A Comparison between Continuous Categorical Emotion Responses and Stimulus Loudness Parameters

Sam Ferguson[*], Emery Schubert[†], Doheon Lee[†], Densil Cabrera [‡] and Gary E. McPherson[§]

[*]*Creativity and Cognition Studios, Faculty of Engineering and IT, University of Technology*
Email: samuel.ferguson@uts.edu.au
[†] *School of Arts and Letters, Faculty of Arts, University of New South Wales*
Email: e.schubert@unsw.edu.au
[‡] *Faculty of Architecture, Design and Planning, The University of Sydney*
Email: densil.cabrera@sydney.edu.au
[§]*Melbourne Conservatorium of Music, The University of Melbourne*
Email: g.mcpherson@unimelb.edu.au

Abstract—This paper investigates the use of psychoacoustic loudness analysis as a method for determining the likely emotional responses of listeners to musical excerpts. 19 excerpts of music were presented to 86 participants (7 randomly chosen excerpts per participant) who were asked to rate the emotion category using the emotion-clock-face continuous response interface. The same excerpts were analysed with a loudness model, and time series results were summarised as both loudness median and standard deviation. Comparisons indicate that the median and standard deviation of loudness plays an important role in determining the emotion category responses.

Keywords-Emotion; loudness; music.

I. INTRODUCTION

A. Continuous Emotion Response

Continuous emotion response methods allow the capture of temporally related self-report information from subjects as opposed to requesting a post-presentation overall rating for a stimulus. Many previous studies have used a two-dimensional response surface for users to continuously describe their emotion, in terms of valence (happy/sad) and arousal (sleepy/excited) [1].

Using these responses it is possible to describe the contribution of various musical features to the emotional characteristics of the music [2].

Recent research has developed methods for continuous responses that employ continuous categorical responses [3]. This means that the response is made, not in terms of underlying emotion dimensions (valence, arousal), but in response to specific, possibly more well-understood emotion categories (sad, happy, angry etc.) (see Juslin & Laukka for a discussion of categorical emotion responses [4]).

B. Emotions in Music and Sound

For the purpose of investigating loudness and its effect on emotion response, we first need to establish methods used in studies that rate multiple emotions in response to musical stimuli that contain some type of putative emotion content. Some studies use stimuli that are drawn from general musical recordings of various pieces that are judged to express a range of putative emotions (usually by expert listeners). Other studies use a single musical piece (often a simple popular melody such as *'Oh, when the saints'*) with a performer directed to communicate the putative emotions. These two approaches could be characterised as (1) an ecological approach, with stimuli drawn from an ecologically valid sample of existing musical repertoire, or (2) a directed approach, where the emotional cues are directed to be applied to an arbitrary stimulus, and the resulting stimuli are less ecologically common. By ecologically common, we mean that the particular putative emotion applied to the particular musical melody may not be a common occurrence in a given listener's experience - for example, a piece that has intrinsically happy features being performed to sound as sad as possible. A 'directed' stimulus, however, will be more rigorously controlled (all excerpts will be roughly the same duration, range, instrumentation, form etc. and should only vary in the parameters used to express emotions).

Gabrielsson and Juslin have conducted significant amounts of research using the 'directed' approach on communicating emotion using artificial stimuli (a single stimulus altered in emotion content to create many stimuli of different characters by a performer) [5]. Juslin discussed the way in which guitarists used musical performance characteristics to perform particular musical compositions to express particular emotions [6].

Juslin has also presented data summarising the importance of emotion based on questionnaire studies [4]. More recently, some researchers (sucha as Zentner et al. [7] and Eerola [8]) have conducted a large set of studies to define emotion categories carefully. They used factor analysis of emotion terms to define 9 different emotion categories (wonder, transcendence, tenderness, nostalgia, peacefulness, power, joyful activation, tension, sadness), which were categorised into 3 main classes of emotion (sublimity, vitality, unease).

Juslin and Laukka investigated the correspondence between vocal expression and musical performance, surveying a number of studies on each category to determine the consensus on the 'code' used for communicating emotions [9]. They found that there was a close correspondence between the cues used for emotion communication in speech and music. Juslin integrates previous research into communication through the GERMS model [10].

