Timbre Features and Music Emotion in Plucked String, Mallet Percussion, and Keyboard Tones

Chuck-jee Chau, Bin Wu, Andrew Horner
Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
chuckjee@cse.ust.hk, bwuaa@cse.ust.hk, horner@cs.ust.hk

ABSTRACT

Music conveys emotions by means of pitch, rhythm, loudness, and many other musical qualities. It was recently confirmed that timbre also has direct association with emotion, for example, that a horn is perceived as sad and a trumpet heroic in even isolated instrument tones. As previous work has mainly focused on sustaining instruments such as bowed strings and winds, this paper presents an experiment with non-sustaining instruments, using a similar approach with pairwise comparisons of tones for emotion categories. Plucked string, mallet percussion, and keyboard instrument tones were investigated for eight emotions: Happy, Sad, Heroic, Scary, Comic, Shy, Joyful, and Depressed. We found that plucked string tones tended to be Sad and Depressed, while harpsichord and mallet percussion tones induced positive emotions such as Happy and Heroic. The piano was emotionally neutral. Beyond spectral centroid and its deviation, which are important features in sustaining tones, decay slope was also significantly correlated with emotion in non-sustaining tones.

1. INTRODUCTION

As one of the oldest art forms, music was developed to convey emotion. All kinds of music from ceremonial to casual incorporate emotional messages. Much work has been done on music emotion recognition using melody [1], harmony [2], rhythm [3, 4], lyrics [5], and localization cues [6].

Different musical instruments produce varied timbres, and timbre is an important feature that shapes the emotional character of an instrument. Previous research has shown that emotion is also associated with timbre. Scherer and Oshinsky [7] found that timbre is a salient factor in the rating of synthetic sounds. Peretz *et al.* [8] showed that timbre speeds up discrimination of emotion categories. Bigand *et al.* [9] reported similar results in their study of emotion similarities between one-second musical excerpts. Timbre has also been found to be essential to music genre recognition and discrimination [10, 11, 12].

Eerola *et al.* [13] worked on the direct connection between emotion and timbre, and confirmed strong correlation between features such as attack time and brightness and the emotion dimensions valence and arousal for one-second isolated instrument sounds. These two dimensions refer to how positive and energetic a music stimulus sounds respectively [14]. Asutay *et al.* [15] also studied the valence and arousal responses from subjects on 18 environmental sounds. Using a different approach than Eerola, Ellermeier *et al.* [16] investigated the unpleasantness of environmental sounds using paired comparisons.

Wu *et al.* [17] studied pairwise emotional correlation among sustaining instruments, such as the clarinet and violin. It was found that emotion correlated significantly with spectral centroid, spectral centroid deviation, and even/odd harmonic ratio.

But, what about sounds that decay immediately after the attack, and do not sustain, such as the piano? This study considers the comparison of naturally decaying sounds and the correlation of spectral features and emotional categories. Eight plucked string, mallet percussion, and keyboard instrument sounds were investigated for eight emotions: Happy, Sad, Heroic, Scary, Comic, Shy, Joyful, and Depressed.

2. SIGNAL REPRESENTATION

The stimuli were analyzed and represented as a sum of sinusoids, with time-varying amplitudes and frequencies:

$$s(t) = \sum_{k=1}^{K} A_k(t) \cos\left(2\pi \int_0^t \left(kf_a + \Delta f_k(\tau)\right) d\tau + \theta_k(0)\right), \tag{1}$$

where
$s(t)$ = sound signal,
t = time in s,
τ = integrand dummy variable representing time,
k = harmonic number,
K = number of harmonics,
$A_k(t)$ = amplitude of the kth harmonic at time t,
f_a = analysis frequency and approximate fundamental frequency (349.2 Hz for our tones),
$\Delta f_k(t)$ = frequency deviation of the kth harmonic, so that $f_k(t) = kf_a + \Delta f_k(t)$ is the total instantaneous frequency of the kth harmonic, and
$\theta_k(0)$ = initial phase of the kth harmonic.

3. SPECTRAL CORRELATION MEASURES

3.1 Frequency Domain Features

In the study by Wu [17], it was found that emotion was affected by spectral variations in the instrument tones. Different measures of spectral variations are possible, and the following are used in this study.

First of all, the instantaneous rms amplitude is given by:

$$A_{rms}\left(t_n\right) = \sqrt{\sum_{k=1}^{K} A_k^2\left(t_n\right)}, \tag{2}$$

where t_n is the analysis frame number. N in the following equations represents the total number of analysis frames for the entire tone (or a portion of the tone for the feature *decay slope*).

