

The Emotional Characteristics of Piano Sounds with Different Pitch and Dynamics

CHUCK-JEE CHAU, RONALD MO, AND ANDREW HORNER, *AES Member*

(chuckjee@cse.ust.hk) (ronmo@cse.ust.hk)

(horner@cs.ust.hk)

The Hong Kong University of Science and Technology, Hong Kong

The piano is an instrument extensively used in classical, jazz, and pop music since its broad pitch range and ample dynamic levels allow the instrument to become self-contained and versatile for various kinds of music. We conducted listening tests to compare the effects of pitch and dynamics on the emotional characteristics of isolated one-second piano sounds. Listeners compared the sounds pairwise over ten emotion categories. The results showed that all ten emotional categories were significantly affected by pitch and nine of them by dynamics. In particular, the emotional characteristics Happy, Romantic, Comic, Calm, Mysterious, and Shy generally increased with pitch but sometimes decreased at the highest pitches. The characteristics Heroic, Angry, and Sad generally decreased with pitch. Scary was strong in the extreme low and high registers. In terms of dynamics, the results showed that the characteristics Heroic, Comic, Angry, and Scary were stronger for loud notes, while Romantic, Calm, Mysterious, Shy, and Sad were stronger for soft notes. Surprisingly, Happy was not affected by dynamics. These results help quantify the emotional characteristics of piano sounds.

0 INTRODUCTION

Previous research has shown that both sustained and non-sustained instrument sounds have strong emotional characteristics. For example, the trumpet, clarinet, and violin were found to be sounding relatively happy compared to other sustained instruments, even in isolated sounds apart from musical context, while the horn sounded relatively sad [1–3]. The sounds of the marimba and xylophone were relatively happier compared to other non-sustained instruments, while the sounds of the harp and guitar sounded relatively sad [4, 5].

Most previous work has only studied single mid-range pitches of instrument, where loudness was equalized to allow consistent comparison. We were curious about the effects of pitch and dynamics in isolated sounds on the emotional characteristics of the piano. Based on this, a number of interesting general questions arise: Which emotional characteristics tend to increase or decrease with increasing pitch, and which are strongest in the mid-register? Which emotional characteristics are stronger for loud or soft notes? Are there any emotional characteristics that change dramatically with pitch or with dynamics, or are relatively unaffected?

We formulated the current study to systematically compare the emotional characteristics of piano notes at different pitches and dynamic levels. The study includes representative pitches ranging from the lowest to the highest octave.

The dynamic levels included loud, medium, and soft (*forte*, *mezzo*, and *piano*). All sounds were isolated with a duration of one second. They were compared pairwise over ten emotional categories: Happy, Heroic, Romantic, Comic, Calm, Mysterious, Shy, Angry, Scary, and Sad.

This work provides a basic overview of the emotional characteristics of the piano across the different octaves at different dynamics. This research can suggest possible ideas to recording engineers, composers, and pianists in manipulating the emotional characteristics of the instrument in live performances and recordings.

1 BACKGROUND

Previous research has investigated emotion recognition in music, especially addressing melody [6], harmony [7, 8], rhythm [9, 10], lyrics [11], and localization cues [12]. Similarly, researchers have found timbre to be useful in a number of applications such as automatic music genre classification [13], automatic song segmentation [14], and song similarity computation [14].

Researchers have considered music emotion and timbre together in a number of studies, and we give an overview below. We also review previous research on the timbre of the piano as well as the effects of pitch and dynamics in musical excerpts.

1.1 Music Emotion and Timbre

Timbre was found to be an important factor in music emotion in the study of Scherer and Oshinsky [15]. They investigated the relationship between emotional attributes and synthetic sounds by manipulating different acoustic parameters such as amplitude, pitch, envelope, and filter cut-off. Further studies of Peretz et al. [16], Krumhansl [17], and Filipic et al. [18] also confirmed that listeners could already discriminate emotions or even identify the artist by using very short musical excerpts as short as 0.25 s. The sounds were so short that factors such as rhythm, melody, and chord progression were largely excluded.

In these and other studies on music emotion, there are two main models for emotion measurement: categorical and dimensional [19]. The categorical model allows listeners to pick emotional words or categories of emotion, while the dimensional model uses numerical ratings of one or more dimensions. Another consideration is whether to study the expressed emotion of the sound (sometimes called perceived emotion) or the experienced emotion of the listener (sometimes called induced or felt emotion). Different studies have used different terms and shades of meaning.

Hevner's early work [20] pioneered the use of the categorical model in music and emotion research. She divided 66 adjectives into 8 groups where adjectives in the same group were related and compatible, measuring tonality, harmony, rhythm, and melody. The work of Hailstone et al. [21] and Baume [22] were also prominent examples of using emotional categories for studying the effect of timbral features and music emotion. The emotional categories used by Scherer and Oshinsky [15] were Anger, Fear, Boredom, Surprise, Happiness, Sadness, and Disgust, and other studies followed their example [23, 21, 17]. In order to determine the best categories to describe emotion, Zentner et al. [24] conducted a series of experiments to compile a list of musically-relevant emotional terms.

Perhaps the most common dimensional model used is the Valence-Arousal plane proposed by Russell [25], but others have also been considered [26]. In the work of Bigand et al. [27], the analysis of emotion ratings of musical excerpts with multidimensional scaling showed that a 3D space provided a good fit with Arousal and Valence as the primary dimensions.

Categorical or dimension models both have important uses and applications. Eerola and Vuoskoski [28, 29] compared categorical and dimensional models and observed a high correspondence between the categorical and dimensional results. They also found that the three dimensions Valence, Energy, and Tension could be reduced to the two dimensions Valence and Arousal without significantly reducing the goodness of fit.

To further investigate the relationship between instrument timbre and emotion, Eerola et al. [30] studied the correlation of perceived emotion with temporal and spectral sound features. They asked listeners to rate the perceived affect qualities of one-second instrument tones using five dimensions: Valence, Energy, Tension, Preference, and Intensity. Orchestral and some exotic instruments were in-

cluded in their collection, pitched at Eb4, at both loud and medium-loud dynamic levels but equalized for loudness to focus attention on spectral differences. They correlated the ratings with acoustic features such as attack time and brightness and found strong correlations between these acoustic features and the emotion dimensions Valence and Arousal.

In a study closely related to the current paper, Wu et al. [3, 1, 2, 31] compared the emotional characteristics of sustaining instruments. Like Ellermeier [32], they used a Bradley-Terry-Luce (BTL) model to rank paired comparisons of eight sounds from eight wind and bowed string instruments such as the trumpet, flute, and bowed violin. Eight emotional categories for expressed emotion were tested including Happy, Sad, Heroic, Scary, Comic, Shy, Joyful, and Depressed. The results showed distinctive emotional characteristics for each instrument. The timbral features spectral centroid and even/odd harmonic ratio were significantly correlated with the emotional characteristics for these sustaining instruments.

And most related to the current paper, in our recent study of non-sustained instrument sounds [4, 5], using the same methodology and emotional categories as Wu [31, 1–3], we tested several musical instruments: plucked violin, guitar, harp, marimba, xylophone, vibraphone, piano, and harpsichord. The results showed that decay slope and the density of significant harmonics were significantly correlated for these non-sustaining instruments. Among the eight instruments, the piano was often ranked in the middle, indicating neutral emotional characteristics relative to the other seven instruments. This was somewhat surprising since the piano has the widest repertoire among all instruments in classical, jazz, and pop music with abundant transcriptions also written for the piano.

Table 1 summarizes the literature on music emotion and timbre. It shows whether a categorical or dimensional model was used, as well as the particular emotional categories or dimensions. It also shows whether the study considered the expressed or perceived emotion of the sound, or the experienced emotion of the listener (felt or induced emotion), or mood in general [33]. The types of stimuli are also included (i.e., whether single tones or musical excerpts). Finally, the type of user input is listed (whether a rating of individual sounds, comparison of paired sounds, or selection from a group of options).