A further complication in emotion and music research is whether the emotion is perceived in the music or felt in the listener [11], the locus of emotion. Investigations of this distinction have found broad similarity between the two loci (perceived and felt), although differences exist and are not insignificant [12].

C. Loudness Modelling

Loudness is the perception of how loud a sound is, as opposed to a physical quantity of energy. Loudness is often discussed in

relation to the auditory periphery, in terms of the sound transforms that occur in the outer and middle ear, and the way that the sound is analysed and integrated in the inner ear [13]. The frequency content, duration and bandwidth all affect loudness, as well as (of course) the sound pressure of the sound.

A loudness model translates a recorded sound into a estimation of a loudness percept, by modelling many auditory processes such as the outer, middle and inner frequency responses and the process of basilar membrane excitation and masking. These processes can result in different values for loudness measurements for sounds that possess exactly the same sound pressure level, due to differences in the temporal and spectral distribution of the energy. These differences can, for instance, lower the loudness of narrow-band or short sounds when compared with longer sounds or broad band sounds. Loudness models incorporate numerous computation stages and a description of the entire process is outside of the scope of this paper, however, the model used is available for use within the free software 'Psysound3' [14], which was used for the computation.

The units used to describe loudness are sones - a ratio scale of loudness - where sound of 10 sones is expected to be perceived to be twice as loud as sound of 5 sones.

In previous research, we investigated continuous loudness responses to musical excerpts [15], and showed that loudness responses are quite robust, even when incorporating the complexity of musical stimuli, as opposed to stationary industrial or background noise. Loudness is strongly related to emotion responses, and previous research has shown the crucial role that loudness plays in music emotion communication, and in particular, the relationship with arousal. Eerola and Vuoskoski [16] for instance, surveyed the literature and found that through both production studies (where musicians are asked to 'produce' an emotional performance), and analytical approaches, particular emotions can be related to parameters of loudness [17].

II. METHOD

A. Participants

Eighty-six participants took part in the subjective experiments. There were 43 female participants and 49 male participants, while 4 participants did not report their gender. Their ages ranged from 18 to 50 years old, with a median age of 20 years old and a standard deviation of 4.89. One participant also did not report their age. Sixty-five participants among the 86 participants had some form of musical training ranging from 1 year to 18 years. A median of their musical training period is 4.5 years with the standard deviation of 5.27 years.

B. Stimuli

Stimuli were selected to represent each of six target emotions; Excited, Happy, Calm, Sad, Scared, and Angry - excerpts that were ambiguous or confusing were avoided. The selection of music in this study was restricted to film music of a mostly orchestral nature that did not have lyrics. Each segment was edited to start and end as close to a phrase boundary as possible, which was to express a single emotion. Film music has an advantage over many other musical styles for our purposes because it is explicitly programmatic, and often lacks both a long-term structure (eg. symphony form), and lyrical content. Three excerpts from each of the 6 target emotions, resulting in 18 excerpts were selected ranging from 7 to 27 seconds

in duration, as shown in Table I. An additional Excited excerpt was selected and presented to all participants. With this extra excerpt, the use of 7 stimuli would reduce the likelihood that the participant matches each of the 6 stimuli with the 6 faces and uses a process of elimination strategy to select some responses, rather than responding to properties of the stimulus. Stimuli were pilot tested for their likelihood of expressing a range of putative emotions [3], [18], [19].

Using stimuli that are from popular films raises the question whether the listeners may respond to the visual and emotional content of the film rather than the musical content. We expected that although some participants may have seen the film before, they would probably not be likely to associate a particular excerpt with a particular visual scene. Nevertheless, we asked participants if they found the music to be familiar, with response options of 'No', 'Not Sure', 'Heard It Somewhere' or 'Yes'. Of the 602 stimulus presentations (86 participants multiplied by 7 stimuli), only 47 of the responses were 'yes' (7%). These positive familiarity responses were generally well spread across the stimuli with no stimulus recording more than one third of respondents stating familiarity to be 'Yes'. It is therefore relatively unlikely that participants were responding to a conscious recollection of the film's content in any significant manner.

