# The Effects of Early-Release on Emotion Characteristics and Timbre in Non-Sustaining Musical Instrument 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

*Timbre evokes emotion in music, as do loudness, pitch, rhythm, and other music qualities. Recent research has confirmed the correlation between timbral features with emotion in isolated musical instrument tones. Attack time, spectral centroid and its deviation have significant correlation with emotional characteristics in tones of sustaining (e.g., trumpet) and non-sustaining instruments (e.g., harp). Previous work has considered instrument tones of one to two seconds duration. This paper presents a similar experiment with pairwise comparison of very short (250 ms) isolated tones of non-sustaining instruments, including plucked string, pitched percussion, and keyboard instruments. The tones were investigated for the emotion categories Happy, Sad, Heroic, Scary, Comic, Shy, Joyful, and Depressed. The results agreed with those for one-second tones, with plucked string tones evoking negative emotional categories and pitched percussion tones evoking positive categories. Surprisingly, the emotional characteristics were even more clearly distinguished for 250 ms tones. We also found that the density of significant harmonics, a feature that has particular relevance to non-sustaining tones, was observed to achieve especially strong correlation for all tested emotion categories in these tones with early release.*

## 1. INTRODUCTION

Music is undoubtedly one influential means to convey emotion. There are emotional messages in any music, be it ceremonial or casual. Previous researchers have examined the recognition of music emotion using melody [1], harmony [2], rhythm [3], etc.

Emotion is also closely related with timbre. Scherer and Oshinsky [4] demonstrated the importance of timbre in the rating of synthetic tones. Timbre was shown useful in discrimination of emotion categories [5], and in music genre recognition and discrimination [6].

Timbre is the characteristic fingerprint of musical instruments. Eerola *et al.* [7] and Wu *et al.* [8] studied instrument tone samples, and confirmed the correlation between various timbre features and emotion for one- or two-second isolated tones for mainly sustaining instruments, such as flute and violin. We also conducted an experiment that showed similar results for non-sustaining instruments [9].

Gjerdingen and Perrott [10] discovered that music genre can be recognized in music excerpts as short as 250 ms. Does the shorter duration affect timbre and emotion? In this study, we compare non-sustaining tones with early release after 250 ms, and measure the correlation of timbral features and emotional categories. Eight plucked string, pitched percussion, and keyboard instrument tones are investigated for eight emotional categories.



## 2. SPECTRAL CORRELATION MEASURES

We analyzed the correlation between music emotion and several common timbral features. They are listed in Table 1. All of them are weighted by the instantaneous rms amplitude.

### 2.1 Density of Significant Harmonics

We propose a new frequency domain feature in addition to those commonly used in studies of musical timbre. By defining a threshold, we can say that certain harmonics are significant if their amplitude is above the threshold. The density is the ratio of the number of significant harmonics over the significant harmonic bandwidth. This is then averaged over all the frames. The results of our tested tones can be seen in Figure 1. The threshold used in this study is 5% of the amplitude of the fundamental frequency.

