics we observed for individual instrument sounds is indicative of how MP3 compression would change different types of music. Is music with positive emotional characteristics such as Happy, Heroic, Romantic, Comic, or Calm more likely to be degraded by MP3 compression, while music that is Mysterious, Shy, Scary, or Sad actually reinforced to some degree in these characteristics? Is Angry music in general much less affected by MP3 compression? It will be interesting to explore these questions in future work. It will also be useful to investigate how other compression methods such as Advanced Audio Coding (AAC) compression change the perceived emotional characteristics of instrument sounds and music.

In conclusion, this study has investigated the impact of MP3 compression on the emotional characteristics of musical instruments, which has not been explored previously. Our work quantifies how much the emotional characteristics of instruments such as the saxophone have been changed by MP3 compression, and gives an indication of whether these changes are acceptable for particular bit rates and instruments. We believe that in addition to subjective quality evaluations [48, 40, 46, 47, 82] and discrimination measures [49, 50], changes in emotional characteristics can provide an additional metric for audio codec evaluation. Other than subjectively evaluating the quality loss of compressed sounds, or the changes in the timbre space, our study gives another perspective in evaluating the effect of lossy audio compression by considering the changes in the space of emotional characteristics.

The current study also helps provide the basis for content-based refinements of audio codecs in the future. As an example, if we know that the trumpet is particularly changed in emotional characteristics by compression at 32 Kbps, if we have a piece by Miles Davis with a prominent trumpet throughout, we may decide to use a higher bit rate to encode it. Or, future research may indicate how the trumpet could be compressed at 32 Kbps without substantially changing its emotional characteristics.

6.2 Artificial Reverberation

The results for Experiment 2 not only show that convolution reverberation and parametric reverberation somewhat effect the emotional characteristic of musical instrument sounds in a similar way (but more pronounced), but also gives a new definition of emotion footprint. Instead of viewing the footprints in instrument perspective as we have seen in our study of instrument space, we can now represent the footprints of emotional characteristics of the halls in a relative manner. Here comes to an interesting question: are the hall footprints instrument-independent? In other words, if we add convolution reverberation (e.g., a Disney Hall setting) to an instrument sound which we have not tested before (e.g., the piano sound), will it give similar result as we found here? This is an area of future work which will have a wide range of implications.

Another interesting thing to look at for Experiment 2 is how the data in the columns of Table 3.2 represents the particular colorings of these halls as compared to the general characteristics of a generic concert hall with reverberation times we picked. In other words, we would like to know how particular or general these footprints of emotional characteristics are. Of course we cannot tell at this stage. This suggests another future work though. One way to find out would be to re-run the same experiment on three or more halls for each of the five levels of reverberation times we tested in the experiment to determine the general trends and isolated differences. Nevertheless, the strong agreement of the convolution and parametric reverberation results already suggests that the trends emerging in the two studies are basically indicative of how the underlying emotional characteristics change with reverberation time. Within these trends, the colorations of the particular halls may bring out individual emotional characteristics such as Comic, Happy, Heroic, or Shy.

Let's take a look at the results we found in investigating the underlying instrument space of reverberation. In Experiments 3 and 4, we have shown that both parametric and convolution reverberation changed the emotional characteristics uniformly in about the same way for different instruments. That means the underlying instrument space does not change much with reverberation in terms of emotional characteristics.This result is contrasting to our first parametric reverberation study [38] and Experiment 2, which is somewhat surprising. In any case, it will be interesting to see if these same overall results hold for other instruments, pitches, and dynamics, as we only tested Eb4-*forte* tones for eight instruments. One of the future work would be to investigate if these results hold for other types of reverberation such as plate reverberation and spring reverberation.

The consistent rankings we found in these two experiments certainly help explain why listeners can identify each instrument in different reverberation environments. This raises an interesting question about instrument identification: When listeners identify an instrument, are they identifying its unique sound, timbre, relative emotional characteristics, or a combination of these? This is another potential area for further work.

This work also has implications for music emotion research of single musical instrument tones, where most studies do not explicitly state whether the tones are anechoic or with light reverberation, and assume it does not matter too much. The results suggest that this is a somewhat safe assumption if the relative emotional characteristics between instruments are the main consideration. Since reverberation changes the emotional characteristics of instruments uniformly in about the same way, then the relative space of emotional characteristics between the instruments will be maintained. So, we can use the numerous samples with light reverberation to compare instruments in terms of their emotional characteristics and expect about the same relative characteristics if they had been recorded in an anechoic chamber or a hall with different reverberation. Of course in other situations it really can make a difference. Since reverberation smears the temporal and spectral envelopes, it changes the timbre of the sound. Similarly, reverberation can greatly change the emotional characteristics of the sound. If changes in timbre or absolute emotional characteristics are the main consideration of the study, reverberation can indeed make a difference, and should be handled with caution and appropriate disclaimers should be included. In any case, it is useful to know which situations are relatively safe and which can be problematic.

Another great area for further work would be in the parameterization of the temporal and spectral envelope smearing of reverberation. With different amounts and lengths of reverberation, how much change can we expect in the temporal and spectral envelopes? Will it be uniform among different instruments as we found here, or instrument-dependent? To our knowledge, the temporal and spectral envelope smearing effects have not been parameterized in detail.

REFERENCES

[1] Kate Hevner, "Experimental studies of the elements of expression in music," *The American Journal of Psychology*, pp. 246–268, 1936.

[2] Isabelle Peretz, Lise Gagnon, and Bernard Bouchard, "Music and Emotion: Perceptual Determinants, Immediacy, and Isolation after Brain Damage," *Cognition*, vol. 68, no. 2, pp. 111–141, 1998.

[3] George Tzanetakis and Perry Cook, "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[4] Wolfgang Ellermeier, Markus Mader, and Peter 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.

[5] J-J Aucouturier, François Pachet, and Mark 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.

[6] Emmanuel Bigand, Sandrine Vieillard, François Madurell, Jeremy Marozeau, and A Dacquet, "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.

[7] Yi-Hsuan Yang, Yu-Ching Lin, Ya-Fan Su, and Homer H. Chen, "A Regression Approach to Music Emotion Recognition," *IEEE TASLP*, vol. 16, no. 2, pp. 448–457, 2008.

[8] Marcel Zentner, Didier Grandjean, and Klaus R Scherer, "Emotions evoked by the sound of music: characterization, classification, and measurement.," *Emotion*, vol. 8, no. 4, pp. 494, 2008.

[9] Julia C Hailstone, Rohani Omar, Susie MD Henley, Chris Frost, Michael G Kenward, and Jason D Warren, "It's not what you play, it's how you play it: Timbre affects perception of emotion in music," *The Quarterly Journal of Experimental Psychology*, vol. 62, no. 11, pp. 2141–2155, 2009.

[10] Suzanne Filipic, Barbara Tillmann, and Emmanuel Bigand, "Judging familiarity and emotion from very brief musical excerpts," *Psychonomic bulletin & review*, vol. 17, no. 3, pp. 335–341, 2010.

[11] Carol L Krumhansl, "Plink:" thin slices" of music," 2010.

[12] Tuomas Eerola and Jonna 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.

[13] Jonna K Vuoskoski and Tuomas Eerola, "Measuring music-induced emotion a comparison of emotion models, personality biases, and intensity of experiences," *Musicae Scientiae*, vol. 15, no. 2, pp. 159–173, 2011.

