Identifying Listener-informed Features for Modeling Time-varying Emotion Perception

Simin Yang[1], Elaine Chew[2], Mathieu Barthet[1]

[1] Centre for Digital Music, Queen Mary University of London, UK
{simin.yang, m.barthet}@qmul.ac.uk
[2] CNRS-UMR9912/STMS IRCAM, Paris, France
elaine.chew@ircam.fr

Abstract. Music emotion perception can be highly subjective and varies over time, making it challenging to find salient explanatory acoustic features for listeners. In this paper, we dig deeper into the reasons listeners produce different emotion annotations in a complex classical music piece in order to gain a deeper understanding of the factors that influence emotion perception in music performance. An initial study collected time-varying emotion ratings (valence and arousal) from listeners of a live performance of a classical trio; a follow-up study interrogates the reasons behind listeners' emotion ratings through the re-evaluation of several preselected music segments of various agreement levels informed from the initial study. Thematic analysis of the time-stamped comments revealed themes pertaining primarily to musical features of loudness, tempo, and pitch contour as the main factors influencing emotion perception. The analysis uncovered features such as instrument interaction, repetition, and expression embellishments, which are less mentioned in computational music emotion recognition studies. Our findings lead to proposals for ways to incorporate these features into existing models of emotion perception and music information retrieval researches. Better models for music emotion provide important information for music recommendation systems and applications in music and music-supported therapy.

Keywords: music and emotion, live performance, human computer interaction, thematic analysis

1 Introduction and Background

Music perception studies show that the same music can communicate a range of emotions that vary over time and across listeners [16, 27]. Time-continuous annotation of music enables to capture detailed localised emotion cues, and inter-rater differences can be studied by involving multiple annotators. Previous music emotion studies have evidenced correlations between musical attributes such as dynamics, tempo, mode, timbre, harmony, articulation, timbre, and emotion judgements [11, 15, 18]. In the Music Emotion Recognition (MER) field, several approaches have been proposed to map acoustic features to time-continuous emotional annotations [17, 26, 22]. Yet, little is known on the relative importance

of these features across listeners. Machine learning approaches for MER yielded improved performances overall through extensive testing of different feature sets (bag of audio words), however, these approaches are facing the issue of confounded model performances [1, 14]. In addition, most of the low-level acoustic features involved such as Mel-frequency cepstral coefficients (MFCCs) do not explain the underlying cognitive mechanisms [2, 5, 31].

The subjective nature of music emotion perception has also been less investigated [11, 30]. Traditional approaches to dynamic emotion recognition typically take the average of multi-rater annotations as "target" and discard inconsistent ratings; however, subjective ratings can make the average prone to reliability issues. The variability in rater agreement with the ground truth data may induce a natural upper bound for any algorithmic approach, thus a bottleneck of the MER system performance [13]; it might also lead to a systematic misrepresentation of emotion perception [10]. Such potential limits have also been discussed by in the context of the largest publicly available emotion dataset to date, (DEAM) [1], which provides multi-rater time-varying emotion annotations on over 1800 tracks. Since relatively low agreement between annotators has been found in this dataset, the authors propose as future perspective that *"instead of taking the average values of the emotional annotations as the ground truth and training a generalised model for predicting them, we might want to have a look at the raw annotations and investigate the difference across the annotators."*. This highlights the importance of inter-rater variability in MER researches. As emotion data acquisition can be really expensive and time-consuming, it would be a loss to ignore subjective information which may already exist in available emotion datasets.

In this paper, we present an empirical study aiming to better understand the factors that influence emotion judgements, by exploring time-varying music emotion ratings in a real classical music performance. After collecting emotional annotations from participants in a live context, we conducted exploratory research to find the most relevant features. This was done by asking participants to re-evaluate time-stamped emotion ratings and explain their choice. This provides us with factors related to emotion ratings that have a cognitive meaning. Initial thematic analysis [8] of the time-stamped explanations revealed themes pertaining primarily to musical features of loudness, tempo, and pitch contour as the main factors influencing emotion perception. The analysis also uncovered features such as instrument interaction, repetition, and expression embellishments which are less employed in computational music emotion models. With the recent advances in music information retrieval e.g. in source separation and instrument recognition, listener-informed features can potentially be incorporated for future MER research.