For the loudness calculations, an association between the digital signal and the likely listening must be decided. This is an arbitrary measure of course, as a particular excerpt may be listened to at various levels, even though the level was controlled for the participants in this particular listening test. However, for the purposes of *relative* comparison between stimuli, one stimulus was chosen as a reference , and all the other stimuli had the same gain applied. LA_{eq} (which is the averaged A-weighted sound pressure level over the given length of stimuli) for the excerpt 'Angry4_ToyStory3' was set to 75 dBA and all the other excerpts were adjusted relatively, so that the relative level differences between the music excerpts was retained. The participants all experienced this relative difference, as the listening level was controlled for the experiment. This calibration was done entirely within the digital realm, after the experiment was completed, and is only a relative adjustment for use as an approximation of a listening level close to that which the excerpts were listened to at.

C. Software Interface

The emotion-face-clock interface was developed using Max/MSP software [1]. Mouse movements in the x and y axis were sampled and stored in an audio buffer that was synchronized with the musical material which had a sampling rate of 44.1kHz. Although the position of the mouse was captured with an extreme resolution, they mostly consisted of constant values, with short periods of movement – the participants would often keep their fingers close to the mouse, but without movement, until they had decided to move, in which case the movement would be rapid, and then the movement would stop on the planned position, until the next rapid movement. Many made movements short movements between faces when they changed their minds, but continuous movement throughout the stimulus was rare. These mouse movements were then downsampled to 25 Hz before being converted into one of eight states: the centre, one of the six emotions represented by

[1] http://www.cycling74.com

Table I

STIMULI USED IN THE EXPERIMENT. THREE EXCERPTS INTENDED TO
SONIFY EACH EMOTION WERE SELECTED, EXCEPT FOR EXCITEMENT,
WHICH HAS FOUR (SEE TEXT FOR RATIONALE).

| Stimulus	Film: Track Name	Start Time	Dur. (s)
Angry1	Up: 52 Chachki Pickup	00"53	17
Angry4	Toy Story 3: Come to Papa	00"38	20
Angry5	Toy Story 3: Cowboy!	03"36	19
Calm1	Finding Nemo: Wow	00"22	16
Calm2	Finding Nemo: Field Trip	00"00	21
Calm3	Finding Nemo: The Turtle Lope	00"48	20
Excited1	Toy Story: Infinity and Beyond	00"15	16
Excited4	Cars: The Piston Cup	00"05	7
Excited5	Cars: The Big Race	01"11	18
Excited3	Up: Memories Can Weigh You Down	00"26	21
Happy1	Cars: McQueen and Sally	00"04	16
Happy2	Monsters Inc.: Monsters, Inc.	00"06	15
Happy3	Up: Up with Titles	00"00	10
Sad1	Cars: Goodbye	00"00	27
Sad6	Toy Story 3: You Got Lucky	01"00	21
Sad7	Toy Story 3: So Long	02"20	23
Scared1	Cars: McQueen's Lost	00"55	11
Scared2	Up: The Explorer Motel	00"34	19
Scared4	Up: Giving Muntz the Bird	00"54	14
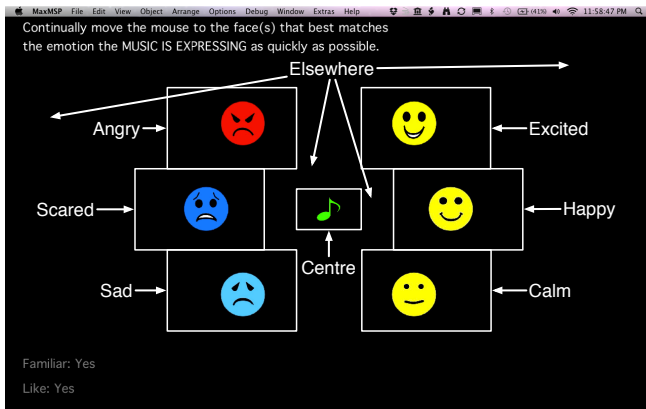
Figure 1. The emotion-clock-face interactive response instrument.

schematic faces, or 'elsewhere' (see Figure 1 for the arrangement of these spatial categories).

