Timbre's function in the perception of affective intentions:
Contextual information and effects of learning

Lena Heng[1†]

Stephen McAdams[2]

[1, 2] Schulich School of Music, McGill University, Montreal, Quebec, Canada

[†] Corresponding author: lena.heng@mail.mcgill.ca

Abstract

Timbre has been identified by music perception scholars as a component in the communication of affect in music. While its function as a carrier of perceptually useful information about sound source mechanics has been established, less is understood about whether and how it functions as a carrier of information for communicating affect in music. To investigate these issues, listeners trained in Chinese and Western musical traditions were presented with phrases, measures, and individual notes of recorded excerpts interpreted with a variety of affective intentions by performers on instruments from the two cultures. These excerpts were analyzed to determine acoustic features that are correlated with timbre characteristics. Analysis revealed consistent use of temporal, spectral, and spectrotemporal attributes in judging affective intent in music, suggesting purposeful use of these properties within the sounds by listeners. Comparison between listeners' perceptions across notes and longer segments also revealed greater accuracy in perception with increased musical context. How timbre is used for musical communication appears to be implicated differently across musical traditions. The important role timbre plays also appears to vary for different positions within a musical phrase, suggesting that patterns of change over time are crucial in emotional communication.

KEYWORDS: *timbre, musical affect, acoustic features, cross-cultural differences*

Introduction

Timbre has been shown to carry perceptually useful information about sound source mechanics, but while it has been identified by music perception scholars as a component in the communication of affect in music, there is still a lot to uncover about how it functions as a carrier of information for such communication. A recent definition by McAdams (2019) considers timbre as a "complex auditory attribute, or as a set of attributes, of a perceptually fused sound event in addition to those of pitch, loudness, perceived duration, and spatial position...[and] is also a perceptual property, not a physical one" (p.23). Even though timbre is a psychophysical attribute, and it is the perception of physical properties that defines timbre, it is undeniable that one cannot underestimate the importance of the physical acoustic properties that give rise to timbre perception. As surface acoustic properties carry important information for perception, a systematic approach to sound analysis that is "oriented towards human perception" (Peeters et al., 2011, p. 2902) and their relation to the communication of affective intentions in music will greatly aid the understanding of timbre's function in musical communication.

As Juslin and Timmers observed, "both the mean level of a cue and its variability throughout the performance may be important for the communicative process" (2010, p. 462). Therefore, it is likely that the amount of information timbre carries across different parts of a phrase varies according to musical context. Cultural noise might also be present because of "disparities which may exist between the habit responses required by the musical style and those which a given individual actually possesses" (Meyer, 1994, p. 16). As a result, how timbre is used for musical communication may also be different across musical traditions.

Previous research

Studies have revealed some differences between listeners from different cultures—Chinese and Western—in the multidimensional space obtained from rating dissimilarities of instrument sounds (e.g., Zhang and Xie, 2017). Although these differences might have been due to different sets of instruments used (Chinese vs. Western instruments), the different dimensions obtained from the multidimensional scaling could also imply a focus on different aspects of a sound by different groups of listeners (McAdams et al., 1995).

With regards to attributing affective intentions to musical sounds, Scherer and Oshinsky (1977) systematically manipulated certain acoustic parameters to study listeners' ratings on both discrete emotions and dimensional affective intentions. Other researchers since then have also found consistent mappings of acoustic cues to affective responses (e.g., Schimmack & Grob, 2000; Eerola et al., 2012; Bowman & Yamauchi, 2016; McAdams et al., 2017). There are also many overlaps in the acoustic dimensions that seem to be involved in carrying these affective intentions, suggesting that affective content is not just related to a single acoustic dimension, but also communicated through the complex combinations and interactions of several acoustic parameters. Thompson and Balkwill (2010) also proposed in their cue redundancy model that listeners who are familiar with a musical style should be able to easily decode meanings in music of that style because "they can draw from both culture-specific and psychophysical cues" (p. 766).

Research questions

This study aims to look into how different aspects of a sound may be implicated in different affective intentions and how musical context provides varying amounts of information in musical communication. In addition, it also attempts to look at whether differences in musical experience and training influence both the ways acoustic cues are used by listeners and the accuracy of their responses.