3.1.1 Spectral Centroid

Spectral centroid is a popular spectral measure, closely related to perceptual brightness. Normalized spectral centroid (*NSC*) is defined as [18]:

$$NSC(t_n) = \frac{\sum_{k=1}^{K} kA_k\left(t_n\right)}{\sum_{k=1}^{K} A_k\left(t_n\right)}. \tag{3}$$

3.1.2 Spectral Centroid Deviation

Spectral centroid deviation was qualitatively described by Krumhansl [19] as the temporal evolution of the spectral components. Krimphoff [20] defined spectral centroid deviation as the root-mean-squared deviation of the normalized spectral centroid (*NSC*) over time given by:

$$SCD = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}\left(NSC\left(t_n\right) - NSC_{xx}\right)^2}, \tag{4}$$

where NSC_{xx} could be the average, rms, or maximum value of NSC. A time-average value is used in this study. Note that Krimphoff used the term "spectral flux" in his original presentation, but other researchers have used the term spectral centroid deviation instead since it is more specific.

3.1.3 Spectral Incoherence

Beauchamp and Lakatos [21] measured spectral fluctuation in terms of spectral incoherence, a measure of how much a spectrum differs from a coherent version of itself. Larger incoherence values indicate a more dynamic spectrum, and smaller values indicate a more static spectrum. A perfectly static spectrum has an incoherence of zero.

A perfectly coherent spectrum is defined to be the average spectrum of the original, but unlike the original, all harmonic amplitudes vary in time proportional to the rms amplitude and, therefore, in fixed ratios to each other. Put another way, the harmonic amplitudes are fixed. The coherent version of the kth harmonic amplitude is defined by:

$$\hat{A}_k\left(t_n\right) = \frac{\bar{A}_k A_{rms}\left(t_n\right)}{\sqrt{\sum_{k=1}^{K}\bar{A}_k^2}}, \tag{5}$$

where $\bar{A}_k$ is the time-averaged amplitude of the kth harmonic. Then, spectral incoherence of the original spectrum is defined as:

$$SI = \sqrt{\frac{\sum_{n=0}^{N-1}\sum_{k=1}^{K}\left(A_k\left(t_n\right) - \hat{A}_k\left(t_n\right)\right)^2}{\sum_{n=0}^{N-1}\left(A_{rms}\left(t_n\right)\right)^2}}. \tag{6}$$

Spectral incoherence (*SI*) varies between 0 and 1 with higher values indicating more incoherence (a more dynamic spectrum).

3.1.4 Spectral Irregularity

Krimphoff [20] introduced the concept of spectral irregularity to measure the jaggedness of a spectrum. Spectral irregularity was redefined by Beauchamp and Lakatos [21] as:

$$SIR = \frac{1}{N}\sum_{n=0}^{N-1}\frac{\sum_{k=2}^{K-1} A_k\left(t_n\right)\left|A_k\left(t_n\right) - \tilde{A}\left(t_n\right)\right|}{A_{rms}\left(t_n\right)\sum_{k=2}^{K-1} A_k\left(t_n\right)}, \tag{7}$$

where

$$\tilde{A}\left(t_n\right) = \left(A_{k-1}\left(t_n\right) + A_k\left(t_n\right) + A_{k-1}\left(t_n\right)\right)/3.$$

This formula defines the difference between a spectrum and a spectrally smoothed version of itself, averaged over both harmonics and time and normalized by rms amplitude.

3.1.5 Even/odd Harmonic Ratio

Even/odd harmonic ratio [22] is another measure of spectral irregularity and jaggedness, and is based on the ratio of even and odd harmonics:

$$E/O = \frac{\sum_{n=0}^{N-1}\sum_{j=1}^{K/2} A_{2j}\left(t_n\right)}{\sum_{n=0}^{N-1}\sum_{j=1}^{(K+1)/2} A_{2j-1}\left(t_n\right)}, \tag{8}$$

This measure is especially important for clarinet tones, which have strong odd harmonics in the lower register. Though a low E/O (e.g., for low clarinet tones) will usually result in a relatively high *SIR*, the reverse is not necessarily true.

3.2 Time Domain Features

Since overall amplitude changes are vital to non-sustaining tones, several time-domain features are included in this study.

3.2.1 Attack Time

Instead of measuring the time to reach the peak rms amplitude, the term *attack time* here measures the time to reach the first local maximum in rms amplitude from the beginning of the tone.

3.2.2 Decay Ratio

We use the term *decay ratio* to define the ratio between the rms amplitude 30 ms before the tone ends against the peak rms amplitude:

$$DR = \frac{A_{rms}(t_{end-30ms})}{A_{rms}(t_{peakRms})}. \tag{9}$$

The numerator time point was chosen since a linear fade-out was applied over 30 ms from 0.97 s to 1.0 s to the tones in this study. A fast decaying instrument such as the plucked violin had a decay ratio of 0 since it had already decayed to zero by 0.97 s.