1.2 The Piano

Researchers have studied emotional responses to piano music in a number of studies. Somewhat related to this study, Wapnick [36] assessed preferences for pitch, tempo, and timbre in solo piano excerpts. Fast tempo and bright timbre were preferred, but there was no preference for pitch. More generally, Nair et al. [37] identified the emotional difference when listening to mechanical and expressive piano playing. They observed different brain activation for different versions of the same piece of piano music.

The timbre of the piano has been investigated in many studies, though without consideration of emotional characteristics. Guevara and Wakefield [38] investigated the

Table 1. Summary of the literature on music emotion and timbre.

Year	Author(s)	Emotion Model	Emotional Categories/Dimensions	Emotion Type	Stimuli Type	Stimuli Evaluation
1936	Hevner [20]	Categorical	8 groups of adjectives	Expressed	Musical excerpts	Rating
1977	Scherer and Oshinsky [15]	Categorical and Dimensional	<i>Dimensions: Pleasantness, Activity, and Potency</i> <i>Categories: Anger, Fear, Boredom, Surprise, Happiness, Sadness, and Disgust</i> Happy/Sad	Expressed	Instrument tones	Rating
1998	Peretz, Gagnon, and Bouchard [16]	Dimensional		Perceived	Musical excerpts	Rating
2005	Bigand, Vieillard, Madurell, Marozeau, and Dacquet [27]	Dimensional	Valence, Arousal, and a third dimension expressing the influence of body gestures	Induced	Musical excerpts	Paired comparison
2006	Lu, Liu, and Zhang [23]	Categorical	Contentment, Depression, Exuberance, and Anxious/Frantic	Mood	Musical excerpts	Selection
2008	Zentner, Grandjean, and Scherer [24]	–	–	Felt and Perceived	–	Rating
2009	Hailstone, Omar, Henley, Forst, Kenward, and Warren [21]	Categorical	Happiness, Sadness, Fear, and Anger	Perceived	Novel melodies	Rating
2009	Yang, Lin, Su, and Chen [34]	Dimensional	Valence and Arousal	Induced	Musical excerpts	Rating
2010	Krumhansl [17]	Categorical	Happiness, Sadness, Anger, Fear, and Tenderness	Perceived	Musical excerpts	Rating
2010	Filipic, Tillmann, and Bigand [18]	Dimensional	Degree of Emotionally Touching	Felt	Musical excerpts	Rating
2011	Panda and Paiva [35]	Dimensional	Valence and Arousal	Conveyed	Musical excerpts	Selection
2011	Eerola and Vuoskoski [28]	Categorical and Dimensional	<i>Categories: Happy, Sad, Tender, Fearful, Angry</i> <i>Dimensions: Valence, Energy, Tension</i>	Perceived	Film music excerpts	Rating
2011	Vuoskoski and Eerola [29]	Categorical, Dimensional, and Geneva Emotional Music Scale	<i>Categories: Sadness, Happiness, Tenderness, Fear, and Anger</i> <i>Dimensions: Valence, Arousal, and Tension</i> Valence, Energy, Tension, Preference, and Intensity	Induced	Film music excerpts	Rating
2012	Eerola, Ferrer, and Alluri [30]	Dimensional	Valence, Energy, Tension, Preference, and Intensity	Perceived	Instrument tones	Rating
2013	Kostek and Plewa [26]	Dimensional	Energy and Arousal	Mood	Musical excerpts	Rating
2015	Baume [22]	Categorical	Terror, Joy, Peace, and Excitement	Induced	Music tracks	–
2014–15	Wu, Horner, and Lee [3, 1, 2, 31]	Categorical	Happy, Sad, Heroic, Scary, Comic, Shy, Joyful, and Depressed	Perceived	Instrument tones	Paired comparison
2014–15	Chau, Wu, and Horner [4, 5]	Categorical	Happy, Sad, Heroic, Scary, Comic, Shy, Joyful, and Depressed	Perceived	Instrument tones	Paired comparison



Fig. 1. Selected pitches on the piano.

dynamics and articulation of the piano using modal distribution. Yuen and Horner [39] used sampling of the inharmonic attack to improve realism together with wavetable synthesis for the simpler decay. Lee and Horner [40] clustered groups of harmonics with similar frequency deviations in the same wavetable in a wavetable synthesis model with disjoint sets of harmonics. Piano sounds were successfully synthesized with perceptual modeling by Hamadicharef and Ifeachor [41], using analysis, filtering, and quality assessment. Cella [42] separated recorded piano sounds into harmonic and noise components. To more efficiently describe the timbre of the piano, Bernays and Traube [43] obtained a four-dimensional space of common piano timbre descriptors derived from professional pianists' suggestions.

So, while several studies have considered the timbre of the piano across its pitch and dynamic range [36, 38–43], no studies have measured the changes in emotional characteristics of piano sounds across its pitch and dynamic range to our knowledge.

1.3 Pitch and Dynamics

Several researchers have considered the effects of different pitch and dynamics on musical excerpts. Kamenetsky et al. [44] investigated the effects of tempo and dynamics on the perception of four MIDI musical excerpts. One version had static tempo and dynamics, another varying tempos, another varying dynamics, and the last varying tempos and dynamics. Participants then rated each excerpt on a seven-point scale for likeability and emotional expressiveness. While tempo was found to have no effect on the ratings, variations in dynamics were found to result in higher ratings for both measurements.

Krumhansl [45] investigated changes in emotion for subjects listening to three-minute musical excerpts and found that large variations in dynamics and pitch resulted in significantly higher ratings for the category Fear. Work by Huron et al. [46] into the perception of sine tone and MIDI synthesized piano melodies found that higher-pitched melodies were considered more submissive than lower-pitched melodies. Kosta et al. [47] ran comprehensive experiments on the meanings of dynamic markings in piano pieces.

So, while there have been a variety of studies that have considered the effects of pitch and dynamics and other musical features such as tempo on listener preferences for musical excerpts, there is no study on the effects of pitch and dynamics on the emotional characteristics of individual piano sounds. The closest is the work by Eerola et al. [30] already mentioned in Sec. 1.1, where they considered multiple dynamic levels (loud and medium-loud) in their study of music emotion and timbre, where the loudness of the tones had been equalized so as to determine the effects

of timbre alone. But, all tones were pitched at E_b4 in the study.

2 EXPERIMENT

We conducted listening tests to compare the effect of pitch and dynamics on the emotional characteristics of piano sounds. For this investigation, we used short sounds isolated from musical context in order to isolate the effects of pitch and dynamics. We also correlated the emotion ratings with several pitch, dynamic, and spectral features in order to better understand how they are related.

2.1 Test Materials

2.1.1 Stimuli

The stimuli used in the listening tests were sounds from a grand piano with different combinations of dynamics and pitch. All sounds were from the *RWC Music Database* [48]. The sounds were played by the same pianist on a Steinway grand piano (Piano #3, “Normal”: 013PFNOF, 013PFNOM, 013PFNOP).

Three different levels of dynamics were used: *forte*, *mezzo*, and *piano*, with *forte* being the loudest, *piano* the softest, and *mezzo* in between. We have listened to all the selected pitches and dynamics and confirmed that the original expression of dynamics by the pianist were consistent and reasonable across different pitches and across different dynamic levels, so the amplitude of the samples was not further adjusted.

To avoid the effect of pitch intervals other than the octave influencing emotional characteristics, we chose only the C pitches of the piano (C1–C8), with C1 the lowest and C8 the highest (as shown in Fig. 1). All sounds were recorded and sampled at 44,100 Hz with 16-bit resolution and played back using the D/A converter with 24-bit resolution at the original sampling rate.

Any silence before the onset of each sound was removed. The sound durations were then truncated to 1.0 second using a 30 ms linear fade-out before the end of each sound. In all cases, the fade-outs sounded like a natural damping or release of the sound.