[14] Erkin Asutay, Daniel Västfjäll, Ana Tajadura-Jiménez, Anders Genell, Penny Bergman, and Mendel Kleiner, "Emoacoustics: A study of the psychoacoustical and psychological dimensions of emotional sound design," *Journal of the Audio Engineering Society*, vol. 60, no. 1/2, pp. 21–28, 2012.

[15] Chris Baume, "Evaluation of acoustic features for music emotion recognition," in *134th Audio Engineering Society Convention*. Audio Engineering Society, 2013.

[16] Judith Liebetrau, Johannes Nowak, Thomas Sporer, Matthias Krause, Martin Rekitt, and Sebastian Schneider, "Paired Comparison as a Method for Measuring Emotions," in *135th Audio Engineering Society Convention*, Oct 2013.

[17] Klaus R Scherer and James S Oshinsky, "Cue Utilization in Emotion Attribution from Auditory Stimuli," *Motivation and Emotion*, vol. 1, no. 4, pp. 331–346, 1977.

[18] Tuomas Eerola, Rafael Ferrer, and Vinoo Alluri, "Timbre and Affect Dimensions: Evidence from Affect and Similarity Ratings and Acoustic Correlates of Isolated Instrument Sounds," *Music Perception: An Interdisciplinary Journal*, vol. 30, no. 1, pp. 49–70, 2012.

[19] Bin Wu, Andrew Horner, and Chung Lee, “Musical Timbre and Emotion: The Identification of Salient Timbral Features in Sustained Musical Instrument Tones Equalized in Attack Time and Spectral Centroid,” in *International Computer Music Conference (ICMC), Athens, Greece*, 14-20 Sept 2014, pp. 928–934.

[20] Bin Wu, Chung Lee, and Andrew Horner, “The Correspondence of Music Emotion and Timbre in Sustained Musical Instrument Tones,” *Journal of the Audio Engineering Society*, vol. 62, no. 10, pp. 663–675, 2014.

[21] Bin Wu, C.W. Wun, Chung Lee, and A. Horner, “Investigating Correlation between Musical Timbres and Emotions,” in *International Society for Music Information Retrieval Conference (ISMIR), Curitiba, Brazil*, 2013, pp. 415–420.

[22] Bin Wu, Andrew Horner, and Chung Lee, “Emotional Predisposition of Musical Instrument Timbres with Static Spectra,” in *International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan*, Nov 2014, vol. 253-258.

[23] Chuck-jee Chau, Bin Wu, and Andrew Horner, “Timbre Features and Music Emotion in Plucked String, Mallet Percussion, and Keyboard Tones,” in *International Computer Music Conference (ICMC), Athens, Greece*, 14-20 Sept 2014, pp. 982–989.

[24] Chuck-jee Chau, Bin Wu, and Andrew Horner, “The Emotional Characteristics and Timbre of Nonsustaining Instrument Sounds,” *Journal of the Audio Engineering Society*, vol. 63, no. 4, pp. 228–244, 2015.

[25] Laura-Lee Balkwill and William Forde Thompson, “A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues,” *Music perception*, pp. 43–64, 1999.

[26] Janto Skowronek, Martin McKinney, and Steven Van De Par, “A Demonstrator for Automatic Music Mood Estimation,” *Proceedings of the International Conference on Music Information Retrieval*, 2007.

[27] Inger Ekman and Raine Kajastila, “Localization Cues Affect Emotional Judgments–Results from a User Study on Scary Sound,” in *Audio Engineer-*

ing Society Conference: 35th International Conference: Audio for Games. Audio Engineering Society, 2009.

[28] Yajie Hu, Xiaoou Chen, and Deshun Yang, "Lyric-Based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method," *Proceedings of ISMIR*, 2009.

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

[30] Magdalena Plewa and Bozena Kostek, "A Study on Correlation between Tempo and Mood of Music," in *133rd Audio Engineering Society Convention*, Oct 2012.

[31] Imre Lahdelma and Tuomas Eerola, "Single chords convey distinct emotional qualities to both naïve and expert listeners," *Psychology of Music*, p. 0305735614552006, 2014.

[32] Chuck-jee Chau and Andrew Horner, "The effect of pitch and dynamics on the emotional characteristics of piano sounds," 2015.

[33] Chuck-jee Chau, Ronald Mo, and Andrew Horner, "The emotional characteristics of piano sounds with different pitch and dynamics," *Journal of the Audio Engineering Society*, vol. 64, no. 11, pp. 918–932, 2016.

[34] Lothar Cremer, Helmut A Müller, and Theodore J Schaultz, *Principles and Applications of Room Acoustics*, vol. 1, Applied Science New York, 1982.

[35] Daniel Västfjäll, Pontus Larsson, and Mendel Kleiner, "Emotion and auditory virtual environments: affect-based judgments of music reproduced with virtual reverberation times," *CyberPsychology & Behavior*, vol. 5, no. 1, pp. 19–32, 2002.

[36] Ana Tajadura-Jiménez, Pontus Larsson, Aleksander Väljamäe, Daniel Västfjäll, and Mendel Kleiner, "When room size matters: acoustic influences on emotional responses to sounds," *Emotion*, vol. 10, no. 3, pp. 416–422, 2010.

[37] Francis Rumsey, "Reverberation and how to remove it," *J. Audio Eng. Soc*, vol. 64, no. 4, pp. 262–266, 2016.

[38] Ronald Mo, Bin Wu, and Andrew Horner, "The effects of reverberation on the emotional characteristics of musical instruments," *J. Audio Eng. Soc*, vol. 63, no. 12, pp. 966–979, 2016.

[39] Duncan Williams, "Affective potential in vocal production," in *Audio Engineering Society Convention 139*. Audio Engineering Society, 2015.

[40] H. Fuchs, W. Hoeg, and D. Meares, "Iso/mpeg subjective tests on multichannel audio systems: design and methodology," in *Broadcasting Convention, 1994. IBC 1994., International*, Sep 1994, pp. 152–157.

[41] Ben Sisario, "Downloads in decline as streamed music soars," *The New York Times (New York edition)*, p. B3, July 2014.

[42] Steven van de Par and Armin Kohlrausch, "Three approaches to the perceptual evaluation of audio compression methods," *The Journal of the Acoustical Society of America*, vol. 107, no. 5, pp. 2875–2875, 2000.

[43] Markus Erne, "Perceptual audio coders" what to listen for"," in *Audio Engineering Society Convention 111*. Audio Engineering Society, 2001.

[44] Chia-Ming Chang, Han-Wen Hsu, Kan-Chun Lee, Wen-Chieh Lee, Chi-Min Liu, Shou-Hung Tang, Chung-Han Yang, and Yung-Cheng Yang, "Compression artifacts in perceptual audio coding," in *Audio Engineering Society Convention 121*. Audio Engineering Society, 2006.

[45] Paulo Marins, "Characterizing the perceptual effects introduced by low bit rate spatial audio codecs," in *Audio Engineering Society Convention 131*. Audio Engineering Society, 2011.

[46] D.G. Kirby, F. Feige, and U. Wustenhagen, "Iso/mpeg subjective tests on multichannel audio coding systems: practical realisation and test results," in *Broadcasting Convention, 1994. IBC 1994., International*, Sep 1994, pp. 132–139.