The facial expressions moving around the clock in a clockwise direction were Excited, Happy, Calm, Sad, Scared and Angry (Figure 1). Note that the verbal labels for the faces are for the convenience of the researcher, and are not necessarily the same as those used by participants. More importantly, the expressions progressed sequentially around the clock such that related emotions were closer together than distant emotions, when considered from the point of view of Hevner's adjective circle [20] or Russell's circumplex model [21]. The quality of these labels were tested against participant data using the explicit labelling of the same stimuli in an earlier study [18].

D. Procedure

Participants wore a set of headphones (Sennheiser HD280), and were positioned in front of an Apple MacBook Pro computer. The level of the headphones was controlled by choosing the same gain setting on the Apple computer, which had a stepped gain control.

The stimuli presentation order was randomised for each presentation. A subset of 6 musical excerpts were selected automatically and at random from the predetermined set of 18 excerpts. One excerpt from the pool of 3 excerpts for each of the six emotions was picked at random, meaning each listener heard all of the emotion categories. In addition one extra excerpt (*Excited1_ToyStory*) was presented to all participants, for a total of of 7 excerpts.

Firstly, some tasks were presented that familiarised the user with the emotion-face-clock interface – listening and responding to spoken words using the interface. The participant was then instructed to click the green icon (quaver) button to commence listening to the musical excerpts, and was asked to 'Continually move the mouse to the face(s) that best matches the emotion the MUSIC IS EXPRESSING as quickly as possible'. The participants were allowed to move the mouse pointer from one face icon to another at any time continuously, for the period of playback and for up to 10 seconds after playback had finished. When the participant moved the mouse over one of the faces, the icon of the face was highlighted (magnified by 10%) to provide feedback. On completion of the continuous response to the musical excerpt, the participant was asked to perform several other short rating tasks. The next stimulus was prepared by pressing the right arrow and then the process began again until all the 7 stimuli were exhausted.

The resultant data encodes the mouse-pointing gestures of the respondent in terms of both: (1) the face pointed to, and (2) the time that of the pointing gesture. Time values were all referenced against the beginning of the audio recording. Due to the green play button being positioned at the centre of the screen, the starting position was controlled and all the responses share exactly the same time axis resolution.

E. Digital Psychoacoustic and Acoustic Analyses

The analyses of the stimuli were undertaken using the Time-varying Loudness Model by Glasberg and Moore [22], which is an update of the work by Moore, Glasberg and Baer [23], as implemented in Psysound3 [14]. In previous research we have found this model to be of a high accuracy with respect to subjective results (see Ferguson et al for details [24]). Rennies et al [25] has discussed the differences and similarities between this model and other models in more detail. Glasberg and Moore's Time-varying Loudness Model has long-term and short-term loudness outputs, and in this study we chose the short-term loudness output for the loudness calculations.

The loudness time-series were downsampled to a temporal resolution of 25 Hz, to match the downsampled continuous emotion responses. This sampling rate is approximately the frame rate of a film (25 frames per second) and under the refresh rate of a computer screen (up to 120 frames per second, but totally dependent on computer), and therefore is likely to capture the movement of the mouse-pointer adequately. The resolution of 25 Hz is much higher than the necessary sampling rate for emotion responses (2 Hz) [26], and considering the temporal smoothing of the model, the resolution of 40ms per sample is unlikely to be a limitation. The loudness timeseries were obtained and lowpass filtered using Matlab's `resample` command.

III. RESULTS

In this section we will investigate and compare the subjective emotion responses and the objective psychoacoustic analyses.
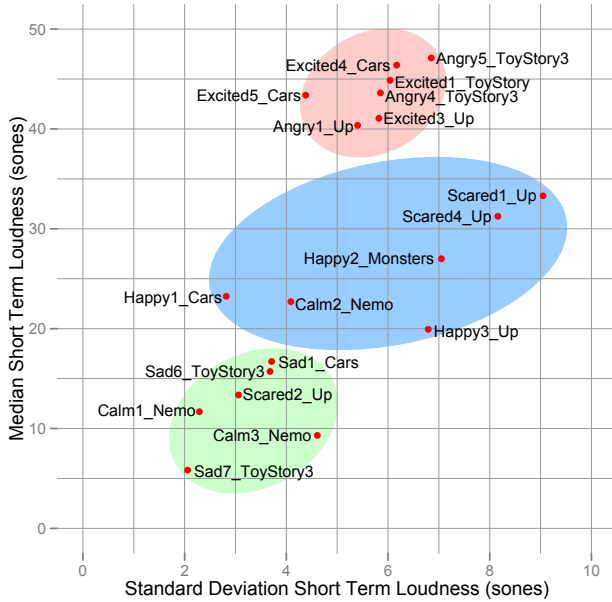
Figure 2. Comparing loudness median and standard deviation for several excerpts. Excited and Angry excerpts fall into one group (red), Happy and Scared another (blue), and calm and sad another again (green). Particular excerpts (Calm2_Nemo and Scared2_Up) do not follow these groupings, but further inspection shows these excerpts have particular characteristics (see text).

