Timbre's function within a musical phrase in the perception of affective intents

Lena Heng[1†] and Stephen McAdams[1]
[1] Schulich School of Music, McGill University, Montreal, Quebec, Canada
[†] Corresponding author: lena.heng@mail.mcgill.ca

Introduction

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, studies of whether and how it functions as a carrier of information for communicating affect in music are still in their infancy. It is a "complex auditory attribute... [and] it is also a perceptual property, not a physical one" (McAdams, 2019). However, even as timbre is a psychophysical attribute, it is the perception of the physical properties of a sound that defines timbre, and therefore the physical acoustic properties are important in timbre perception.

If timbre functions as a carrier of affective content, different aspects of the acoustic property of a sound may be implicated for different affective intents. The amount of information timbre carries across different parts of a phrase may vary according to musical context. In addition, how timbre is used for musical communication may also be different across musical traditions. 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 the 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). How these acoustic features are used may also have been learned differently across different musical traditions.

This study therefore aims to find out if specific acoustic features or combinations of them are related to affective intents communicated in a performance, and how increasing musical context might influence this process of understanding. In addition, it also attempts to look at whether differences in musical experience play a role in this decoding process for performances by instruments from a musical tradition listeners have different familiarity with, and whether different acoustic features within a sound are being used differently by each of the groups of listeners in the understanding of affective intents in music.

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 criteria for musicians during participant recruitment was to have more than five years of formal musical training in either the Chinese (mean = 12.00, SD = 2.98) or Western (mean = 12.13, SD = 7.40) music tradition, and the criteria for nonmusicians (mean = 0.2, SD = 0.41) was less than a year of formal training in any type of music. 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. All the CHM listeners however self-identified as being more proficient in Chinese music than Western music. All the 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 participants 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 judgements about each stimulus' perceived affective intent within a two-dimensional affective space of valence and arousal. Experiment B involved participants listening to measures and phrases of these same recorded excerpts and making judgements of the affective intents. 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

The first set of analyses attempts to address the question of how acoustic features may be related to listeners' decoding of perceived affective intents, and whether formal training in different musical traditions influences the ways in which these listeners use the acoustic features. The interest in this study is focused 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 of this communication process. 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 on the affective space, regardless of whether they were correct in their judgement of the performer's affective intent. These four quadrants are: low valence and arousal, low valence high arousal, high valence and arousal, and high valence low arousal. 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. The Kruskal-Wallis test on ranks was used to test if the acoustic descriptors that are perceived as expressing different affective intents were significantly different between the groups of listeners, 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. Further post-hoc pairwise comparisons were performed using the Mann-Whitney test, and the Bonferroni-Holm method was used to adjust the critical alpha for the multiple pair-wise comparisons. Due to space limitations, graphic representations of the results can be found on the internet at the following address: http://132.206.14.109/supplementaryMaterials/HengTIMBRE2020/1.pdf. With increasing musical context (note to measure to phrase), there was increasing differentiation between the different affective intents, indicating an important function of contextual information in understanding perceived affective intents. It also appears that even when the notes are presented individually in a random order to listeners, listeners are able to use certain acoustic features quite consistently in their attempts at understanding affective intents. 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 intents even at the note level. CHM listeners were generally more consistent in the acoustic features they used to determine the perceived affective intents. There was also greater differentiation between the different affective intents in the CHM listeners, followed by the WM listeners, whereas NM listeners were the least consistent and had the least differentiation between the different affective intents. 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.

Ordinary least squares linear multiple regression is next used to look at the relationships between the acoustic features and the dimensions of valence and arousal rated by listeners. The 13 acoustic descriptors for each note were regressed onto the valence and arousal values, which ranged from –1 to +1. A high degree of collinearity was present between these acoustic descriptors, and some were excluded from the regression equations in different analyses. The collinearity values were different for the different notes which resulted in different sets of acoustic descriptors being represented in each regression equation. As the sample sizes for each note were the same, a meaningful comparison could be made with the *t*-values of the regression coefficients. Comparing across listener groups, there appears to be more significant (and more highly significant $p < .001$) *t*-values for the CHM listeners as can be seen for the iqr of RMS energy envelope: http://132.206.14.109/supplementaryMaterials/HengTIMBRE2020/2.pdf, followed by WM, and then the NM listeners. This trend can be also found in the other acoustic features. These different groups of listeners also appear to utilize the acoustic features differently for the affective intents. The weight of the acoustic features used in each note over the excerpt also varies across the different listener groups, as well as across different instruments. Spectral centroid median, for instance, is consistently negatively correlated with valence in the CHM listeners for stimuli played by the *dizi*, whereas the relationship is not as consistent for the other two groups of listeners: http://132.206.14.109/supplementaryMaterials/HengTIMBRE2020/3.pdf. However, the correlation of spectral centroid with valence is less strong in all the groups of listeners for stimuli played by the flute. Similarly, when instruments with similar sound-producing mechanisms are compared (*dizi* and flute; *erhu* and violin; *pipa* and guitar), the temporal centroid appears to be implicated more in communicating affective intents in instruments of the Chinese music tradition, although the few notes with temporal centroid that correlated with affective intents for the violin appeared to have higher significance. CHM listeners also appeared to make greater use of spectral flatness iqr and RMS envelope iqr in their judgements of perceived arousal.

