

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

RONALD MO¹, GA LAM CHOI², CHUNG LEE³, AND ANDREW HORNER⁴, *AES Member*
(ronmo@cse.ust.hk) (glchoi@ust.hk) (im.lee.chung@gmail.com) (horner@cse.ust.hk)

¹*Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

²*Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

³*United Overseas Bank, Singapore*

⁴*Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

Musical instrument sounds have distinct timbral and emotional characteristics, and when audio processes are applied to them their timbral and emotional characteristics are changed. This paper investigates the effects of MP3 compression on the emotional characteristics of eight sustained instrument sounds using a listening test. We compared the compressed sounds pairwise over ten emotional categories at several bit rates. The results showed that MP3 compression strengthened neutral and negative emotional characteristics such as Mysterious, Shy, Scary, and Sad; and weakened positive emotional characteristics such as Happy, Heroic, Romantic, Comic, and Calm. Interestingly, Angry was relatively unaffected by MP3 compression. Probably, the background “growl” artifacts added by MP3 compression decreased positive emotional characteristics and increased negative characteristics such as Mysterious and Scary. For instruments, MP3 compression effected some instruments more and others less. The trumpet was the most effected and the horn by far the least.

0 INTRODUCTION

Different musical instruments have different sounds, timbres, and 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. A variety of previous research has considered different aspects of music emotion [1–8], including many papers on the connection between timbre and music emotion [9–17]. In particular, several studies have focused on individual musical instrument sounds where emotional characteristics were compared among various instruments, sometimes at different

pitch and dynamic levels [18–26]. This work has found distinctive emotional characteristics for each instrument and strong correlations between particular emotional characteristics and timbral features.

In addition to aspects directly manipulated by the performer on the instrument (such as pitch), any audio processing applied to a sound will potentially change the sound, its timbre, and emotional characteristics [27–29]. For example, artificial reverberation added to a sound can make it more romantic and mysterious [29].

Other 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 [30]. 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 [31, 32]. 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 [33–40]. These artifacts change the timbre of the sound more and more with higher compression rates [41]. 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 [41].

The topic of the current paper 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.

1 RESEARCH METHODOLOGY

1.1 Overview

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. [20–23], Chau et al. [24–26], and Mo et al. [29] but using MP3 compressed stimuli. Paired comparisons were chosen for sim-

ilarity. This section gives further details about the listening test.

1.2 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 [40–48]. 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 [49]. 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 artifacts. 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 [40], 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.

1.3 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 [50]. For this study, we used the same categories we have used in our previous research on musical instruments [20–26, 29]. The ratings of the emotional categories according to the “Affective Norms for English Words” [51] are shown in Fig. 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.

1.4 Listening Test

There were 20 undergraduate students hired for the listening test. All subjects were fluent in English and were

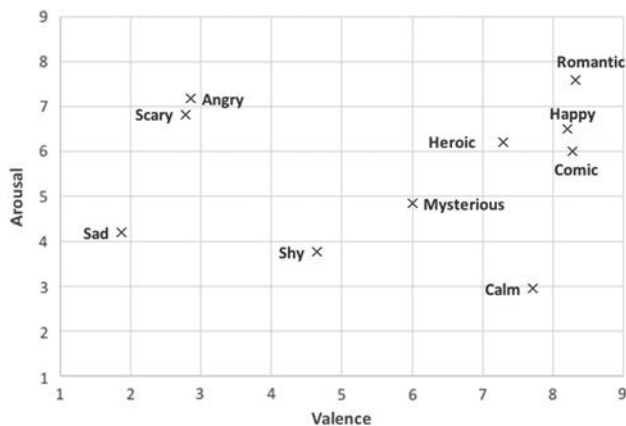


Fig. 1. Distribution of the emotional characteristics in the dimensions Valence and Arousal. The Valence and Arousal values are given in the 9-point rating in ANEW [51]. Valence shows the positiveness of an emotional category; Arousal shows the energy level of an emotional category.



Fig. 2. Paired comparison listening test interface

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 (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* [52]. The dictionary definitions we used in our experiment are shown in Table 1. Every subject made paired comparisons between the sounds (see user interface in Fig. 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 [24, 43]. 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.

The subjects were seated in a “quiet room” with 39 dB SPL background noise level (mostly due to computers and ventilation). The noise level was reduced further with headphones. Sound signals were converted to analog by a Sound Blaster X-Fi Xtreme Audio sound card and then presented through Sony MDR-7506 headphones at a level of approximately 78 dB SPL. 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 sounds for this test. 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 LISTENING TEST 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:

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

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 [53].