We chose 1 s as the duration as it was long enough to allow listeners to hear a representative portion of the decay but not too long or else the overall length of the listening test would become too long. In our previous study [5], we found that emotional characteristics were clear for very short sounds of 0.25 s duration, but only for mid-register pitches. To allow listeners to hear the details of the attack and early decay well for the lowest pitches, we could not use a duration too short, and 1 s tones were the best compromise in length. We felt that listeners would judge 1 s or 2 s sounds (or longer) with similar results.

Table 2. Ten chosen emotional categories and related music expression markings commonly used by classical composers.

Emotional Category	Commonly-used Italian musical expression marks
Happy	<i>allegro, gustoso, gioioso, giocoso, contento</i>
Heroic	<i>eroico, grandioso, epico</i>
Romantic	<i>romantico, appassionato, affetto, affettuoso, passionato</i>
Comic	<i>capriccio, ridicolosamente, spiritoso, comico, buffo</i>
Calm	<i>calmato, tranquillo, pacato, placabile, sereno</i>
Mysterious	<i>misterioso, misteriosamente</i>
Shy	<i>timido, riservato, timoroso</i>
Angry	<i>adirato, stizzito, furioso, furibondo, rabbioso, irato</i>
Scary	<i>sinistro, terribile, allarmante, feroce</i>
Sad	<i>dolore, lacrimoso, lagrimoso, mesto, triste, mesto, freddo</i>

2.1.2 Emotional Categories

The subjects compared the stimuli in terms of 10 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 as shown in Table 1 (e.g., Happy, Sad, Scary/Fear/Calm, Tender/Calm/Romantic), roughly corresponding to the four quadrants of the Valence–Arousal plane, but there are lots of variations beyond that [19].

We selected these 10 categories because composers often use these terms in tempo and expression markings in their scores (e.g., *mysteriously, shyly*, etc.). We chose to use simple English emotional categories so that they would be familiar and self-apparent to lay people rather than Italian music expression markings traditionally used by classical composers to specify the character of the music. The chosen emotional categories and related Italian expression markings [49–52] are listed in Table 2. We tried to include a well-balanced group of emotional categories, and these 10 categories are similar to the 8 adjective groups of Hevner [20]. Other researchers have also used some of these (or related) emotional categories [15, 24, 21]. Our previous research showed the statistical significance of these terms for comparing single instrument sounds [31, 1–5].

One advantage of using a categorical instead of a dimensional emotional model is that it allows faster decision-making by listening test subjects. However, these emotional categories can still be represented in a dimensional model, such as the Valence–Arousal model [25]. Their ratings according to the *Affective Norms for English Words (ANEW)* [53] are shown in Fig. 2 using the Valence–Arousal model. Valence indicates the positivity of an emotional category; Arousal indicates the energy level of an emotional category. Though Scary and Angry are similar in terms of Valence and Arousal, they have distinctly different meanings. Romantic, Happy, Comic, and Heroic are also similar in Valence and Arousal but distinctly different in meaning.

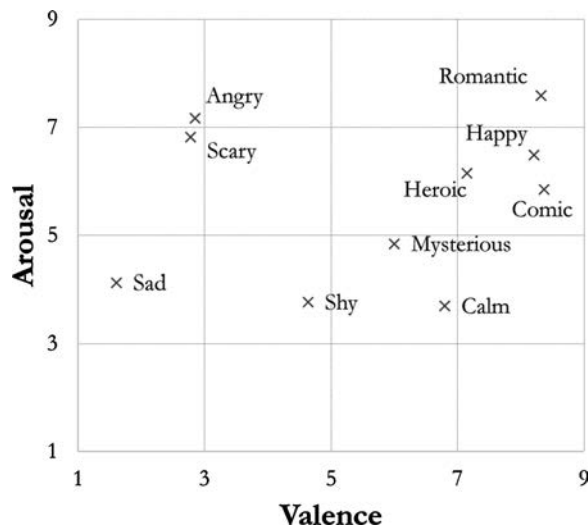


Fig. 2. Distribution of the emotional characteristics in the dimensions Valence and Arousal. The Valence and Arousal values are given by the 9-point rating in ANEW [53].

2.2 Test Procedure

There were 26 subjects hired for the listening test, ranging in age from 19 to 24 with an average age of 20.8. All subjects were fluent in English. All subjects were undergraduate students at the Hong Kong University of Science and Technology, where all courses are taught in English. All stated that they had no known hearing problems. Subjects did not have highly-trained ears (e.g., recording engineers, professional musicians, or music conservatory students) but were average attentive listeners.

The subjects were seated in a quiet room with 39 dB SPL background noise level. Residual noise was mostly due to computers and air conditioning. The noise level was further reduced with headphones. Sound signals were converted to analog with a Sound Blaster X-Fi Xtreme Audio sound card with 24-bit resolution with a sampling rate of 44,100 Hz and a 108 dB S/N ratio, and then presented through Sony MDR-7506 headphones. We felt that basic-level professional headphones were adequate in representing the piano sounds for this test as the sounds were readily distinguishable. A big advantage of the Sony MDR-7506 headphones is their relative comfort in a relatively long listening test such as this one, especially for subjects not used to tight-fitting studio headphones. The loudness on all computers were calibrated manually so that the C4 forte sound sample sounded the same as judged by the authors.

The subjects were provided with an instruction sheet containing definitions of the 10 emotional categories from the *Cambridge Academic Content Dictionary* [54]. The dictionary definitions we used in our experiment are shown in Table 3.

Every subject made pairwise comparisons on a computer among all the 24 combinations of pitches and dynamics for each emotional category. During each trial, subjects heard a pair of piano sounds of different pitches/dynamics and were prompted to choose the sound that represented the given emotional category more strongly. Each trial was a

Table 3. The dictionary definitions of the emotional categories used in our experiment.

Emotional Category	Definition [54]
Happy	Glad, pleased
Heroic	Exhibiting or marked by courage and daring
Romantic	Making someone think of love
Comic	Causing laughter or amusement
Calm	A quiet and peaceful state or condition
Mysterious	Exciting wonder, curiosity, or surprise while baffling efforts to comprehend or identify
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

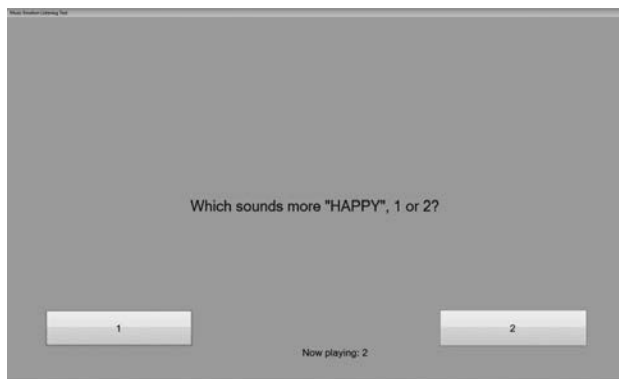


Fig. 3. Listening test interface.

single paired comparison requiring minimal memory from the subjects. In other words, subjects did not need to remember all of the tones, just the two in each comparison. One big advantage of using paired comparisons of emotional categories is that it allows faster decision-making by the subjects. Paired comparison is also a simple decision and is easier than absolute rating. In the beginning of the test subjects were given one practice pair for each emotional category in which results were discarded.

Fig. 3 shows a screenshot of the listening test interface. Each combination of two different pitch or dynamic was presented once for each emotional category, and the listening test totaled $\binom{24}{2} \text{ combinations} \times 10 \text{ emotional categories} = 2760$ 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). However, the emotional categories were presented in order to avoid confusing and fatiguing the subjects. As with any listening test, there can be learning at the beginning and fatigue at the end. For this test, there were 10 test trials at the start of each emotional category that were not used in calculations to minimize the effect of learning.

The listening test took about 3 hours, with a short break of 5 minutes after every 30 minutes to help minimize listener fatigue and maintain consistency.

3 EXPERIMENT RESULTS

We obtained pairwise voting results from the subjects. The Bradley–Terry–Luce (BTL) model [55, 56] was then used to derive rankings based on the number of positive votes each sound received for each emotional category. For each emotional category, the BTL scale values for all the combinations of dynamic and pitch sum up to 1. The BTL value for each sound is the probability that listeners will choose that sound when considering a certain emotional category. For example, if all 24 combinations were judged equally happy, the BTL scale values would be $1/24 = 0.0417$.

