The Effects of MP3 Compression on Emotional Characteristics

Ronald Mo
Department of Computer Science and Engineering,
Hong Kong University of Science and Technology,
Hong Kong
ronmo@cse.ust.hk

Ga Lam Choi
Department of Computer Science and Engineering,
Hong Kong University of Science and Technology,
Hong Kong
glchoi@cse.ust.hk

Chung Lee
The Information Systems Technology and Design Pillar,
Singapore University of Technology and Design,
20 Dover Drive, Singapore 138682
chung_lee@sutd.edu.sg

Andrew Horner
Department of Computer Science and Engineering,
Hong Kong University of Science and Technology,
Hong Kong
horner@cse.ust.hk

ABSTRACT

Previous research has shown that MP3 compression changes the similarities of musical instruments, while other research has shown that musical instrument sounds have strong emotional characteristics. This paper investigates the effect of MP3 compression on music emotion. We conducted listening tests to compare the effect of MP3 compression on the emotional characteristics of eight sustained instrument sounds. We compared the compressed sounds pairwise over ten emotional categories. The results show that MP3 compression strengthened the emotional characteristics Sad, Scary, Shy, and Mysterious, and weakened Happy, Heroic, Romantic, Comic, and Calm. Interestingly, Angry was relatively unaffected by MP3 compression.

1. INTRODUCTION

Though most listeners know that extreme MP3 compression degrades audio quality, many are willing to compromise quality for convenience. This is reflected to the current portable music consumption trend where consumers are using internet music streaming services more frequently than buying CDs or downloads [1]. Major streaming services use MP3 compression.

As previous research has shown that musical instrument sounds have strong and distinctive emotional characteristics [2, 3, 4, 5, 6], it would be interesting to know how MP3 compression affects the emotional characteristics of musical instruments. In particular, we will address the following questions: 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?

2. BACKGROUND

2.1 MP3 Compression

MP3 compression reduces the size of audio files by discarding less audible parts of the sound. When an instrument sound is encoded using an MP3 codec, due to the lossy nature of MP3 compression, the sound is altered. The perceptual quality of lossy compression is a longstanding subject of digital audio research. Zwicker found a number of characteristics of the human auditory system including simultaneous masking and temporal masking formed a part of the psychoacoustic model of MP3 encoders [7]. Van de Par and Kohlrausch proposed a number of methods to evaluate different audio compression codecs [8].

Various studies have investigated the perceptual artifacts generated by low bit rate audio codecs. Erne produced a CD-ROM that demonstrates some of the most common coding artifacts in low bit rate codecs. They explained and presented audio examples for each of the coding artifacts separately using different degrees of distortion [9]. Chang et al. constructed models of the audible artifacts generated by temporal noise shaping and spectral band replication, which are far more difficult to model using existing encoding systems [10]. Marins carried out a series of experiments aiming to identify the salient dimensions of the perceptual artifacts generated by low bit rate spatial audio codecs [11].

Previous studies have also subjectively evaluated the perceptual quality loss in MP3 compression [12, 13, 14, 15]. A recent study evaluated the discrimination of musical instrument tones after MP3 compression using various bit rates [16]. A following study [17] compared dissimilarity scores for instrument tone pairs after MP3 compression to determine whether instrument tones sound more or less similar after MP3 compression, and found that MP3 can change the timbre of musical instruments.

2.2 Music Emotion and Timbre

Researchers have considered music emotion and timbre together in a number of studies, which are well-summarized in [6].

3. METHODOLOGY

3.1 Overview

We conducted listening tests to compare pairs of original and compressed instrument sounds over different emotional categories. Paired comparisons were chosen for simplicity. This section gives further details about the listening test.

3.2 Listening Test

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 [16, 17, 18, 19, 20, 21, 22, 23, 24]. Using the same samples makes it easier to compare results.