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 [54, 55]. 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

Table 1. The dictionary definitions of the emotional categories used in our experiment.

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

derived the corresponding 95% confidence intervals for the compressed sounds using the method proposed by Bradley [54].

Figs. 3 to 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 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. [40], which found low, medium, and high discrimination rates for the same bit rates.

To understand which instruments and emotional categories were most and least affected by MP3 compression, Table 3 shows the number of compressed sounds that were significantly different from the original sound for each instrument and emotional category. Based on Table 3, the trumpet was the most effected instrument, while the horn was by far the least effected instrument. Lee et al. [40] also found the MP3-compressed horn the most difficult to discriminate compared to the other instruments. Among the emotional categories in Table 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.

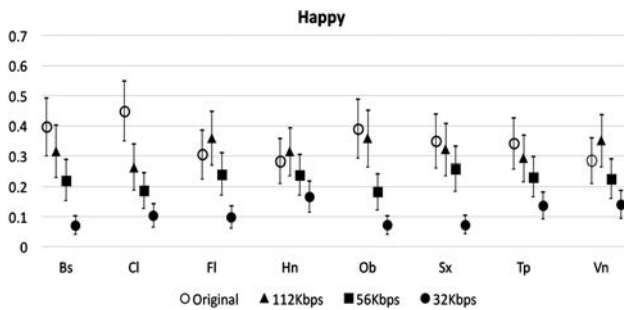


Fig. 3. BTL scale values and the corresponding 95% confidence intervals for the emotional category Happy.

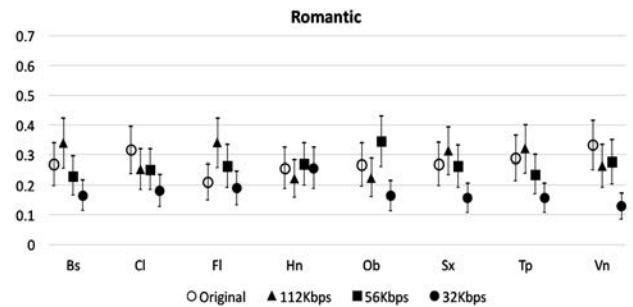


Fig. 5. BTL scale values and the corresponding 95% confidence intervals for Romantic.

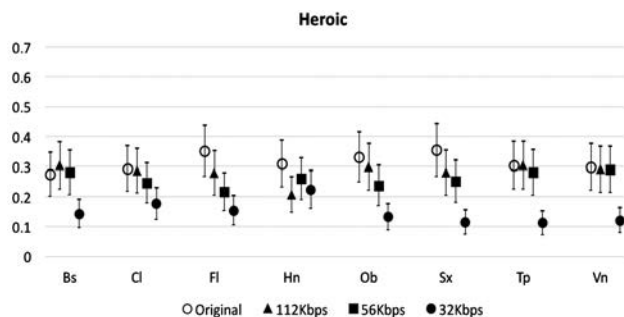


Fig. 4. BTL scale values and the corresponding 95% confidence intervals for Heroic.

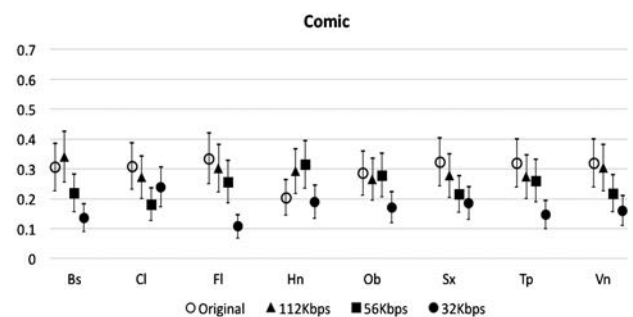


Fig. 6. BTL scale values and the corresponding 95% confidence intervals for Comic.

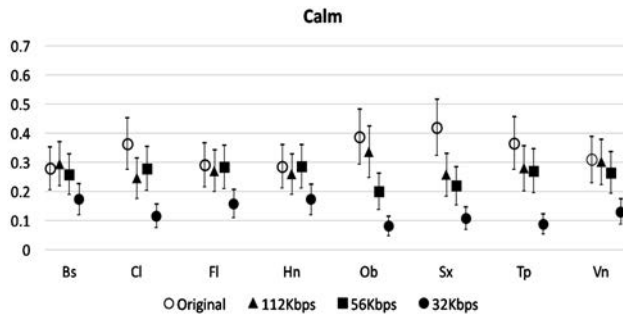


Fig. 7. BTL scale values and the corresponding 95% confidence intervals for Calm.