3.2.3 Decay Slope

All tones used in this study had natural decays, and there was no sustain. *Decay slope* is the average difference in rms amplitude between adjacent analysis frames. The slope was averaged from the peak rms amplitude until the rms amplitude reached zero.

$$DS = \frac{1}{N}\sum_{n=1}^{N}\left(A_{rms}\left(t_n\right) - A_{rms}\left(t_{n-1}\right)\right) \tag{10}$$

3.3 Local Spectral Features

Many spectral features are more relevant to sustaining tones than decaying tones. Therefore, an amplitude weighting was also tested on the spectral features based on the instantaneous rms amplitude, as defined in Eq. 2. This helped emphasize high-amplitude parts of the tone near the end of the attack and beginning of the decay, and thus de-emphasized the noisy transients. The amplitude-weighted features are denoted by "AW" in our feature tables.

4. EXPERIMENT

Our experiment consisted of a listening test where subjects compared pairs of instrument tones for different emotions.

4.1 Stimuli

4.1.1 Prototype Instrument Tones

The stimuli used in the listening test were tones of non-sustaining instruments (i.e., decaying tones). There were eight instruments in three categories:

- Plucked string instruments: guitar (Gt), harp (Hp), plucked violin (Vn)
- Mallet percussion instruments: marimba (Ma), vibraphone (Vb), xylophone (Xy)
- Keyboard instruments: harpsichord (Hd), piano (Pn)

The tones were from the McGill [23], and RWC [24] sample libraries. All tones had fundamental frequencies (f_0) close to 349.2 Hz (F4) except the harp, which was 329.6 Hz (E4). The harp tone was pitch-shifted to 349.2 Hz using the software Audacity. All tones used a 44,100 Hz sampling rate.

The loudness of the eight tones was equalized by a two-step process to avoid loudness affecting emotion. The initial equalization was by peak rms amplitude. It was further refined manually until the tones were judged of equal loudness by the authors.

4.1.2 Duration of Tones

The original recorded tones were of various lengths, with some as long as 5.6 s including room reverberation, and some as short as 0.9 s. They were processed so that the tones were of the same duration.

First, silence before each tone was removed. The tone durations were then truncated to 1 second, and a 30 ms linear fade-out was introduced before the end of each tone. Some of the original tones were less than 1 second long (e.g., the plucked violin and the xylophone), and were padded with silence.

4.1.3 Method for Spectral Analysis

A phase-vocoder algorithm was used in the analysis of the instrument tones. Unlike normal Fourier analysis, the window size was chosen according to the fundamental frequency so that frequency bins aligned with the harmonics of the input signal. Beauchamp gives more details of the phase-vocoder analysis process [25].

4.2 Subjects

There were 34 subjects hired for the listening test, aged from 19 to 26. All subjects were undergraduate students at the Hong Kong University of Science and Technology.

4.2.1 Consistency

Subject responses were first screened for inconsistencies. Consistency was defined based on the four comparisons of a pair of instruments A and B for a particular emotion as follows:

$$consistency_{A,B} = \frac{max\left(v_A, v_B\right)}{4} \tag{11}$$

where v_A and v_B are the number of votes a subject gave to each of the two instruments. A consistency of 1 represents perfect consistency, whereas 0.5 represents random guessing. The mean average consistency of all subjects was 0.755.

Predictably subjects were only fairly consistent because of the emotional ambiguities in the stimuli. We assessed the quality of responses further using a probabilistic approach. A probabilistic model [26], successful in image labelling, was adapted for our purposes. The model takes the difficulty of labelling and the ambiguities in image categories into account, and estimates annotators' expertise and the quality of their responses. Those making low-quality responses are unable to discriminate between image categories and are considered random pickers. In our study, we verified that the three least consistent subjects made responses of the lowest quality. They were excluded from the results, leaving 31 subjects.

4.3 Emotion Categories

The subjects compared the stimuli in terms of eight emotion categories: Happy, Sad, Heroic, Scary, Comic, Shy, Joyful, and Depressed. These terms were selected by the authors for their relevance to composition and arrangement. Their ratings according to the Affective Norms for English Words [27] are shown in Figure 1 using the Valence-Arousal model. Happy, Joyful, Comic, and Heroic form one cluster, and Sad and Depressed another.

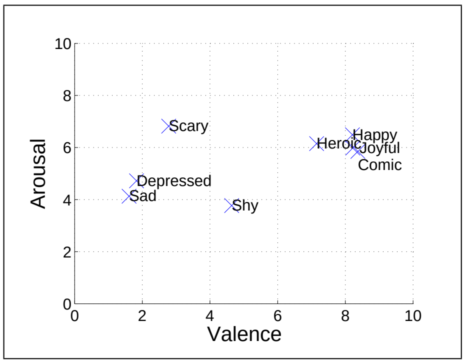

Figure 1. Russel's Valence-Arousal emotion model. Valence is how positive an emotion is. Arousal is how energetic an emotion is.