Method

To investigate these issues, three groups of listeners with different musical backgrounds (Chinese musicians (CHM) and Western musicians (WM) and nonmusicians (NM), n = 30 per group) from Singapore were recruited for listening experiments. The criterion for musicians during participant recruitment was to have more than five years of formal musical training in either the Chinese (M = 12.00, SD = 2.98) or Western (M = 12.13, SD = 7.40) music tradition, and the criterion for nonmusicians was less than a year of formal training in any type of music (M = 0.2, SD = 0.41). There was no significant difference between the number of years of musical training between the CHM and WM listeners, $F(1, 58)$ = 1.41, p = .24. None of the WM listeners had any prior training in Chinese music while some CHM listeners had received formal instruction in Western music. However, all CHM listeners self-identified as being more proficient in Chinese music than Western

music. All participants had casual exposure to both Chinese and Western art music, both being ubiquitous musical forms found in Singapore.

One professional musician for each instrument (dizi, flute, erhu, violin, pipa, and guitar) was recruited for the recording. The two-dimensional model of valence and arousal (Russell, 1980) was explained to the performers, and they were asked to interpret the excerpt of music in performance with five different affective intents: low valence and arousal, low valence and high arousal, high valence and arousal, high valence and low arousal, and neutral.

All of the listeners took part in two experimental sessions conducted at least a week apart. As the stimuli used for both experiments were obtained from the same recordings, this delay between the first and second experiments was to reduce any memory effects. Experiment A involved participants listening to individual notes extracted from the recorded excerpts, which were interpreted with a variety of affective intents by performers on Western and Chinese instruments, and then making judgments about each stimulus' perceived affective intent within a two-dimensional affective space of valence and arousal (Russell, 1980). Experiment B involved participants listening to measures and phrases of these same recorded excerpts and again making global judgments of the affective intent. Half of the participants were randomly assigned to experiment A first while the other half were assigned to experiment B first.

Using the Timbre Toolbox implemented in the MATLAB environment, individual notes were analyzed for their temporal, spectral, and spectrotemporal descriptors. Based on hierarchical clustering analyses done by Peeters and colleagues (2011), 13 acoustic descriptors that represent each cluster were selected. These acoustic descriptors included median and interquartile range of spectral centroid, spectral flatness, and RMS envelope, as well as the median for noisiness, harmonic spectral deviation, spectrotemporal variation, temporal centroid, frequency and amplitude modulations, and log attack time.

Results

To look into how much agreement there was between the listeners' perception of the affective intentions with what was intended by the performers, the average ratings of perceived affective intentions for each group of listeners were plotted with respect to their valence and arousal responses as shown in Figure 1. Due to

space limitations, graphic representations of Figure 1 can be found at the following address: http://132.206.14.109/supplementaryMaterials/HengFDoMC2021/Fig1.pdf. Although each instrument has a slightly different pattern, arousal appears generally accurate with the data points more or less staying on the correct side of the space. Valence however is somewhat more ambiguous; positive-valence/low-arousal is often confused as negative-valence/low-arousal. With increasing musical context, valence appears to become more differentiated and becomes more accurate for the high-arousal conditions. The points also start to spread out more over the valence-arousal space.

The Kruskal-Wallis test on ranks was used to look at the main effects of listener groups with accuracy. Post-hoc comparisons between the listener groups were done using the Mann-Whitney test with the Holm method used to adjust the alpha-level for multiple pairwise comparisons. Figure 2 shows the differences in accuracy for each affective intention, over notes, measures, and phrases. CHM listeners appear to be the most accurate. WM listeners generally perform more accurately than NM listeners with the exception of negative-valence, low arousal stimuli. All listeners also appear to fare badly for positive-valence, low-arousal stimuli. As can be seen in Figure 1, they tend to confuse this affective intention for negative-valence, low-arousal. Figure 2 can be accessed through this link: http://132.206.14.109/supplementaryMaterials/HengFDoMC2021/Fig2.pdf.

Next, the acoustic features listeners use to decode perceived affective intents are explored. This set of analyses focuses on whether listeners fluent in a particular musical tradition converge on a similar set of acoustic features they use in their decoding process, rather than on the accuracy. Instead of looking at the number of "correct" responses from the listeners, all the responses of the listeners in each group were coded into one of the four quadrants in the affective space, regardless of whether they were correct in their judgment of the performer's affective intent. The values of each acoustic descriptor for the notes in a particular quadrant are averaged. From this, four different sets of values for each acoustic descriptor are obtained over the 30-note excerpt. Similar procedures are used for listeners' responses from individual notes, measures, and phrases. Given that the sample size for each group of perceived affective intent can be very different and that consequently the assumptions for parametric tests might be violated, the Kruskal-Wallis test was used to test if the acoustic descriptors that are perceived as

expressing different affective intents were significantly different between the groups of listeners, and post-hoc pairwise comparisons were performed using the Mann-Whitney test with Holm corrections.