Fig. 4 displays the Bradley–Terry–Luce (BTL) scale values of the sounds. The corresponding 95% confidence intervals are also shown for each sample.

For the emotional category Happy, Fig. 4 shows similar curves for the three dynamic levels. There was a clear upward trend in pitch, with the exception of the highest pitches. The overall trend was an arching curve peaking at C6, with the lowest pitches least Happy. Contrastingly, the result for Sad shows wider differences between the dynamic levels. There were consistently sadder responses for soft notes. There was a general downward trend in pitch with the lowest pitch range the saddest.

The responses for Heroic and Angry were similar. The loud notes were most Heroic and Angry, with a downward trend in pitch. The biggest difference between Heroic and Angry was that Angry was strictly monotonic and Heroic was not.

The responses for Romantic, Calm, and Shy were basically similar. The soft notes were most Romantic, Calm, and Shy, with an upward trend in pitch. The difference between dynamic levels was large. Romantic decreased at the highest notes (C7 and C8).

For Comic Fig. 4 shows an arch with the peak at mid-range pitch C5. The response was strongest for louder sounds. The response curve for Scary is an inverted arch, almost a mirror image of Comic. The lowest and highest pitches were scarier, a result that indirectly agrees with Krumhansl [45] who found that large variations in pitch resulted in significantly higher ratings for Fear. The differences between dynamic levels were not as large as in most other emotional categories.

For Mysterious there was an upward trend in pitch and relatively similar curves for the three dynamic levels.

For each emotional category, a two-way ANOVA without replication was performed on the continuous BTL scale values to test for the effects of pitch and dynamics. As a preliminary step, a Shapiro–Wilk test was performed to check whether the BTL data were normally distributed, and only Comic–C1 and Angry–*mezzo* were not normally distributed (see Tables 8 and 9 in the Appendix).

Also, since there is only one BTL value for each pitch–dynamic pair, the variance in each factor combination (i.e., pitch–dynamic pair) is zero. Since two pitch–dynamic pairs were not normally distributed, we performed both parametric ANOVAs (i.e., two-way ANOVA without

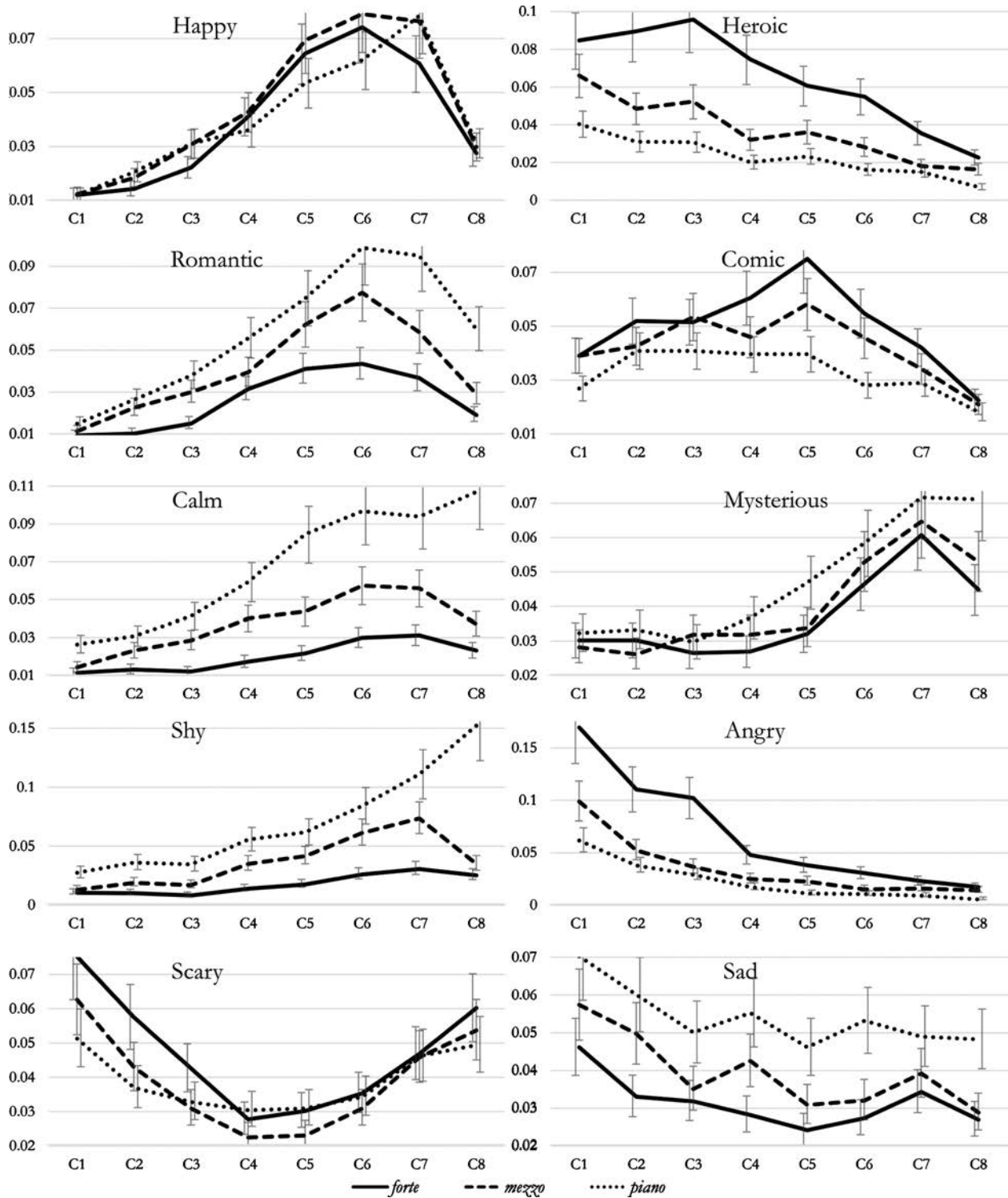


Fig. 4. Emotional characteristics of piano sounds based on the BTL scale values and the corresponding 95% confidence intervals.

replication) and non-parametric tests (i.e., Friedman ANOVA by ranks).

Table 4 shows the parametric and non-parametric results, which are in agreement. The effect of pitch was significant for all ten emotional categories at the $p < 0.05$ level, and the effect of dynamics was significant for nine of the ten emotional categories at the $p < 0.05$ level. The effect of dynamics was not significant for Happy.

3.1 Correlation with Acoustic Features

The continuous BTL scale values were then correlated with extracted acoustic features using linear Pearson correlation and partial correlation. The acoustic features we tested are listed in Table 5. The feature values of the piano sounds are given in Tables 10 and 11 in the Appendix.

Since we previously found that correlation was stronger for non-sustaining instrument sounds when

Table 4. p -values from ANOVAs and Friedman tests for the effects of pitch and dynamics. Values that were not significant ($p \geq 0.05$) are shown in bold.

	<i>Two-Way ANOVA</i>		<i>Friedman Test</i>	
	Pitch	Dynamics	Pitch	Dynamics
Happy	0.0000	0.1770	0.0051	0.0724
Heroic	0.0000	0.0000	0.0058	0.0003
Romantic	0.0000	0.0000	0.0051	0.0003
Comic	0.0000	0.0000	0.0203	0.0015
Calm	0.0060	0.0000	0.0089	0.0003
Mysterious	0.0000	0.0010	0.0086	0.0046
Shy	0.0260	0.0010	0.0084	0.0003
Angry	0.0000	0.0010	0.0041	0.0003
Scary	0.0000	0.0140	0.0044	0.0208
Sad	0.0000	0.0000	0.0106	0.0003

Table 5. Features used in analysis of the piano sounds.