## 3. EXPERIMENT

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

### 3.1 Stimuli

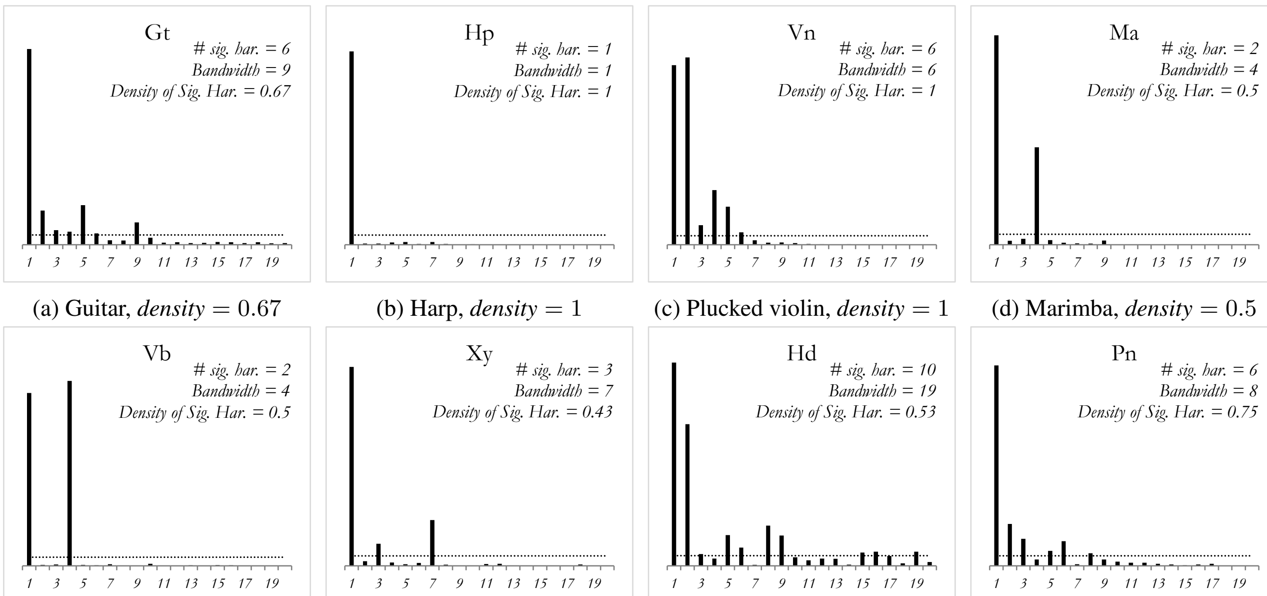
#### 3.1.1 Prototype Instrument Tones

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)
- Pitched percussion instruments: marimba (Ma), vibraphone (Vb), xylophone (Xy)
- Keyboard instruments: harpsichord (Hd), piano (Pn)

| *Feature* | *Meaning* |
|---|---|
| Attack Time | Time to reach the first local maximum in rms amplitude from the beginning of the tone |
| Decay Ratio | Ratio between the rms amplitude 30 ms before the tone ends against the peak rms amplitude |
| Decay Slope | Average difference in rms amplitude between adjacent analysis frames |
| Spectral Centroid | A quantitative measure of perceptual brightness based on the centre of gravity of a harmonic spectrum |
| Spectral Centroid Deviation | Rms deviation of the normalized spectral centroid [11] |

**Table 1**: Timbral features used in analysis




(a) Guitar, *density* $= 0.67$ (b) Harp, *density* $= 1$ (c) Plucked violin, *density* $= 1$ (d) Marimba, *density* $= 0.5$

(e) Vibraphone, *density* $= 0.5$ (f) Xylophone, *density* $= 0.43$ (g) Harpichord, *density* $= 0.53$ (h) Piano, *density* $= 0.75$

**Figure 1**: Average spectrum of of the tested instruments, with the threshold represented by the dotted line

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

The silence before each tone was removed. To study the effect of early-release, the tone durations were then truncated to 250 ms. A 30 ms linear fade-out was introduced before the end of each tone. Loudness was first equalized by peak rms amplitude, and manual refinement was made until the tones were judged of equal loudness by the authors.

#### *3.1.2 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. More details of the phase-vocoder analysis process is given by Beauchamp [12].

### 3.2 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.

A consistency test was done based on the four comparisons of every pair of instruments. A consistency of 1 represents perfect consistency, whereas 0.5 represents random guessing. The mean average consistency of all subjects was 0.774. The three least consistent subjects were excluded from the results, leaving 31 subjects.

### 3.3 Emotional Categories

The subjects compared the stimuli in terms of eight emotional categories: Happy, Sad, Heroic, Scary, Comic, Shy, Joyful, and Depressed. These terms and their counterparts in other languages are commonly used by composers for tempo and expressions in their pieces. Simple English words are used to avoid confusion for the subjects who are not native English speakers.

### 3.4 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 making the stronger impression of a given emotional category. Each combination of two different instruments was presented in four trials for each emotional category, and the listening test totaled $\binom{8}{2} \times 4 \times 8 = 896$ trials. For each emotional category, 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 [13]. The listening test took about 40 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. The Sound Blaster sound card utilized 24-bit depth with a maximum sampling rate of 96 kHz and a 108 dB S/N ratio.

## 4. EXPERIMENT RESULTS

### 4.1 Voting Results

The raw results were pairwise votes for each instrument pair and each emotion. The instruments were then ranked using the Bradley-Terry-Luce (BTL) model [14]. For each emotional category, the BTL scale values for the instruments sum up to 1. The BTL value for each instrument is the probability that listeners will choose that instrument when considering a certain emotional category. The 95% confidence intervals of the BTL values were obtained to test the significance of the instrument ranks, as shown in Figure 2.