A. Excerpt Objective Parameter Results

Some clear groupings appear when we investigate the objective loudness results on their own. Loudness median and standard deviation appear to be an important component of the stimulus, at least within this small set of musical excerpts.

In Figure 2 we present the loudness results for the 19 excerpts as a scatter plot. These excerpts were carefully selected by expert listeners (but not using any acoustic analysis) so that they only expressed one particular emotion, but it can be seen that there is still significant variance between them. A defined grouping is seen of the angry and excited excerpts, with a mid-scale standard deviation of loudness and a high median loudness. Scared and happy excerpts showed mid-scale median loudness and sometimes high standard deviations also, although this was a less clear grouping. Sad and calm excerpts showed quite consistent grouping in the lower corner of the graph with low values of both loudness median and standard deviation. This is consistent with a lower median loudness, as the variance would therefore be likely to be in a smaller range.

B. Excerpt Emotion Response Results

There were clear patterns in regard to reported emotions (see Table II and Figure 3).

In many cases the emotion that was chosen the most was the same as the intended emotion of the excerpt, except for *Angry1_Up*, *Angry5_ToyStory3*, *Calm2_Nemo* and *Sad1_Cars* – 4 of the 19 excerpts presented. Angry and scared faces were reported for both the angry and the scared excerpts. Calm and Happy emotions were reported for calm excerpts. For happy excerpts and excited excerpts

Excerpt	Excited	Happy	Calm	Sad	Scared	Angry
Excited1_ToyStory	**0.71**	**0.22**	0.01	0.01	0.04	0.01
Excited3_Up	**0.60**	**0.32**	0.01	0.00	0.03	0.04
Excited4_Cars	**0.81**	**0.10**	0.06	0.00	0.03	0.00
Excited5_Cars	**0.44**	**0.32**	0.03	0.01	0.13	0.07
Happy1_CarsAUD	0.10	**0.63**	**0.25**	0.03	0.00	0.00
Happy2_Monsters	0.19	**0.74**	0.07	0.00	0.00	0.00
Happy3_Up	0.11	**0.85**	0.03	0.00	0.00	0.00
Calm1_Nemo	0.00	**0.18**	**0.77**	0.05	0.00	0.00
Calm2_Nemo	0.08	**0.49**	**0.38**	0.02	0.04	0.00
Calm3_Nemo	0.01	**0.11**	**0.82**	0.05	0.00	0.00
Sad1_Cars	0.04	0.15	**0.51**	**0.31**	0.00	0.00
Sad6_ToyStory3	0.00	0.10	**0.13**	**0.72**	0.04	0.01
Sad7_ToyStory3	0.00	0.07	**0.37**	**0.56**	0.00	0.00
Scared1_Up	0.03	0.00	0.03	0.02	**0.73**	**0.19**
Scared2_Up	0.00	0.00	0.00	0.03	**0.89**	**0.08**
Scared4_Up	0.01	0.00	0.04	0.07	**0.59**	**0.28**
Angry1_Up	0.13	0.03	0.02	0.00	**0.48**	**0.34**
Angry4_ToyStory3	0.04	0.03	0.00	0.04	**0.38**	**0.51**
Angry5_ToyStory3	0.07	0.03	0.00	0.01	**0.47**	**0.42**

the happy and excited faces were reported. For the sad excerpts, sad and calm were reported.

C. Comparing Emotion Response Against Loudness Results

Comparing the emotion responses against loudness results shows reasonable groupings, but also some interesting outlier excerpts that break these groupings. Importantly, there seems to be systematic relationships between emotion responses and loudness parameters.