2 Data Acquisition and Statistical Analyses

2.1 Stimulus: Babajanian Piano Trio

In a previous study [35], we collected time-based emotion annotations in a live music performance setting. We chose the piece *Piano Trio in F# minor by Arno Babajanian* which was performed by a professional pianist, cellist and violinist. This piece contains widely disparate characters; as a result, it might express various emotion to participants over time and enable us to capture more explanations from different listeners' perspectives; also this piece is rarely known to the public, thus avoiding familiarity bias. 15 participants provided ratings of valence (degree of pleasantness) and arousal (degree of excitation) [25] which were collected using our web-based and smartphone-friendly app Mood Rater based on a previous framework for audience participation in live music [12]. The audio recording of the concert and emotion data logged on the server-side were synchronised thanks to timestamps. Previous analyses showed varied levels of inter-rater agreement [29], from very low agreement to significant agreement. These results lead us to conduct a follow-up study, which is described in the present paper, in order to better understand the factors influencing listeners judgements of valence and arousal in response to music.

In the follow-up study, we used the video recording of the first two movements of the performance, resulting in stimuli of 17 minutes in length. According to the score provided by the performers, the first movement is marked *Largo-Allegro expressivo-Maestoso*, meaning it is largely in a slow tempo with a faster middle part; the second is marked *Andante*, meaning it is performed at a walking pace. The piece could be segmented into 25 segments based on rehearsal marks on the score[3], lasting from 38 to 72 seconds. Considering the duration of the study for participants, we selected seven excerpts (Segment 5, 7, 12, 13, 14, 17) within the recording for reflective feedback. These excerpts last from 38 to 67 seconds and last 6 minutes in total. Selection of these seven excerpts was based on the diversity of music attributes represented by the stimuli (e.g., instrumentation, loudness, tempo), the diversity of agreement levels of emotion ratings among listeners to cover both commonalities and divergences in music emotion perception. The ICCs of these seven selected experts range from ICC=-0.13, $p > 0.05$ to ICC=0.67, $p < 0.05$ in both arousal and valence.

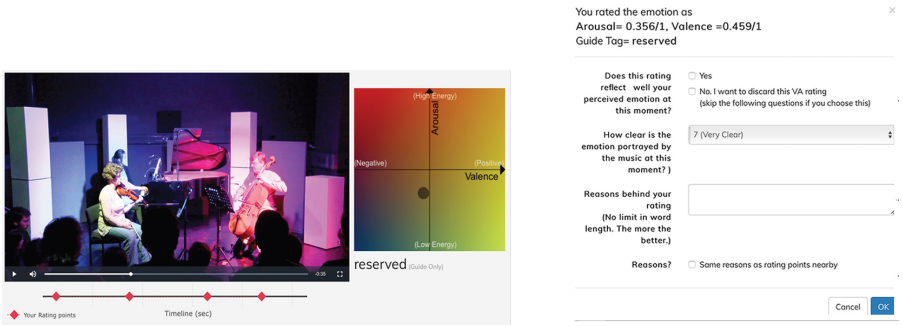
2.2 Procedure

The follow-up study consisted of a rating task followed by a reflective feedback task. Each participant was seated in front of a computer in a quiet sound-proofed room and interacted with a web-based application for stimulus delivery and data

[3] Rehearsal marks are used to identify specific points in a score to facilitate rehearsing. Many scores and parts have bar numbers, every five or ten bars, or at the beginning of each page or line. But as pieces and individual movements of works became longer (extending to several hundred bars), rehearsal marks became more practical in rehearsal, which provides a guideline to segment music.

acquisition. Sound stimuli were presented through headphones with the same sound level (Beyerdynamic DT 770 Pro). Participants were first introduced to the goal of the study, the Valence and Arousal (VA) space, and the self-report framework. Participants then followed a rating trial.