Figures 3 to 5 shows the percentage of notes for each acoustic feature that are significantly different over the different affective intentions for each group of listeners. Figure 3 (CHM listeners): http://132.206.14.109/supplementaryMaterials/HengFDoMC2021/Fig3.pdf. Figure 4 (WM listeners): http://132.206.14.109/supplementaryMaterials/HengFDoMC2021/Fig4.pdf. Figure 5 (NM listeners): http://132.206.14.109/supplementaryMaterials/HengFDoMC2021/Fig5.pdf.

As can be seen from these figures, differentiation between the different affective intentions increased with more musical context. However, it also appears that even when the notes are presented individually in a random order, listeners are quite consistent in their understanding of perceived affective intentions. This effect is even more pronounced for the CHM listeners where the values for several acoustic descriptors such as the spectral centroid median for dizi stimuli were all significantly different between the different affective intentions even at the note level. CHM listeners were generally more consistent in the acoustic features they used to determine the perceived affective intentions. There was also greater differentiation between the different affective intentions in the CHM listeners, followed by the WM listeners, whereas NM listeners were the least consistent and had the least differentiation. This trend was seen regardless of the musical tradition of the performer: CHM listeners performed with the greatest consistency in excerpts played by both Chinese and Western instruments.

Discussion

Increasing musical context provides listeners with more information for decoding affective intentions expressed by the performers and there was greater accuracy and clearer divergence between the different affective intentions. Listeners trained in Chinese music appear to be the most consistent, and this could be a result of differences in musical training. There may be differences in the emphasis on the use of timbre in the musical training of the Chinese, as compared to the Western musical tradition, and because of that, listeners trained in the Chinese music tradition may be more sensitive to minute changes in the way timbre is being manipulated in expressing an affective intent. Musical understanding is dependent on the performer, the

stylistic characteristics of the composer, the musical tradition, and also on the experience and expertise of the listener and the listening process, just to name a few of these complex factors that might play a part in communicating musical intentions.

Conclusion

The function of timbre in communicating musical information is a highly complex process with many interactions involving not only different musical parameters, but also interactions between the multitude of acoustic features that make up the quality of a sound. Listeners with different musical backgrounds also appear to utilize the acoustic features to different extents which suggests that conventions regarding timbre function in musical communication are learned.

No continuous response was elicited from listeners with respect to changes in affective intents over the course of the excerpt in this experiment. Although the comparisons across responses for notes, measures, and phrases provide an indication of musical context providing increasing cues for understanding, future studies could attempt to look at continuous responses to better understand the function of timbre over the course of the music.

References

Bowman, C., & Yamauchi, T. (2016). Perceiving categorical emotion in sound: The role of timbre. Psychomusicology: Music, Mind, and Brain, 26(1), 15–25. https://doi.org/10.1037/pmu0000105

Eerola, T., Ferrer, R., & Alluri, V. (2012). Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. Music Perception: An Interdisciplinary Journal, 30(1), 49–70. https://doi.org/10.1525/mp.2012.30.1.49

McAdams, S. (2019). The perceptual representation of timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. Popper, & R. Fay (Eds.), Timbre: Acoustics, perception, and cognition (pp. 23–57). Springer. https://doi.org/10.1007/978-3-030-14832-4_2

McAdams, S., Douglas, C., & Vempala, N. N. (2017). Perception and modeling of affective qualities of musical instrument sounds across pitch registers. Frontiers in Psychology, 8, 1–19. https://doi.org/10.3389/fpsyg.2017.00153

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychological Research, 58(3), 177–192. https://doi.org/10.1007/BF00419633

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The timbre toolbox: Extracting audio descriptors from musical signals. The Journal of the Acoustical Society of America, 130(5), 2902–2916. https://doi.org/10.1121/1.3642604

Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178. https://doi.org/10.1037/h0077714

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. Motivation and Emotion, 1(4), 331–346. https://doi.org/10.1007/BF00992539

Schimmack, U., & Grob, A. (2000). Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. European Journal of Personality, 14(4), 325–345. https://doi.org/10.1002/1099-0984(200007/08)14:4<325::AID-PER380>3.0.CO;2-I

Thompson, W. F., & Balkwill, L. (2010). Cross-cultural similarities and differences. In P. N. Juslin & J. A. Sloboda (Eds.), Handbook of Music and Emotion: Theory, Research, Applications (pp. 755–788). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230143.003.0027

Zhang, J., & Xie, L. (2017). Analysis of timbre perceptual discrimination for Chinese traditional musical instruments. 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 1–4.