Do You Hear What I Hear?

Simin Yang and Mathieu Barthet, Queen Mary University of London
Courtney Reed and Elaine Chew, King's College London

This installment of Computer's *series highlighting the work published in IEEE Computer Society journals comes from* IEEE Transactions on Affective Computing.

Musical performance is often described as expressing emotion. However, the human perception of emotion in music is not well understood. The studies by Yang et al.[1] examine listeners' emotional perception over time to a performance of a single musical piece experienced in live concert conditions, and in the lab, through video recordings. The authors aimed to find out the following:

1. What level of agreement exists between listeners of the same performance?
2. How are perceived emotions related to the semantic features of the music (expressible in linguistic terms) and to machine-extractable music features?
3. What aspects of the music itself and of the listener, like music expertise, influence perceived emotions?

The (Western) classical piece of music chosen was "Piano Trio in F# Minor" (1952) by Babajanian. This violin–cello–piano trio was performed live twice by some of the authors. There were also lighting effects. Videos of the performances can be viewed at https://bit.ly/BabajanianTrio, and Figure 1 shows a still from a video of the performance. There are three movements in the 23-min piece, which span a wide expressive range. The piece is not well known, which minimized bias in the listeners due to familiarity.

There were two phases in this research. First, 15 participants attended a live concert and rated their perceived emotions throughout the performance using their smartphones with an emotion-rating mobile application previously developed by some of the authors. In the second phase, involving 21 different participants, the participants watched a video of the same performance, rated the emotions they perceived, and shared their reasoning for the ratings in open comments, specifically focusing on segments that had received divergent ratings during the first phase.

In both phases, emotion is described in a 2D space, referred to as the *valence/arousal* (VA) space. First, the *valence*

Digital Object Identifier 10.1109/MC.2023.3315470
Date of current version: 13 November 2023

 0018-9162/23©2023IEEE

consists of evaluating the emotion (for example, how positive or negative or how pleasurable or aversive it is), while *arousal* describes the degree of activation associated with it (for example, being active or passive or being excited or sleepy). The smartphone application shows a colored representation of the VA space that the participants can click on to record their perceived emotions. An emotion word is also displayed as the closest in the VA space to where they clicked to help the participants locate the emotion they would like to report in the 2D space. Figure 1 shows the emotion-rating interface, showing the selection of an emotion that is slightly negative in valence and arousal and corresponds to the emotion description "reserved." A total of 949 annotations were collected during the live performance phase, and 3,176 were collected during the video-viewing phase. The second phase also allowed participants to review and modify their emotion ratings and included a measure of musical expertise.

In answer to the first question addressed by the study, there was a wide variation in ratings across participants, demonstrating the individuality of musical emotion perception. However, the agreement was found to be much higher in the video viewing phase than in the live performance phase. This was likely due to the extra attention and focus the participants had when in a lab setting as opposed to in a live performance.

A thematic analysis of the open-ended reflections on emotion ratings resulted in seven themes that were found consistently across participants. First and most obvious were perceptual acoustic features, linked to musical attributes (such as pitch and loudness) and their variations (including harmonic progression, expressive timing, and timbral variation). Second, participants referred to how the instruments were arranged together. For example, while a solo violin part sounded sad, the addition of the cello increased the valence of the perceived emotion. Third, metaphorical language was often used to describe the music, such as "wailing" for the long violin notes. Fourth, a theme emerged linked to the creation and resolution of tension in music, which are well known as strong emotional elements. If expectations are violated,