After the rating trial, participants rated the perceived emotion while watching the video recording. They could rate the perceived emotion whenever they perceived a change by clicking on the VA space presented next to the video. In particular, participants were informed that ratings were assumed constant until a change was made. Participants were allowed to pause or rewind the music as needed. For each click on the VA space, both the corresponding UTC timestamp and the corresponding time position of the video were recorded. In addition, corresponding emotion tags were shown below the VA space along upon clicking to help participants to use the VA space. These tags were selected based on [7, 36], which provide a set of normative emotional ratings for a large number of words in English. Participants were informed that these tags were only a guide and they could have their own interpretations of the VA space.



(a) Rating points review

(b) Pop-up window for rating confirmation

Fig. 1: Interfaces for reflective feedback task in the follow-up study (reflective condition)

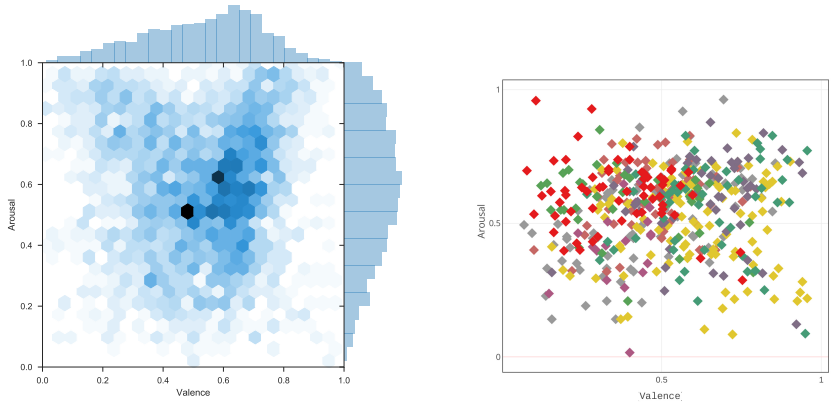
After the rating task, participants started the reflective feedback task. This task was designed for participants to review and confirm each emotion rating they had just given. As shown in Figure 1a, the emotion rating points (shown as red diamonds) were automatically displayed under the video on a synchronised timeline with the video time-slider. By hovering over the rating points, the corresponding VA ratings would be presented in the VA space on the right panel for reflective feedback. By clicking on each rating point, a pop-up window (Figure 1b) appeared for participants to confirm their rating and assess how clearly the emotion was perceived (from 1, very unclear, to 7, very clear). A comment box was provided to allow participants to provide reasons for their ratings using free descriptions. Participants were made aware that there were no right or wrong

answers and they were invited to report as much as possible. After the two tasks, participants completed a questionnaire to collect demographic information, as well as information such as music experience (Goldsmith Music Sophistication Index [24]). The duration of the whole study for each participant ranged from 1.5 to 2.5 hours.

2.3 Participants

21 participants (11 males and 10 females; age M= 28.8, SD=5.5; age range: 23-46 years) participated in the study. One participant stopped after reviewing the first 2 excerpts. Participants had varying degrees of music training (years of engagement in regular, daily practice of a musical instrument: >10 year: 11; 6-9 years: 1; 4-5 years: 1; 1-2 years: 3; 0 year: 5). All participants were current residents in the United Kingdom.

2.4 Explanatory Statistics of Collected Data



(a) Distribution of 3181 VA ratings over 17 minutes with histogram using hexagonal bins (Rating task)

(b) Distribution of 483 explanation ratings, rating points fell on separate segments are shown in different colours (Reflective feedback task)

Fig. 2: Distribution of collected data in the follow-up study (reflective condition)

Rating Task: Over the course of the live performance recording (17 minutes, 25 segments), 3181 VA emotion ratings were collected in total from the 21 participants (151± 96 per participant). Figure 2a depicts the distribution of all 3181 collected VA ratings. This figure shows that the collected data span all four quadrants of the VA space, which is in line with the varied expression within the piece. By comparing the time differences in UTC timestamp as well as those

of video recording timestamps, we found that 10 people rewound or paused the video during the rating process, and no one skipped or fast-forwarded the video. **Reflective feedback task:** 21 participants re-evaluated the 1098 VA ratings they have given on seven pre-selected segments. Among 1098 reviewed ratings, the participants gave explanations and clarity levels towards 471 ratings and categorised another 605 ratings as transition ratings, owing the same reasons than others. 8 participants discarded 23 previous ratings and 7 participants provided 12 new ratings. We collected 483 explanations (23±9 explanations per participant, 7000+ words in total) in total. From Figure 2b we can see that the ratings cover a fairly wide span of the VA space. Hence the explanations represent a broad coverage of emotional responses for the recorded live music performance.