[47] WH Schmidt and E Steffen, "Iso/mpeg subjective tests on multichannel audio coding systems: statistical analysis," in *International Broadcasting Convention - IBC '94*. 1994, pp. p. 158 – 163, IET.

[48] G Stoll and F Kozamernik, "Ebu subjective listening tests on low-bitrate audio codecs," 2003.

[49] Chung Lee and Andrew Horner, "Discrimination of mp3-compressed musical instrument tones," *Journal of the Audio Engineering Society*, vol. 58, no. 6, pp. 487–497, 2010.

[50] Chung Lee, Andrew Horner, and Bin Wu, "The effect of mp3 compression on the timbre space of sustained musical instrument tones," *Journal of the Audio Engineering Society*, vol. 61, no. 11, pp. 840–849, 2013.

[51] Ronald Mo, Bin Wu, and Andrew Horner, "The effects of reverberation time and amount on the emotional characteristics," in *42nd International Computer Music Conference, Utrecht, The Netherlands*, 2016, p. 12.

[52] McGill University, "The mcgill university master samples collection on dvd (3 dvds)," .

[53] "Prosonus shop," .

[54] "Rwc music database," .

[55] "University of iowa musical instrument samples," *University of Iowa*, 2004, http://theremin.music.uiowa.edu/MIS.html.

[56] Stephen McAdams, James W Beauchamp, and Suzanna Meneguzzi, "Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters," *The Journal of the Acoustical Society of America*, vol. 105, no. 2, pp. 882–897, 1999.

[57] Andrew Horner, James Beauchamp, and Richard So, "Detection of random alterations to time-varying musical instrument spectra," *The Journal of the Acoustical Society of America*, vol. 116, no. 3, pp. 1800–1810, 2004.

[58] James W Beauchamp, Andrew B Horner, Hans-Friedrich Koehn, and Mert Bay, "Multidimensional scaling analysis of centroid-and attack/decay-normalized musical instrument sounds," *The Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 3276–3276, 2006.

[59] Andrew B Horner, James W Beauchamp, and Richard HY So, "Detection of time-varying harmonic amplitude alterations due to spectral interpolations between musical instrument tones," *The Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 492–502, 2009.

[60] Andrew B Horner, James W Beauchamp, and Richard HY So, "Evaluation of mel-band and mfcc-based error metrics for correspondence to discrimination of spectrally altered musical instrument sounds," *Journal of the Audio Engineering Society*, vol. 59, no. 5, pp. 290–303, 2011.

[61] Marina Bosi and Richard E Goldberg, *Introduction to digital audio coding and standards*, vol. 721, Springer Science & Business Media, 2012.

[62] Chung Lee, Andrew Horner, and James Beauchamp, "Discrimination of musical instrument tones resynthesized with piecewise-linear approximated harmonic amplitude envelopes," *J. Audio Eng. Soc*, vol. 60, no. 11, pp. 899–912, 2012.

[63] "Lame mp3 encoder," .

[64] Patrik N Juslin and John Sloboda, *Handbook of music and emotion: Theory, research, applications*, Oxford University Press, 1993.

[65] Margaret M Bradley and Peter J Lang, "Affective norms for english words (anew): Instruction manual and affective ratings," Tech. Rep., Citeseer, 1999.

[66] "Cambridge academic content dictionary," .

[67] Fleiss L Joseph, "Measuring Nominal Scale Agreement among Many Raters," *Psychological Bulletin*, vol. 76, no. 5, pp. 378–382, 1971.

[68] Ralph A Bradley, "Paired comparisons: Some basic procedures and examples," *Nonparametric Methods*, vol. 4, pp. 299–326, 1984.

[69] Florian Wickelmaier and Christian 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.

[70] James W Beauchamp, "Analysis and synthesis of musical instrument sounds," in *Analysis, Synthesis, and Perception of musical sounds*, pp. 1–89. Springer, 2007.

[71] Takayuki Hidaka and Leo L Beranek, "Objective and subjective evaluations of twenty-three opera houses in europe, japan, and the americas," *The Journal of the Acoustical Society of America*, vol. 107, no. 1, pp. 368–383, 2000.

[72] Leo Beranek, *Concert halls and opera houses: music, acoustics, and architecture*, Springer Science & Business Media, 2004.

[73] "Altiverb 7," *Audio Ease*, 2016, https://www.audioease.com/altiverb/.

[74] Chuck-jee Chau and Andrew Horner, "The emotional characteristics of mallet percussion instruments with different pitches and mallet hardness," in *Proceedings of the International Computer Music Conference*, 2016, p. 401.

[75] Chuck-jee Chau, Bin Wu, and Andrew Horner, "The effects of early-release on emotion characteristics and timbre in non-sustaining musical instrument tones," in *Proc. 41st Int. Comp. Music Conf.(ICMC)*, 2015, pp. 138–141.

[76] Samuel JM Gilburt, Chuck-jee Chau, and Andrew Horner, "The effects of pitch and dynamics on the emotional characteristics of bowed string instruments," in *Proceedings of the International Computer Music Conference*, 2016, p. 405.

[77] Ronald Mo, Ga Lam Choi, Chung Lee, and Andrew Horner, "The effects of mp3 compression on perceived emotional characteristics in musical instruments," *Journal of the Audio Engineering Society*, vol. 64, no. 11, pp. 858–867, 2016.

[78] Ronald Mo, Richard HY So, and Andrew Horner, "An investigation into how reverberation effects the space of instrument emotional characteristics," *Journal of the Audio Engineering Society*, vol. 64, no. 12, pp. 988–1002, 2016.

[79] Ronald Mo, Ga Lam Choi, Chung Lee, and Andrew Horner, "The effects of mp3 compression on emotional characteristics," in *42nd International Computer Music Conference, Utrecht, The Netherlands*, 2016, p. 411.

[80] "Cool edit," *Adobe Systems*, 2000, https://creative.adobe.com/products/audition.

[81] Ronald Mo and Andrew Horner, "The effects of convolution reverberation on the emotional characteristics of musical instrument sounds," *Journal of the Audio Engineering Society*, 2017.

[82] Thilo Thiede, William C Treurniet, Roland Bitto, Christian Schmidmer, Thomas Sporer, John G Beerends, and Catherine Colomes, "Peaq-the itu standard for objective measurement of perceived audio quality," *Journal of the Audio Engineering Society*, vol. 48, no. 1/2, pp. 3–29, 2000.

APPENDIX A

NORMALITY TEST FOR EXPERIMENT 2

Anechoic Chamber										
	Angry	Calm	Comic	Happy	Heroic	Mysterious	Romantic	Sad	Scary	Shy
Bs	0.004	0.000	0.002	0.001	0.000	0.000	0.000	0.000	0.003	0.000
Cl	0.000	0.000	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Fl	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.002	0.000
Hn	0.002	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.004	0.000
Ob	0.000	0.000	0.004	0.002	0.000	0.000	0.000	0.000	0.001	0.001
Sx	0.001	0.000	0.006	0.008	0.000	0.000	0.000	0.000	0.006	0.001
Tp	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.012	0.000
Vn	0.000	0.000	0.005	0.004	0.000	0.000	0.000	0.000	0.003	0.000

Table A.1: Results from the Shapiro-Wilk test showing the significance values of the data for each instrument and emotional characteristic for an Anechoic Chamber. Values less than 0.05 indicate that the data were not normally distributed for that case.