These rankings are compared with those for tones of 1.0 s duration in our previous experiment [9]. The general rankings were similar, with pitched percussion (i.e., xylophone, marimba, and vibraphone) highly ranked for positive emotional categories and plucked string instruments (i.e., harp, guitar, and plucked violin) highly ranked for negative emotion categories. The most notable difference was the ranking of the plucked violin, which was not as Sad for the longer duration. Several emotional categories, including Joyful, Sad, and Comic, had a noticeable range expansion for the shorter tones. This means that for these categories listeners found it easier to differentiate between instruments on shorter tones.

### 4.2 Correlation Results

Pearson correlation between these features and the emotional categories are given in Table 2.

The density of significant harmonics was strongly correlated with all the eight emotional categories. A sparse spectrum (i.e., low density) is more correlated to positive emotional categories such as Joyful and Comic, and vice versa. This was especially obvious for Sad and Depressed with the high correlation values.

There was significant correlation between the decay slope and most emotional categories as well. A fast-decaying tone (i.e., more negative slope) evoked darker feelings, such as Sad and Shy. This was similar to the results for longer tones.

In contrast, attack time, spectral centroid and spectral centroid deviation were less important for emotion differentiation in shorter tones. Only a few emotional categories were correlated with these features. This differs from the conventional wisdom that a brighter sound makes an impression of more positive emotional characteristics.

## 5. DISCUSSION

Our results confirmed that timbre evokes emotions in even very short tones of 250 ms duration. This duration is so short that it is only half a beat at a fast tempo of 120 beats per minute. The non-sustaining instruments were recognized in emotion-clusters of similar physical properties (e.g., pitched percussion evoked positive emotions, and plucked strings evoked negative emotions). A few timbral features were observed to be significantly correlated with emotions (e.g., density of significant harmonics and decay slope).

We found the density of significant harmonics to be a robust feature to differentiate emotions among musical instruments, regardless of the tone duration. It had a very strong impact on the perception of all emotions. In general, most string instruments have a dense spectrum sloping downwards from the fundamental frequency (e.g., see Figure 1c). Some winds instruments, such as the clarinet, can have unbalanced odd and even harmonics, while pitched percussion instruments have only a few non-adjacent harmonics (e.g., see Figure 1d).

Decay slope also had strong correlations with many emotions for non-sustaining tones of different durations. The decay slopes we tested were generally steeper at the beginning of the decay. The average slope was more gentle for longer tones. Surprisingly, decay slope was significant for most emotions in both situations. We are currently investigating why this is the case.

In the non-sustaining tones that were released early, spectral centroid and spectral centroid deviation did not have as significant an effect on emotion as in longer non-sustaining tones of 1.0 s duration. The correlation between these features and emotion was much weaker. It is probable that the tones were too short for listeners to really notice the brightness very much.

We observed the difference in emotion perception as the tone evolved. The difference in perception may be a result of the change in the listeners' focus, or the aural process of the human brain. This is yet to be determined.

The correlation between emotion and timbre provides a point of reference for composers and arrangers. By tweaking the timbre, musicians can manipulate the emotional message in the music, and create and explore new possibilities of music expression.

### Acknowledgments

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

## 6. 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] 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.

[4] 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.

[5] I. Peretz, L. Gagnon, and B. Bouchard, "Music and emotion: perceptual determinants, immediacy, and isolation

