

EMOTIONAL PREDISPOSITION OF MUSICAL INSTRUMENT TIMBRES WITH STATIC SPECTRA

Bin Wu
Department of Computer
Science and Engineering
Hong Kong University
of Science and Technology
Hong Kong
bwuaa@cse.ust.hk

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

Chung Lee
The Information Systems
Technology and Design Pillar
Singapore University
of Technology and Design
20 Dover Drive
Singapore 138682
im.lee.chung@gmail.com

ABSTRACT

Music is one of the strongest triggers of emotions. Recent studies have shown strong emotional predispositions for musical instrument timbres. They have also shown significant correlations between spectral centroid and many emotions. Our recent study on spectral centroid-equalized tones further suggested that the even/odd harmonic ratio is a salient timbral feature after attack time and brightness. The emergence of the even/odd harmonic ratio motivated us to go a step further: to see whether the spectral shape of musical instruments alone can have a strong emotional predisposition. To address this issue, we conducted follow-up listening tests of static tones. The results showed that the even/odd harmonic ratio again significantly correlated with most emotions, consistent with the theory that static spectral shapes have a strong emotional predisposition.

1. INTRODUCTION

Music is one of the most effective media for conveying emotion. A lot of work has been done on emotion recognition in music, especially addressing melody [4], harmony [18], rhythm [23, 25], lyrics [15], and localization cues [11].

Some recent studies have shown that emotion is also closely related to timbre. Scherer and Oshinsky found that timbre is a salient factor in the rating of synthetic tones [24]. Peretz *et al.* showed that timbre speeds up discrimination of emotion categories [22]. Bigand *et al.* reported similar results in their study of emotion similarities between one-second musical excerpts [7]. It was also found that timbre is essential to musical genre recognition and discrimination [3, 5, 27].

Even more relevant to the current study, Eerola carried

out listening tests to investigate the correlation of emotion with temporal and spectral sound features [10]. The study confirmed strong correlations between features such as attack time and brightness and the emotion dimensions valence and arousal for one-second isolated instrument tones. Valence and arousal are measures of how pleasant and energetic the music sounds [31]. Asutay *et al.* also studied valence and arousal responses to 18 environmental sounds [2]. Despite the widespread use of valence and arousal in music research, composers may find them rather vague and difficult to interpret for composition and arrangement, and limited in emotional nuance. Using a different approach than Eerola, Ellermeier *et al.* investigated the unpleasantness of environmental sounds using paired comparisons [12].

Recently, we investigated the correlations between emotion and timbral features [30]. In our previous study, listening test subjects compared tones in terms of emotion categories such as Happy and Sad. We equalized the stimuli attacks and decays so that temporal features would not be factors. This modification isolated the effects of spectral features such as spectral centroid. Average spectral centroid significantly correlated for all emotions, and spectral centroid deviation significantly correlated for all emotions. This correlation was even stronger than average spectral centroid for most emotions. The only other correlation was spectral incoherence for a few emotions.

However, since average spectral centroid and spectral centroid deviation were so strong, listeners did not notice other spectral features much. This raised the question: if average spectral centroid was equalized in the tones, would spectral incoherence be more significant? Would other spectral characteristics emerge as significant? We tested this idea on spectral centroid normalized tones, and found that even/odd harmonic ratio was significant. This made us even more curious: if musical instruments tones only differed from one another in their spectral shapes, would they still have strong emotional predispositions? To answer this question, we conducted the follow-up experiment described in this paper using emotion responses for static spectra tones.



© Bin Wu, Andrew Horner, Chung Lee.

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** Bin Wu, Andrew Horner, Chung Lee. "Emotional Predisposition of Musical Instrument Timbres with Static Spectra", 15th International Society for Music Information Retrieval Conference, 2014.

2. LISTENING TEST

In our listening test, listeners compared pairs of eight instruments for eight emotions, using tones that were equalized for attack, decay, and spectral centroid.

2.1 Stimuli

2.1.1 Prototype instrument sounds

The stimuli consisted of eight sustained wind and bowed string instrument tones: bassoon (Bs), clarinet (Cl), flute (Fl), horn (Hn), oboe (Ob), saxophone (Sx), trumpet (Tp), and violin (Vn). They were obtained from the McGill and Prosonus sample libraries, except for the trumpet, which had been recorded at the University of Illinois at Urbana-Champaign School of Music. The original of all these tones were used in a discrimination test carried out by Horner *et al.* [14], six of them were also used by McAdams *et al.* [20], and all of them used in our emotion-timbre test [30].

