

Continuous Response to Music using Discrete Emotion Faces

Emery Schubert¹, Sam Ferguson², Natasha Farrar¹, David Taylor¹ and Gary E. McPherson³,

¹ Empirical Musicology Group, University of New South Wales, Sydney, Australia

² University of Technology, Sydney, Australia

³ Melbourne Conservatorium of Music, University of Melbourne, Melbourne, Australia
E.Schubert@unsw.edu.au

Abstract. An interface based on expressions in simple graphics of faces were aligned in a clock-like distribution with the aim of allowing participants to quickly and easily rate emotions in music continuously. We developed the interface and tested it using six extracts of music, one targeting each of the six faces: ‘Excited’ (at 1 o’clock), ‘Happy’ (3), ‘Calm’ (5), ‘Sad’ (7), ‘Scared’ (9) and ‘Angry’ (11). 30 participants rated the emotion expressed by these excerpts on our ‘emotion-face-clock’. By demonstrating how continuous category selections (votes) changed over time, we were able to show that (1) more than one emotion-face could be expressed by music at the same time and (2) the emotion face that best portrayed the emotion the music conveyed could change over time, and that the change could be attributed to changes in musical structure.

Keywords: Emotion in music, continuous response, discrete emotions, time-series analysis, film music.

1 Introduction

Research on continuous ratings of emotion expressed by music (that is, rating the music while it is being heard) has led to improvements in understanding and modeling music’s emotional capacity. This research has produced time series models where musical features such as loudness, tempo, pitch profiles and so on are used as input signals which are then mapped onto emotional response data using least squares regression and various other strategies [1-4].

One of the criticisms of self-reported continuous response however, is the rating response format. During their inception in the 1980s and 1990s [5, 6] such measures have mostly consisted of participants rating one dimension of emotion (such as the happiness, or arousal, or the tension, and so on) in the music. This approach could be viewed as so reductive that a meaningful conceptualization of emotion is lost. For

example, Russell's [7, 8] work on the structure of emotion demonstrated that a large amount of variance in emotion can be explained by two fairly independent dimensions, frequently labeled valence and arousal. The solution to measuring emotion continuously can therefore be achieved by rating the stimulus twice (that is, in two passes), once along a valence scale (with poles of the scale labeled positive and negative), and once along an arousal scale (with poles labeled active and sleepy) [for another multi-pass approach see 9]. In fact, some researchers have combined these scales at right angles to form an 'emotion space' so as to allow a good compromise between reductive simplicity (the rating scale), and the richness of emotional meaning (applying what were thought to be the two most important dimensions in emotional structure simultaneously and at right angles) [e.g. 10, 11, 12].

The two dimensional emotion space has provided an effective approach to help untangle some of the relations between musical features and emotional response, as well as providing a deepening understanding of how emotions ebb and flow during the unfolding of a piece of music. However, the model has been placed under scrutiny on several occasions. The most critical matter that is of concern in the present research is theory and subsequent labeling of the emotion dimensions and ratings. For example, the work of Schimmack [13, 14] has reminded the research community that there are different ways of conceptualizing the key dimensions of emotion, and one dimension may have other dimensions hidden within it. Several researchers have proposed three key dimensions of emotion [15-17]. Also, dimensions used in the 'traditional' two dimensional emotion space may be hiding one or more dimensions. Schimmack demonstrated that the arousal dimension is more aptly a combination of underlying 'energetic arousal' and 'tense arousal'. Consider, for instance, the emotion of 'sadness'. On a single 'activity' rating scale with poles labeled active and sleepy, sadness will most likely occupy low activity (one would not imagine a sad person jumping up and down). However, in a study by Schubert [12] some participants consistently rated the word 'sad' in the high arousal region of the emotion space (all rated sad as being a negative valence word). The work of Schimmack and colleagues suggests that those participants were rating sadness along a 'tense arousal' dimension, because sadness does contain conflicting information about these two kinds of arousal – high tension arousal but low activity arousal.

Some solutions to the limitation of two dimensions are to have more than two passes when performing a continuous response (e.g. valence, tense arousal and activity arousal), or to apply a three dimensional GUI with appropriate hardware (such as a three dimensional mouse). However, in this paper we take the dilemma of dimensions as a point of departure and apply what we believe is the first attempt to use a discrete emotion response interface for continuous self-reported emotion ratings.

Discrete emotions are those that we think of in day-to-day usage of emotions, such as happy, sad, calm, energetic and so forth. They can each be mapped onto the emotional dimensions discussed above, but can also be presented as independent, meaningful conceptualizations of emotion [18-22]. An early continuous self-reported rating of emotion in music that demonstrated an awareness of this discrete structure was applied by Namba *et al.* [23], where a computer keyboard was labeled with fifteen different discrete emotions. As the music unfolded, participants pressed the key representing the emotion that the music was judged to be expressing at that time. The study has to our knowledge not been replicated, and we believe it is because the complexity of learning to decode a number of single letters and their intended emotion-word meaning. It seems likely that participants would have to shift focus between decoding the emotion represented on the keyboard, or finding the emotion and then finding its representative letter before pressing. And this needed to be done on the fly, meaning that by the time the response was ready to be made, the emotion in the music may have changed. The amount of training (about 30 minutes reported in the study) needed to overcome this cognitive load can be seen as an inhibiting factor.