2.5 Measuring Agreement in Participants Emotion Ratings

To quantify the agreement between participants, we computedthe Intra-class Correlation (ICC) [29] at rehearsal segment-level for participants' Valence and Arousal emotion ratings. Specifically, the case of two-way mixed, agreement, average-measures (ICC(2,k)) was adopted for estimating the reliability of the averaged ratings among listeners. Higher ICC values correspond to higher degrees of agreement among listeners, an ICC value of 1 indicates total agreement, while an ICC value of 0 represents random agreement. Negative ICC values are also possible, indicating systematic disagreement. As participants were informed that their emotion will be assumed unchanged until they sent a new rating, we re-sampled individual emotion ratings using a step function at 1Hz for the ICC calculation. The ICC results from both the initial study (live condition) [35] and the current study (reflective condition) are presented in Figure 3.

The ICC of both Arousal and Valence in reflective condition are higher than in the live condition. Possible reasons include: a higher focus and concentration for such an emotion rating task in the lab setting as a single participant compared to real-world live performance setting involving social interactions; the possibility to pause and rewind the videos; differences between groups of participants and larger sample size for ICC calculation in the present study.

3 Listener-informed Features for Music Emotions

3.1 Initial Thematic Analysis on Explanations towards Emotion Ratings

We examined participants' explanations using inductive (bottom-up) thematic analyses [8], a qualitative content analysis approach aiming to look closely into the text in order to find patterns of similar meaning, more than just using a simple count for frequencies of text occurrence.

483 time-stamped explanation data (comments) towards all seven music segments were imported into NVIVO 12 for analysis. Each of the explanation comment was first assigned one or multiple "codes" that identified a feature of the

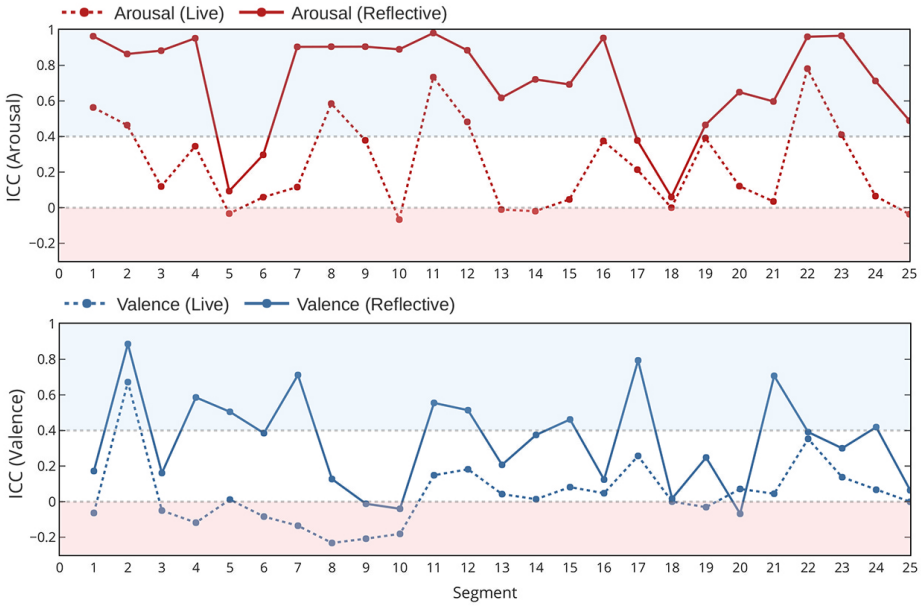


Fig. 3: Intra-class correlation (ICC) for Arousal (top) and Valence (bottom) in both the Live (dotted lines) and Reflective (plain lines) conditions

comment. Broader themes, which were not predetermined, were then obtained by refocusing the analysis at a broader perspective and collating all the relevant coded data within the identified themes.