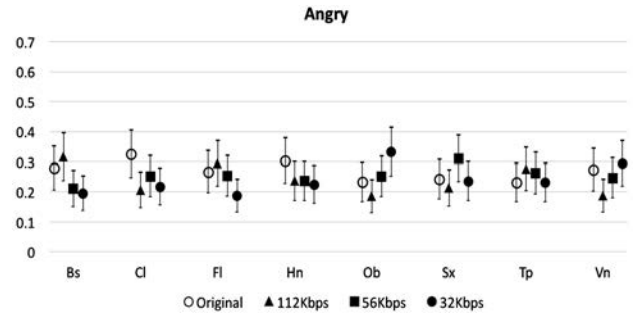


Fig. 10. BTL scale values and the corresponding 95% confidence intervals for Angry.

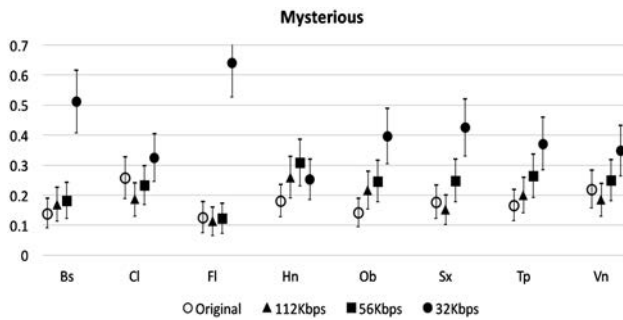


Fig. 8. BTL scale values and the corresponding 95% confidence intervals for Mysterious.

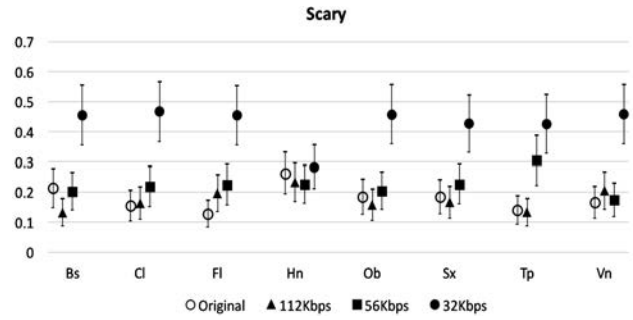


Fig. 11. BTL scale values and the corresponding 95% confidence intervals for Scary.

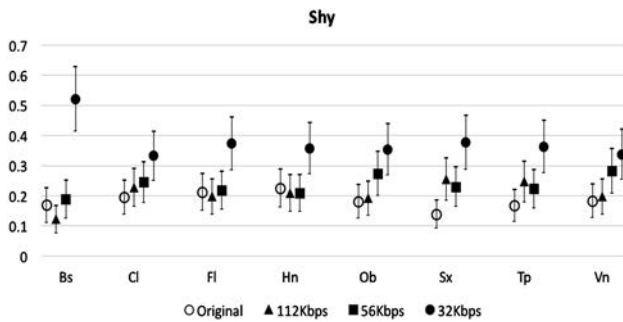


Fig. 9. BTL scale values and the corresponding 95% confidence intervals for Shy.

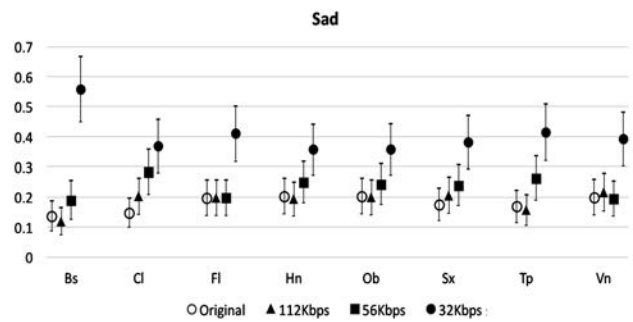


Fig. 12. BTL scale values and the corresponding 95% confidence intervals for Sad.

Table 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.

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 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.