Royal National Theatre										
	Angry	Calm	Comic	Happy	Heroic	Mysterious	Romantic	Sad	Scary	Shy
Bs	0.008	0.002	0.018	0.096	0.007	0.000	0.003	0.004	0.002	0.022
Cl	0.066	0.000	0.023	0.002	0.008	0.011	0.000	0.007	0.006	0.009
Fl	0.000	0.057	0.008	0.088	0.004	0.001	0.000	0.01	0.046	0.002
Hn	0.005	0.003	0.012	0.005	0.007	0.005	0.015	0.000	0.012	0.021
Ob	0.006	0.018	0.000	0.109	0.014	0.000	0.000	0.000	0.008	0.024
Sx	0.094	0.011	0.003	0.04	0.079	0.000	0.000	0.007	0.014	0.021
Tp	0.065	0.013	0.104	0.028	0.039	0.000	0.000	0.000	0.026	0.067
Vn	0.021	0.043	0.005	0.078	0.008	0.001	0.001	0.001	0.022	0.005

Table A.2: Results from the Shapiro-Wilk test showing the significance values of the data for each instrument and emotional characteristic for the Royal National Theatre.

Empire Hall										
	Angry	Calm	Comic	Happy	Heroic	Mysterious	Romantic	Sad	Scary	Shy
Bs	0.009	0.002	0.000	0.053	0.007	0.003	0.003	0.004	0.026	0.002
Cl	0.017	0.009	0.004	0.004	0.014	0.018	0.03	0.03	0.001	0.028
Fl	0.000	0.001	0.006	0.001	0.005	0.011	0.001	0.006	0.006	0.014
Hn	0.016	0.002	0.064	0.002	0.031	0.008	0.006	0.097	0.018	0.003
Ob	0.005	0.008	0.023	0.001	0.011	0.002	0.008	0.000	0.019	0.029
Sx	0.044	0.006	0.001	0.036	0.001	0.001	0.001	0.014	0.013	0.01
Tp	0.025	0.029	0.026	0.002	0.000	0.000	0.001	0.003	0.106	0.006
Vn	0.098	0.000	0.017	0.01	0.006	0.002	0.016	0.039	0.019	0.01

Table A.3: Results from the Shapiro-Wilk test showing the significance values of the data for each instrument and emotional characteristic for the Empire Hall.

Disney Hall										
	Angry	Calm	Comic	Happy	Heroic	Mysterious	Romantic	Sad	Scary	Shy
Bs	0.003	0.000	0.013	0.004	0.015	0.004	0.011	0.001	0.059	0.015
Cl	0.003	0.000	0.017	0.022	0.036	0.013	0.000	0.068	0.044	0.022
Fl	0.016	0.013	0.012	0.027	0.002	0.003	0.01	0.001	0.027	0.002
Hn	0.01	0.029	0.012	0.027	0.014	0.019	0.023	0.003	0.006	0.004
Ob	0.032	0.002	0.002	0.019	0.044	0.003	0.011	0.029	0.001	0.013
Sx	0.001	0.016	0.009	0.004	0.016	0.013	0.000	0.003	0.004	0.008
Tp	0.01	0.029	0.000	0.002	0.024	0.017	0.000	0.008	0.000	0.013
Vn	0.021	0.001	0.002	0.025	0.006	0.000	0.002	0.001	0.109	0.001

Table A.4: Results from the Shapiro-Wilk test showing the significance values of the data for each instrument and emotional characteristic for the Disney Hall.

Concertgebouw Hall										
	Angry	Calm	Comic	Happy	Heroic	Mysterious	Romantic	Sad	Scary	Shy
Bs	0.049	0.007	0.019	0.063	0.083	0.001	0.001	0.006	0.015	0.078
Cl	0.021	0.002	0.011	0.002	0.009	0.01	0.002	0.006	0.004	0.022
Fl	0.008	0.015	0.012	0.004	0.062	0.003	0.003	0.000	0.013	0.034
Hn	0.012	0.018	0.019	0.049	0.05	0.015	0.015	0.002	0.031	0.004
Ob	0.004	0.007	0.051	0.003	0.073	0.001	0.015	0.009	0.118	0.026
Sx	0.047	0.014	0.014	0.001	0.021	0.001	0.011	0.002	0.043	0.001
Tp	0.000	0.004	0.002	0.007	0.03	0.008	0.002	0.001	0.091	0.008
Vn	0.018	0.012	0.009	0.063	0.039	0.000	0.015	0.006	0.077	0.038

Table A.5: Results from the Shapiro-Wilk test showing the significance values of the data for each instrument and emotional characteristic for the Concertgebouw Hall.

King's College Chapel										
	Angry	Calm	Comic	Happy	Heroic	Mysterious	Romantic	Sad	Scary	Shy
Bs	0.002	0.015	0.001	0.012	0.003	0.000	0.000	0.004	0.037	0.04
Cl	0.017	0.007	0.019	0.056	0.032	0.000	0.000	0.003	0.006	0.004
Fl	0.000	0.011	0.001	0.028	0.002	0.000	0.000	0.001	0.014	0.014
Hn	0.003	0.001	0.003	0.019	0.023	0.001	0.019	0.005	0.015	0.006
Ob	0.003	0.003	0.056	0.021	0.02	0.000	0.003	0.02	0.008	0.011
Sx	0.013	0.036	0.000	0.034	0.017	0.001	0.005	0.001	0.008	0.017
Tp	0.002	0.003	0.001	0.002	0.003	0.000	0.001	0.000	0.007	0.008
Vn	0.009	0.034	0.001	0.008	0.03	0.000	0.000	0.003	0.006	0.021

Table A.6: Results from the Shapiro-Wilk test showing the significance values of the data for each instrument and emotional characteristic for the King's College Chapel.

APPENDIX B

WILCOXON SIGNED-RANK TEST FOR EXPERIMENT 2

	Anechoic RT_{60}=0	Royal National Theatre RT_{60}=0.94	Empire Hall RT_{60}=1.31	Disney Hall RT_{60}=1.80	Concertgebouw Hall RT_{60}=2.32	King's College Chapel RT_{60}=5.44
Angry	**22**	17	17	7	3	1
Calm	0	7	8	11	**19**	16
Comic	0	**15**	4	9	2	1
Happy	0	8	16	**18**	8	3
Heroic	0	4	**11**	3	5	1
Mysterious	0	4	16	16	27	**33**
Romantic	0	7	14	16	25	**27**
Sad	0	7	12	12	**25**	**25**
Scary	0	0	4	1	**9**	7
Shy	0	5	5	**13**	8	1

Table B.1: Based on Wilcoxon signed-rank tests, how often each instrument was statistically significantly greater (for $p < 0.05$) than the others for each hall impulse response and emotional characteristic. The maximum possible value is 40 and the minimum possible value is 0. The maximum for each emotional characteristic is shown in bold and shaded.