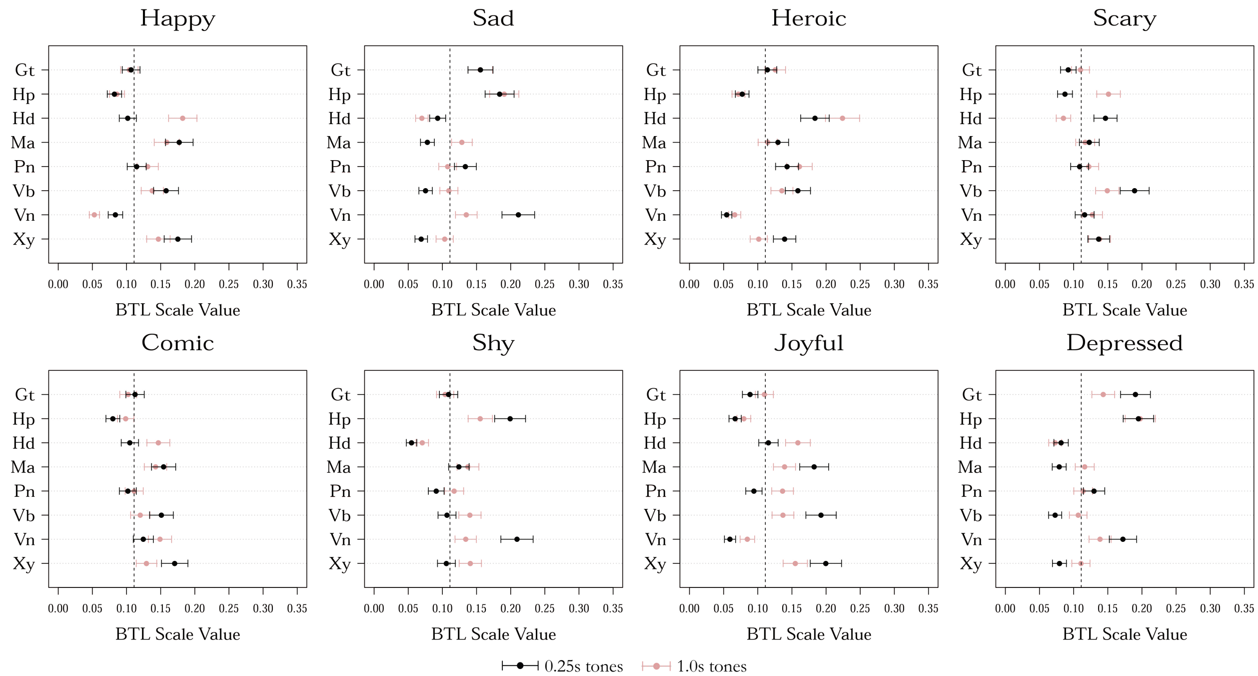



**Figure 2**: BTL scale values and the corresponding 95% confidence intervals. The dotted line represents no preference.

| Features \ Emotional Categories | Happy | Sad | Heroic | Scary | Comic | Shy | Joyful | Depressed | Number of emotional categories with significant correlation | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | *0.25s tones* | *1.0s tones* |
| Attack Time | 0.48 | **–0.73**$^{**}$ | **0.68**$^{*}$ | 0.43 | 0.35 | –0.60 | 0.56 | **–0.79**$^{**}$ | 3 | **5** |
| Decay Ratio | **0.69**$^{*}$ | **–0.79**$^{**}$ | **0.65**$^{*}$ | 0.42 | 0.47 | –0.59 | **0.74**$^{**}$ | –0.62 | 4 | 0 |
| Decay Slope | **0.67**$^{*}$ | **–0.82**$^{**}$ | **0.76**$^{**}$ | 0.54 | 0.52 | **–0.73**$^{**}$ | **0.75**$^{**}$ | **–0.68**$^{*}$ | **6** | **5** |
| Spectral Centroid | –0.15 | –0.32 | **0.67**$^{*}$ | 0.36 | –0.15 | **–0.69**$^{*}$ | 0.01 | –0.39 | 2 | **6** |
| Spectral Centroid Deviation | –0.41 | 0.05 | 0.31 | 0.11 | –0.26 | –0.39 | –0.29 | –0.06 | 0 | **5** |
| Even/odd Harmonic Ratio (AW) | –0.10 | 0.02 | 0.10 | **0.62**$^{*}$ | 0.14 | –0.01 | 0.00 | –0.26 | 1 | 1 |
| Density of Significant Harmonics | **–0.82**$^{**}$ | **0.96**$^{***}$ | **–0.83**$^{**}$ | **–0.76**$^{**}$ | **–0.70**$^{*}$ | **0.73**$^{**}$ | **–0.90**$^{***}$ | **0.94**$^{***}$ | **8** | **5** |

**Table 2**: Pearson correlation between emotional categories and the timbral features for the 0.25 s tones.
$^{***}$: $p \leq 0.01$; $^{**}$: $0.01 < p \leq 0.05$; $^{*}$: $0.05 < p < 0.1$.

after brain damage," *Cognition*, vol. 68, no. 2, pp. 111–141, 1998.

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

[7] 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.

[8] 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. (ISMIR)*, November 4-8 2013.

[9] C.-j. Chau, B. Wu, and A. Horner, "Timbre features and music emotion in plucked string, mallet percussion, and keyboard tones." in *Proc. 40th Int. Comp. Music Conf. (ICMC)*, 2014, pp. 982–989.

[10] R. O. Gjerdingen and D. Perrott, "Scanning the dial: The rapid recognition of music genres," *J. New Music Research*, vol. 37, no. 2, pp. 93–100, 2008.

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

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

[13] Cambridge University Press. Cambridge academic content dictionary. [Online]. Available: http://dictionary.cambridge.org/dictionary/american-english

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