Emotional Category	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	

To see which emotional categories were strengthened or weakened by MP3 compression, Fig. 13 shows how often the original instruments sounds were statistically significantly greater than the three compressed sounds. The values in Fig. 13 are different from the sum in the final column of Table 3 which counts any significant difference as +1 for both those significantly greater and those significantly less. In Fig. 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 than 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 unaf-

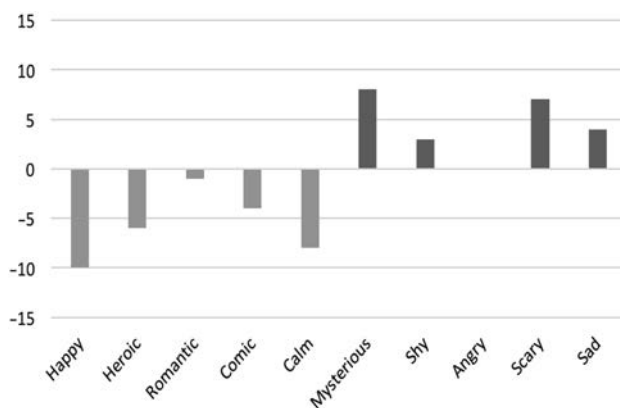


Fig. 13. The number of significant differences between the original and compressed sounds, where strengthened emotional categories are positive and weakened emotional categories are negative.

ected by MP3 compression for the compression rates we tested.

3 DISCUSSION

The goal of our work was to understand how emotional characteristics of instruments vary with MP3 compression. Based on Table 3 and Fig. 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 Fig. 13).
2. Positive emotional characteristics (Happy, Heroic, Romantic, Comic, and Calm) decreased with more MP3 compression in the samples we tested (see Fig. 13).
3. Angry was relatively unaffected by MP3 compression for the rates we tested (see Fig. 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 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. [41] 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 [40], 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 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 Fig. 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.

Based on these results, 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 that 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 [31, 37–39, 56] and discrimination measures [40, 41], 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.

4 ACKNOWLEDGMENTS

Thanks to the anonymous reviewers for their valuable time in reviewing this paper.

REFERENCES

- [1] L.-L. Balkwill and W. F. Thompson “A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues,” *Music Perception: An Interdisciplinary J.*, vol. 17, no. 1, pp. 43–64 (1999 Fall). DOI: <http://dx.doi.org/10.2307/40285811>.
- [2] J. Liebetrau et al., “Paired Comparison as a Method for Measuring Emotions,” presented at the *135th Convention of the Audio Engineering Society* (2013 Oct.), convention paper 9016.
- [3] I. Lahdelma and T. Eerola “Single Chords Convey Distinct Emotional Qualities to Both Naïve and Expert Listeners,” *Psychology of Music*, vol. 44, no. 1, p. 0305735614552006 (2014).
- [4] J. Skowronek, M. McKinney, and S. Van De Par “A Demonstrator for Automatic Music Mood Estimation,” *Proceedings of the International Conference on Music Information Retrieval* (2007).
- [5] M. Plewa and B. Kostek “A Study on Correlation between Tempo and Mood of Music,” presented at the *133rd Convention of the Audio Engineering Society* (2012 Oct.), convention paper 8800.
- [6] Y. Hu, X. Chen, and D. Yang “Lyric-Based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method,” *Proceedings of ISMIR* (2009).
- [7] I. Ekman and R. Kajastila “Localization Cues Affect Emotional Judgments—Results from a User Study on Scary Sound,” presented at the *AES 35th International Conference: Audio for Games* (2009 Feb.), conference paper 23.
- [8] G. Tzanetakis and P. Cook “Musical Genre Classification of Audio Signals,” *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302 (2002). DOI: <http://dx.doi.org/10.1109/tsa.2002.800560>.
- [9] K. Hevner “Experimental Studies of the Elements of Expression in Music,” *Amer. J. Psychology*, vol. 48, no. 2, pp. 246–268 (1936). DOI: <http://dx.doi.org/10.2307/1415746>.
- [10] I. Peretz, L. Gagnon, and B. Bouchard “Music and Emotion: Perceptual Determinants, Immediacy, and

Isolation after Brain Damage,” *Cognition*, vol. 68, no. 2, pp. 111–141 (1998). DOI: [http://dx.doi.org/10.1016/S0010-0277\(98\)00043-2](http://dx.doi.org/10.1016/S0010-0277(98)00043-2).