The tones were presented in their entirety. The tones were nearly harmonic and had fundamental frequencies close to 311.1 Hz (Eb4). The original fundamental frequencies deviated by up to 1 Hz (6 cents), and were synthesized by additive synthesis at 311.1 Hz.

Since loudness is potential factor in emotion, amplitude multipliers were determined by the Moore-Glasberg loudness program [21] to equalize loudness. Starting from a value of 1.0, an iterative procedure adjusted an amplitude multiplier until a standard loudness of 87.3 ± 0.1 phons was achieved.

2.2 Stimuli Analysis and Synthesis

2.2.1 Spectral Analysis Method

Instrument tones were analyzed using a phase-vocoder algorithm, which is different from most in that bin frequencies are aligned with the signal's harmonics (to obtain accurate harmonic amplitudes and optimize time resolution) [6]. The analysis method yields frequency deviations between harmonics of the analysis frequency and the corresponding frequencies of the input signal. The deviations are approximately harmonic relative to the fundamental and within $\pm 2\%$ of the corresponding harmonics of the analysis frequency. More details on the analysis process are given by Beauchamp [6].

2.2.2 Spectral Centroid Equalization

Different from our previous study [30], the average spectral centroid of the stimuli was equalized for all eight instruments. The spectra of each instrument was modified to an average spectral centroid of 3.7, which was the mean average spectral centroid of the eight tones. This modification was accomplished by scaling each harmonic amplitude by its harmonic number raised to a to-be-determined power:

$$A_k(t) \leftarrow k^p A_k(t) \quad (1)$$

For each tone, starting with $p = 0$, p was iterated using Newton's method until an average spectral centroid was obtained within ± 0.1 of the 3.7 target value.

2.2.3 Static Tone Preparation

The static tones were 0.5s in duration and were generated using the average steady-state spectrum of each spectral centroid equalized tone with linear 0.05s attacks and decays, and 0.4 sustains.

2.2.4 Resynthesis Method

Stimuli were resynthesized from the time-varying harmonic data using the well-known method of time-varying additive sinewave synthesis (oscillator method) [6] with frequency deviations set to zero.

2.3 Subjects

32 subjects without hearing problems were hired to take the listening test. They were undergraduate students and ranged in age from 19 to 24. Half of them had music training (that is, at least five years of practice on an instrument).

2.4 Emotion Categories

As in our previous study [30], the subjects compared the stimuli in terms of eight emotion categories: Happy, Sad, Heroic, Scary, Comic, Shy, Joyful, and Depressed.

2.5 Listening Test Design

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 which tone more strongly aroused a given emotion. Each combination of two different instruments was presented in four trials for each emotion, and the listening test totaled $C_2^8 \times 4 \times 8 = 896$ 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, ...).

Before the first trial, the subjects read online definitions of the emotion categories from the Cambridge Academic Content Dictionary [1]. The listening test took about 1.5 hours, with breaks every 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 further reduced with headphones. Sound signals were converted to analog by a Sound Blaster X-Fi Xtreme Audio sound card, and then presented through Sony MDR-7506 headphones at a level of approximately 78 dB SPL, as measured with a sound-level meter. The Sound Blaster DAC utilized 24 bits with a maximum sampling rate of 96 kHz and a 108 dB S/N ratio.

3. RESULTS

3.1 Quality of Responses

The subjects' responses were first screened for inconsistencies, and two outliers were filtered out. Consistency was defined based on the four comparisons of a pair of instruments A and B for a particular emotion the same with our previous work [30]:

$$\text{consistency}_{A,B} = \frac{\max(v_A, v_B)}{4} \quad (2)$$

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 approximately random guessing. The mean average consistency of all subjects was 0.74. Also, as in our previous work [30], we found that the two least consistent subjects had the highest outlier coefficients using White *et al.*'s method [28]. Therefore, they were excluded from the results.