4.4 Listening Test

Every subject made pairwise comparisons of all eight instruments. During each trial, subjects heard a pair of tones from different instruments and were prompted to choose the tone arousing a given emotion more strongly. Each combination of two different instruments was presented in four trials for each emotion, and the listening test totaled $\binom{8}{2} \times 4 \times 8 = 896$ trials. For each emotion, the overall trial presentation order was randomized (i.e., all the Happy comparisons were first in a random order, then all the Sad comparisons were second, and so on).

Before the first trial, the subjects read online definitions of the emotion categories from the Cambridge Academic Content Dictionary [28]. The listening test took about 1 hour, with a short break of 5 minutes after 30 minutes.

The subjects were seated in a "quiet room" with less than 40 dB SPL background noise level. Residual noise was mostly due to computers and air conditioning. The noise level was reduced further with headphones. Sound signals were converted to analog with a Sound Blaster X-Fi Xtreme Audio sound card, and then presented through Sony MDR-7506 headphones at a level of approximately 78 dB SPL, as measured with a sound-level meter. The Sound Blaster DAC utilized 24-bit depth with a maximum sampling rate of 96 kHz and a 108 dB S/N ratio.

5. EXPERIMENT RESULTS

5.1 Voting Results

The raw results were pairwise votes for each instrument pair and each emotion, and are illustrated in Figure 2 in greyscale. The rows show the percentage of positive votes each instrument received, compared the other instruments. The lighter the cell color, the more positive votes the "row instrument" received when compared against the "column instrument". Taking the Heroic emotion as an example, the harpsichord was judged to be more Heroic than all the other instruments.

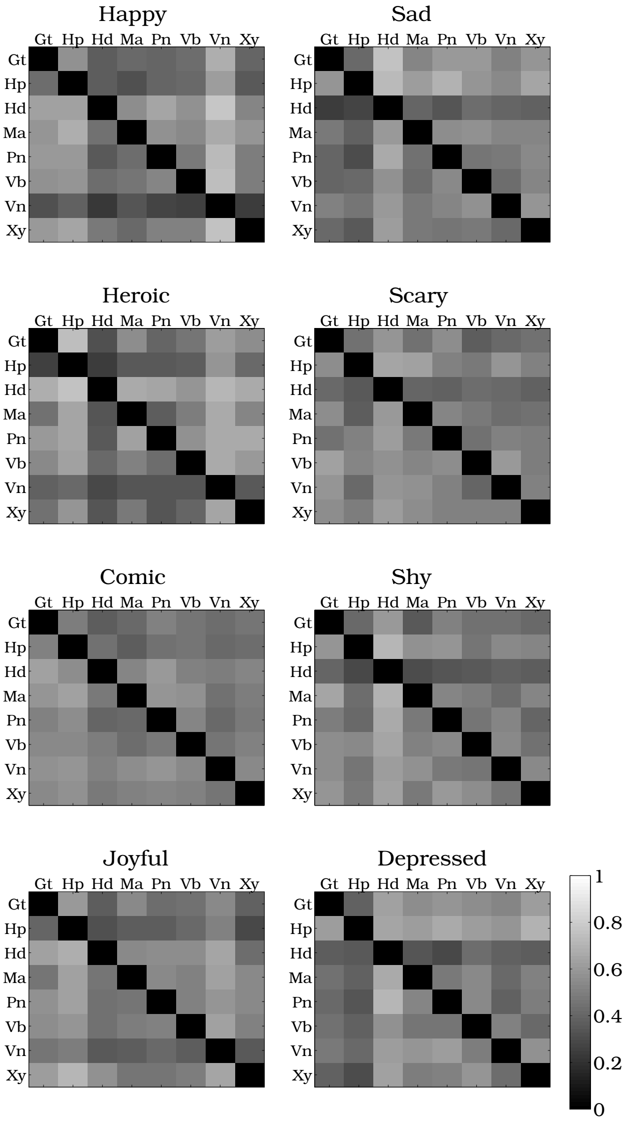

Figure 2. Comparison between instrument pairs. Lighter color indicates more positive votes for the "row instrument" compared to the "column instrument".

The greyscale charts give a basic idea of the emotional-distinctiveness of an instrument. Most emotions were distinctive with a mix of lighter and darker blocks, but Comic, Scary, and Joyful were more difficult to distinguish as shown by the nearly uniform grey color.