Feature	Meaning
Fundamental frequency	Fundamental frequency of the sound
RMS amplitude	Amplitude of the sound
Attack Time (ms)	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 over the peak RMS amplitude
Attack/Decay Slope (ms^{-1})	Average difference in RMS amplitude (linear) between adjacent analysis frames
Density of Significant Harmonics	Ratio of the number of significant harmonics over the significant harmonic bandwidth [5]
Spectral Centroid	A quantitative measure of perceptual brightness based on the center of gravity of a harmonic spectrum, defined here as $spectral\ centroid = \frac{\sum_k k a_k}{\sum_k a_k}$, where k is the k^{th} harmonic, a_k is the amplitude of the k^{th} harmonic [57]
Spectral Centroid Deviation	RMS deviation of the normalized spectral centroid [58]
Spectral Irregularity	Jaggedness of a spectrum from harmonic to harmonic [59]
Even/Odd Harmonic Ratio	Amplitude ratio of even and odd harmonics [60]
Tristimulus	Relative weights of of harmonic groups [61]

amplitude-weighted averages were used [4, 5], all the spectral features in the current experiment were amplitude weighted. Amplitude weighting gives more weight to high-amplitude parts of the sound such as the later attack and early decay, and de-emphasizes the noisy soft late decay. The exact weighting is the ratio of the current linear amplitude over the peak linear amplitude at any point. We also included pitch- and dynamic-related features.

Pearson correlation between the features and the BTL scale values for each emotional category are given in Table 6. Perhaps not surprisingly, most of the features showed significant correlation with most of the emotional categories

(see last column in Table 6) since changes in pitch and dynamics are accompanied by corresponding changes in timbre. In particular, both pitch (logarithm of fundamental frequency) and dynamics (peak RMS amplitude in dB) correlated with eight emotional categories. Seven of the emotional categories correlated with nearly all the features (see bottom row in Table 6). The three exceptions were: Comic, Scary, and Sad. It is easy to see why, since the arching shape of Comic, the bowl-like shape of Scary, and the somewhat oscillatory curves of Sad in Fig. 4 contrast with the monotonic (or near-monotonic) changes in the feature values in Tables 10 and 11.

Partial correlation analysis was also done to study the effects of pitch and dynamics. Table 7 shows the correlation between emotional categories and features when the effects of pitch and dynamics were removed. Different from Table 6, most of the remaining features only showed significant correlation with half or less of the emotional categories (see last column in Table 7). Attack slope was the only feature that increased in Table 7 compared to Table 6. Only four of the emotional categories correlated with half or more of the features (see bottom row in Table 7). The biggest changes were for Heroic and Shy, which decreased from nearly all significant features in Table 6 to only one in Table 7 when the effects of pitch and dynamics were removed. This indicates Heroic and Shy were strongly affected by pitch and dynamics and less affected by temporal and spectral envelope features than other emotional categories. Conversely, Scary and Comic more than doubled in the number of significant features when the effects of pitch and dynamics were removed, revealing their relative sensitivity to temporal and spectral envelope features.

4 DISCUSSION

The main goal of our work was to determine how the emotional characteristics of piano sounds vary with pitch and dynamics. With respect to the original motivating questions of this paper, from Table 4 and Fig. 4 we can observe the following regarding pitch:

- All ten emotional categories were significantly affected by pitch.
- Happy, Romantic, Calm, Mysterious, and Shy generally increased with pitch but sometimes decreased at the highest pitches.
- Heroic, Angry, and Sad generally decreased with pitch, though with different slopes.
- Comic was strongest in the mid-register and weaker in the highest and lowest registers.
- Scary was strongest in the lowest and highest registers.

Regarding dynamics:

- Nine of the emotional categories were significantly affected by dynamics.
- Heroic, Comic, Angry, and Scary were stronger for loud notes.

Table 6. Pearson correlation between emotional categories and the piano features with significant correlation values shown in bold (for $p < 0.05$).

Features \ Emotional Categories	Emotional Categories										No. of emotional categories with significant correlations
	Happy	Heroic	Romantic	Comic	Calm	Mysterious	Shy	Angry	Scary	Sad	
Log of Fundamental Frequency	0.64	-0.62	0.54	-0.32	0.56	0.82	0.59	-0.74	-0.13	-0.47	8
Peak RMS Amplitude (dB)	-0.43	0.88	-0.67	0.63	-0.86	-0.86	-0.89	0.81	0.13	-0.08	8
Attack time (ms)	0.64	-0.34	0.58	-0.14	0.36	0.61	0.32	-0.46	-0.15	-0.41	5
Attack slope (linear amplitude/ms)	0.28	-0.08	-0.14	0.01	-0.24	0.25	-0.19	-0.19	0.23	-0.62	1
Decay ratio	-0.80	0.55	-0.65	0.10	-0.58	-0.72	-0.54	0.74	0.36	0.50	8
Decay slope (linear amplitude/ms)	0.41	-0.78	0.56	-0.31	0.63	0.50	0.57	-0.89	-0.54	0.04	8
Density of Significant Harmonics	0.82	-0.60	0.65	-0.08	0.56	0.66	0.49	-0.79	-0.39	-0.42	8
Spectral Centroid	-0.72	0.50	-0.59	-0.05	-0.51	-0.55	-0.45	0.78	0.58	0.55	9
Spectral Centroid Deviation	-0.64	0.49	-0.54	-0.06	-0.47	-0.48	-0.40	0.77	0.61	0.50	8
Spectral Irregularity	0.52	-0.72	0.59	-0.58	0.67	0.90	0.70	-0.70	0.06	-0.17	8
Even/Odd Harmonic Ratio	-0.57	0.50	-0.50	0.35	-0.51	-0.74	-0.50	0.47	-0.08	0.27	7
Tristimulus T1 (harmonic 1)	0.72	-0.65	0.65	-0.31	0.63	0.84	0.61	-0.75	-0.18	-0.42	8
Tristimulus T2 (harmonics 2-4)	-0.47	0.52	-0.49	0.54	-0.54	-0.87	-0.58	0.46	-0.29	0.25	8
Tristimulus T3 (harmonics 5+)	-0.78	0.63	-0.66	0.12	-0.60	-0.70	-0.54	0.83	0.45	0.46	9
No. of features with significant correlations	13	12	13	3	12	13	11	13	4	9	

- Romantic, Calm, Mysterious, Shy, and Sad were stronger for soft notes.
- Surprisingly, Happy was not affected by dynamics.
- Heroic, Romantic, Comic, Calm, Shy, Angry, and Sad showed consistently large changes with dynamics.
- The medium dynamic level (*mezzo*) was nearly always in the middle among the three tested dynamic levels.

The results show that pitch generally had a similar effect on emotional categories with similar Valence. The high-Valence characteristics Happy, Romantic, Comic, and Calm

had broadly similar shapes in Fig. 4 (mostly increasing, usually arching), while the low-Valence characteristics Angry and Sad were decreasing. Heroic however was an outlier, since it is high in Valence but with a downward pitch trend in Fig. 4. Scary was another exception, showing a downward trend like Angry and Sad in the lowest register, but with an upward trend in the highest register. Heroic and Scary were also exceptions in the partial correlations in Tables 6 and 7, indicating different sensitivities to temporal and spectral envelope features than other emotional categories.

Dynamics had a similar effect on most emotional categories with similar Arousal. The high-Arousal

Table 7. Pearson partial correlation between emotional categories and the piano features with the effects of pitch and dynamics removed. Significant correlation values shown in bold (for $p < 0.05$).