We measured the level of agreement among the remaining 30 subjects with an overall Fleiss' Kappa statistic [16]. Fleiss' Kappa was 0.026, indicating a slight but statistically significant agreement among subjects. From this, we observed that subjects were self-consistent but less agreed in their responses than our previous study [30], since spectral shape was the only factor that could possibly affect emotion.

We also performed a χ^2 test [29] to evaluate whether the number of circular triads significantly deviated from the number to be expected by chance alone. This turned out to be insignificant for all subjects. The approximate likelihood ratio test [29] for significance of weak stochastic transitivity violations [26] was tested and showed no significance for all emotions.

3.1.1 Emotion Results

Same with our previous work, we ranked the spectral centroid equalized instrument tones by the number of positive votes they received for each emotion, and derived scale values using the Bradley-Terry-Luce (BTL) model [8, 29] as shown in Figure 1. The likelihood-ratio test showed that the BTL model describes the paired-comparisons well for all emotions. We observe that: 1) The distribution of emotion ratings were much narrower than the original tones in our previous study [30]. The reason is that spectral shape was the only factor that could possibly affect emotion, which made it more difficult for subjects to distinguish. 2) Opposite of our previous study [30], the horn evoked positive emotions. It was ranked as the least Shy and Depressed, and among the most Heroic and Comic. 3) The clarinet and the saxophone were contrasting outliers for all emotions (except Scary).

Figure 2 shows BTL scale values and the corresponding 95% confidence intervals of the instruments for each emotion. The confidence intervals cluster near the line of indifference since it was difficult for listeners to make emotional distinctions. Table 1 shows the spectral characteristics of the static tones (time-domain spectral char-

acteristics are omitted since the tones are static). With all time-domain spectral characteristics removed, spectral shape features such as even/odd harmonic ratio became more salient. Specifically, even/odd ratio was calculated according to Caclin *et al.*'s method [9]. Pearson correlation between emotion and spectral characteristics are shown in Table 2. Both spectral irregularity and even/odd harmonic ratio are measures of spectral jaggedness, where even/odd harmonic ratio measures a particular, extreme type of spectral irregularity that is typical of the clarinet. In Table 2, even/odd harmonic ratio significantly correlated with nearly all emotions. The correlations were much stronger than in the original tones [30], and indicate that spectral shape by itself can arouse strong emotional responses.

4. DISCUSSION

These results are consistent with our previous results [30] and Eerola's Valence-Arousal results [10]. All these studies indicate that musical instrument timbres carry cues about emotional expression that are easily and consistently recognized by listeners. They show that spectral centroid/brightness is a significant component in music emotion. Beyond Eerola's and our previous findings, we have found that spectral shape by itself can have strong emotional predispositions, and even/odd harmonic ratio is the most salient timbral feature after attack time and brightness in static tones.

In hindsight, perhaps it is not so surprising that static spectra tones have emotional predispositions just as dynamic musical instrument tones do. It is somewhat analogous to viewers' emotional dispositions to primary colors [13, 17, 19].

Of course, just because static tones have emotional predispositions, it does not mean they are interesting to listen to. The dynamic spectra of real acoustic instruments are much more natural and life-like than any static tones, regardless of emotional predisposition. This is reflected in the wider range of emotion rankings of the original dynamic tones compared to the static tones.

For future work, it will be fascinating to see how emotion varies with pitch, dynamic level, brightness, articulation, and cultural backgrounds.

5. ACKNOWLEDGMENT

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

6. REFERENCES

- [1] happy, sad, heroic, scary, comic, shy, joyful and depressed. *Cambridge Academic Content Dictionary*, 2013. Online: <http://goo.gl/v5xJZ> (17 Feb 2013).
- [2] Erkin Asutay, Daniel Västfjäll, Ana Tajadura-Jiménez, Anders Genell, Penny Bergman, and Mendel Kleiner. *Emoacoustics: A Study of the Psychoacoustical and Psychological Dimensions of Emotional Sound*