Figure 4 presents the main themes and the associated codes with their number of occurrence. The occurrence of each code, counted in terms of the number of comments which referred to it, are attached next to each code. As we can see from Figure 4, ten key themes were obtained: **Dynamics, Rhythm, Melody, Harmony, Timbre, Instrument, Structure, Expression, Visuals cues**. It should be noted that some of the themes which emerged overlap as explanations are often multifaceted, such as between **Dynamics** and **Instrument**. In the following discussion, the following notation is used: N refers to the total number of codes for a (sub)theme, and C refers to the number of comments in which a code is found.

As shown in Figure 4, **Dynamics (N=209)** is the most frequently mentioned theme. In this piece, *loudness* (N=169) seems to have been the most salient feature behind participants' music emotion perception. References to **Rhythm (N=114), Harmony (N=114), Melody (N=113)** are also frequently made. Under these three themes, *tempo* (N=74), *pitch contour* (N=67), *mode(major, minor)* (N=50) emerged as three salient factors for music emotion perception. These themes are in line with previous music emotion perception studies which have shown the importance of dynamics, tempo, mode in music emotion perception. In addition, the following themes were found: **Instrument (N=177),**

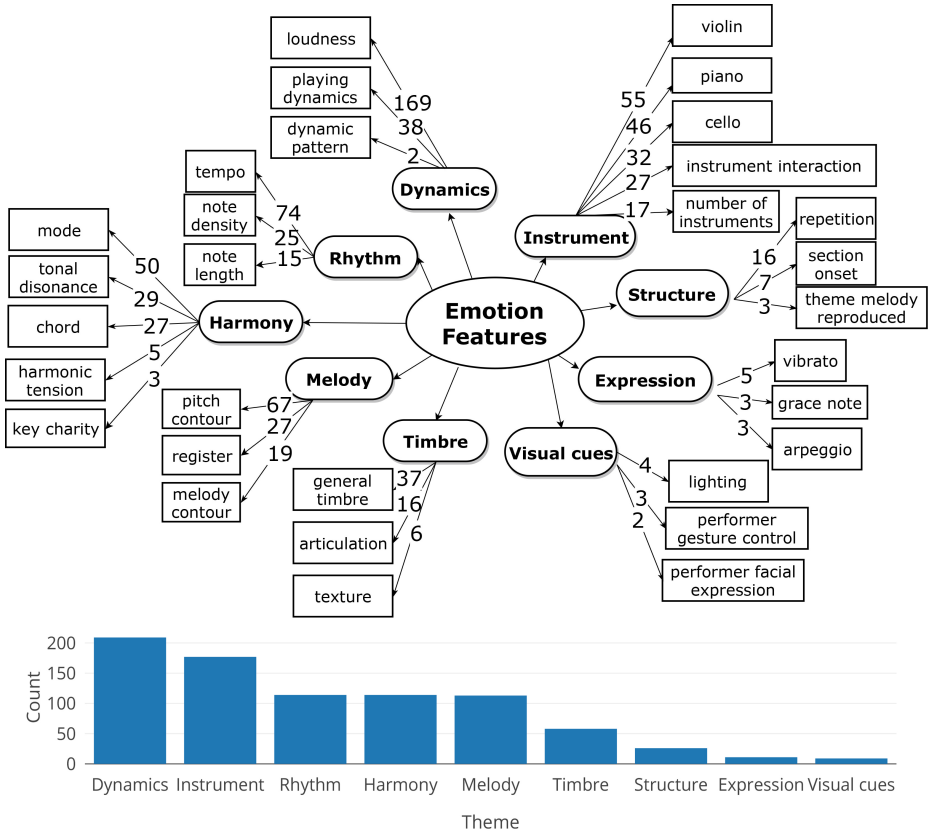


Fig. 4: Thematic analysis of audiences explanation comments

Structure (N=26), Expression (N=11), Visuals cues (N=9). Since these factors are less mentioned in computational emotion research, we discuss them into more details in the following.