Features \ Emotional Categories	Emotional Categories										No. of emotional categories with significant correlations
	Happy	Heroic	Romantic	Comic	Calm	Mysterious	Shy	Angry	Scary	Sad	
Attack time (ms)	0.32	-0.08	0.52	-0.04	0.30	0.33	0.14	-0.05	-0.10	0.16	1
Attack slope (linear amplitude/ms)	-0.39	-0.35	-0.50	-0.38	-0.60	-0.08	-0.51	0.06	0.47	0.42	4
Decay ratio	-0.74	0.09	-0.63	-0.66	-0.61	0.17	-0.22	0.39	0.70	0.10	5
Decay slope (linear amplitude/ms)	0.19	-0.42	0.13	0.33	-0.03	-0.41	-0.32	-0.77	-0.67	-0.11	3
Density of Significant Harmonics	0.69	-0.39	0.52	0.51	0.45	-0.17	0.04	-0.57	-0.57	0.05	6
Spectral Centroid	-0.41	0.14	-0.43	-0.7	-0.40	0.45	-0.05	0.62	0.89	0.41	5
Spectral Centroid Deviation	-0.30	0.19	-0.31	-0.6	-0.27	0.42	0.02	0.63	0.80	0.37	3
Spectral Irregularity	-0.14	-0.06	0.02	-0.62	-0.04	0.48	0.00	0.32	0.58	0.63	4
Even/Odd Harmonic Ratio	-0.19	0.03	-0.16	0.18	-0.20	-0.38	-0.12	-0.25	-0.28	-0.18	0
Tristimulus T1 (harmonic 1)	0.57	-0.26	0.65	0.13	0.63	0.33	0.22	-0.17	-0.21	0.14	3
Tristimulus T2 (harmonics 2-4)	0.11	-0.12	-0.05	0.56	-0.1	-0.65	-0.20	-0.53	-0.70	-0.32	4
Tristimulus T3 (harmonics 5+)	-0.65	0.36	-0.57	-0.66	-0.51	0.31	-0.02	0.67	0.87	0.17	6
No. of features with significant correlations	4	1	7	7	5	3	1	6	9	1	

characteristics Heroic, Comic, Angry, and Scary were strongest for loud notes, while the low-Arousal characteristics Calm, Mysterious, Shy, and Sad were strongest for soft notes. Romantic was an exception. It seems listeners were interpreting Romantic as tender-Romantic more than passionate-Romantic in judging the piano sounds. The high-Arousal category Happy was another exception since it was not affected by dynamics.

The results demonstrate how categorical emotional models can give more nuance and detail than the dimensional model with only Valence and Arousal, since they can distinguish between emotional categories with similar Valence and Arousal values. For example, Angry and Scary have similar Valence and Arousal values, but are very different in Fig. 4. These results suggest that they are distinctly different emotional characteristics. The same idea holds for Happy, Heroic, and Romantic.

The above results can give possible ideas to musicians in instrumentation, pianists in blending and balancing notes, and recording engineers in mixing recordings and live performances. Emotional characteristics can be manipulated in a recording, performance, or composition by emphasizing pitches and dynamics that are comparatively stronger in representing these characteristics. The results confirm some existing common practices for emotional emphasis (e.g., using loud and low piano chords for angry passages) and expose some less well-known ones such as the potential Comic character of the loud mid-range.

5 ACKNOWLEDGMENTS

This work has been supported by Hong Kong Research Grants Council grant HKUST613112. Thanks very much to the anonymous reviewers for their careful and insightful comments to improve the clarity, presentation, and analysis in the the paper.

6 REFERENCES

- [1] 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," *Proc. 40th Int. Comput. Music Conf. (ICMC)* (2014), pp. 928–934.
- [2] B. Wu, A. Horner, and C. Lee, "Emotional Predisposition of Musical Instrument Timbres with Static Spectra," *Proc. 15th Int. Soc. Music Inform. Retrieval Conf. (ISMIR)* (2014 Nov.), pp. 253–258.
- [3] B. Wu, A. Horner, and C. Lee, "The Correspondence of Music Emotion and Timbre in Sustained Musical Instrument Sounds," *J. Audio Eng. Soc.*, vol. 62, pp. 663–675 (2014 Oct.). doi: <http://dx.doi.org/10.17743/jaes.2014.0037>
- [4] C.-J. Chau, B. Wu, and A. Horner, "Timbre Features and Music Emotion in Plucked String, Mallet Percussion, and Keyboard Tones," *Proc. 40th Int. Comput. Music Conf. (ICMC)* (2014), pp. 982–989.
- [5] 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>
- [6] 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). doi: <http://dx.doi.org/10.2307/40285811>
- [7] J. Liebetrau, S. Schneider, and R. Jezierski, "Application of Free Choice Profiling for the Evaluation of Emotions Elicited by Music," *Proc. 9th Int. Symp. Comput. Music Modeling and Retrieval (CMMR 2012): Music and Emotions* (2012), pp. 78–93.
- [8] I. Lahdelma and T. Eerola, "Single Chords Convey Distinct Emotional Qualities to Both Naïve and Expert Listeners," *Psychology of Music*, vol. 44, no. 1 (2014).
- [9] J. Skowronek, M. F. McKinney, and S. Van De Par, "A Demonstrator for Automatic Music Mood Estimation," *Proc. Int. Soc. Music Inform. Retrieval Conf. (ISMIR)* (2007), pp. 345–346.
- [10] 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.
- [11] Y. Hu, X. Chen, and D. Yang, "Lyric-Based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method," *Proc. Int. Soc. Music Inform. Retrieval Conf. (ISMIR)* (2009), pp. 123–128.
- [12] 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.
- [13] G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293–302 (2002). doi: <http://dx.doi.org/10.1109/TSA.2002.800560>
- [14] 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). doi: <http://dx.doi.org/10.1109/TMM.2005.858380>
- [15] 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>
- [16] 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)
- [17] C. L. Krumhansl, "Plink: 'Thin Slices' of Music," *Musical Perception*, vol. 7, pp. 337–354 (2010). doi: <http://dx.doi.org/10.1525/mp.2010.27.5.337>
- [18] S. Filipic, B. Tillmann, and E. Bigand, "Judging Familiarity and Emotion from Very Brief Musical Excerpts," *Psychonomic Bulletin & Rev.*, vol. 17, no. 3, pp. 335–341 (2010). doi: <http://dx.doi.org/10.3758/PBR.17.3.335>
- [19] J. A. Sloboda and P. N. Juslin, "At the Interface between the Inner and Outer World," *Handbook of Music and Emotion*, ch. 4, pp. 73–97 (Oxford Scholarship Online

2010).

[20] K. Hevner, "Experimental Studies of the Elements of Expression in Music," *The Amer. J. Psychology*, vol. 48, no. 2, pp. 246–268 (1936 Apr.). doi: <http://dx.doi.org/10.2307/1415746>

[21] J. C. Hailstone, R. Omar, S. M. Henley, C. Frost, M. G. Kenward, and J. D. Warren, "It's Not What You Play, It's How You Play It: Timbre Affects Perception of Emotion in Music," *The Quart. J. Experimental Psychology*, vol. 62, no. 11, pp. 2141–2155 (2009). doi: <http://dx.doi.org/10.1080/17470210902765957>

[22] 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.

[23] L. Lu, D. Liu, and H.-J. Zhang, "Automatic Mood Detection and Tracking of Music Audio Signals," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 1, pp. 5–18 (2006). doi: <http://dx.doi.org/10.1109/TSA.2005.860344>

[24] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement," *Emotion*, vol. 8, no. 4, p. 494 (2008). doi: <http://dx.doi.org/10.1037/1528-3542.8.4.494>

[25] J. A. Russell, "A Circumplex Model of Affect," *J. Personality & Social Psychology*, vol. 39, no. 6, p. 1161 (1980). doi: <http://dx.doi.org/10.1037/h0077714>

[26] B. Kostek and M. Plewa, "Parametrisation and Correlation Analysis Applied to Music Mood Classification," *Int. J. Computational Intell. Studies*, vol. 2, no. 1, pp. 4–25 (2013). doi: <http://dx.doi.org/10.1504/IJCISTUDIES.2013.054734>

[27] 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). doi: <http://dx.doi.org/10.1080/02699930500204250>

[28] T. Eerola and J. K. Vuoskoski, "A Comparison of the Discrete and Dimensional Models of Emotion in Music," *Psychology of Music*, vol. 39, no. 1, pp. 18–49 (2011). doi: <http://dx.doi.org/10.1177/0305735610362821>

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

[30] 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). doi: <http://dx.doi.org/10.1525/mp.2012.30.1.49>

[31] B. Wu, S. Wun, C. Lee, and A. Horner, "Spectral Correlates in Emotion Labeling of Sustained Musical Instrument Tones," *Proc. 14th Int. Soc. Music Inform. Retrieval Conf. (ISMIR)* (November 4–8 2013).