Inspired by Namba *et al.*'s pioneering work, we wanted to develop a way of measuring emotional response continuously but one which captured the benefits of discrete emotion rating, while applying a simple, intuitive user interface.

2 Using discrete facial expressions as a response format

By applying the work of some of the key research of emotion in music who have used discrete emotion response tools [24-26], and based on our own investigation [27], we devised a system of simple,

schematic facial expressions intended to represent a range of emotions that are known to be evoked by music. Further, we wanted to recover the topology of semantic relations, such that similar emotions were positioned beside one another, whereas distant emotions were physically more distant. This approach was identified in Hevner’s [28-31] adjective checklist. Her system consisted of groups of adjectives, arranged in a circle in such a way as to place clusters of words near other clusters of similar meaning. For example, the cluster of words containing ‘bright, cheerful, joyous ...’ was adjacent to the cluster of words containing ‘graceful, humorous, light...’, but distant from the cluster containing the words ‘dark, depressing, doleful...’. Eventually, the clusters would form a circle, from which it derived its alternative names ‘adjective clock’ [32] and ‘adjective circle’ [31]. Modified version of this approach, using a smaller number of words, are still in use [33]. Our approach also used a circular form, but using faces instead of words. Consequently, we named the layout an ‘emotion-face-clock’. Literate and non-literate cultures have become adept at speedy interpretation of emotional expression in faces [34, 35], making them more suitable for emotion rating tasks than words. Further, several emotional expressions are universal [36, 37] making the reliance on a non-verbal, non-language specific format appealing [38-40].

Selection of faces to be used for our response interface were based on the literature of commonly used emotion expressions to describe music [41], the recommendations made on a review of the literature by Schubert and McPherson [42] but also such that the circular arrangement was plausible. The faces selected corresponded roughly with the emotions from top moving clockwise (see Fig. 1): Excited (at 1 o’clock), Happy (3), Calm (5), Sad (7), Scared (9) and Angry (11 o’clock), with the bottom of the circle separated by Calm and Sad. The words used to describe the faces are selected for the convenience of the researchers. Although a circle arrangement was used, a small gap between the positive emotion faces and the negative emotion faces was imposed, because a spatial gap angry and excited, and between calm and sad reflected a semantic distance (Fig. 1). We did not impose our labels of the emotion-face expressions onto the participants. Pilot testing using retrospective ratings of music using the verbal expressions are reported in Schubert *et al.* [27].

3 Aim

The aim of the present research was to develop and test the emotion-face-clock as a means of continuously rating the emotion expressed by extracts of music.

4 Method

4.1 Participants

Thirty participants were recruited from a music psychology course that consisted of a range of students including some specializing in music. Self-reported years of music lessons ranged from 0 to 16 years, mean 6.6 years ($SD = 5.3$ years) with 10 participants reporting no music lessons ('0' years). Ages ranged from 19 to 26 years (mean 21.5 years, $SD = 1.7$ years). Twenty participants were male.

4.2 Software realisation

The emotion-face-clock interface was prepared, and controlled by MAX/MSP software, with musical extracts selected automatically and at random from a predetermined list of pieces. Mouse movements were converted into one of eight states: centre, one of the six emotions represented by schematic faces, and 'elsewhere' (Fig. 1). The eight locations were then stored in a buffer that was synchronized with the music, with a sampling rate of 44.1kHz. Given the redundancy of this sampling rate for emotional responses to music [which are in the order of 1 Hz – see 43], down-sampling to 25Hz was performed prior to analysis. The facial expressions moving around the clock in a clockwise direction were Excited, Happy, Calm, Sad, Scared and Angry. Note that the verbal labels for the faces are for the convenience of the researcher, and do not have to be the same as those used by participants. More important was that the expressions progressed sequentially around the clock such that related emotions were closer together than distant emotions, as described above. However, the quality of our labels were tested against participant data using the explicit labeling of the same stimuli in an earlier study [27].