Discussion

There appears to be increasing differentiation in judgement of the different affective intents from participants' responses of notes to measures to phrases, indicating an important function of contextual information in understanding perceived affective intents. While this is expected, it also appears that even when the notes are presented individually and in a random order to listeners, listeners who have had musical training (CHM and WM) are able to use certain acoustic features quite consistently in their attempts at understanding affective intents. Although this is not the same as compared with an approach in which participants rate continuous changes in affective intents for an excerpt of music, the relationships emerging from this current experiment suggests that contextual changes to timbre manipulations by performers contribute a certain extent to the timbre quality of the produced sound, and these subtle changes can provide enough information for quite consistent decoding of affective intents, albeit with less and different information as compared to when a listener can hear an entire excerpt in the right sequence. Musical listening therefore appears to be a complex combination of decoding acoustic information from each individual sound and a complimentary ensemble of contextual cues for an increasingly nuanced understanding.

Results also show that listeners trained in the Chinese music tradition are the most consistent in decoding affective intent of a musical performance, for both Chinese and Western instruments, and nonmusicians fared the worst. Differences in musical training could perhaps have led to an increased focus on timbre characteristics in comprehending affective intent in music. There may be differences in the emphasis on timbre use in the musical training of different musical traditions. The divergent ways in which musical parameters are used and interpreted within the Chinese music tradition compared to the Western music tradition may mean that CHM listeners are much more sensitive to minute changes in the way timbre is being manipulated in expressing an affective intent. While this analysis does not reflect whether CHM listeners are more accurate than WM or NM listeners, it does indicate the increased consistency with

which listeners make use of each acoustic feature, as shown in the greater divergences in the acoustic features between the affective intents.

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 rich multitude of acoustic features that make up the quality of a sound. It appears that how timbre is used in communicating affective intents in music is also different across different musical traditions, with listeners having experiences in different musical traditions making use of different sets of acoustic features to different extents. Once again, this suggests the importance of learning with respect to conventions regarding timbre manipulations in musical communication. Musical understanding is therefore dependent on the performer, the stylistic characteristics of the composer, the musical tradition, and also on the experience of the listener and the listening process, to name just a few of these complex factors.

Only listeners from Singapore were recruited for this study. Such a sampling means that besides the difference in musical backgrounds, confounding variables from other socio-cultural factors of the listeners were reduced. However, another interesting question will be whether CHM, WM, and NM listeners from other localities might have different responses from those in Singapore. If so, this might imply very subtle differences in musical communication in different parts of the world, even within the same type of musical tradition. This study focuses only on whether listeners make consistent use of particular acoustic features in understanding perceived affective intents, and if differences in musical experience might relate to differences in this process. Future work will attempt to look into how performers utilize these timbre manipulations in expressing their intents, and also at the amount of convergence between the performers' intents and the listeners' comprehension of them. No continuous response was elicited from listeners with respect to changes in affective intents over the course of the excerpt. While the comparisons across responses for notes, measures, and phrases provide an indication of musical context providing increasing cues for understanding, future studies could also attempt to look at continuous responses to better understand the function of timbre over the course of an excerpt of music.

Acknowledgments

We thank CIRMMT for funding the Inter-Centre Research Exchange. We also thank Dr. Dorien Herremans (Singapore University of Technology and Design) for hosting LH for the Inter-Centre Research Exchange and for providing support in data collection. This research was supported by grants to SMc from the Canadian Social Sciences and Humanities Research Council (895-2018-1023) and the Fonds de recherche Québec—Société et culture (017-SE-205667), as well as a Canada Research Chair (950-223484).

References

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 Handbook of Auditory Research, vol 69. Springer, Cham.

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.

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.

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.

Zhang, J., & Xie, L. (2017). Analysis of timbre perceptual discrimination for Chinese traditional musical instruments. *Proceedings of 10th International Congress on Image and Signal Processing.* Shanghai: China., 1–4.