Figure 3 displays the ranking of instruments derived using the Bradley-Terry-Luce (BTL) model [29, 16]. The rankings are based on the number of positive votes each instrument received for each emotion. The values represent the scale value of each instrument compared to the base instrument (i.e., the one with the lowest ranking). For example, for Happy, the ranking of the harpsichord was

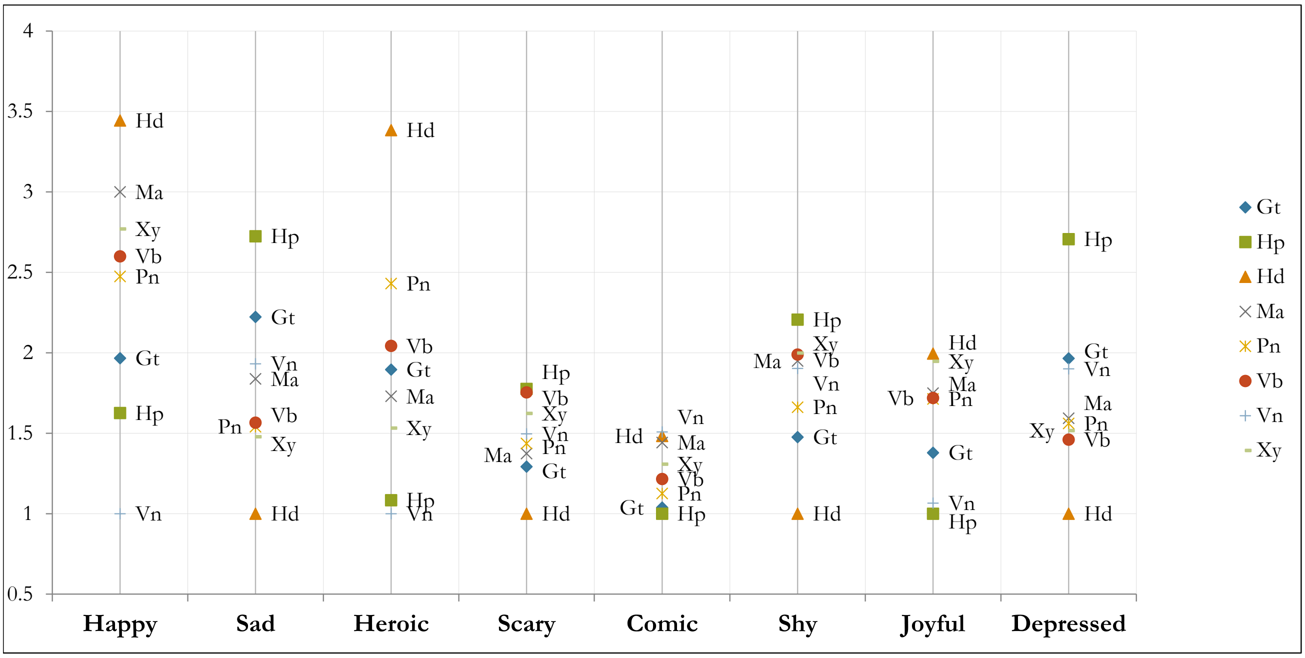

Figure 3. Bradley-Terry-Luce scale values of the instruments for each emotion

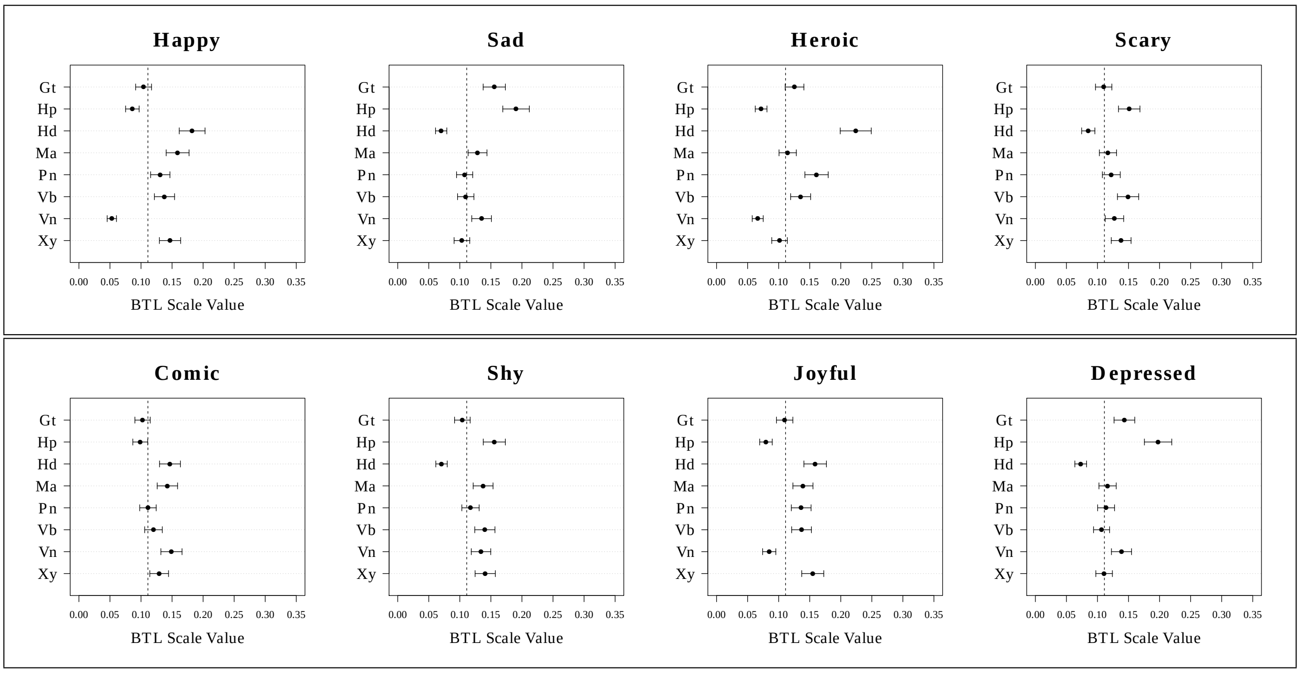

Figure 4. BTL scale values and the corresponding 95% confidence intervals. The dotted line represents no preference.

3.5 times that of the violin. The figure presents a more effective comparison of the magnitude of the differences between instruments. The wider the spread of the instruments along the y-axis, the more divergent and distinguishable they are.

The harpsichord stood out as the most Heroic and Happy instrument, and was ranked highly for other high-valence emotions such as Comic and Joyful. The mallet percussion (marimba, xylophone, and vibraphone) also ranked highly for the same emotions. The harp stood out for Sad and Depressed, with the guitar second. The harp was also top-ranked instrument for Shy, and perhaps surprisingly Scary. The mallet percussion were collectively ranked second Shy.

The plucked violin was at or near the bottom for Happy, Heroic, and Joyful (through on the top for the other high-valence emotion Comic). This is opposite the bowed violin, which was highly-ranked for Happy in Wu's study [17].

The ranges for Comic and Scary were rather compressed, representing listeners' difficulty in differentiating instruments for these emotions.

The instruments were often in clusters by instrument type. The plucked string instruments including harp, gui-

Instruments Features	Gt	Hp	Hd	Ma	Pn	Vb	Vn	Xy
Attack Time	0.003	0.009	0.062	0.049	0.014	0.019	0.006	0.039
Decay Ratio	0.137	0.074	0.069	0.056	0.133	0.248	0.000	0.019
Decay Slope	-1.448	-2.402	-0.634	-1.261	-0.860	-0.531	-1.868	-0.986