[32] 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).

[33] P. N. Juslin and P. Laukka, "Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening," *J. New Music Research*, vol. 33, no. 3, pp. 217–238 (2004). doi: <http://dx.doi.org/10.1080/0929821042000317813>

[34] 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). doi: <http://dx.doi.org/10.1109/TASL.2007.911513>

[35] R. Panda and R. P. Paiva, "Using Support Vector Machines for Automatic Mood Tracking in Audio Music," presented at the *130th Convention of the Audio Engineering Society* (2011 May), convention paper 8378.

[36] J. Wapnick, "Pitch, Tempo, and Timbral Preferences in Recorded Piano Music," *J. Research in Music Educ.*, vol. 28, no. 1, pp. 43–58 (1980). doi: <http://dx.doi.org/10.2307/3345052>

[37] D. G. Nair, E. W. Large, F. Steinberg, and J. S. Kelso, "Perceiving Emotion in Expressive Piano Performance: A Functional MRI Study," *Proc. 7th Int. Conf. on Music Perception and Cognition: Adelaide: Causal Productions* (2002), pp. 627–630.

[38] R. Guevara and G. Wakefield, "A Modal Distribution Approach to Piano Analysis and Synthesis," in *Proc. 1996 Int. Comput. Music Conf. (ICMC)* (1996), pp. 350–351.

[39] J. Yuen and A. Horner, "Hybrid Sampling-Wavetable Synthesis with Genetic Algorithms," *J. Audio Eng. Soc.*, vol. 45, pp. 316–330 (1997 May).

[40] K. Lee and A. Horner, "Modeling Piano Tones with Group Synthesis," *J. Audio Eng. Soc.*, vol. 47, pp. 101–111 (1999 Mar.).

[41] B. Hamadicharef and E. Ifeakor, "Perceptual Modeling of Piano Tones," presented at the *119th Convention of the Audio Engineering Society* (2005 Oct.), convention paper 6525.

[42] C. E. Cella, "Harmonic Components Extraction in Recorded Piano Tones," presented at the *128th Convention of the Audio Engineering Society* (2010 May), convention paper 8111.

[43] M. Bernays and C. Traube, "Verbal Expression of Piano Timbre: Multidimensional Semantic Space of Adjectival Descriptors," *Proc. Int. Symposium on Performance Science (ISPS2011)*, European Assoc. of Conservatoires (AEC) Utrecht, Netherlands (2011), pp. 299–304.

[44] S. B. Kamenetsky, D. S. Hill, and S. E. Trehub, "Effect of Tempo and Dynamics on the Perception of Emotion in Music," *Psychology of Music*, vol. 25, pp. 149–160 (1997). doi: <http://dx.doi.org/10.1177/0305735697252005>

[45] C. L. Krumhansl, "An Exploratory Study of Musical Emotions and Psychophysiology," *Canadian J. Experimental Psychology/Revue canadienne de psychologie expérimentale*, vol. 51, no. 4, p. 336 (1997).

[46] D. Huron, D. Kinney, and K. Precoda, "Relation of Pitch Height to Perception of Dominance/Submissiveness

in Musical Passages,” *Music Perception*, vol. 10, no. 1, pp. 83–92 (2000). doi: <http://dx.doi.org/10.2307/40285540>

[47] K. Kosta, O. F. Bandtlow, and E. Chew, “Practical Implications of Dynamic Markings in the Score: Is Piano Always Piano?” presented at the *AES 53rd International Conference: Semantic Audio* (2014 Jan.), conference paper 1-3.

[48] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, “RWC Music Database: Music Genre Database and Musical Instrument Sound Database,” *Proc. Int. Soc. Music Inform. Retrieval Conf. (ISMIR)*, vol. 3, pp. 229–230 (2003).

[49] M. Kennedy and K. Joyce Bourne, *The Oxford Dictionary of Music* (Oxford University Press, 2012).

[50] Dolmetsch Organization, Dolmetsch Online—Music Dictionary [Online]. Available: <http://www.dolmetsch.com/musictheorydefs.htm>

[51] Classical.dj. Classical Musical Terms [Online]. Available: http://classical.dj/musical_terms.html

[52] Connect for Education Inc., Onmusic Dictionary. [Online]. Available: <http://dictionary.onmusic.org/>

[53] 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).

[54] Cambridge University Press, Cambridge Academic Content Dictionary [Online]. Available: <http://dictionary.cambridge.org/dictionary/american-english>

[55] 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)

[56] 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>

[57] 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). doi: <http://dx.doi.org/10.1121/1.1778741>

[58] J. Krimphoff, “*Analyse Acoustique et Perception du Timbre*,” unpublished DEA thesis, Université du Maine, Le Mans, France (1993).

[59] 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,” *Proc. 7th Int. Conf. Music Percept. Cognition* (2002), pp. 592–595.

[60] 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). doi: <http://dx.doi.org/10.1121/1.1929229>

[61] H. F. Pollard and E. V. Jansson, “A Tristimulus Method for the Specification of Musical Timbre,” *Acta Acustica united with Acustica*, vol. 51, no. 3, pp. 162–171 (1982).

APPENDIX

Table 8. Results of a Shapiro–Wilk test to check the normality of the BTL data (for populations of pitches). Entries in bold were not normally distributed.

		Statistic	df	Sig.
BTL Values for Happy	C1	0.790	3	0.090
	C2	0.970	3	0.665
	C3	0.777	3	0.060
	C4	0.923	3	0.462
	C5	0.954	3	0.588
	C6	0.943	3	0.539
	C7	0.835	3	0.200
	C8	0.979	3	0.721
BTL Values for Heroic	C1	0.992	3	0.825
	C2	0.949	3	0.565
	C3	0.963	3	0.631
	C4	0.904	3	0.399
	C5	0.969	3	0.659
	C6	0.954	3	0.586
	C7	0.860	3	0.268
	C8	0.987	3	0.782
BTL Values for Romantic	C1	0.971	3	0.670
	C2	0.915	3	0.434
	C3	0.969	3	0.661
	C4	0.963	3	0.629
	C5	0.979	3	0.722
	C6	0.983	3	0.754
	C7	0.979	3	0.724
	C8	0.920	3	0.452
BTL Values for Comic	C1	0.750	3	0.000
	C2	0.865	3	0.281
	C3	0.865	3	0.280
	C4	0.952	3	0.579
	C5	0.999	3	0.944
	C6	0.967	3	0.651
	C7	0.986	3	0.774
	C8	0.976	3	0.701
BTL Values for Calm	C1	0.890	3	0.353
	C2	0.994	3	0.848
	C3	0.994	3	0.858
	C4	0.998	3	0.911
	C5	0.971	3	0.672
	C6	0.989	3	0.801
	C7	0.985	3	0.767
	C8	0.872	3	0.300

Table 8. Continued.

		Statistic	df	Sig.
BTL Values for Mysterious	C1	1.000	3	0.967
	C2	0.994	3	0.850
	C3	0.985	3	0.767
	C4	1.000	3	0.996
	C5	0.836	3	0.203
	C6	0.998	3	0.923
	C7	0.974	3	0.688
	C8	0.952	3	0.580
BTL Values for Shy	C1	0.853	3	0.249
	C2	0.965	3	0.641
	C3	0.964	3	0.635
	C4	1.000	3	0.989
	C5	0.997	3	0.903
	C6	0.985	3	0.765
	C7	0.999	3	0.926
	C8	0.806	3	0.130
BTL Values for Angry	C1	0.969	3	0.662
	C2	0.892	3	0.361
	C3	0.826	3	0.177
	C4	0.924	3	0.466
	C5	0.992	3	0.829
	C6	0.912	3	0.423
	C7	1.000	3	0.973
	C8	0.934	3	0.505
BTL Values for Scary	C1	0.999	3	0.941
	C2	0.944	3	0.542
	C3	0.869	3	0.293
	C4	0.965	3	0.639
	C5	0.815	3	0.152
	C6	0.913	3	0.429
	C7	0.963	3	0.631
	C8	0.986	3	0.771
BTL Values for Sad	C1	0.998	3	0.925
	C2	0.982	3	0.745
	C3	0.878	3	0.319
	C4	0.999	3	0.935
	C5	0.952	3	0.579
	C6	0.881	3	0.328
	C7	0.964	3	0.638
	C8	0.817	3	0.155

Table 9. Results of a Shapiro–Wilk test to check the normality of the BTL data (for populations of dynamics). Entries in bold were not normally distributed.