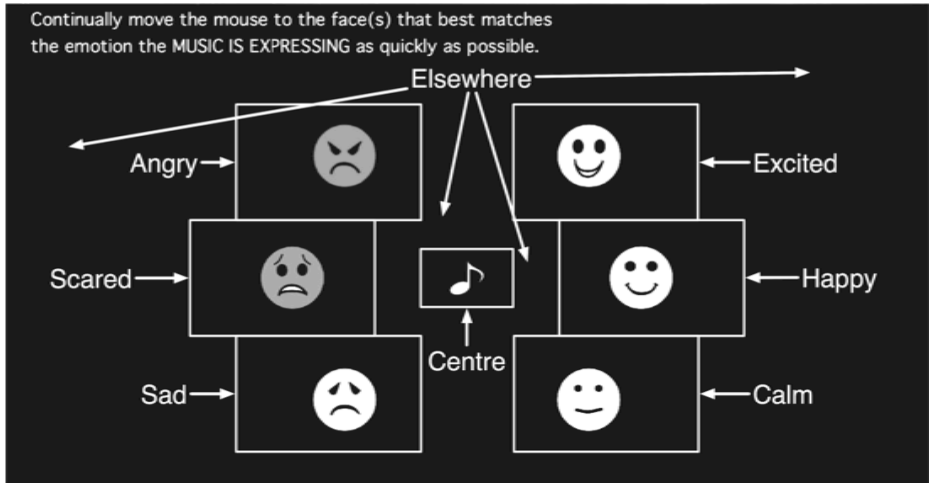


Fig. 1. Emotion-face-clock graphic user interface. This is a grayscale version. Face colours were yellow shades for right three faces (Excited [bright yellow], Happy and Calm), red for Angry, dark blue for Scared and light blue for Sad, based on [27]. Crotchet icon in Centre was green when ready to play, and grayed out, opaque when excerpt was playing. Text in top two lines provided instructions for the participant. White boxes, arrows and labels were not visible to the participants. These indicate the regions used to determine the eight response categories.

4.3 Procedure

Participants were tested one at a time. The participant sat at the computer display and wore headphones. After introductory tasks and instructions, the emotion-face-clock interface was presented, with a green icon (quaver) in the centre (Fig. 1). The participant was instructed to click the green button to commence listening, and to track the emotion that the music was expressing by selecting the facial expression that best matched the response. They were asked to make their selection as quickly as possible. When the participant moved the mouse over one of the faces, the icon of the face was highlighted to provide feedback. The participant was asked to perform several other tasks. The focus of the present report is on continuous rating over time of emotion that six extracts of music were expressing.

4.4 Stimuli

Because the aim of this study is to examine our new continuous response instrument, we selected six musical excerpts for which we had emotion ratings made using tradition post-performance ratings scales from a previous study [27]. The pieces were taken from Pixar animated movies, based on the principle that the music would be written to stereotypically evoke a range of emotions. The excerpts selected were 11 to 21 seconds long with the intention of primarily depicting each of the emotions of the six faces on the emotion-face-clock. In our reference to the stimuli in this report, they were labeled according to their target emotion: *Angry*, *Scared*, *Sad*, *Calm*, *Happy* and *Excited*. More information about the selected excerpts is shown in **Table 1**. When referring to a musical stimulus the emotion label is capitalized and italicised.

Table 1. Stimuli used in the study.

| Stimulus code (target emotion) | Film music excerpt | Start time within CD track (MM'SS elapsed) | Duration of excerpt (s) |
|-----------------------------------|-----------------------------------|--|----------------------------|
| <i>Angry</i> | Up: 52 Chachki Pickup | 00"53 | 17 |
| <i>Calm</i> | Finding Nemo: Wow | 00"22 | 16 |
| <i>Excited</i> | Toy Story: Infinity and Beyond | 00"15 | 16 |
| <i>Happy</i> | Cars: McQueen and Sally | 00"04 | 16 |
| <i>Sad</i> | Toy Story 3: You Got Lucky | 01"00 | 21 |
| <i>Scared</i> | Cars: McQueen's Lost | 00"55 | 11 |

5 Results and Discussion

Responses were categorized into one of eight possible responses (one of the six emotions, the centre location, and any other space on the emotion-face-clock labeled 'elsewhere' – see Fig. 1) based on mouse positions recorded during the response to each piece of music. This process was repeated for each sample (25 per second). Two main analyses were conducted. First, the relationships between the collapsed continuous ratings against rating scale results from a previous study using the same stimuli, and then an analysis of the time series responses for each of the six stimuli.

5.1 Summary responses

In a previous study, 26 participants provided ratings of each of the six stimuli used in the present study (for more details, see [27] for details) along 11 point rating scales from ‘0 (not at all)’ to ‘10 (a lot)’. The scales were labeled Angry, Scared, Sad, Calm, Happy and Excited. No faces were used in the response interface for that study.

The continuous responses from the current study were collapsed so that the number of votes a face received as the piece unfolded was tallied, producing a proportional representation of faces that were selected as indicating the emotion expressed by each face for a particular stimulus. The plots of these results are shown in Fig. 2. Take for example the responses made to the Angry excerpt. All participants first ‘votes’ were for the ‘Centre’ category because they had to click the icon at the centre of the emotion-face-clock to commence listening. As participants decided which face represented the emotion expressed, they moved the mouse to cover the appropriate face. So, as the piece unfolded, at any given time, some of the 30 participants might have the cursor on the Angry face, while some on the Scared face, and another who may not yet have decided remains in the centre or has moved the mouse, but not to a face (‘elsewhere’). With a sampling rate of 25 Hz it was possible to see how these votes changes over time (the focus of the next analysis). At each sample, the votes were tallied into the eight categories. Hence each sample had a total of 30 votes (one per participant). At any sample it was possible to determine whether participants were or were not in agreement about the face that best represented the emotion expressed by the music.