APPENDIX C

NORMALITY TEST FOR EXPERIMENT 3

	Happy		Heroic		Comic		Sad	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	0.95	-0.29	1.51	0.35	0.96	-0.24	-1.33	0.06
Cl	2.36	1.07	0.52	-1.56	0.27	-0.92	-0.52	-1.19
Fl	0.82	-0.87	0.93	-0.50	1.31	1.03	-1.50	0.29
Hn	0.11	-1.26	1.36	-0.30	0.97	-1.20	-2.23	1.28
Ob	-1.53	-1.19	-1.73	0.52	-0.10	-0.37	0.56	-0.42
Sx	0.01	-0.50	-0.61	-1.24	-0.17	-0.26	-0.95	-0.67
Tp	-0.24	-1.05	-2.09	-0.75	-1.80	-0.37	1.09	-1.51
Vn	0.08	-0.54	-0.25	-0.34	-1.11	0.21	3.13	2.55
	Scary		Shy		Romantic		Mysterious	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	1.12	-1.10	-1.09	1.16	0.50	-0.44	-1.18	-0.18
Cl	-0.44	-0.69	-2.65	1.70	-1.53	0.76	-1.01	-0.98
Fl	1.61	0.01	-0.34	0.36	-1.21	-0.23	-1.37	-0.01
Hn	0.54	-1.27	-1.99	0.13	-0.75	-1.38	0.07	-0.92
Ob	0.25	-0.72	1.03	-0.41	1.71	-0.10	-0.01	-1.05
Sx	-0.26	-0.84	-0.10	-0.25	-0.87	0.25	-0.16	-1.05
Tp	-0.35	-1.06	5.68	9.00	3.15	1.58	2.10	-0.65
Vn	0.18	-0.81	1.06	-1.33	1.36	-0.07	-0.67	-1.12

Table C.1: The normality of the data for each instrument and emotional characteristics for Anechoic. An absolute $Z_{skewness}$ greater than 1.96 indicates a significant skew (i.e., either positively or negatively skewed) at $p < 0.05$. An absolute $Z_{kurtosis}$ greater than 1.96 indicates a significant kurtosis (i.e., either leptokurtic or platykurtic) at $p < 0.05$.

	Happy		Heroic		Comic		Sad	
	Z_{skewness}	Z_{kurtosis}	Z_{skewness}	Z_{kurtosis}	Z_{skewness}	Z_{kurtosis}	Z_{skewness}	Z_{kurtosis}
Bs	0.56	-1.05	0.05	-0.62	0.34	0.58	0.01	-0.29
Cl	1.80	3.61	1.03	-0.97	2.16	0.65	-1.40	-0.42
Fl	0.04	-1.22	0.55	-1.00	-0.49	-0.07	-0.07	-0.64
Hn	1.97	0.71	1.15	-1.20	1.56	0.26	-2.50	1.44
Ob	-0.43	-0.96	-1.80	-0.66	-2.43	3.45	0.68	-0.83
Sx	-1.13	2.24	0.33	-0.36	1.62	-0.24	-0.89	-0.04
Tp	-2.07	-0.36	-1.93	-0.76	-1.18	-1.01	0.94	-1.33
Vn	-0.79	-0.66	-0.25	0.04	-0.61	0.31	1.15	-0.71
	Scary		Shy		Romantic		Mysterious	
	Z_{skewness}	Z_{kurtosis}	Z_{skewness}	Z_{kurtosis}	Z_{skewness}	Z_{kurtosis}	Z_{skewness}	Z_{kurtosis}
Bs	0.11	-1.05	-0.31	-0.80	-1.20	-0.35	-0.63	0.73
Cl	-0.62	-0.42	-1.60	0.12	-2.15	0.85	-1.70	0.38
Fl	0.68	-1.05	0.67	-1.19	-1.20	-0.24	0.38	-0.79
Hn	-0.63	-0.92	-1.22	-0.58	-0.27	-1.44	0.12	-1.67
Ob	-0.79	-0.75	1.87	0.06	1.50	-0.98	0.52	-0.23
Sx	0.23	-0.84	-0.14	-0.29	-0.14	-0.80	0.50	-0.84
Tp	-0.53	-0.75	2.02	0.61	3.53	3.01	2.57	0.10
Vn	0.75	-0.97	0.46	-0.87	2.08	2.89	-0.27	-0.69

Table C.2: The normality of the data for each instrument and emotional characteristics for Small Hall Front.

	Happy		Heroic		Comic		Sad	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	-0.01	-1.31	0.56	-0.51	0.24	-0.35	-0.30	-0.53
Cl	0.84	-0.38	1.71	0.23	2.15	1.27	-2.02	-0.26
Fl	-1.02	-0.26	0.43	0.90	0.42	-0.18	-0.15	0.42
Hn	2.05	0.40	0.25	-1.09	1.45	0.36	-1.36	0.34
Ob	-2.72	1.16	-1.75	-0.03	-0.94	-0.72	2.09	0.01
Sx	0.76	-1.31	0.68	-0.34	-1.29	0.87	-1.22	0.10
Tp	-0.36	-1.24	-2.67	0.12	-2.61	1.57	2.25	0.06
Vn	-0.11	-0.24	-0.20	-1.04	0.48	-0.41	2.12	-0.07
	Scary		Shy		Romantic		Mysterious	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	-0.13	-1.05	-1.85	-0.47	0.37	-1.00	0.74	-0.32
Cl	-0.70	-0.31	-1.57	-0.01	-1.30	-0.62	-1.38	-0.66
Fl	-0.17	-0.42	0.85	-1.21	0.06	-1.29	0.78	-0.83
Hn	0.14	-0.98	-0.85	-0.94	-0.25	-1.53	-0.72	-0.43
Ob	1.07	-0.71	0.80	-0.32	1.76	-0.81	2.87	0.73
Sx	-0.50	-1.01	1.52	0.19	0.64	-1.01	1.34	-0.10
Tp	0.05	-1.70	1.04	-1.43	1.47	-1.27	0.57	-1.29
Vn	-0.52	-0.12	1.44	-0.48	1.91	0.70	-0.06	-0.71

Table C.3: The normality of the data for each instrument and emotional characteristics for Small Hall Back.

	Happy		Heroic		Comic		Sad	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	-0.67	0.03	0.54	-0.12	0.70	-0.50	-0.34	-1.18
Cl	2.22	-0.19	1.57	0.65	1.36	-0.89	-1.26	-0.03
Fl	0.60	0.37	2.57	3.54	-0.69	-0.85	-0.42	0.61
Hn	1.60	-0.46	0.63	-0.84	2.11	-0.07	-2.12	0.58
Ob	-2.34	0.31	-0.97	-0.45	-0.05	-1.16	0.25	-1.32
Sx	-0.12	1.17	0.77	-0.40	-1.01	0.00	0.74	-1.52
Tp	-1.08	-1.22	-2.81	0.50	-2.42	0.33	3.44	2.36
Vn	-1.30	-0.40	-0.91	-0.35	-0.32	-0.59	0.16	-0.46
	Scary		Shy		Romantic		Mysterious	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	-1.18	-0.93	-2.19	1.53	0.19	0.60	-0.17	-0.87
Cl	-0.03	-1.49	0.06	-1.33	-2.60	0.51	-0.78	-1.36
Fl	0.61	-1.23	-1.69	1.16	0.79	-1.15	-0.48	-0.27
Hn	0.58	-0.75	-0.87	-0.84	-1.43	0.74	-1.02	-0.49
Ob	-1.32	-1.29	0.20	-1.33	1.34	1.07	0.52	-1.32
Sx	-1.52	-1.44	0.15	-0.88	0.55	-0.80	-0.10	0.29
Tp	2.36	-1.92	2.70	0.41	2.26	3.06	2.67	-0.03
Vn	-0.46	-1.31	0.48	-0.51	0.44	-0.74	1.03	0.51

Table C.4: The normality of the data for each instrument and emotional characteristics for Large Hall Front.