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

[12] J. C. Hailstone et al., “It’s Not What You Play, It’s How You Play It: Timbre Affects Perception of Emotion in Music,” *Quarterly J. Experimental Psychology*, vol. 62, no. 11, pp. 2141–2155 (2009). DOI: <http://dx.doi.org/10.1080/17470210902765957>.

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

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

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

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

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

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

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

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

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

[22] B. Wu et al., “Investigating Correlation between Musical Timbres and Emotions,” *International Society for*

Music Information Retrieval Conference (ISMIR), Curitiba, Brazil (2013), pp. 415–420.

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

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

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

[26] C.-j. Chau and A. Horner “The Effect of Pitch and Dynamics on the Emotional Characteristics of Piano Sounds,” *41st International Computer Music Conference: Looking Back, Looking Forward, University of North Texas, Denton, United States* (25 Sep.–1 Oct. 2015).

[27] D. Västfjäll, P. Larsson, and M. 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). DOI: <http://dx.doi.org/10.1089/109493102753685854>.

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

[29] R. Mo, B. Wu, and A. Horner “The Effects of Reverberation on the Emotional Characteristics of Musical Instruments,” *J. Audio Eng. Soc.*, vol. 63, pp. 966–979 (2015 Dec.). DOI: <http://dx.doi.org/10.17743/jaes.2015.0082>.

[30] D. Williams “Affective Potential in Vocal Production,” presented at the *139th Convention of the Audio Engineering Society* (2015 Oct.), convention paper 9446.

[31] H. Fuchs, W. Hoeg, and D. Meares “ISO/MPEG Subjective Tests on Multichannel Audio Systems: Design and Methodology,” *Broadcasting Convention, 1994. IBC 1994., International* (1994), pp. 152–157. DOI: <http://dx.doi.org/10.1049/cp:19940744>.

[32] B. Sisario “Downloads in Decline as Streamed Music Soars,” *The New York Times (New York edition)* (July 3, 2014), B3.

[33] S. van de Par and A. Kohlrausch “Three Approaches to the Perceptual Evaluation of Audio Compression Methods,” *J. Acous. Soc. Amer.*, vol. 107, no. 5, pp. 2875–2875 (2000). DOI: <http://dx.doi.org/10.1121/1.428678>.

[34] M. Erne “Perceptual Audio Coders: What to Listen For,” presented at the *111th Convention of the Audio Engineering Society* (2001 Nov.), convention paper 5489.

[35] C.-M. Chang et al., “Compression Artifacts in Perceptual Audio Coding,” presented at the *121st Convention of the Audio Engineering Society* (2006 Oct.), convention paper 6872.