The face by face tallies for each of these samples were accumulated and divided by the total number of samples for the excerpt. This provided a summary measure of the time-series to approximate the typical response profile for the stimulus in question. These profiles are reported in Fig. 2 in the right hand column. Returning to the *Angry* example we see that participants spent most time on the Angry face, followed by Scared and then the Centre. This suggests that the piece selected indeed best expressed anger according to the accumulated summary of the time series. The second highest votes belonging to the Scared face can be interpreted as a ‘near miss’ because of all the emotions on the clock, the scared face is semantically closest to the Angry face, despite obvious differences (for a discussion, see [27]). In fact, when comparing the accumulated summary with the post-

performance rating scale profile (from the earlier study), the time series produces a profile more in line with the proposed target emotion. The post-performance ratings demonstrate that Angry is only the third highest scored scale, after Scared and Excited. The important point, however, is that Scared and Excited are located on either side of the emotion-face-clock, making them the most semantically related alternatives to angry of the available faces. For each of the other stimuli, the contour of the profiles for post-performance ratings and accumulated summary of continuous response are identical.

These profiles matches are evidence for the validity of the emotion-face-clock because they mean that the faces are used to provide a similar meaning to the emotion words used in the post-performance verbal ratings. We can therefore be reasonably confident that at least five of the faces selected can be represented verbally by the five verbal labels we have used (the sixth – Anger, being confused occasionally with Scared). The similarity of the profile pairs in Fig. 2 is also indicative of the reliability of the emotion-face-clock because it more-or-less reproduces the emotion profile of the post-performance ratings.

Two further observations are made about the summary data. Participants spend very little time away from a face or the centre of the emotion-face-clock (the elsewhere region is selected infrequently for all six excerpts). While there is the obvious explanation that the six faces and the screen centre occupy the majority of the space on the response interface (see Fig. 1) the infrequent occurrence of the Elsewhere category also may indicate that participants are fairly certain about the emotion that the music is conveying. That is, when an emotion face is selected by a participant, they are likely to believe that to be the best selection, even if it is in disagreement with the majority of votes, or with the *a priori* proposed target emotion. If this were not the case, we might expect participants to hover in ‘no mans land’ of the emotion-face-clock—Elsewhere and Centre.

The ‘no-mans-land’ response may be reflected by the accumulated time spent on the centre category. As mentioned, time spent in the centre category is biased because participants always commence their responses from that region (in order to click the play button). The centre category votes can therefore be viewed as indicating two kinds of systematic responses: (1) initial response time and (2) response uncertainty. Initial response time is the time required for a participant to orient to the required task just as the temporally unfolding stimulus

commences. The orienting process generally takes several seconds to complete, prior to ratings becoming more ‘reliable’ [44-46]. So stimuli in Figure 2 with large bars for ‘Centre’ may require more time before an unambiguous response is made.