	Happy		Heroic		Comic		Sad	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	0.07	-1.29	-0.74	0.96	0.22	-0.55	-0.87	-0.24
Cl	2.18	0.35	1.38	-0.79	2.16	1.22	-1.91	0.69
Fl	0.79	-1.29	0.54	0.28	0.56	-0.91	-1.30	1.94
Hn	0.54	-0.71	0.25	-1.11	0.11	-1.03	-1.80	0.56
Ob	-1.26	-1.42	-0.82	-0.33	-0.91	-0.66	1.50	-0.27
Sx	-0.54	-0.15	-1.33	1.52	-1.07	-0.93	-0.12	0.56
Tp	-0.97	-0.35	-1.80	-0.08	-1.37	-0.22	1.42	-0.33
Vn	-1.95	1.25	-1.50	0.06	0.03	-0.76	0.34	-0.37
	Scary		Shy		Romantic		Mysterious	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	0.77	-0.60	-0.54	-0.25	-1.35	0.06	-0.40	-1.01
Cl	0.33	-1.03	-0.91	-0.05	-2.07	0.90	-1.81	-0.19
Fl	0.22	-1.30	0.71	-0.34	0.79	-0.79	-1.90	2.56
Hn	0.76	-0.80	-0.78	-0.07	-0.58	-0.82	-1.21	-0.02
Ob	0.14	-0.72	0.78	-0.51	1.78	0.52	0.81	-0.31
Sx	0.73	-1.07	-0.36	0.00	0.93	-0.59	0.34	-0.95
Tp	-0.92	-1.15	3.12	2.96	1.46	-0.54	2.45	-0.03
Vn	-0.84	0.92	0.96	-0.98	1.47	0.23	0.49	-0.25

Table C.5: The normality of the data for each instrument and emotional characteristics for Large Hall Back.

APPENDIX D

WILCOXON SIGNED-RANK TEST FOR EXPERIMENT 3

Anechoic	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	2	0	0	0	**4**	3	**4**	3
Heroic	0	0	0	0	5	0	**7**	3
Comic	0	0	0	0	4	2	**7**	2
Sad	3	**6**	3	5	1	2	0	1
Scary	0	1	0	0	1	1	1	**2**
Shy	4	**6**	4	4	1	3	0	1
Romantic	3	**5**	3	3	0	3	0	1
Mysterious	0	**6**	5	1	0	0	0	0

Small Hall Front	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	0	0	0	0	5	1	**7**	5
Heroic	1	0	0	0	4	0	**7**	3
Comic	0	0	0	0	**6**	3	**6**	4
Sad	2	**6**	4	5	1	1	0	1
Scary	0	0	**1**	0	0	0	0	**1**
Shy	2	**6**	4	4	1	1	0	1
Romantic	4	**5**	3	3	1	3	0	1
Mysterious	1	**5**	1	1	0	1	0	1

Small Hall Back	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	1	0	0	0	**4**	3	**4**	3
Heroic	1	0	0	1	6	2	**7**	2
Comic	1	0	2	0	4	2	**5**	3
Sad	3	**5**	**5**	**5**	1	2	0	1
Scary	0	0	**4**	0	0	0	0	0
Shy	4	**6**	5	1	1	1	0	1
Romantic	3	**5**	3	3	1	3	0	1
Mysterious	0	**7**	1	1	0	0	0	1

Large Hall Front	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	3	0	0	0	**5**	2	**5**	**5**
Heroic	2	0	1	1	3	2	**7**	2
Comic	0	0	0	0	5	0	**7**	5
Sad	**4**	**4**	**4**	**4**	1	3	0	1
Scary	0	0	**2**	0	0	0	0	0
Shy	3	**7**	4	3	1	3	0	1
Romantic	3	**5**	3	4	0	3	0	0
Mysterious	1	2	**4**	1	1	1	0	1

Large Hall Back	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Happy	1	0	1	0	4	1	**6**	1
Heroic	2	0	1	2	2	2	**7**	3
Comic	0	0	0	1	5	1	**7**	4
Sad	4	**6**	4	**6**	0	2	0	0
Scary	0	0	**2**	0	0	0	0	0
Shy	4	**7**	4	4	1	1	0	1
Romantic	4	**5**	3	4	1	2	0	1
Mysterious	2	**6**	2	2	2	0	0	2

Table D.1: Based on Wilcoxon signed-rank tests, how often each instrument was statistically significantly greater (for $p < 0.05$) than the others for Experiment 3. The maximum possible value is 7 and the minimum possible value is 0. The maximum for each reverberation type and emotional characteristic is shown in bold and shaded.

APPENDIX E

NORMALITY TEST FOR EXPERIMENT 4

	Angry		Calm		Comic		Happy		Heroic	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	0.26	-0.56	-1.79	0.71	0.43	-0.68	0.89	-1.17	-0.44	-0.49
Cl	1.40	-0.83	-3.05	2.58	1.85	1.49	2.57	1.14	1.57	-0.60
Fl	-0.11	-1.63	0.88	-0.15	0.17	0.08	-0.57	-0.69	0.36	-0.83
Hn	1.07	-1.47	-1.56	-0.23	3.09	0.94	2.28	0.86	1.03	-1.30
Ob	-1.26	0.15	1.71	0.02	-1.04	-1.00	-0.85	-0.90	-0.69	-1.13
Sx	-0.56	-0.74	1.42	-0.84	0.48	0.72	-0.39	-1.14	1.42	0.86
Tp	-3.54	2.38	3.77	2.20	-2.52	0.36	-2.93	0.59	-1.87	-0.93
Vn	-0.11	-0.88	1.35	0.80	-0.96	0.14	-0.89	0.03	0.25	-1.23
	Mysterious		Romantic		Sad		Scary		Shy	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	-1.29	-0.55	-0.89	-1.38	-2.05	-0.39	0.60	-0.78	-0.41	-1.02
Cl	-1.15	0.33	-1.81	0.36	-1.27	-0.85	0.09	-0.72	-0.37	-1.10
Fl	0.26	-0.41	0.37	-0.83	-0.69	0.12	0.97	0.52	0.80	-0.68
Hn	-2.02	0.62	-1.66	-0.55	-1.44	-0.92	0.78	-0.93	-2.59	1.98
Ob	2.07	0.93	0.40	-0.71	1.07	-0.30	-0.81	-1.30	3.10	2.04
Sx	0.74	-0.03	0.85	0.12	-0.24	-1.16	0.78	-1.10	0.80	1.24
Tp	3.41	2.33	2.80	0.44	3.58	2.82	-0.98	-1.34	3.91	2.90
Vn	0.03	-1.28	1.58	0.39	0.85	-0.50	-0.41	-0.13	1.64	0.18

Table E.1: The normality of the data for each instrument and emotional characteristics for Anechoic. An absolute $Z_{skewness}$ greater than 1.96 indicates a significant skew (i.e., either positively or negatively skewed) at $p < 0.05$. An absolute $Z_{kurtosis}$ greater than 1.96 indicates a significant kurtosis (i.e., either leptokurtic or platykurtic) at $p < 0.05$.