Most of the sad and calm excerpts appear to be grouped together with low loudness medians and standard deviations. Generally, they communicate their emotion as predicted, and are reported consistently. The happy and scared excerpts appear to have a mid-range loudness median, but be often characterised by a high-range loudness standard deviation. The grouping is not as strong though, as one happy excerpt appeared to be lower in loudness standard deviation (*Happy1_Cars*), and another excerpt has both a lower loudness median and standard deviation (*Scared2_Up*).

Finally, the angry and excited excerpts are very closely grouped, but although the excited excerpts are consistently reported as excited, there appears to be a between the scared and angry excerpts. There is only a small amount of complementary confusion, however, for a scared excerpt (*Scared4_Up*) showing both scared and angry emotion reports.

Calm2_Nemo is the only calm excerpt response where the calm emotion was not reported, in favour of the happy emotion. We can also see that this excerpts has a higher median loudness than the other calm excerpts 2. It seems that this excerpt is 'pushed' by the higher median loudness towards the happy excerpt grouping, and the emotion responses reflect this shift. *Happy1_Cars*, although recognised adequately, has a higher rating for Calm than it does for Excited (unlike *Happy2_Monsters* and *Happy3_Up*). It also shows a lower value for loudness standard deviation (ie. less 'bouncy'). *Scared2_Up* is one example of an excerpt that was well-recognised, but showed loudness parameters that were not similar to the rest of
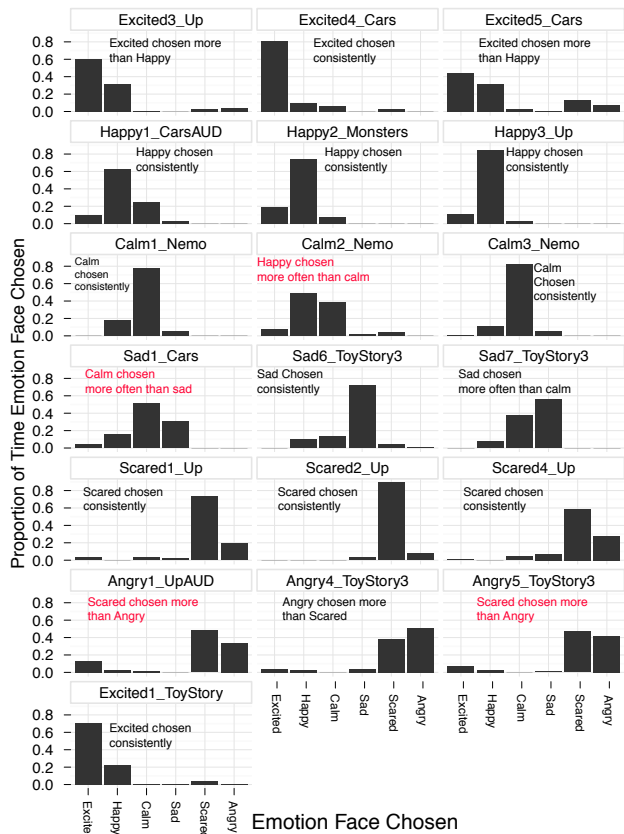
Figure 3. Emotion report results presented as a bar chart.

the scared or happy excerpts. This example is the only one where the loudness parameters and the emotion responses appear not to match in a figurative way. This may be due to the factor of musical expectation - on listening to both *Scared4_Up* and *Scared1_Up* they both contain louder sections (i.e. a 'scary' section), while the *Scared2_Up* excerpt only presents the anticipation section before a loud section, accounting for the lower values of both loudness parameters.

IV. DISCUSSION

This study has compared parameters of loudness extracted from several musical excerpts against the continuous emotion responses captured from respondents who listened to them. The purpose of this comparison is to test computational models of detecting emotion against subjective emotion responses based on emotion categories. In most cases it seems that using loudness parameters gives good performance for clustering excerpts of music that have similar emotion characteristics.