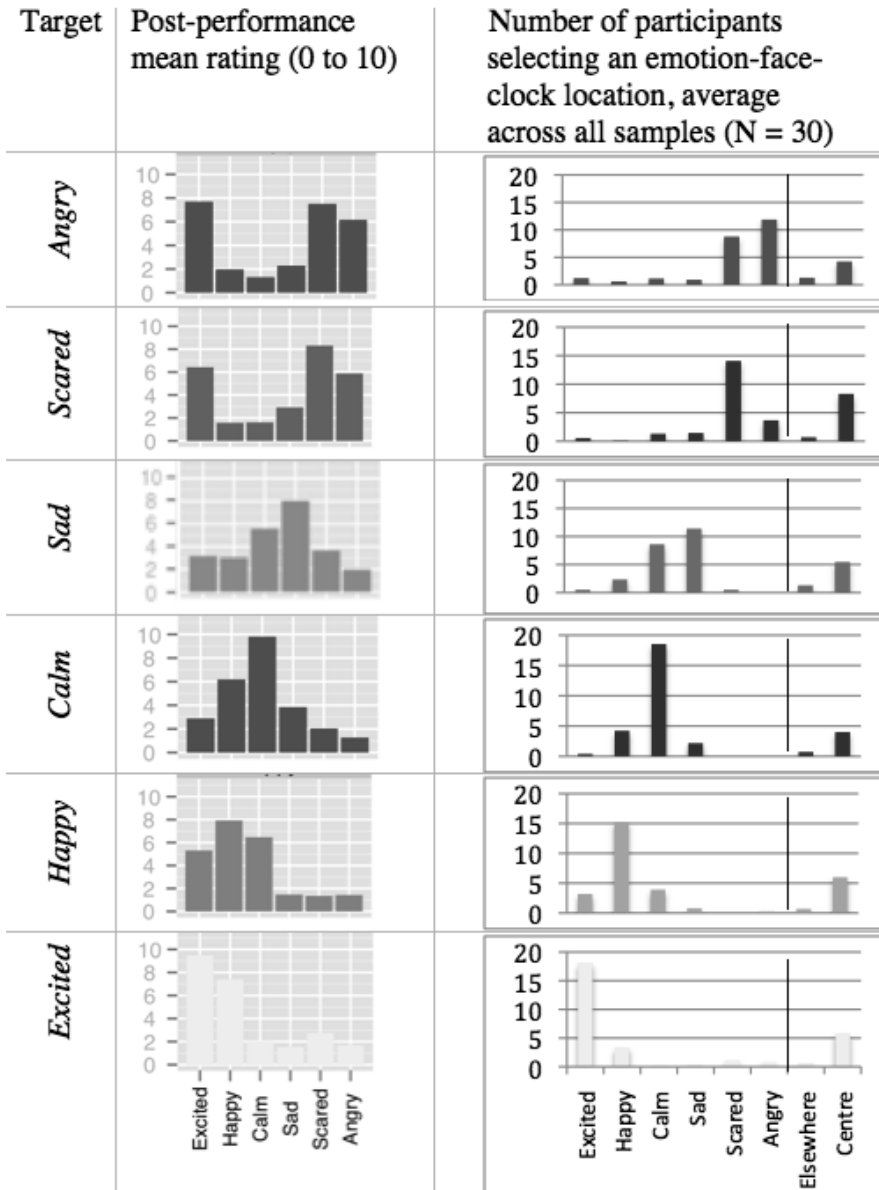


Fig. 2. Comparison of post performance ratings [from 27] (left column of charts) with sample averaged continuous response face counts for thirty participants (right column of charts) for the six stimuli, each with a target emotion shown in the leftmost column.

The relatively large amount of time spent in the Centre for this piece may, also, be an indicator of uncertainty of response. Well after a typical orientation period has passed, for this excerpt, uncertainty in rating remains (as will become clear in the next sub-section). The *Scared* stimulus has the largest number of votes for the Centre location (on average, at any single sample, eight out of thirty participants were in the centre of the emotion-face-clock). Without looking at the time series data, we may conclude that the *Scared* excerpt produced the least ‘confident’ rating, or that the faces provided were unable to produce satisfactory alternatives for the participants.

Using this logic (long time spent in the Centre and Elsewhere), we can conclude that the most confident responses were for those pieces where accumulated time spent in the Centre and Elsewhere were the lowest. The *Calm* stimulus had the highest ‘confidence’ rating (an average of about 4 participants at the Centre or Elsewhere combined). Interestingly, the *Calm* example also had the highest number of accumulated votes for any single category (the target, Calm emotion) — which was selected on average by 18 participants at any given time.

The analysis of summary data provides a useful, simple interpretation of the continuous responses. However, to appreciate the richness of the time-series responses, we now examine the time-series data for each stimulus.

5.2 Continuous responses

Fig. 3 shows the plots of the stacked responses from the 30 participants at each sample, for each stimulus. The beginning of each time series, thus, demonstrates that all participants commenced their response at the Centre (the first, left-most vertical ‘line’ of each plot is all black, indicating the Centre). By scanning for black regions for each of the plots in Fig. 2 some of the issues raised in the accumulated summary analysis, above, are addressed. We can see that the black and grey disappears for the *Calm* plot after 6 seconds have elapsed. For each of the other stimulus a small amount of doubt remains at certain times – in some cases a small amount of uncertainty is reported throughout (there are no samples in the *Scared* and *Excited* stimuli where all participants have selected a face). Further, the largest area of black and grey occurs in the *Scared* plot.

The time taken for ‘most’ participants to make a decision about the selection of a first face is fairly stable across stimuli. Inspection of Fig. 3 reveals that in the range of 0.5 seconds through to 5 seconds most participants have selected a phase. This provides a rough estimate of the initial orientation time for emotional response using categorical data (for more information, see [44]).

Another important observation of the time-series of Fig. 3 is the ebb and flow of face frequencies. In the summary analysis it was possible to see when more than one emotion face was selected to identify the emotion expressed by the music. However, here we can see *when* these ‘ambiguities’ occur. The *Angry* and *Sad* stimuli provide the clearest examples of more than one dominant emotion. For the *Angry* excerpt, the ‘Scared’ face is frequently reported in addition to Angry. And the number of votes for the Scared face slightly increase toward the end of the excerpt. Thus, it appears that the music is expressing two emotions at the same time, or that the precise emotion was not available on the emotion-face-clock.