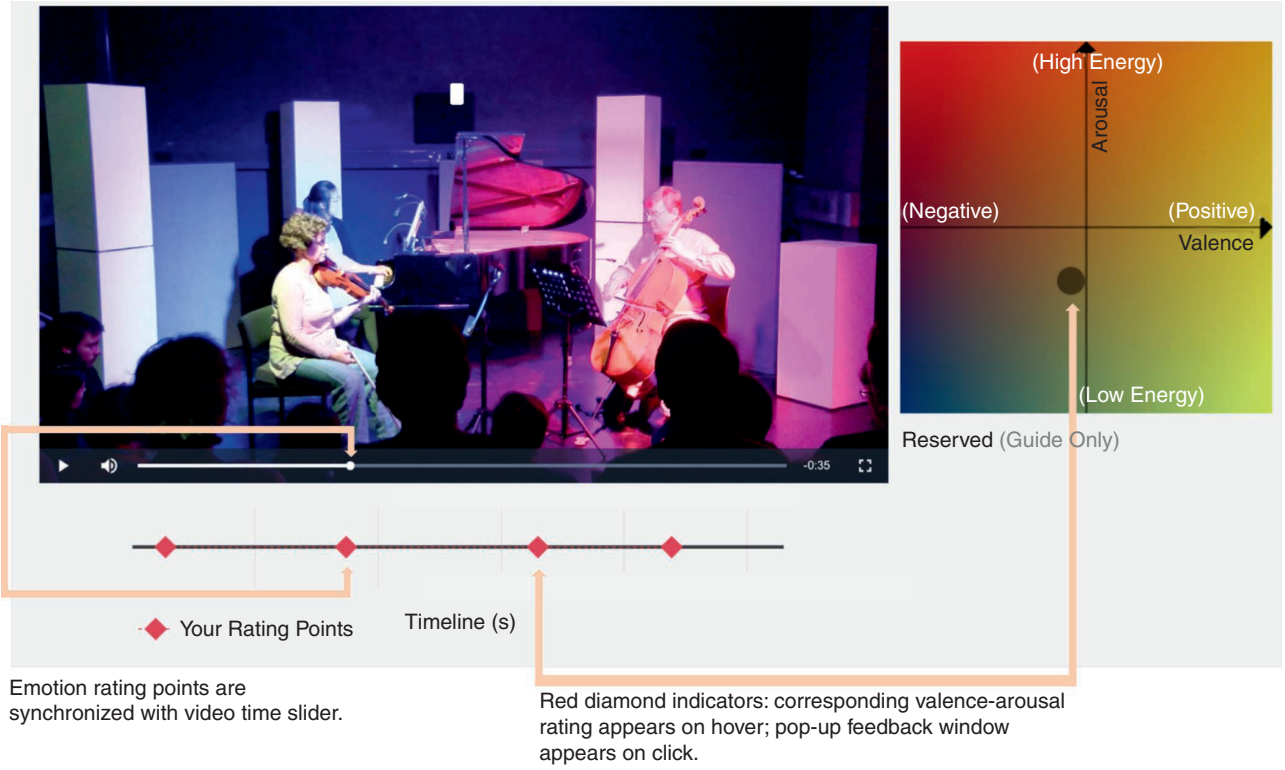

FIGURE 1. Emotion reflection task. The performance video is shown on the left, along with the emotion-rating interface on the right as well as the time points where the participant rated emotions.

listeners feel more uncertain and insecure. The fifth theme focused on repetition and boundaries in the music. The last two themes focused on the expressions of the individual performers (including embodied expressions) and stage visuals.

Musical expertise did not appear to have an effect on rater agreement; it did, however, have an effect on the thematic analysis.

Musical expertise did not appear to have an effect on rater agreement; it did, however, have an effect on the thematic analysis. The final analysis involves finding correlations between the themes derived from the open-ended comments and acoustical features that could be derived from existing audio analysis tools. These tools extract features from audio signals such as root mean square energy, dynamic change, pitch variability, and spectral flatness, among others. The article presents a comprehensive analysis of these connections, potentially allowing researchers to use these off-the-shelf tools to extract features indicative of the seven themes described previously.

In summary, the human perception of emotion in music is a very subtle process, which is subject to individual interpretation and can exhibit significant individual variability. The affective computing research community strives to better understand human emotion and endow artificial intelligence with some of the same (fundamentally human) abilities. Conducting a detailed two-phase examination of listeners' perceived emotions for a specific musical work in live and lab settings, the methods used give significant information about music perception and cognition as well as very interesting leads for future research in affective computing.

REFERENCE

1. Y. Simin, C. N. Reed, E. Chew, and M. Barthet, "Examining emotion perception agreement in live music performance," *IEEE Trans. Affective Comput.*, vol. 14, no. 2, pp. 1442–1460, Apr./Jun. 2023, doi: 10.1109/TAFFC.2021.3093787.

SIMIN YANG is a Ph.D. candidate at the School of Electronic Engineering and Computer Science at Queen Mary University of London, E1 4NS London, U.K. Contact her at simin.yang@qmul.ac.uk.

MATHIEU BARTHET is a senior lecturer in digital media at the School of Electronic Engineering and Computer Science at Queen Mary University of London, E1 4NS London, U.K. He is the director of the U.K. Research and Innovation (UKRI) Centre for Doctoral Training (CDT) in Media and Arts Technology, and co-investigator of the UKRI CDT in AI and Music. Contact him at m.barthet@qmul.ac.uk.

COURTNEY REED is a postdoctoral research associate in design analytics and music physiology at the Department of Engineering and the School of Biomedical Engineering & Imaging Sciences at King's College London, WC2R 2LS London, U.K. Contact her at courtney.reed@kcl.ac.uk.

ELAINE CHEW is a pianist and professor of engineering in the Department of Engineering and the School of Biomedical Engineering & Imaging Sciences at King's College London, WC2R 2LS London, U.K. She is a principal investigator with the European Research Council (ERC) project COSMOS and with the recent ERC POC HEART.FM. Contact her at elaine.chew@kcl.ac.uk.