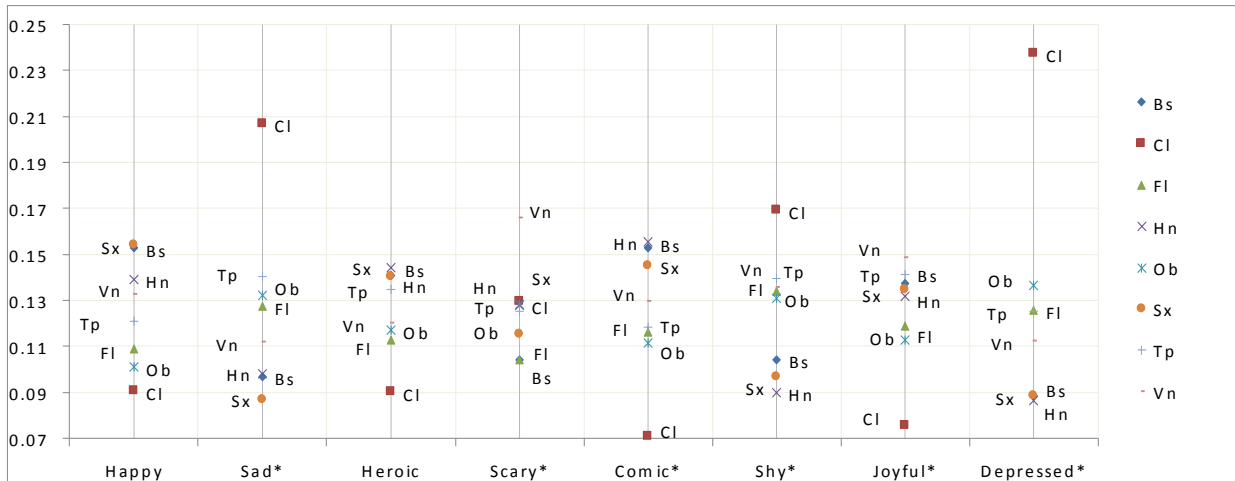


Figure 1. Bradley-Terry-Luce scale values of the static tones for each emotion.

- Design. *Journal of the Audio Engineering Society*, 60(1/2):21–28, 2012.
- [3] Jean-Julien Aucouturier, François Pachet, and Mark Sandler. The Way it Sounds: Timbre Models for Analysis and Retrieval of Music Signals. *IEEE Transactions on Multimedia*, 7(6):1028–1035, 2005.
- [4] Laura-Lee Balkwill and William Forde Thompson. A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues. *Music Perception*, 17(1):43–64, 1999.
- [5] Chris Baume. Evaluation of Acoustic Features for Music Emotion Recognition. In *Audio Engineering Society Convention 134*. Audio Engineering Society, 2013.
- [6] James W Beauchamp. Analysis and Synthesis of Musical Instrument Sounds. In *Analysis, Synthesis, and Perception of Musical Sounds*, pages 1–89. Springer, 2007.
- [7] 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 and Emotion*, 19(8):1113–1139, 2005.
- [8] Ralph A Bradley. Paired Comparisons: Some Basic Procedures and Examples. *Nonparametric Methods*, 4:299–326, 1984.
- [9] Anne Caclin, Stephen McAdams, Bennett K Smith, and Suzanne Winsberg. Acoustic Correlates of Timbre Space Dimensions: A Confirmatory Study Using Synthetic Tones. *Journal of the Acoustical Society of America*, 118:471, 2005.
- [10] Tuomas Eerola, Rafael Ferrer, and Vinoo Alluri. Timbre and Affect Dimensions: Evidence from Affect and Similarity Ratings and Acoustic Correlates of Isolated Instrument Sounds. *Music Perception*, 30(1):49–70, 2012.
- [11] Inger Ekman and Raine Kajastila. Localization Cues Affect Emotional Judgments—Results from a User Study on Scary Sound. In *Audio Engineering Society Conference: 35th International Conference: Audio for Games*. Audio Engineering Society, 2009.
- [12] Wolfgang Ellermeier, Markus Mader, and Peter Daniel. Scaling the Unpleasantness of Sounds According to the BTL Model: Ratio-scale Representation and Psychoacoustical Analysis. *Acta Acustica United with Acustica*, 90(1):101–107, 2004.
- [13] Michael Hemphill. A note on adults’ color–emotion associations. *The Journal of genetic psychology*, 157(3):275–280, 1996.
- [14] Andrew Horner, James Beauchamp, and Richard So. Detection of Random Alterations to Time-varying Musical Instrument Spectra. *Journal of the Acoustical Society of America*, 116:1800–1810, 2004.
- [15] Yajie Hu, Xiaou Chen, and Deshun Yang. Lyric-Based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method. *Proceedings of ISMIR*, 2009.
- [16] Fleiss L Joseph. Measuring Nominal Scale Agreement among Many Raters. *Psychological Bulletin*, 76(5):378–382, 1971.
- [17] Naz Kaya and Helen H Epps. Relationship between color and emotion: A study of college students. *College student journal*, 38(3), 2004.
- [18] Judith Liebetrau, Sebastian Schneider, and Roman Jezierski. Application of Free Choice Profiling for the Evaluation of Emotions Elicited by Music. In *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012): Music and Emotions*, pages 78–93, 2012.
- [19] Banu Manav. Color-emotion associations and color preferences: A case study for residences. *Color Research & Application*, 32(2):144–150, 2007.