The *Sad* excerpt appears to be mixed with Calm for the same reasons (co-existence of emotions or precision of the measure). While the Calm face received fewer votes than the Sad face, the votes for Calm peak at around the 10th second (15 votes received over the time period 9.6 to 10.8s) of the *Sad* excerpt. The excerpt is in a minor mode, opening with an oboe solo accompanied by sustained string chords and harp arpeggios. At around the 15th second (peaking at 18 votes over the time period 15.0 to 15.64s) the number of votes for Calm face begin to decrease and the votes for the Sad face peak. Hence, some participants find the orchestration and arch shaped melody in the oboe more calm than sad, until some additional information is conveyed in the musical signal (at around the 14th second), they remain on Calm. At the 10th second of this excerpt the oboe solo ends, and strings alone play, with cello and violin coming to the fore, with some *portamento* (sliding between pitches). These changes in instrumentation may have provided cues for participants to make the calm to sad shift after a delay of a few seconds [43].

Thus a plausible interpretation of the mixed responses is that participants have different interpretations of the various emotions expressed, *and* the emotion represented by the GUI faces. However, the changes in musical structure are sufficient to explain a change in response. What is important here, and as we have argued elsewhere, is that the difference between emotions is (semantically) small [27], and

that musical features could be modeled to predict the overall shift away from calmness and further toward sadness in this example.

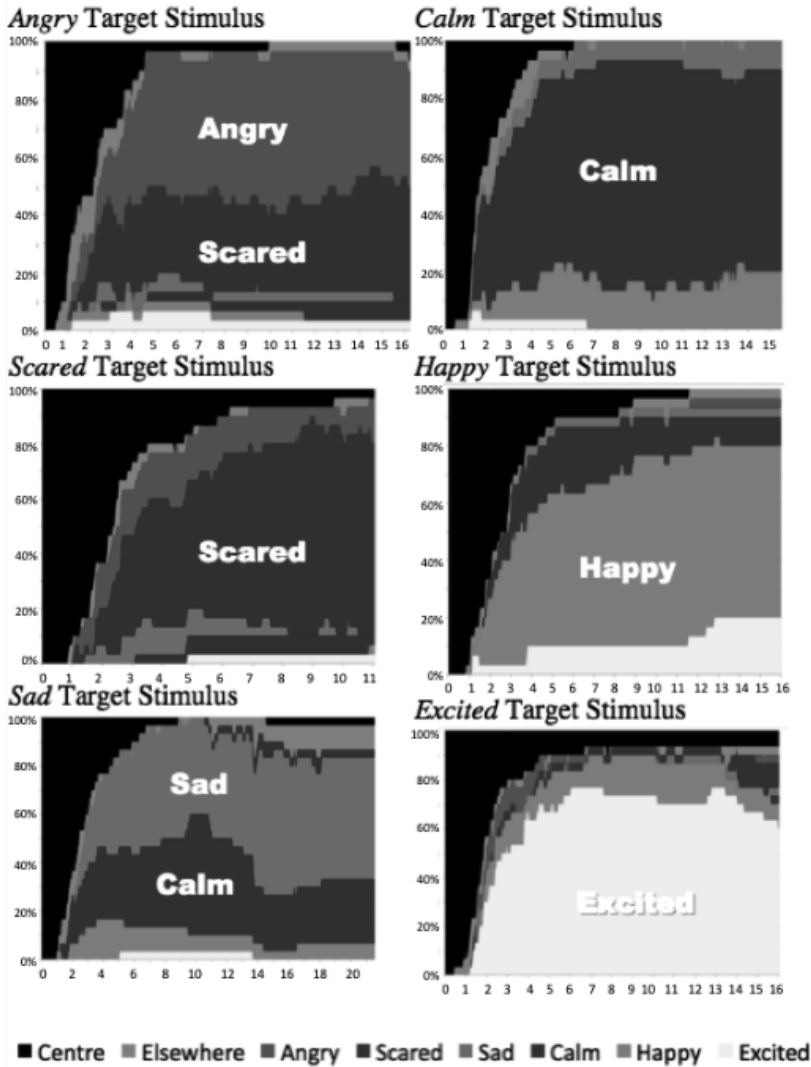


Fig. 3. Time series plots for each stimulus showing stacked frequency of faces selected over time (see **Table 1** for duration on x-axis) for the 30 participants (y-axis), with face selected represented by the colour code shown. Black and grey representing centre of emotion-face-clock (where all participants commence continuous rating task) and anywhere else respectively. Note that the most dominant colour (the most frequently selected face across participants and time) match with the target emotion of the stimulus.

6 Conclusions

In this paper we reported the development and testing of a categorical response interface consisting of a small number of salient emotional expressions upon which participants can rate emotions as a piece of music or other stimulus unfolds. We developed a small set of key emotional expression faces found in music research, and arranged them into a circle such that they were meaningfully positioned in space, and such that they resembled traditional valence-arousal rating scale interfaces (positive emotions toward the right, high arousal emotions toward the top). We called the response space an emotion-face-clock because the faces progressed around a clock in such a way that the expressions changed in a semantically related and plausible manner.