Compressed sounds were encoded and decoded using the LAME MP3 encoder [25]. Instrument sounds were compressed with three different bit rates (32, 56, and 112 Kbps). These three bit rates gave near-perfect (for 32 Kbps), intermediate (for 56 Kbps), and near-random discrimination (for 112 Kbps) in a previous discrimination study of these MP3-compressed musical instrument sounds [16].

The subjects compared the stimuli in terms of ten emotional categories: Happy, Heroic, Romantic, Comic, Calm, Mysterious, Shy, Angry, Scary, and Sad. We carefully picked the emotional categories based on terms we felt composers were likely to write as expression marks for performers (e.g., *mysteriously*, *shyly*, etc.) and at the same time would be readily understood by lay people. The subjects were provided with an instruction sheet containing definitions of the ten emotional categories from the Cambridge Academic Content Dictionary [26]. Every subject made paired comparisons between the sounds.

The test asked listeners to compare four types of compressed sounds for each instrument over ten emotion categories. During each trial, subjects heard a pair of sounds from the same instrument with different types of compression (no compression, 112Kbps, 54Kbps, and 32Kbps) 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 [19, 4, 27].

Each combination of two different compressions was presented for each instrument and emotion 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.

4. RANKING RESULTS FOR THE EMOTIONAL CHARACTERISTICS WITH DIFFERENT OF MP3 BIT RATES

We ranked the compressed sounds by the number of positive votes they received for each instrument and emotion, and derived scale values using the Bradley-Terry-Luce (BTL) statistical model [28, 29]. For each instrument-emotion 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 emotion 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% confidence intervals for the compressed sounds using the method proposed by Bradley [28].

Fig. 1 to 6 show the BTL values and corresponding 95% confidence intervals for each emotional category. Based on the data in Fig. 1 - 6, Table 1 shows the number of instruments that were significantly different from the original sound (i.e., the 95% confidence intervals of the original and compressed sounds did not overlap) for each compression rate and emotional category. The table shows that there were relatively few differences for 112 and 56Kbps, but most of the instruments were significantly different for 32Kbps in nearly every category. This agrees with the results of Lee et al. [16], which found very good discrimination between the original and compressed sounds at 32Kbps, but poor discrimination at 56 and 112Kbps.

To help understand which instruments and emotional categories were most and least affected by MP3 compression, Table 2 shows the number of compressed sounds that were significantly different from the original sound for each instrument and emotional category. Based on the data, the clarinet was the most affected instrument (closely followed by the oboe and saxophone), while the horn was by far the least affected instrument. Lee et al. [16] also found the MP3-compressed horn relatively more difficult to discriminate from the original compared to other instruments. Among emotional categories in Table 2, Happy and Calm were the most affected, and Angry was by far the least affected.

Fig. 7 shows how often the original instruments sounds were statistically significantly greater than the three compressed sounds (This is different than the sum in the final column of Table 2 which counts any significant difference - both those significantly greater and those significantly less). Positive values indicate an increase in the emotional characteristics, and negative values a decrease. Again, Happy and Calm were the most affected emotional characteristics. 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., Sad, Scary, Shy, Mysterious) tended to increase with more MP3 compression. As an exception, Angry was relatively unaffected by MP3 compression for the compression rates we tested.