Instrument (N=177) Under this theme, many people associated their emotion judgements with one specific instrument or multiple instruments. Violin (N=55), piano (N=46) and cello (N=32) were all frequently referred to for participants' emotional judgements. It provides an indication that some people pay attention to different instruments, which influence their perception of emotion. There are many parameters that performers can control and shape depending on the instrument, from loudness, tempo, timing, articulation to complex continuous aspects such as intonation, instrument timbrel control, and ornaments. Although similar levels of loudness can be reached with different instruments, they can be discriminated by their timbre, and timbre variations have been shown to be an important factor of expressiveness [4]. It can be assumed that performers' timbrel variations also influence the perception of emotion. Other than this, we also extracted codes relating to *instrument interaction* (N=27)

when participants referred to the music with a specific collaboration between multiple performers with multiple instruments, which is sensible as for much of the time music is played in ensembles for instance, the following cases that were mentioned by participants: 1. Multiple instruments are playing the same music melody, which affects the perception of arousal and valence (C=4) 2. The appearance of an instrument can lead to changes of emotion perception, e.g. *"First the violin and cello melty start more warm, then, the piano starts playing and energy increase."* (C=5) 3. Two instruments were responding to each other e.g. *"strong notes alternated between piano and violin"* (C=3).

Structure (N=26) This theme refers to participants' comments on emotion referring to music structure. Supporting codes are *repetition, section onset, theme melody reproduced*. Participants associated their emotion with *repetition* (N=16) of specific music patterns, e.g. *"repetition of same melody accompanying increasing loudness and pitch build up the emotion"*. Transition points, or onsets of a new section within the music, are also possibly lead to the emotion change(N=7). Participants also associated emotion change with the reappearance of theme melody at a given point within the performance (N=3).

Expression (N=11) We categorised supporting codes that were referring to specific music embellishments under this theme. Music embellishments can be obtained by adding notes or producing particular variations to decorate the main music line (or harmony). In particular, people associated the emotion changes with vibratos in violin (N=5), grace notes in piano (N=3) and arpeggios in piano (N=3). Interestingly, these specific factors are mentioned by people with over ten years' music training in violin and piano respectively, and it indicates that people might pay more attention to the instrument they have expertise in playing for emotion perception.

Visual cues (N=9) As participants rated video recordings, some actively reported reasons from the visual perspective, even if this was not mentioned in the task. Participants mentioned the lighting influenced their emotion perceptions. In particular, participants associated the decrease of arousal as the lights turned dark in the final examined segment (N=4). People also referred to the motions of performer gesture, such as bow movement on cello and violin, as reasons for emotion judgements (N=5). Besides, participants mentioned the facial expressions they observed from the performers as reasons, e.g. *"cellist's face looks very expressive, face screws up"*.

3.2 Insights for Building MER Models and MIR

The identification of appropriate and well-functioning features is one of the most important targets in Music information retrival (MIR) researches. Based on our current findings derived from participants' comments, we discuss some insights for the developing better MER systems in the following.

From the **instrument** theme, as participants distinguished between instrumentation and were impacted in an emotional sense by instrumental roles and interactions within the performance, it indicates that using separate instrumental tracks or combinations of them for building music emotion recognition models

might help to improve the prediction accuracy, comparing to modelling emotion directly from the mixed/mastered audio. Previous work by [28] has achieved a better emotion recognition results using multi-track audio of a small group of rock music. With more multi-track datasets [6, 19] open to public nowadays, this is an interesting avenue to explore further. Also, as people associated their emotion judgements to specific patterns of *instrument interaction*, a better detection of numbers of instruments playing at a given time, a better understanding of long-term interaction between instruments as well as the role of each instrument through the audio analysis may benefit emotion prediction. From the **structure** theme, as "repetition in music" has been reported to influence participants' emotion judgement such as building up emotion, being able to detect repetitions from music may also benefit MER. From the **expression** theme, as people have associated emotion judgements with specific music embellishments, it would help to incorporate the automatic detection of vibrato or other music ornaments into building MER systems especially for time-varying music emotion recognition. Recent advances in the MIR field on playing technique detection may provide such opportunities, such as works of detection of vibrato in violin and erhu [20, 34], arpeggios in multiple instruments [3], pedalling in piano [21] and representative playing techniques in guitar [32, 9] and bamboo flute [33]. Moreover, finally the **visual cues** theme indicated that dynamics of emotional perception in live performance could be a multimodal phenomenon, and multimodal emotion sensing using computer vision [23] and audio can also be promising in the future design of music emotion studies.