- [20] Stephen McAdams, James W Beauchamp, and Suzanna Meneguzzi. Discrimination of Musical Instrument Sounds Resynthesized with Simplified Spectrotemporal Parameters. *Journal of the Acoustical Society of America*, 105:882, 1999.
- [21] Brian CJ Moore, Brian R Glasberg, and Thomas Baer. A Model for the Prediction of Thresholds, Loudness, and Partial Loudness. *Journal of the Audio Engineering Society*, 45(4):224–240, 1997.
- [22] Isabelle Peretz, Lise Gagnon, and Bernard Bouchard. Music and Emotion: Perceptual Determinants, Immediacy, and Isolation after Brain Damage. *Cognition*, 68(2):111–141, 1998.
- [23] Magdalena Plewa and Bozena Kostek. A Study on Correlation between Tempo and Mood of Music. In *Audio Engineering Society Convention 133*, Oct 2012.
- [24] Klaus R Scherer and James S Oshinsky. Cue Utilization in Emotion Attribution from Auditory Stimuli. *Motivation and Emotion*, 1(4):331–346, 1977.
- [25] Janto Skowronek, Martin McKinney, and Steven Van De Par. A Demonstrator for Automatic Music Mood Estimation. *Proceedings of the International Conference on Music Information Retrieval*, 2007.
- [26] Amos Tversky. Intransitivity of Preferences. *Psychological Review*, 76(1):31, 1969.
- [27] George Tzanetakis and Perry Cook. Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.
- [28] Jacob Whitehill, Paul Ruvolo, Tingfan Wu, Jacob Bergsma, and Javier Movellan. Whose Vote Should Count More: Optimal Integration of Labels from Labelers of Unknown Expertise. *Advances in Neural Information Processing Systems*, 22(2035-2043):7–13, 2009.
- [29] Florian Wickelmaier and Christian Schmid. A Matlab Function to Estimate Choice Model Parameters from Paired-comparison Data. *Behavior Research Methods, Instruments, and Computers*, 36(1):29–40, 2004.
- [30] Bin Wu, Simon Wun, Chung Lee, and Andrew Horner. Spectral Correlates in Emotion Labeling of Sustained Musical Instrument Tones. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*, November 4-8 2013.
- [31] Yi-Hsuan Yang, Yu-Ching Lin, Ya-Fan Su, and Homer H. Chen. A Regression Approach to Music Emotion Recognition. *IEEE TASLP*, 16(2):448–457, 2008.