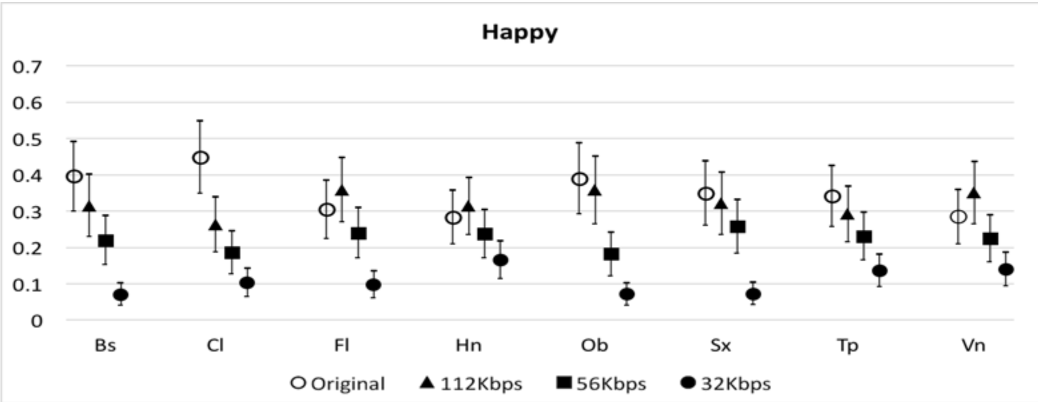

Figure 1. BTL scale values and the corresponding 95% confidence intervals for the emotional category Happy.

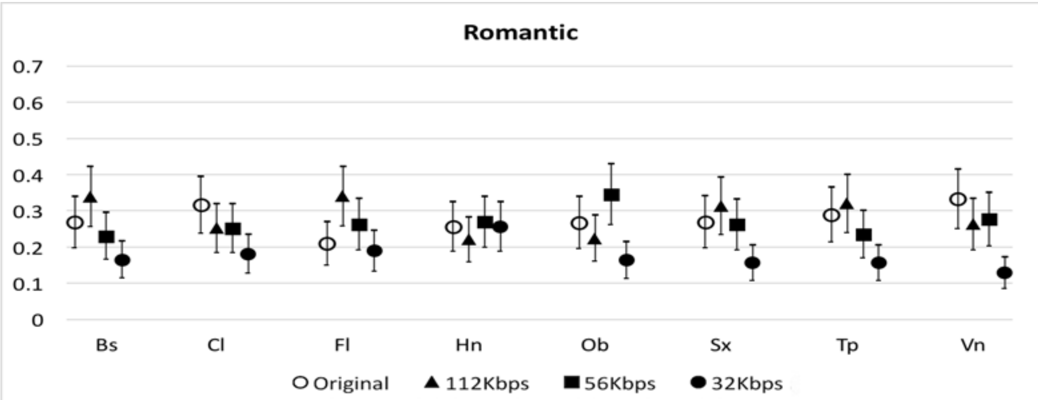

Figure 2. BTL scale values and the corresponding 95% confidence intervals for Romantic.

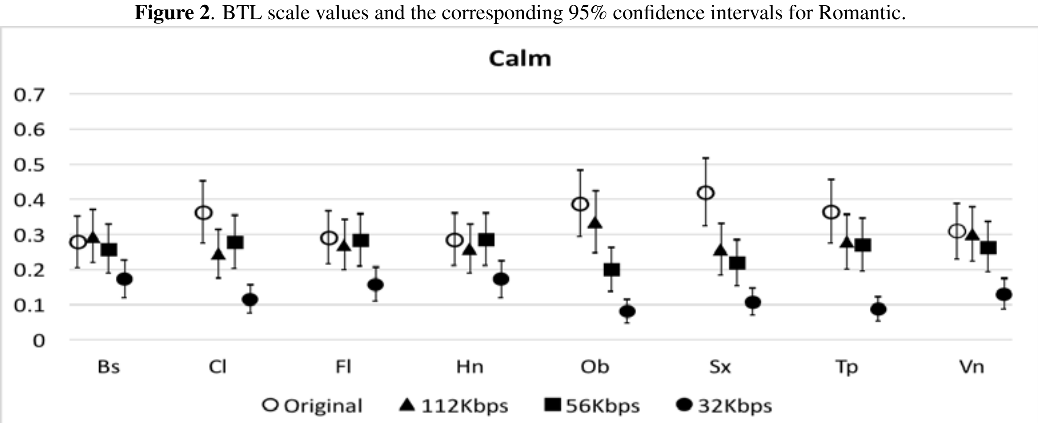

Figure 3. BTL scale values and the corresponding 95% confidence intervals for Calm.