Spectral Centroid	2.210	1.747	6.661	2.336	3.146	2.947	23.087	9.498
Spectral Centroid (AW)	2.389	1.501	7.442	2.003	3.018	2.894	2.762	3.674
Spectral Centroid Deviation	0.954	0.927	2.066	1.687	1.089	0.824	17.656	8.919
Spectral Centroid Deviation (AW)	1.826	1.093	5.887	1.461	1.987	1.504	2.603	2.335
Spectral Incoherence	0.142	0.020	0.283	0.258	0.072	0.165	0.205	0.177
Spectral Incoherence (AW)	0.223	0.024	0.310	0.318	0.083	0.200	0.226	0.186
Spectral Irregularity	0.160	0.283	0.084	0.223	0.135	0.254	0.129	0.233
Spectral Irregularity (AW)	0.137	0.283	0.067	0.217	0.122	0.263	0.164	0.242
Even/odd Harmonic Ratio	0.170	0.038	0.968	0.215	0.301	0.749	0.927	0.046
Even/odd Harmonic Ratio (AW)	0.208	0.033	0.864	0.328	0.340	0.832	1.079	0.035

Table 1. Features of the instrument tones. *AW* indicates amplitude-weighted features (see Section 3.3).

Emotion Features	Happy	Sad	Heroic	Scary	Comic	Shy	Joyful	Depressed	Number of emotions with significant correlation
Attack Time	**0.86****	**-0.69***	0.59	-0.52	**0.62***	-0.41	**0.78****	**-0.70***	**5**
Decay Ratio	0.19	-0.06	0.33	0.21	-0.50	-0.04	0.16	-0.12	0
Decay Slope	**0.78****	**-0.89****	**0.76****	-0.34	0.28	-0.48	**0.90****	**-0.92****	**5**
Spectral Centroid	-0.50	-0.14	-0.35	-0.03	0.62	0.04	-0.28	-0.06	0
Spectral Centroid (AW)	0.60	**-0.81****	**0.81****	**-0.69***	0.50	**-0.81****	**0.62***	**-0.76****	**6**
Spectral Centroid Deviation	-0.53	-0.04	-0.49	0.10	0.56	0.20	-0.30	0.03	0
Spectral Centroid Deviation (AW)	0.45	**-0.71***	**0.72****	**-0.77****	0.57	**-0.83****	0.46	**-0.66***	**5**
Spectral Incoherence	0.50	**-0.65***	0.40	**-0.62***	**0.88****	-0.48	0.53	**-0.73****	4
Spectral Incoherence (AW)	0.47	-0.53	0.38	**-0.65***	**0.76****	-0.49	0.48	**-0.67***	3
Spectral Irregularity	-0.10	0.53	-0.58	**0.83****	-0.44	**0.84****	-0.20	0.52	2
Spectral Irregularity (AW)	-0.23	0.51	**-0.69***	**0.90****	-0.31	**0.92****	-0.28	0.53	3
Even/odd Harmonic Ratio	0.03	-0.53	0.39	-0.37	**0.63***	-0.46	0.09	-0.52	1
Even/odd Harmonic Ratio (AW)	-0.08	-0.45	0.26	-0.28	**0.64***	-0.34	0.00	-0.45	1