		Statistic	df	Sig.
BTL Values for Happy	Forte	0.904	8	0.315
	Mezzo	0.888	8	0.224
	Piano	0.951	8	0.721
BTL Values for Heroic	Forte	0.942	8	0.63
	Mezzo	0.952	8	0.731
	Piano	0.977	8	0.949
BTL Values for Romantic	Forte	0.888	8	0.223
	Mezzo	0.950	8	0.713
	Piano	0.952	8	0.730
BTL Values for Comic	Forte	0.978	8	0.952
	Mezzo	0.965	8	0.857
	Piano	0.847	8	0.088
BTL Values for Calm	Forte	0.900	8	0.288
	Mezzo	0.960	8	0.812
	Piano	0.895	8	0.261
BTL Values for Mysterious	Forte	0.838	8	0.072
	Mezzo	0.845	8	0.084
	Piano	0.854	8	0.105
BTL Values for Shy	Forte	0.900	8	0.291
	Mezzo	0.915	8	0.394
	Piano	0.895	8	0.260
BTL Values for Angry	Forte	0.858	8	0.115
	Mezzo	0.765	8	0.012
	Piano	0.844	8	0.083
BTL Values for Scary	Forte	0.949	8	0.697
	Mezzo	0.929	8	0.511
	Piano	0.854	8	0.106
BTL Values for Sad	Forte	0.865	8	0.134
	Mezzo	0.919	8	0.418
	Piano	0.876	8	0.172

Table 10. Features of the piano sounds of the lower register (C1–C4).

Features	Sounds											
	c1forte	c1mezzo	c1piano	c2forte	c2mezzo	c2piano	c3forte	c3mezzo	c3piano	c4forte	c4mezzo	c4piano
Log of Fundamental Frequency	5.03	5.03	5.03	6.03	6.03	6.03	7.03	7.03	7.03	8.03	8.03	8.03
Peak RMS Amplitude (dB)	-1.98	-6.32	-13.41	-3.07	-9.25	-14.60	0.00	-6.72	-11.69	-5.91	-13.43	-19.64
Attack time (ms)	79.27	63.42	79.27	93.41	108.98	93.41	81.01	46.29	115.73	21.12	15.36	-17.28
Attack slope (<i>linear amplitude/ms</i>)	1.39	0.68	0.21	1.05	0.30	0.18	4.50	1.44	0.62	6.41	4.81	2.40
Decay ratio	0.20	0.20	0.19	0.25	0.26	0.26	0.30	0.29	0.31	0.21	0.21	0.21
Decay slope (<i>linear amplitude/ms</i>)	-0.25	-0.15	-0.07	-0.27	-0.13	-0.07	-0.51	-0.24	-0.13	-0.56	-0.24	-0.12
Density of Significant Harmonics	0.60	0.61	0.63	0.62	0.61	0.60	0.59	0.71	0.71	0.83	0.92	0.95
Spectral Centroid	20.56	19.67	19.62	10.88	9.79	9.92	6.21	5.21	4.81	3.47	3.03	3.07
Spectral Centroid Deviation	9.81	8.85	8.73	4.03	3.24	3.16	2.97	2.39	1.96	2.10	1.77	1.76
Spectral Irregularity	0.09	0.09	0.11	0.08	0.10	0.10	0.07	0.08	0.09	0.15	0.16	0.16
Even/Odd Harmonic Ratio	0.69	0.66	0.60	1.10	1.15	1.15	0.49	0.47	0.42	1.14	1.08	1.02
Tristimulus T1 (<i>harmonic 1</i>)	0.01	0.01	0.01	0.11	0.12	0.13	0.35	0.38	0.43	0.30	0.34	0.39
Tristimulus T2 (<i>harmonics 2–4</i>)	0.31	0.32	0.37	0.35	0.39	0.42	0.31	0.34	0.36	0.43	0.47	0.47
Tristimulus T3 (<i>harmonics 5+</i>)	0.69	0.67	0.62	0.54	0.50	0.45	0.34	0.28	0.21	0.28	0.20	0.14

Table 11. Features of the piano sounds of the higher register (C5–C8).

Features	Sounds											
	c5forte	c5mezzo	c5piano	c6forte	c6mezzo	c6piano	c7forte	c7mezzo	c7piano	c8forte	c8mezzo	c8piano
Log of Fundamental Frequency	9.03	9.03	9.03	10.03	10.03	10.03	11.03	11.03	11.03	12.03	12.03	12.03
Peak RMS Amplitude (dB)	-3.54	-9.53	-16.40	-5.84	-12.76	-20.92	-10.15	-16.40	-22.13	-15.22	-18.94	-39.30
Attack time (ms)	21.06	17.23	16.27	18.18	16.26	14.35	5.26	2.87	2.15	3.11	4.42	1.20
Attack slope (<i>linear amplitude/ms</i>)	69.72	35.04	8.69	67.81	29.28	12.24	344.87	150.33	65.14	768.90	483.22	40.45
Decay ratio	0.04	0.04	0.04	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00
Decay slope (<i>linear amplitude/ms</i>)	-1.58	-0.72	-0.31	-2.31	-1.05	-0.40	-3.44	-1.62	-0.83	-3.93	-2.46	-0.25
Density of Significant Harmonics	0.97	1.00	1.00	1.00	1.00	0.97	0.98	0.98	0.97	0.98	0.98	0.91
Spectral Centroid	2.27	2.07	1.96	1.50	1.43	1.53	1.18	1.17	1.21	1.10	1.10	1.52
Spectral Centroid Deviation	1.82	1.62	1.50	1.35	1.29	1.37	1.12	1.10	1.13	1.07	1.08	1.48
Spectral Irregularity	0.14	0.17	0.20	0.16	0.20	0.28	0.26	0.29	0.31	0.30	0.31	0.31
Even/Odd Harmonic Ratio	0.31	0.24	0.15	0.25	0.17	0.05	0.07	0.04	0.02	0.06	0.04	0.03
Tristimulus T1 (<i>harmonic 1</i>)	0.58	0.66	0.77	0.75	0.82	0.92	0.89	0.94	0.96	0.95	0.96	0.91
Tristimulus T2 (<i>harmonics 2–4</i>)	0.30	0.26	0.19	0.24	0.16	0.06	0.10	0.05	0.03	0.05	0.04	0.05
Tristimulus T3 (<i>harmonics 5+</i>)	0.12	0.09	0.05	0.02	0.01	0.02	0.01	0.01	0.02	0.00	0.00	0.04

THE AUTHORS



Chuck-jee Chau



Ronald Mo



Andrew Horner

Chuck-jee Chau is a Ph.D. student in the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology (HKUST). His research focuses on timbre analysis and music emotion. During his master studies he developed the timbre visualization tool *pvan+* for phase vocoder analysis. He obtained his B.Eng. in computer engineering from the Chinese University of Hong Kong (CUHK) with a minor in music. Besides computer music research, he is also a versatile collaborative pianist and mallet percussionist active in chamber music performances.

Ronald Mo is pursuing his Ph.D. in the Department of Computer Science and Engineering at the Hong Kong

University of Science and Technology. His research interests include timbre of musical instruments, music emotion recognition, and digital 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.

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.