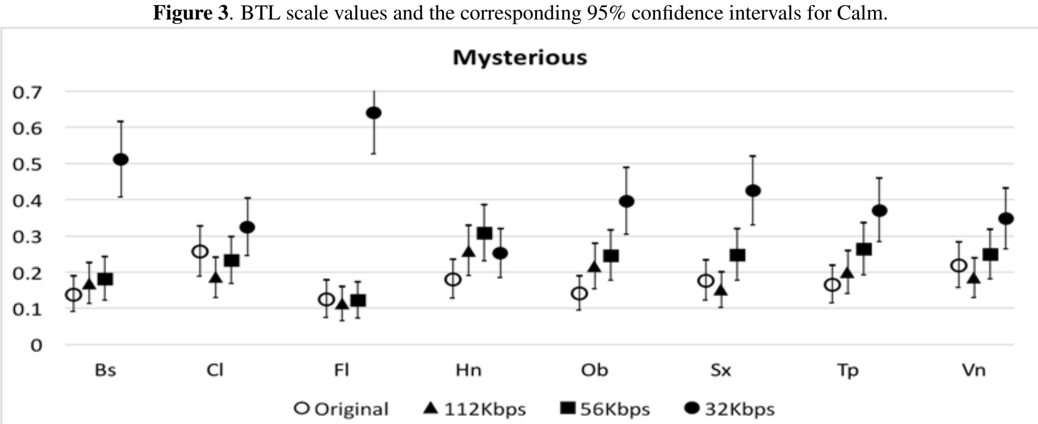

Figure 4. BTL scale values and the corresponding 95% confidence intervals for Mysterious.

5. DISCUSSION

The goal of our work was to understand how emotional characteristics of instruments vary with MP3 compression. Based on the Table 2 and Figure 7, our main findings are as follows:

1. Negative and neutral emotional characteristics (Sad, Scary, Shy, and Mysterious) increased with more MP3 compression in the samples we tested (see Figure 7).
2. Positive emotional characteristics (Happy, Heroic, Romantic, Comic, and Calm) decreased with more MP3 compression in the samples we tested (see Figure 7).
3. Angry was relatively unaffected by MP3 compression for the rates we tested (see Figure 7).
4. MP3 compression affected some instruments more and others less. The clarinet, oboe, and saxophone were most affected, and the horn by far the least affected (see Table 2).

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 and Calm while increasing others such as Mysterious by changing the quality of sounds to be somewhat different and unnatural. The above results demonstrate how a categorical emotional model can give more emotional nuance and detail than a 2D dimensional model with only Valence and Arousal. For example, Scary and Angry are very close to each one another in terms of Valence and Arousal, yet Scary was significantly increased with more compression while Angry was relatively unaffected. The results suggest that they are distinctively different emotional characteristics.

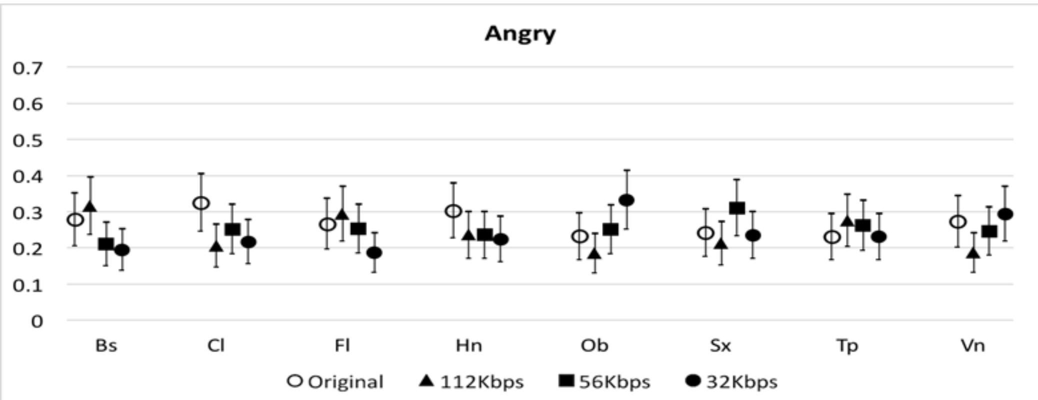

Figure 5. BTL scale values and the corresponding 95% confidence intervals for Angry.