- [36] P. Marins “Characterizing the Perceptual Effects Introduced by Low Bit Rate Spatial Audio Codecs,” presented at the *131st Convention of the Audio Engineering Society* (2011 Oct.), convention paper 8531.
- [37] D.G. Kirby, F. Feige, and U. Wustenhagen “ISO/MPEG Subjective Tests on Multichannel Audio Coding Systems: Practical Realization and Test Results,” *Broadcasting Convention, 1994. IBC 1994., International* (1994), pp. 132–139. DOI: <http://dx.doi.org/10.1049/cp:19940741>.
- [38] W. H. Schmidt and E. Steffen “ISO/MPEG Subjective Tests on Multichannel Audio Coding Systems: Statistical Analysis,” *International Broadcasting Convention – IBC '94* (1994), pp. 158–163. DOI: <http://dx.doi.org/10.1049/cp:19940745>.
- [39] G. Stoll and F. Kozamernik *EBU Subjective Listening Tests on Low-Bitrate Audio Codecs* (2003).
- [40] C. Lee and A. Horner “Discrimination of MP3-Compressed Musical Instrument Tones,” *J. Audio Eng. Soc.*, vol. 58, pp. 487–497 (2010 Jun.). DOI: <http://dx.doi.org/10.1121/1.3508003>.
- [41] C. Lee, A. Horner, and B. Wu “The Effect of MP3 Compression on the Timbre Space of Sustained Musical Instrument Tones,” *J. Audio Eng. Soc.*, vol. 61, pp. 840–849 (2013 Nov.).
- [42] S. McAdams, J. W. Beauchamp, and S. Meneguzzi “Discrimination of Musical Instrument Sounds Resynthesized with Simplified Spectrotemporal Parameters,” *J. Acous. Soc. Amer.*, vol. 105, no. 2, pp. 882–897 (1999). DOI: <http://dx.doi.org/10.1121/1.426277>.
- [43] A. Horner, J. Beauchamp, and R. So “Detection of Random Alterations to Time-Varying Musical Instrument Spectra,” *J. Acous. Soc. Amer.*, vol. 116, no. 3, pp. 1800–1810 (2004). DOI: <http://dx.doi.org/10.1121/1.1778741>.
- [44] J. W. Beauchamp et al., “Multidimensional Scaling Analysis of Centroid-and Attack/Decay-Normalized Musical Instrument Sounds,” *J. Acous. Soc. Amer.*, vol. 120, no. 5, pp. 3276–3276 (2006). DOI: <http://dx.doi.org/10.1121/1.4777232>.
- [45] A. B. Horner, J. W. Beauchamp, and R. H. Y. So “Detection of Time-Varying Harmonic Amplitude Alterations Due to Spectral Interpolations between Musical Instrument Tones,” *J. Acous. Soc. Amer.*, vol. 125, no. 1, pp. 492–502 (2009). DOI: <http://dx.doi.org/10.1121/1.3025916>.
- [46] A. B. Horner, J. W. Beauchamp, and R. H. Y. So “Evaluation of Mel-Band and MFCC-Based Error Metrics for Correspondence to Discrimination of Spectrally Altered Musical Instrument Sounds,” *J. Audio Eng. Soc.*, vol. 59, pp. 290–303 (2011 May).
- [47] M. Bosi and R. E. Goldberg *Introduction to Digital Audio Coding and Standards*, Vol. 721 (Springer Science & Business Media, 2012). DOI: <http://dx.doi.org/10.1007/978-1-4615-0327-9w>.
- [48] 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, pp. 899–912 (2012 Nov.). DOI: <http://dx.doi.org/10.1121/1.3588534>.
- [49] *LAME MP3 Encoder*, URL: <http://lame.sourceforge.net/>.
- [50] P. N. Juslin and J. Slobodan *Handbook of Music and Emotion: Theory, Research, Applications* (Oxford University Press, 1993). DOI: <http://dx.doi.org/10.1093/acprof:oso/9780199230143.001.0001>.
- [51] M. M. Bradley and P. J. Lang *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings* (Tech. rep. Citeseer, 1999).
- [52] *Cambridge Academic Content Dictionary*, URL: <http://dictionary.cambridge.org/dictionary/american-english>.
- [53] F. L. Joseph “Measuring Nominal Scale Agreement among Many Raters,” *Psychological Bulletin*, vol. 76, no. 5, pp. 378–382 (1971). DOI: <http://dx.doi.org/10.1037/h0031619>.
- [54] R. A. Bradley “Paired Comparisons: Some Basic Procedures and Examples,” *Nonparametric Methods*, vol. 4, pp. 299–326 (1984). DOI: [http://dx.doi.org/10.1016/s0169-7161\(84\)04016-5](http://dx.doi.org/10.1016/s0169-7161(84)04016-5).
- [55] F. Wickelmaier and C. Schmid “A Matlab Function to Estimate Choice Model Parameters from Paired-comparison Data,” *Behavior Research Methods, Instruments, and Computers*, vol. 36, no. 1, pp. 29–40 (2004). DOI: <http://dx.doi.org/10.3758/bf03195547>.
- [56] T. Thiede et al., “PEAQ—The ITU Standard for Objective Measurement of Perceived Audio Quality,” *J. Audio Eng. Soc.*, vol. 48, pp. 3–29 (2000 Jan./Feb.).

THE AUTHORS



Ronald Mo



Choi Ga Lam



Chung Lee



Andrew Horner

Ronald Mo is a Ph.D. candidate in the department of computer science and engineering at the Hong Kong University of Science and Technology. His research interests include music emotion recognition, music information retrieval, and audio signal processing. He received his B. Eng. of computer science and M. Phil. of computer science and engineering from the Hong Kong University of Science and Technology in 2007 and 2015 respectively.

Choi Ga Lam recently graduated from Hong Kong University of Science and Technology, with B. Eng. in computer science. Her research interests include music synthesis and music emotion. Personally, she is deeply interested in recording and audio post-processing.

Chung Lee got his Ph.D. degree in computer science at the Hong Kong University of Science and Technology in 2011. His main research interests were timbre discrimination, non-linear spectral alteration, and audio-aid for disable people. He is now a senior information technology analyst in United Overseas Bank, Singapore.

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