4 Conclusion

Understanding how music affects listeners perception of emotion facilitates creating fair and unbiased music information retrieval systems. In this paper, we examined the time-varying music emotion perception from the participants in a complementary way: The collection of time-varying emotion ratings enabled a quantitative measure of emotion responses and retroactive rating reflection; while explanations from participants helped to highlight the reasons behind such emotion judgements. However, we did not give an exhaustive answer regarding listener-informed features but present the current state and experimental data that have been collected so far within this ongoing project. In the future work, we plan to re-conduct the thematic analysis with more coders to increase the validity and reliability of the results. We also plan to investigate the individual differences on time-varying music emotion perception involving music expertise and demographic information, as well as to investigate the reasons behind the varied levels of agreement in perceived emotion agreement over the performance. As one of the most important issue in MIR tasks is the identification of appropriate and well-functioning features, our current findings of listener-informed music features underpin the previous emotion studies, in addition, the identification of less employed music features such as instrumentation and ornaments also generate some insight for the improvement of MER systems.

References

1. Aljanaki, A., Yang, Y.H., Soleymani, M.: Developing a benchmark for emotional analysis of music. PloS one **12**(3), e0173392 (2017)
2. Aucouturier, J.J., Bigand, E.: Mel cepstrum & ann ova: The difficult dialog between mir and music cognition. In: ISMIR. pp. 397–402 (2012)
3. Barbancho, I., Tzanetakis, G., Barbancho, A.M., Tardón, L.J.: Discrimination between ascending/descending pitch arpeggios. IEEE/ACM Transactions on Audio, Speech, and Language Processing **26**(11), 2194–2203 (2018)
4. Barthet, M., Depalle, P., Kronland-Martinet, R., Ystad, S.: Acoustical correlates of timbre and expressiveness in clarinet performance. Music perception: An interdisciplinary journal **28**(2), 135–154 (2010)
5. Barthet, M., Fazekas, G., Sandler, M.: Multidisciplinary perspectives on music emotion recognition: Implications for content and context-based models. Proc. CMMR pp. 492–507 (2012)
6. Bittner, R.M., Salamon, J., Tierney, M., Mauch, M., Cannam, C., Bello, J.P.: Medleydb: A multitrack dataset for annotation-intensive mir research. In: ISMIR. pp. 155–160 (2014)
7. Bradley, M.M., Lang, P.J.: Affective norms for english words (anew): Instruction manual and affective ratings. Tech. rep., Citeseer (1999)
8. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qualitative research in psychology **3**(2), 77–101 (2006)
9. Chen, Y.P., Su, L., Yang, Y.H., et al.: Electric guitar playing technique detection in real-world recording based on f0 sequence pattern recognition. In: ISMIR. pp. 708–714 (2015)
10. Cowie, R., McKeown, G., Douglas-Cowie, E.: Tracing emotion: an overview. International Journal of Synthetic Emotions (IJSE) **3**(1), 1–17 (2012)
11. Eerola, T., Vuoskoski, J.K.: A review of music and emotion studies: approaches, emotion models, and stimuli. Music Perception: An Interdisciplinary Journal **30**(3), 307–340 (2013)
12. Fazekas, G., Barthet, M., Sandler, M.B.: The mood conductor system: Audience and performer interaction using mobile technology and emotion cues. In: 10th International Symposium on Computer Music Multidisciplinary Research (CMMR'13). pp. 15–18 (2013)
13. Flexer, A., Grill, T.: The problem of limited inter-rater agreement in modelling music similarity. Journal of new music research **45**(3), 239–251 (2016)
14. Friberg, A., Schoonderwaldt, E., Hedblad, A., Fabiani, M., Elowsson, A.: Using listener-based perceptual features as intermediate representations in music information retrieval. The Journal of the Acoustical Society of America **136**(4), 1951–1963 (2014)
15. Gabrielsson, A., Lindström, E.: The role of structure in the musical expression of emotions. Handbook of music and emotion: Theory, research, applications **367400** (2010)
16. Hiraga, R., Matsuda, N.: Graphical expression of the mood of music. In: Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on. vol. 3, pp. 2035–2038. IEEE (2004)
17. Imbrasaitė, V., Baltrušaitis, T., Robinson, P.: Emotion tracking in music using continuous conditional random fields and relative feature representation. In: Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on. pp. 1–6. IEEE (2013)