Nevertheless, with respect to loudness level and deviation, there are some complexities. The clustering tends to result in groupings of excerpts that contain one negative and one positive valence emotion. For instance, angry excerpts and excited excerpts are grouped together, happy excerpts and scared excerpts are generally grouped together (although the grouping is more fragile), and calm and sad excerpts are grouped together. The harmonic content of the excerpts

is perhaps the cue that can be used to distinguish the emotion of these excerpts, and indeed for most of these groupings there is little confusion between the two sets of emotion evident from the emotion responses. It could be imagined that a measure of harmony or dissonance could distinguish these groups along a third dimension. However, our results demonstrate that the standard deviation and loudness are able to distinguish the sometimes confused emotions of fear (scared) and anger. We believe this is the first time this measure has been used in a continuous response context. Angry and Scared excerpts were differentiated by their loudness in both median loudness and standard deviation, and so there appeared to be some important differences in the excerpts that were not reflected clearly in the results obtained from the respondents. It may be that the selection of a scared face may have multiple meanings - referring to both the induction of the scared emotion as a *response* to an angry emotion expression from the excerpt, and a scared response to an excerpt expressing the emotion scared. This may account for the confusion in response to the angry excerpts, and indeed, it is the scared excerpt with the lowest median loudness that is recognised most consistently as scared.

The use of simple loudness parameters to model the emotion content of excerpts of music that are short and strongly representative of a single emotion has been found to be reasonably robust. Generally, the excerpts formed specific groupings that tended to be well separated from the other emotion categories within the same arousal category (eg. sad/calm, scared/happy, angry/excited). Where the excerpt's emotion label did not match well to the loudness parameters we extracted we also found a change in the emotion response that seemed to be generally systematic. An exception to this pattern was the *Scared1_Up* excerpt, which had a consistent scared response, but which showed a loudness median and standard deviation much lower than the other excerpts – we assume this may be due to the lack of the 'fright' in the final part of the excerpt, that both the other scared excerpts included. It appears possible that the listener may be able to anticipate this factor in the excerpt, and respond accordingly.

V. CONCLUSION AND FURTHER RESEARCH

This research investigated loudness parameters in 19 musical excerpts in relation to the emotion expressed by those musical excerpts. The excerpts were analysed with a loudness model, and time series results were summarised as both loudness median and standard deviation. The emotion responses indicate that the median and standard deviation of the excerpt's loudness plays an important role in determining the emotion category responses.

The innovation in this study is that:

1) it confirms the literature regarding loudness standard deviation, in comparison to self-reported variation in dynamics
2) it provides a more subtle shading with regard to the excerpts selected - Scared excerpts have very high standard deviation, while Angry (and Excited and Happy) have moderate amounts of standard deviation. Angry and Scared (and the other emotions) are differentiated by loudness, along the arousal dimension lines reported in the literature (again, as summarised by Gabrielsson & Lindstrom [17]).
3) although other studies have shown effects of standard deviation in loudness on emotion, this study is the first (we believe) to do so using continuous ratings.

More work needs to investigate whether temporal responses can be incorporated into this simple model, and whether a running emotion categorisation is possible and could be based on a parameter such as loudness. Defining the way in which a longer sample of music would be analysed, without the luxury of editing the sample to a single representative phrase of music, would add significant complexity. Also, given that changes in loudness parameters can affect the emotion responses in listeners, future research could also focus on the extent to which emotion responses can be manipulated by manipulating the loudness parameters within the stimuli. This reversed approach is likely to offer many challenges, but does hold out the possibility of manipulating emotion responses to music using parameters of emotion rather than parameters of sound.

REFERENCES

[1] E. Schubert, "Continuous self-report methods." in *Handbook of Music and Emotion: Theory, Research, Applications.*, P. N. Juslin and J. A. Sloboda, Eds. Oxford: OUP, 2010, pp. 223–253.

[2] ——, "Modelling perceived emotion with continuous musical features," *Music Perception*, vol. 21, no. 4, pp. 561–585, 2004.

[3] E. Schubert, S. Ferguson, N. Farrar, D. Taylor, and G. E. McPherson, "Continuous response to music using discrete emotion faces," in *9th International Symposium on Computer Music Modelling and Retrieval*, London, UK., 2012.

[4] P. N. Juslin and P. Laukka, "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening," *Journal Of New Music Research*, vol. 33, no. 3, pp. 217–238, 2004.