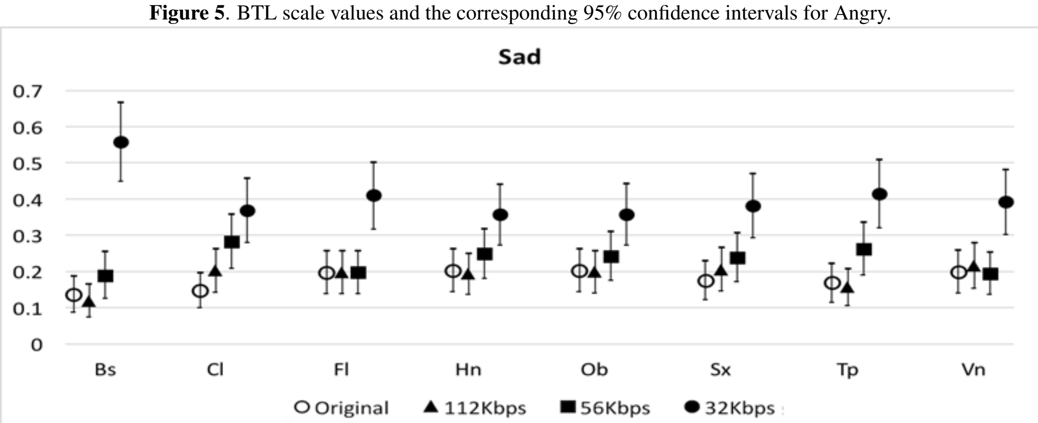

Figure 6. BTL scale values and the corresponding 95% confidence intervals for Sad.

Emotional Category	112Kbps	56Kbps	32Kbps
Happy	1	3	8
Heroic	0	1	7
Romantic	1	0	6
Comic	0	2	5
Calm	2	2	8
Mysterious	0	2	6
Shy	1	0	8
Angry	1	0	1
Scary	0	2	7
Sad	0	1	8
Avg.	0.6	1.3	6.4

Table 1. The number of instruments that were significantly different from the original sound (i.e., the 95% confidence intervals of the original and compressed sounds did not overlap) for each compression rate and emotional category

Emotional Category	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn	Total
Happy	2	3	1	1	2	1	1	1	12
Heroic	1	1	2	0	1	1	1	1	8
Romantic	1	1	1	0	1	1	1	1	7
Comic	1	1	1	1	1	1	1	1	8
Calm	1	2	1	1	2	3	1	1	12
Mysterious	1	0	1	0	2	1	1	1	7
Shy	1	1	1	1	1	2	1	1	9
Angry	0	1	0	0	0	0	0	0	1
Scary	1	1	1	0	1	1	2	1	8
Sad	1	2	1	1	1	1	1	1	9
Total	10	13	10	5	12	12	10	9	

Table 2. The number of compressed sounds that were significantly different from the original sound (i.e., the 95% confidence intervals of the original and compressed sounds did not overlap) for each instrument and emotional category.