Table 2. Pearson correlation between emotion and features of the instrument tones. **: $p \leq 0.05$; *: $0.05 < p < 0.1$.

tar, and plucked violin were similarly ranked. The mallet percussion including marimba, xylophone, and vibraphone were another similarly ranked group. On the other hand, the piano was the most neutral instrument in the rankings, while the harpsichord was consistently an outlier.

The BTL scale values and 95% confidence intervals of the instruments for each emotion are shown in Figure 4, using the method proposed by Bradley [29]. The dotted line for each emotion represents the line of indifference. The confidence intervals are generally uniformly small.

5.2 Correlation Results

The features of the instrument tones are given in Table 1. Pearson correlation between these features and the emotions are given in Table 2. Amplitude-weighted spectral centroid was significantly correlated with six of the eight emotions, and amplitude-weighted spectral centroid deviation with five emotions. Both spectral centroid features significantly correlated for all four low-valence emotions.

By contrast, the same features without amplitude weighting were not correlated with any emotion. Emphasizing the high-amplitude parts of the tone made a big difference.

Decay slope was also significantly correlated to most emotions, but not the more ambiguous emotions Comic, Scary, and Shy. Tones with more negative slopes (i.e., faster decays) were considered more Sad and Depressed. Tones with slower decays were considered more Happy, Heroic, and Joyful.

Our results of decaying tones agreed with results in Eerola [13], where attack time and spectral centroid deviation showed strong correlation with emotion. However, unlike the results in Wu [17], even/odd harmonic ratio was not significantly correlated with emotion for decaying tones.

6. DISCUSSION

Similar to sustaining tones [17], we found spectral centroid and spectral centroid deviation to have a strong impact on emotion perception. In addition, we observed that attack time and decay slope had a strong correlation with many emotions for decaying tones.

Our stimuli included decaying musical instruments of different types. The guitar, violin, and harp are plucked strings, while the mallet percussion are struck wood or metal. The vibrations are resonated by a cavity or tube respectively. The different acoustic structures contribute to evoking different emotions. Our experiment showed that decay slope affects emotion, and decay slope depends in part on the material of the instrument.

The harpsichord makes its sound by plucking multiple strings of the same pitch using a plectrum. It had the opposite emotional effect as other plucked string instruments. While the spectra of the harp and guitar had very few harmonics in a fast decay, the harpsichord had a much more brilliant spectrum and decayed slower. Though the piano is also a keyboard instrument like the harpsichord, the strings are struck by hammers instead of plucked. The piano was emotionally-neutral. Perhaps this is why the piano is so versatile at playing arrangements of orchestral scores, since it can let the emotion of the music shine through its emotionally-neutral timbre.

These findings give music composers and arrangers a basic reference for emotion in decaying tones. Performers, audio engineers, and designers can manipulate these sounds to tweak the emotional effects of the music. Of course, timbre is only one aspect that contributes to the overall drama of the music.

7. FUTURE DEVELOPMENT

In this study, we measured decay slope with a relatively simple approach. A refinement might be to use only significant harmonics rather than all harmonics. A more sophisticated metric will likely increase the robustness of decay slope, though it is obviously relatively effective already.

We only considered one representative tone for each instrument in our study. Of course, in practice percussionists use many types of mallets and striking techniques to make different sounds. Similarly, string players produce different timbres with different plucking positions and finger gestures. It would be great to determine the range of emotion that an instrument can produce using different performance methods.

Our instrument tones were deliberately cut short to allow a uniform-duration comparison in this study. However, in our preliminary preparations some of the instrument tones seemed to give a different emotional impression for different lengths. It would be interesting to re-run the same experiment with shorter tones (e.g., 0.25 s tones or 0.5 s tones). This will reveal even more information about the relationship between emotion and the perception of decaying musical tones of different durations. Our emotional impression of decaying tones may change with time, depending on when the performer stops the note.

Acknowledgments

This work has been supported by Hong Kong Research Grants Council grant HKUST613112.

8. 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*, vol. 17, no. 1, pp. 43–64, 1999.