	Angry		Calm		Comic		Happy		Heroic	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	1.67	1.82	-0.39	-0.28	2.13	0.72	1.46	0.09	1.64	-0.64
Cl	1.55	-0.25	-0.68	0.07	3.01	2.38	1.05	-0.76	2.08	1.57
Fl	-0.54	-0.45	0.56	-0.52	0.17	-0.98	1.75	-0.25	0.95	-0.59
Hn	0.13	-0.98	-1.20	-0.24	1.68	-0.01	2.29	-0.17	0.85	-0.36
Ob	-1.41	-0.58	2.05	-0.51	-0.55	-1.30	-2.38	0.47	-2.10	-0.50
Sx	0.15	-1.02	0.76	-0.56	0.28	-0.88	2.34	1.59	-0.63	-0.57
Tp	-0.41	-1.17	2.88	2.05	-2.06	-0.21	-2.54	0.52	-1.35	-0.49
Vn	-0.07	-0.44	2.72	2.07	-0.13	-0.95	-0.37	-0.71	-1.66	0.10
	Mysterious		Romantic		Sad		Scary		Shy	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	-0.65	-1.00	-0.95	1.82	-1.44	0.60	-0.40	-0.89	-0.78	-1.08
Cl	-0.79	-1.42	-1.08	-1.20	-1.44	-0.14	0.53	-0.91	-1.44	0.94
Fl	0.88	-0.24	0.68	-0.61	-0.60	0.22	-1.32	0.42	-0.42	-0.61
Hn	-2.04	-0.19	-1.84	-0.11	-1.56	-0.32	0.54	-1.01	-0.76	-0.69
Ob	2.25	0.70	1.91	0.32	1.54	-0.48	-2.52	0.63	0.77	-1.03
Sx	1.49	0.14	0.62	-0.94	0.39	-0.05	-0.15	-0.60	0.83	-0.83
Tp	1.43	-1.12	3.23	3.75	1.79	-0.26	-0.87	-1.12	4.93	5.77
Vn	0.12	-0.40	0.45	-1.18	2.04	0.60	-0.77	-1.38	-0.84	0.18

Table E.2: The normality of the data for each instrument and emotional characteristics for the Royal National Theatre.

	Angry		Calm		Comic		Happy		Heroic	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	0.26	-0.85	-1.45	0.20	0.53	-0.60	0.81	0.19	0.47	-0.65
Cl	0.90	-1.17	-1.21	0.10	0.36	0.43	1.87	-0.44	1.23	-0.08
Fl	-0.21	-0.92	-1.01	1.50	0.73	1.27	0.38	-0.63	0.82	-1.03
Hn	1.51	-0.88	-1.98	0.60	3.45	2.99	1.61	-0.01	0.39	-0.94
Ob	-2.96	1.45	1.36	-0.75	-3.02	2.57	-0.30	-0.93	-0.78	-0.57
Sx	-0.94	-0.30	0.94	1.10	-0.49	0.30	0.71	-0.67	0.16	-1.01
Tp	-2.86	1.23	2.31	1.01	-3.79	4.94	-2.77	0.59	-1.44	-1.21
Vn	-0.31	-0.59	0.36	-0.86	1.19	1.82	0.78	-1.12	-1.77	0.86
	Mysterious		Romantic		Sad		Scary		Shy	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	-2.83	0.59	-1.05	0.24	-0.79	-1.22	-0.88	-0.63	-0.21	-0.38
Cl	-1.47	-0.55	-0.33	-1.17	-1.24	0.73	0.30	-0.86	-2.43	2.63
Fl	-0.25	0.15	0.27	-0.91	0.01	0.23	2.45	0.45	-0.81	0.99
Hn	-2.72	1.38	-1.86	-0.16	-3.01	1.39	0.31	-1.67	-1.17	-0.07
Ob	1.43	-0.82	0.65	-0.40	0.93	-0.42	-0.35	-1.00	3.83	3.38
Sx	1.80	0.63	1.25	-0.35	-0.47	-0.87	-0.57	0.24	0.31	-0.86
Tp	2.28	0.63	2.46	0.09	2.07	-0.35	-0.54	-1.58	4.79	6.74
Vn	0.57	-0.44	1.74	0.58	1.89	1.29	-0.33	-1.28	1.89	-0.22

Table E.3: The normality of the data for each instrument and emotional characteristics for the Empire Hall.

	Angry		Calm		Comic		Happy		Heroic	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	1.75	0.40	-1.67	0.18	2.65	2.50	0.34	-1.20	0.12	-0.74
Cl	0.81	-0.76	-0.33	-0.35	1.51	-0.49	0.01	-0.88	0.38	0.18
Fl	-0.19	1.19	-0.43	0.17	-0.11	-0.28	-0.35	-0.67	1.00	0.31
Hn	-0.17	-1.59	-1.57	0.02	2.25	0.50	1.95	0.08	0.88	-0.37
Ob	0.27	-0.80	0.76	-0.68	-1.14	0.08	-1.04	-0.24	-0.18	0.07
Sx	-1.53	0.24	-1.15	0.96	0.72	-0.30	-0.34	-0.81	-0.33	-1.26
Tp	-1.99	-0.18	4.31	6.27	-1.68	-0.11	-1.44	-0.58	-1.05	-0.88
Vn	-0.28	-0.91	0.72	-0.82	0.32	-0.41	-1.59	0.11	0.29	-0.43
	Mysterious		Romantic		Sad		Scary		Shy	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	-1.77	-0.44	-0.54	-0.36	-2.17	1.40	0.05	-1.15	-1.57	1.41
Cl	-2.02	0.00	-0.72	-0.95	-0.56	-0.03	-0.06	-1.45	-1.90	0.01
Fl	-2.16	1.07	-0.07	-0.23	1.44	-0.43	0.31	-0.11	1.50	-0.73
Hn	-1.63	0.10	-1.42	-0.86	-2.29	1.14	-0.34	-0.62	-2.60	1.90
Ob	1.14	-0.40	-0.96	-0.36	2.93	1.88	0.00	-0.80	1.74	0.80
Sx	0.79	-0.48	0.25	0.12	-0.51	-1.43	-0.66	-0.44	0.51	-1.27
Tp	2.38	0.29	2.10	0.58	1.42	-0.98	-0.54	-1.49	3.25	1.62
Vn	0.57	-1.47	0.52	-0.96	0.69	-0.07	1.57	-0.04	0.19	-0.32

Table E.4: The normality of the data for each instrument and emotional characteristics for the Disney Hall.

	Angry		Calm		Comic		Happy		Heroic	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	1.12	0.20	0.23	-1.30	1.83	0.27	1.61	-0.22	0.89	-0.15
Cl	1.61	-0.17	0.11	-0.88	2.72	0.86	3.44	1.72	0.25	-1.31
Fl	1.37	0.11	0.14	-0.31	0.45	-0.03	-0.26	-1.29	2.20	0.51
Hn	-0.14	-1.53	0.04	-1.29	1.15	-0.53	1.56	-0.24	0.69	-1.48
Ob	-1.46	-0.60	2.99	2.40	-3.95	4.19	-2.22	0.62	-0.81	-0.60
Sx	-0.19	0.36	-0.45	-0.31	2.03	-0.09	1.16	-0.32	-0.02	-1.01
Tp	-1.15	-1.05	2.49	1.06	-2.14	-0.64	-2.17	-0.21	-2.46	0.53
Vn	-0.38	-0.85	1.68	0.77	0.32	-0.09	-1.11	-0.08	-1.18	-0.50
	Mysterious		Romantic		Sad		Scary		Shy	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	0.24	-1.63	0.14	-0.58	0.42	-0.50	-0.04	-0.29	-1.68	-0.13
Cl	-1.47	-0.88	-0.90	0.13	-0.80	-1.17	0.72	-0.29	-2.01	0.51
Fl	0.00	-0.18	-0.04	-0.89	-0.49	0.15	-1.57	0.11	0.70	-0.42
Hn	-0.87	-0.40	0.32	-2.01	-1.18	-1.11	1.19	-1.21	-0.15	-0.82
Ob	2.23	0.16	3.33	2.34	1.51	0.08	-1.08	-0.99	2.83	1.88
Sx	0.88	-0.32	0.27	-0.56	-0.41	0.23	1.63	0.49	0.28	-0.74
Tp	3.14	1.13	1.72	-0.44	2.10	0.41	-0.68	-1.43	3.49	1.43
Vn	1.51	-0.22	0.82	-0.74	-0.09	-0.24	0.38	-0.51	1.27	-0.77

Table E.5: The normality of the data for each instrument and emotional characteristics for the Concertgebouw Hall.