The interface was then tested using particular pieces that expressed the emotions intended to represent each of the six faces. The system was successful in measuring emotional ratings in the manner expected. The post-performance ratings used in an earlier study had profile contours that matched the profile contours of the accumulated summary of continuous response in the new device for all but the *Angry* stimulus. We took this as evidence for the reliability and validity of the emotion-face-clock as a self-report continuous measure of emotion.

Continuous response plots allowed investigation of the ebb and flow of ratings, demonstrating that for some pieces two emotions were dominant (the target *Angry* and target *Sad* excerpts in particular), but that the composition of the emotions changed over time, and that the change could be attributed to changes in musical features.

Further analysis will reveal whether musical features can be used to predict categorical emotions in the same way that valence/arousal models do (for a review, see [4]), or whether six emotion faces is optimal. Given the widespread use of categorical emotions in music metadata [47, 48], the categorical, discrete approach to measuring continuous emotional response is bound to be a fruitful tool for researchers interested in automating emotion in music directly into categorical representations.

Acknowledgments. This research was funded by the Australian Research Council (DP1094998).

References

1. Yang, Y.H., et al., *A regression approach to music emotion recognition*. Audio, Speech, and Language Processing, IEEE Transactions on, 2008. **16**(2): p. 448-457.
2. Schmidt, E.M., D. Turnbull, and Y.E. Kim. *Feature selection for content-based, time-varying musical emotion regression*. in *MIR '10 Proceedings of the international conference on Multimedia information retrieval*. 2010. ACM New York, NY.
3. Korhonen, M.D., D.A. Clausi, and M.E. Jernigan, *Modeling emotional content of music using system identification*. IEEE Transactions on Systems Man and Cybernetics Part B- Cybernetics, 2006. **36**(3): p. 588-599.
4. Schubert, E., *Continuous self-report methods*, in *Handbook of Music and Emotion: Theory, Research, Applications.*, P.N. Juslin and J.A. Sloboda, Editors. 2010, OUP: Oxford. p. 223-253.
5. Madsen, C.K. and W.E. Frederickson, *The experience of musical tension: A replication of Nielsen's research using the continuous response digital interface*. Journal of Music Therapy, 1993. **30**(1): p. 46-63.
6. Nielsen, F.V., *Musical tension and related concepts*, in *The semiotic web '86. An international year-book*, T.A. Sebeok and J. Umiker-Sebeok, Editors. 1987, Mouton de Gruyter: Berlin: .
7. Russell, J.A., *Affective space is bipolar*. Journal of Personality and Social Psychology, 1979. **37**(3): p. 345-356.
8. Russell, J.A., *A circumplex model of affect*. Journal of Social Psychology, 1980. **39**: p. 1161-1178.
9. Krumhansl, C.L., *An exploratory study of musical emotions and psychophysiology*. Canadian Journal of Experimental Psychology, 1997. **51**(4): p. 336-352.
10. Cowie, R., et al., *FEELTRACE: An instrument for recording perceived emotion in real time*, in *Speech and Emotion: Proceedings of the ISCA workshop*, R. Cowie, E. Douglas-Cowie, and M. Schroeder, Editors. 2000, Co. Down.: Newcastle, UK. p. 19-24.
11. Nagel, F., et al., *EMuJoy: Software for continuous measurement of perceived emotions in music*. Behavior Research Methods, 2007. **39**(2): p. 283-290.
12. Schubert, E., *Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space*. Australian Journal of Psychology, 1999. **51**(3): p. 154-165.
13. Schimmack, U. and R. Rainer, *Experiencing activation: Energetic arousal and tense arousal are not mixtures of valence and activation*. Emotion, 2002. **2**(4): p. 412-417.
14. Schimmack, U. and A. Grob, *Dimensional models of core affect: A quantitative comparison by means of structural equation modeling*. European Journal Of Personality, 2000. **14**(4): p. 325-345.
15. Wundt, W., *Grundzüge der physiologischen Psychologie*. 1905, Leipzig: Engelmann.
16. Plutchik, R., *The emotions: Facts, theories and a new model*. 1962, New York: Random House. 204.
17. Russell, J.A. and A. Mehrabian, *Evidence for a 3-factor theory of emotions*. Journal of Research in Personality, 1977. **11**(3): p. 273-294.
18. Barrett, L.F. and T.D. Wager, *The Structure of Emotion: Evidence From Neuroimaging Studies*. Current Directions in Psychological Science, 2006. **15**(2): p. 79-83.
19. Barrett, L.F., *Discrete emotions or dimensions? The role of valence focus and arousal focus*. Cognition & Emotion, 1998. **12**(4): p. 579-599.
20. Lewis, M., J.M. Haviland-Jones, and L.F. Barrett, eds. *Handbook of emotions* (3rd ed.). 2008, The Guilford Press: New York, NY.
21. Izard, C.E., *The psychology of emotions*. 1991, NY: Plenum Press.
22. Izard, C.E., *Organizational and motivational functions of discrete emotions*, in *Handbook of emotions*, M. Lewis and J.M. Haviland, Editors. 1993, The Guilford Press: New York, NY. p. 631-641.