[2] J. Liebetrau, S. Schneider, and R. Jezierski, "Application of free choice profiling for the evaluation of emotions elicited by music," in *Proc. 9th Int. Symp. Comput. Music Modeling and Retrieval (CMMR 2012): Music and Emotions*, 2012, pp. 78–93.

[3] J. Skowronek, M. F. McKinney, and S. Van De Par, "A demonstrator for automatic music mood estimation." in *Proc. Int. Soc. Music Inform. Retrieval Conf.*, 2007, pp. 345–346.

[4] M. Plewa and B. Kostek, "A study on correlation between tempo and mood of music," in *Audio Eng. Soc. Conv. 133*. Audio Eng. Soc., 2012.

[5] Y. Hu, X. Chen, and D. Yang, "Lyric-based song emotion detection with affective lexicon and fuzzy clustering method." in *Proc. Int. Soc. Music Inform. Retrieval Conf.*, 2009, pp. 123–128.

[6] I. Ekman and R. Kajastila, "Localization cues affect emotional judgments–results from a user study on scary sound," in *Audio Eng. Soc. Conf.: 35th Int. Conf.: Audio for Games*. Audio Eng. Soc., 2009.

[7] 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.

[8] 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.

[9] E. Bigand, S. Vieillard, F. Madurell, J. 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.

[10] J.-J. Aucouturier, F. Pachet, and M. Sandler, ““The way it sounds”: timbre models for analysis and retrieval of music signals,” *IEEE Trans. Multimedia*, vol. 7, no. 6, pp. 1028–1035, 2005.

[11] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293–302, 2002.

[12] C. Baume, “Evaluation of acoustic features for music emotion recognition,” in *Audio Eng. Soc. Conv. 134*. Audio Eng. Soc., 2013.

[13] 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*, vol. 30, no. 1, pp. 49–70, 2012.

[14] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. H. Chen, “A regression approach to music emotion recognition,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 2, pp. 448–457, 2008.

[15] E. Asutay, D. Västfjäll, A. Tajadura-Jiménez, A. Genell, P. Bergman, and M. Kleiner, “Emoacoustics: A study of the psychoacoustical and psychological dimensions of emotional sound design,” *J. Audio Eng. Soc.*, vol. 60, no. 1/2, pp. 21–28, 2012.

[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] B. Wu, S. Wun, C. Lee, and A. Horner, “Spectral correlates in emotion labeling of sustained musical instrument tones,” in *Proc. 14th Int. Soc. Music Inform. Retrieval Conf.*, November 4-8 2013.

[18] A. Horner, J. Beauchamp, and R. So, “Detection of random alterations to time-varying musical instrument spectra,” *J. Acoust. Soc. Amer.*, vol. 116, no. 3, pp. 1800–1810, 2004.

[19] C. L. Krumhansl, “Why is musical timbre so hard to understand,” *Structure and Perception of Electroacoustic Sound and Music*, vol. 9, pp. 43–53, 1989.

[20] J. Krimphoff, “Analyse acoustique et perception du timbre,” *unpublished DEA thesis, Université du Maine, Le Mans, France*, 1993.

[21] J. Beauchamp and S. Lakatos, “New spectro-temporal measures of musical instrument sounds used for a study of timbral similarity of rise-time-and centroid-normalized musical sounds,” in *Proc. 7th Int. Conf. Music Percept. Cognition*, 2002, pp. 592–595.

[22] A. Caclin, S. McAdams, B. K. Smith, and S. Winsberg, “Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones,” *J. Acoust. Soc. Amer.*, vol. 118, no. 1, pp. 471–482, 2005.

[23] F. J. Opolko and J. Wapnick, *MUMS: McGill University master samples*. Faculty of Music, McGill University, 1987.

[24] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, “RWC music database: Music genre database and musical instrument sound database.” in *Proc. Int. Soc. Music Inform. Retrieval Conf.*, vol. 3, 2003, pp. 229–230.

[25] J. W. Beauchamp, “Analysis and synthesis of musical instrument sounds,” in *Analysis, Synthesis, and Perception of musical sounds*. Springer, 2007, pp. 1–89.

[26] J. Whitehill, P. Ruvolo, T. Wu, J. Bergsma, and J. Movellan, “Whose vote should count more: Optimal integration of labels from labelers of unknown expertise,” in *Advances in Neural Inform. Process. Syst.*, vol. 22, no. 2035-2043, 2009, pp. 7–13.

[27] M. M. Bradley and P. J. Lang, “Affective norms for english words (ANEW): Instruction manual and affective ratings,” *Psychology*, no. C-1, pp. 1–45, 1999.

[28] “happy, sad, heroic, scary, comic, shy, joyful, and depressed,” *Cambridge Academic Content Dictionary*, 2013, online: http://goo.gl/v5xJZ (17 Feb 2013).

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