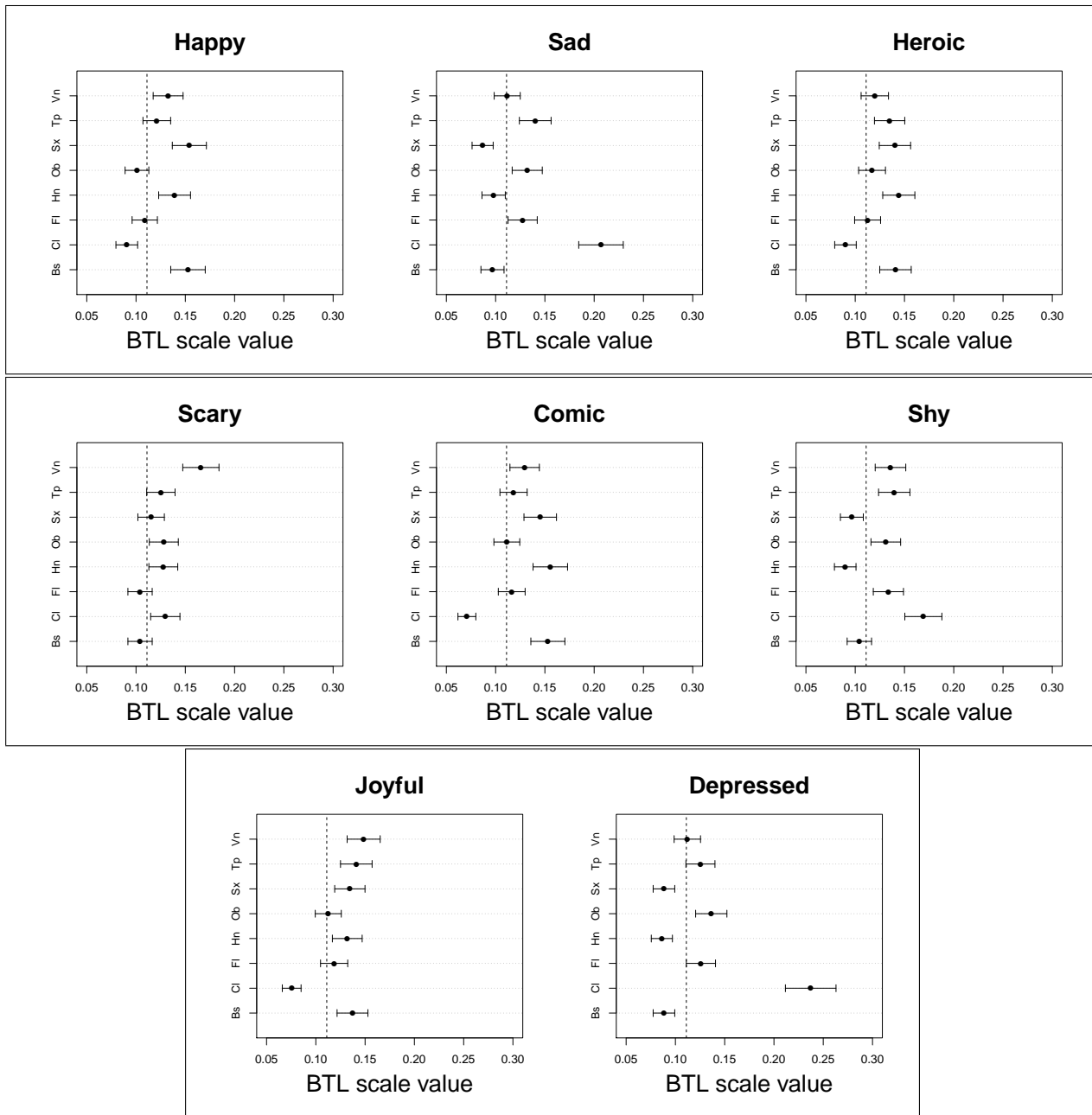


Figure 2. BTL scale values and the corresponding 95% confidence intervals of the static tones for each emotion. The dotted line represents no preference.

Features \ Instruments	Instruments							
	Bs	Cl	Fl	Hn	Ob	Sx	Tp	Vn
Spectral Irregularity	0.0971	0.1818	0.143	0.0645	0.119	0.1959	0.0188	0.1176
Even/odd Ratio	1.2565	0.1775	0.9493	0.9694	0.4308	1.7719	0.7496	0.8771

Table 1. Spectral characteristics of the static instrument tones.

Features \ Emotion	Emotion							
	Happy	Sad	Heroic	Scary	Comic	Shy	Joyful	Depressed
Spectral Irregularity	-0.1467	0.1827	-0.4859	-0.0897	-0.3216	0.1565	-0.509	0.3536
Even/odd Ratio	0.8901**	-0.8441**	0.7468**	-0.3398	0.8017**	-0.7942**	0.6524*	-0.7948**

Table 2. Pearson correlation between emotion and spectral characteristics for static tones. **: $p < 0.05$; *: $0.05 < p < 0.1$.