	Angry		Calm		Comic		Happy		Heroic	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	0.15	-0.12	-0.36	0.23	2.04	0.93	1.39	-0.24	0.09	-0.40
Cl	4.04	5.40	-1.70	0.26	1.99	2.00	1.92	-0.59	0.68	-0.04
Fl	-0.35	-1.00	-0.36	-0.26	0.01	1.58	0.64	-0.17	1.49	0.89
Hn	-0.21	-0.98	-1.27	0.16	2.31	0.11	1.97	-0.47	0.24	-1.67
Ob	-2.19	-0.01	2.11	1.08	-1.71	0.56	-1.15	-1.37	-0.37	-0.78
Sx	0.09	-0.92	0.93	0.06	-0.43	-0.12	-0.37	-0.26	-0.56	0.04
Tp	-0.27	-1.75	3.11	1.36	-3.29	2.92	-2.66	0.72	-1.44	0.16
Vn	-0.98	-0.87	0.79	-0.24	-0.58	-1.01	-0.24	-0.34	-0.41	-0.47
	Mysterious		Romantic		Sad		Scary		Shy	
	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$	$Z_{skewness}$	$Z_{kurtosis}$
Bs	-0.77	-0.77	-1.54	-0.13	0.01	-0.69	-0.20	-1.30	-0.98	-0.64
Cl	-1.82	0.19	-0.63	-0.59	-0.48	-0.45	0.26	-1.15	-4.12	5.43
Fl	-0.26	-0.70	-2.22	2.11	1.23	-0.19	0.65	1.74	-0.33	-0.23
Hn	-0.58	-0.29	-0.73	-0.77	-0.98	-0.87	0.53	-1.62	-2.14	0.64
Ob	0.50	-1.02	0.78	-1.04	3.26	1.95	-1.70	-0.62	3.81	3.34
Sx	0.57	-0.57	1.47	0.86	1.19	-0.73	-0.24	0.39	-1.62	0.55
Tp	2.74	0.65	2.62	0.97	2.28	0.73	-0.52	-1.41	3.25	2.05
Vn	0.06	-1.11	1.36	-0.75	-0.65	-0.84	-1.20	-1.31	1.75	1.34

Table E.6: The normality of the data for each instrument and emotional characteristics for the King's College Chapel.

APPENDIX F

WILCOXON SIGNED-RANK TEST FOR EXPERIMENT 4

Anechoic Chamber (RT_{60}=0)								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Angry	0	0	2	0	4	3	**7**	3
Calm	5	**6**	3	**6**	1	3	0	2
Comic	1	0	2	0	5	3	**6**	3
Happy	1	1	1	0	**6**	2	**6**	3
Heroic	1	0	0	0	1	1	**4**	**4**
Mysterious	3	4	3	**5**	0	2	0	2
Romantic	3	**6**	3	3	1	3	0	2
Sad	3	5	3	**6**	1	2	0	2
Scary	0	0	0	0	3	0	**4**	0
Shy	4	**6**	4	**6**	1	2	0	2

Royal National Theatre (RT_{60}=0.94)								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Angry	1	0	1	1	1	1	**4**	1
Calm	4	**5**	2	4	1	2	0	2
Comic	0	0	2	0	4	3	**6**	4
Happy	0	0	3	0	5	3	**7**	3
Heroic	1	0	1	0	3	2	**5**	2
Mysterious	2	**4**	3	**4**	0	1	0	2
Romantic	3	**6**	2	3	1	2	0	2
Sad	4	4	3	**5**	0	2	0	3
Scary	0	0	1	0	**6**	0	4	0
Shy	4	**6**	4	**6**	0	2	0	2

Empire Hall (RT_{60}=1.31)								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Angry	0	0	1	0	6	2	**7**	1
Calm	**5**	**5**	4	**5**	1	2	0	2
Comic	1	1	1	0	**6**	3	**6**	3
Happy	0	1	3	0	**5**	2	4	2
Heroic	0	0	1	0	5	2	**6**	0
Mysterious	3	**6**	3	**6**	0	2	0	1
Romantic	3	**5**	4	**5**	0	2	0	2
Sad	4	4	4	**6**	1	1	0	1
Scary	0	**1**	0	0	0	0	0	0
Shy	4	**5**	4	4	1	2	0	2

Disney Hall (RT_{60}=1.80)								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Angry	0	0	1	0	1	2	**7**	1
Calm	**5**	**5**	2	**5**	1	2	0	1
Comic	0	1	2	0	**6**	3	**6**	3
Happy	1	1	1	0	2	1	**4**	**4**
Heroic	2	0	0	0	0	3	**6**	1
Mysterious	4	**5**	3	**5**	0	0	0	0
Romantic	4	**5**	2	3	1	2	0	1
Sad	4	4	3	**7**	1	1	0	1
Scary	0	**1**	**1**	0	0	0	0	**1**
Shy	5	4	2	**6**	1	2	0	1

Concertgebouw Hall (RT_{60}=2.32)								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Angry	1	0	1	1	**4**	1	2	1
Calm	**5**	**5**	2	4	1	2	0	2
Comic	0	0	3	0	**6**	4	**6**	4
Happy	2	0	1	0	4	4	**7**	4
Heroic	1	0	1	0	2	1	**6**	1
Mysterious	2	**6**	3	3	1	2	0	1
Romantic	2	3	**4**	2	0	3	0	2
Sad	4	**5**	4	**5**	1	2	0	2
Scary	0	0	0	0	**2**	0	**2**	0
Shy	4	**5**	4	**5**	1	2	0	2

King's College Chapel (RT_{60}=5.44)								
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Angry	1	0	1	1	4	1	**7**	1
Calm	4	4	4	**5**	1	2	0	2
Comic	1	1	1	0	4	4	**7**	4
Happy	0	0	1	0	5	3	**6**	3
Heroic	0	0	2	0	4	2	**6**	2
Mysterious	1	**6**	2	**6**	1	1	0	1
Romantic	2	**7**	2	2	1	2	0	2
Sad	**5**	**5**	4	**5**	0	1	0	1
Scary	0	0	0	0	**1**	0	0	0
Shy	4	**6**	4	**6**	1	2	0	2

Table F.1: Based on Wilcoxon signed-rank tests, how often each instrument was statistically significantly greater (for $p < 0.05$) than the others for each hall and emotional characteristic. The maximum possible value is 7 and the minimum possible value is 0. The maximum for each hall and emotional characteristic is shown in bold and shaded.