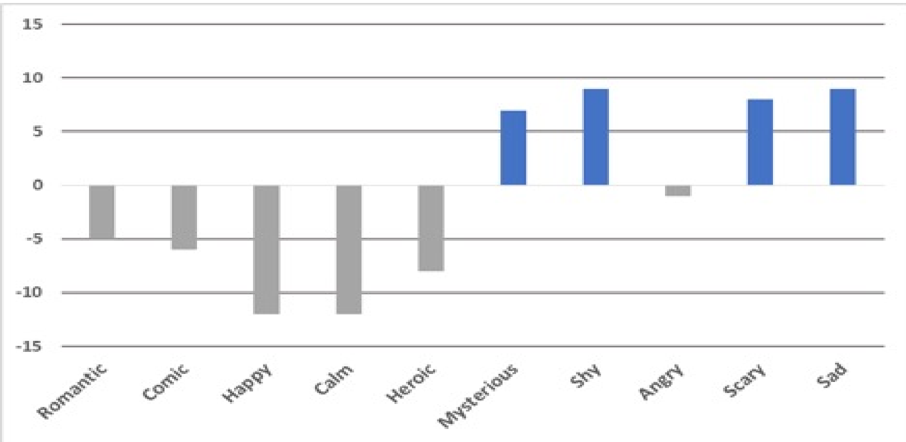

Figure 7. The number of significant differences between the original and compressed sounds, where strengthened emotional categories are positive, and weakened emotional categories are negative.

6. REFERENCES

[1] B. Sisario, "Downloads in Decline as Streamed Music Soars," *The New York Times (New York edition)*, p. B3, July 2014.

[2] T. Eerola, R. Ferrer, and V. 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.

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

[4] C.-j. Chau, B. Wu, and A. 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.

[5] B. Wu, C. Lee, and A. 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.

[6] C.-j. Chau, B. Wu, and A. Horner, "The Emotional Characteristics and Timbre of Nonsustaining Instru-

ment Sounds," *Journal of the Audio Engineering Society*, vol. 63, no. 4, pp. 228–244, 2015.

[7] T. Zwicker, "Psychoacoustics as the basis for modern audio signal data compression," *The Journal of the Acoustical Society of America*, vol. 107, no. 5, pp. 2875–2875, 2000.

[8] S. van de Par and A. 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.

[9] M. Erne, "Perceptual Audio Coders" What to listen for"," in *Audio Engineering Society Convention 111*. Audio Engineering Society, 2001.

[10] C.-M. Chang, H.-W. Hsu, K.-C. Lee, W.-C. Lee, C.-M. Liu, S.-H. Tang, C.-H. Yang, and Y.-C. Yang, "Compression artifacts in perceptual audio coding," in *Audio Engineering Society Convention 121*. Audio Engineering Society, 2006.

[11] P. Marins, "Characterizing the Perceptual Effects Introduced by Low Bit Rate Spatial Audio Codecs," in *Audio Engineering Society Convention 131*. Audio Engineering Society, 2011.

[12] 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.

[13] D. 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.

[14] W. Schmidt and E. Steffen, "ISO/MPEG subjective tests on multichannel audio coding systems: statistical analysis," 1994.

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

[16] C. Lee and A. Horner, "Discrimination of MP3-Compressed Musical Instrument Tones," *Journal of the Audio Engineering Society*, vol. 58, no. 6, pp. 487–497, 2010.

[17] C. Lee, A. Horner, and B. 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.

[18] S. McAdams, J. W. Beauchamp, and S. 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.

[19] A. Horner, J. Beauchamp, and R. 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.

[20] J. W. Beauchamp, A. B. Horner, H.-F. Koehn, and M. 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.

[21] A. B. Horner, J. W. Beauchamp, and R. H. 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.

[22] ——, "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.

[23] M. Bosi and R. E. Goldberg, *Introduction to digital audio coding and standards*. Springer Science & Business Media, 2012, vol. 721.

[24] C. Lee, A. Horner, and J. 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.

[25] LAME MP3 Encoder. [Online]. Available: http://lame.sourceforge.net/

[26] Cambridge Academic Content Dictionary. [Online]. Available: http://dictionary.cambridge.org/dictionary/american-english

[27] R. Mo, B. Wu, and A. Horner, "The Effects of Reverberation on the Emotional Characteristics of Musical Instruments," *J. Audio Eng. Soc*, vol. 63, no. 12, pp. 966–979, 2016. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18055

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

[29] F. Wickelmaier and C. Schmid, "A Matlab Function to Estimate Choice Model Parameters from Paired-comparison Data," *Behavior Research Methods, Instruments, and Computers*, vol. 36, no. 1, pp. 29–40, 2004.