18. Juslin, P.N., Lindström, E.: Musical expression of emotions: Modelling listeners' judgements of composed and performed features. Music Analysis **29**(1-3), 334–364 (2010)
19. Li, B., Liu, X., Dinesh, K., Duan, Z., Sharma, G.: Creating a multitrack classical music performance dataset for multimodal music analysis: Challenges, insights, and applications. IEEE Transactions on Multimedia **21**(2), 522–535 (2019)
20. Li, P.C., Su, L., Yang, Y.H., Su, A.W., et al.: Analysis of expressive musical terms in violin using score-informed and expression-based audio features. In: ISMIR. pp. 809–815 (2015)
21. Liang, B., Fazekas, G., Sandler, M.: Piano sustain-pedal detection using convolutional neural networks. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 241–245. IEEE (2019)
22. Lu, L., Liu, D., Zhang, H.J.: Automatic mood detection and tracking of music audio signals. IEEE Transactions on audio, speech, and language processing **14**(1), 5–18 (2006)
23. Mou, W., Gunes, H., Patras, I.: Alone versus in-a-group: A multi-modal framework for automatic affect recognition. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) **15**(2), 47 (2019)
24. Müllensiefen, D., Gingras, B., Stewart, L., Musil, J.J.: Goldsmiths musical sophistication index (gold-msi) v1. 0: Technical report and documentation revision 0.3. London: Goldsmiths, University of London. (2013)
25. Rusell, J.: A circumplex model of affect. Personality and Social Psychology pp. 1161–1178 (1980)
26. Schmidt, E.M., Kim, Y.E.: Modeling musical emotion dynamics with conditional random fields. In: ISMIR. pp. 777–782 (2011)
27. Schubert, E.: Modeling perceived emotion with continuous musical features. Music Perception: An Interdisciplinary Journal **21**(4), 561–585 (2004)
28. Scott, J., Schmidt, E.M., Prockup, M., Morton, B., Kim, Y.E.: Predicting time-varying musical emotion distributions from multi-track audio. CMMR **6**, 8 (2012)
29. Shrout, P.E., Fleiss, J.L.: Intraclass correlations: uses in assessing rater reliability. Psychological bulletin **86**(2), 420 (1979)
30. Soleymani, M., Aljanaki, A., Yang, Y.H., Caro, M.N., Eyben, F., Markov, K., Schuller, B.W., Veltkamp, R., Weninger, F., Wiering, F.: Emotional analysis of music: A comparison of methods. In: Proceedings of the 22nd ACM international conference on Multimedia. pp. 1161–1164. ACM (2014)
31. Sturm, B.L.: Evaluating music emotion recognition: Lessons from music genre recognition? In: 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). pp. 1–6. IEEE (2013)
32. Su, L., Yu, L.F., Yang, Y.H.: Sparse cepstral, phase codes for guitar playing technique classification. In: ISMIR. pp. 9–14 (2014)
33. Wang, C., Benetos, E., Lostanlen, X., Chew, E.: Adaptive time–frequency scattering for periodic modulation recognition in music signals. In: ISMIR (2019)
34. Yang, L., Rajab, K.Z., Chew, E.: The filter diagonalisation method for music signal analysis: frame-wise vibrato detection and estimation. Journal of Mathematics and Music **11**(1), 42–60 (2017)
35. Yang, S., Barthet, M., Chew, E.: Multi-scale analysis of agreement levels in perceived emotion ratings during live performance. In: Extended abstracts for the Late-Breaking Demo Session of ISMIR (2017)
36. Yang, Y.H., Liu, J.Y.: Quantitative study of music listening behavior in a social and affective context. IEEE Transactions on Multimedia **15**(6), 1304–1315 (2013)