[5] A. Gabrielsson and P. N. Juslin, "Emotional expression in music performance: Between the performer's intention and the listener's experience," *Psychology of Music*, vol. 24, no. 1, pp. 68–91, 1996.

[6] P. N. Juslin, "Cue utilization in communication of emotion in music performance: Relating performance to perception," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26, no. 6, pp. 1797–1813, 2000.

[7] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: Characterization, classification, and measurement," *Emotion*, vol. 8, no. 4, pp. 494–521, 2008.

[8] T. Eerola, "Modeling listeners' emotional response to music," *Topics in Cognitive Science*, vol. 4, no. 4, pp. 607–24, 2012.

[9] P. N. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance: Different channels, same code?" *Psychological Bulletin*, vol. 129, no. 5, pp. 770–814, 2003.

[10] P. N. Juslin, "Five facets of musical expression: A psychologist's perspective on music performance," *Psychology of Music*, vol. 31, no. 3, pp. 273–302, 2003.

[11] A. Gabrielsson, "Emotion perceived and emotion felt: Same or different?" *Musicae Scientiae*, vol. 2001-2002, no. Spec. Issue, pp. 123–147, 2002.

[12] P. Evans and E. Schubert, "Relationships between expressed and felt emotions in music," *Musicae Scientiae*, vol. 12, no. 1, pp. 75–99, 2008.

[13] B. C. J. Moore, *An Introduction to the Psychology of Hearing*. San Diego, California; London: Academic Press, 1997.

[14] D. Cabrera, S. Ferguson, and E. Schubert, "Psysound3: Software for acoustical and psychoacoustical analysis of sound recordings," in *Proceedings of the 13th International Conference on Auditory Display*, Montreal, Canada, 2007.

[15] S. Ferguson, E. Schubert, and R. T. Dean, "Continuous subjective loudness responses to reversals and inversions of a sound recording of an orchestral excerpt," *Musicae Scientiae*, vol. 15, no. 3, pp. 387–401, 2011.

[16] T. Eerola and J. K. Vuoskoski, "A review of music and emotion studies: Approaches, emotion models and stimuli," *Music Perception*, vol. 30, no. 3, pp. 307–340, 2013.

[17] A. Gabrielsson and E. Lindstrm, "The role of structure in the musical expression of emotions," in *Handbook of Music and Emotion*, P. Juslin and J. A. Sloboda, Eds. Oxford: OUP, 2010, pp. 367–400.

[18] E. Schubert, S. Ferguson, N. Farrar, and G. E. McPherson, "Sonification of emotion i: Film music," in *The 17th International Conference on Auditory Display (ICAD-2011)*, Budapest, Hungary, 2011.

[19] S. Ferguson, D. Taylor, E. Schubert, N. Farrar, and G. E. McPherson, "Emotion locus in continuous emotional responses to music," in *'Power of Music' - the 34th National Conference of the Musicological Society of Australia, in conjunction with the 2nd International Conference on Music and Emotion*. Perth, Australia: The University of Western Australia, 2011.

[20] K. Hevner, "Experimental studies of the elements of expression in music," *American Journal of Psychology*, vol. 48, pp. 246–268, 1936.

[21] J. A. Russell, "A circumplex model of affect," *Journal of Social Psychology*, vol. 39, pp. 1161–1178, 1980.

[22] B. R. Glasberg and B. C. J. Moore, "A model of loudness applicable to time-varying sounds," *Journal of the Audio Engineering Society*, vol. 50, no. 5, pp. 331–342, 2002.

[23] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *Journal of the Audio Engineering Society*, vol. 45, no. 4, pp. 224–240, 1997.

[24] S. Ferguson, E. Schubert, and D. Cabrera, "Comparing continuous subjective loudness responses and computational models of loudness for temporally varying sounds," in *129th AES Convention*, San Francisco, USA, 2010.

[25] J. Rennies, J. L. Verhey, and H. Fastl, "Comparison of loudness models for time-varying sounds," *Acta Acustica united with Acustica*, vol. 96, pp. 383–396, 2010.

[26] E. Schubert and W. Dunsmuir, "Regression modelling continuous data in music psychology." in *Music, Mind, and Science*, S. Yi, Ed. Seoul: Seoul National University, 1999, pp. 298–352.