23. Namba, S., et al., *Assessment of musical performance by using the method of continuous judgment by selected description*. Music Perception, 1991. **8**(3): p. 251-275.
24. Juslin, P.N. and P. Laukka, *Communication of emotions in vocal expression and music performance: Different channels, same code?* Psychological Bulletin, 2003. **129**(5): p. 770-814.
25. Laukka, P., A. Gabrielsson, and P.N. Juslin, *Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion*. International Journal of Psychology, 2000. **35**(3-4): p. 288-288.
26. Juslin, P.N., *Communicating emotion in music performance: A review and a theoretical framework.*, in *Music and emotion: Theory and research*, P.N. Juslin and J.A. Sloboda, Editors. 2001, Oxford University Press: London. p. 309-337.
27. Schubert, E., et al. *Sonification of Emotion I: Film Music*. in *The 17th International Conference on Auditory Display (ICAD-2011)*. 2011. Budapest, Hungary: International Community for Auditory Display (ICAD).
28. Hevner, K., *Expression in music: a discussion of experimental studies and theories*. Psychological Review, 1935. **42**: p. 187-204.
29. Hevner, K., *The affective character of the major and minor modes in music*. American Journal of Psychology, 1935. **47**: p. 103-118.
30. Hevner, K., *Experimental studies of the elements of expression in music*. American Journal of Psychology, 1936. **48**: p. 246-268.
31. Hevner, K., *The affective value of pitch and tempo in music*. American Journal of Psychology 49 1937, 621-630 Univ of Illinois Press, US, 1937.
32. Rigg, M.G., *The mood effects of music: A comparison of data from four investigators*. The journal of psychology, 1964. **58**(2): p. 427-438.
33. Han, B., et al. *SMERS: Music emotion recognition using support vector regression*. in *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*. 2009. Kobe International Conference Center, Kobe, Japan: Kobe International Conference Center, Kobe, Japan, October 26-30, 2009.
34. Dimberg, U. and M. Thunberg, *Rapid facial reactions to emotional facial expressions*. Scandinavian Journal of Psychology, 1998. **39**(1): p. 39-45.
35. Britton, J.C., et al., *Facial expressions and complex IAPS pictures: common and differential networks*. Neuroimage, 2006. **31**(2): p. 906-919.
36. Waller, B.M., J.J. Cray Jr, and A.M. Burrows, *Selection for universal facial emotion*. Emotion, 2008. **8**(3): p. 435.
37. Ekman, P., *Facial expression and emotion*. American Psychologist, 1993. **48**(4): p. 384-392.
38. Lang, P.J., *Behavioral treatment and bio-behavioral assessment: Computer applications*, in *Technology in Mental Health Care Delivery Systems*, J.B. Sidowski, J.H. Johnson, and T.A. Williams, Editors. 1980, Ablex: Norwood, NJ. p. 119-137.
39. Bradley, M.M. and P.J. Lang, *Measuring Emotion - The Self-Assessment Mannequin And The Semantic Differential*. Journal Of Behavior Therapy And Experimental Psychiatry, 1994. **25**(1): p. 49-59.
40. Ekman, P. and E.L. Rosenberg, eds. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Series in affective science. 1997, Oxford University Press.: London.
41. Eerola, T. and J.K. Vuoskoski, *A comparison of the discrete and dimensional models of emotion in music*. Psychology of Music, 2011. **39**(1): p. 18-49.
42. Schubert, E. and G.E. McPherson, *The perception of emotion in music*, in *The child as musician: A handbook of musical development* G.E. McPherson, Editor. 2006, Oxford University Press: Oxford. p. 193-212.
43. Schubert, E., *Continuous measurement of self-report emotional response to music*, in *Music and emotion: Theory and research*, P.N. Juslin and J.A. Sloboda, Editors. 2001, Oxford University Press: Oxford. p. 393-414.

44. Schubert, E., *Reliability issues regarding the beginning, middle and end of continuous emotion ratings to music*. Psychology of Music, 2012.
45. Bachorik, J.P., et al., *Emotion in motion: Investigating the time-course of emotional judgments of musical stimuli*. Music Perception, 2009. **26**(4): p. 355-364.
46. Schubert, E. and W. Dunsmuir, *Regression modelling continuous data in music psychology.* , in *Music, Mind, and Science*, S.W. Yi, Editor. 1999, Seoul National University: Seoul. p. 298-352.
47. Trohidis, K., et al. *Multilabel classification of music into emotions*. in *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*. 2008. Philadelphia, PA.
48. Levy, M. and M. Sandler. *A semantic space for music derived from social tags*. in *In Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*. 2007. Vienna, Austria.