

Music emotion classification by audio signal analysis: analysis of self-selected music during game play

Knox, Don; Cassidy, Gianna; Beveridge, Scott; Macdonald, Raymond A.R.

Published in:

Proceedings of the 10th International Conference on Music Perception and Cognition

Publication date:

2008

[Link to publication in ResearchOnline](#)

Citation for published version (Harvard):

Knox, D, Cassidy, G, Beveridge, S & Macdonald, RAR 2008, Music emotion classification by audio signal analysis: analysis of self-selected music during game play. in *Proceedings of the 10th International Conference on Music Perception and Cognition*. pp. 581-587. <<http://www.jsmpc.info/archives/icmpc10archive/index.html>>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please view our takedown policy at <https://edshare.gcu.ac.uk/id/eprint/5179> for details of how to contact us.

Music emotion classification by audio signal analysis: Analysis of self-selected music during game play

Don Knox,^{*1} Gianna Cassidy,^{*2} Scott Beveridge^{*3} Raymond MacDonald^{#4}

^{*}*School of Engineering and Computing, Glasgow Caledonian University, Scotland*

[#]*Department of Psychology, Glasgow Caledonian University, Scotland*

¹d.knox@gcal.ac.uk, ²gianna.cassidy@gcal.ac.uk, ³scott.beveridge@gcal.ac.uk,
⁴raymond.macdonald@gcal.ac.uk

ABSTRACT

Music emotion classification algorithms seek to classify music files automatically by means of audio signal analysis. An overview of these methods is given, and an emotion classification algorithm is applied to the preferred music choices made by test subjects during a game play experiment. Results from the experiment are presented, in which test subjects were exposed to 3 sound conditions: preferred music, game soundtrack and experimenter-selected music. Obtained measures are heart rate, pedometer rate, game score, completion time and enjoyment. The preferred music choices from the experiment are analysed and classified according to mood cluster, valence and arousal. Obtained measures for these music classifications are discussed, as are the implications for automatic mood classification in choosing music for future experiments. It is noted that such mood classification schemes are nascent. A means by which these schemes may be made more robust is proposed, and initial results toward this goal are presented.

I. INTRODUCTION

Recent research studying music preference and performance during game play indicates that exposure to self-selected music may optimise both game play performance and experience when compared to experimenter-chosen music (Cassidy and MacDonald 2008). It is desirable to investigate whether particular characteristics of music play a part in influencing listener choice when selecting music for everyday tasks such as game play. If such features can be identified reliably, recommendations for similar music choices may be made to the listener based on automatic analysis of large music collections such as those found on personal music players.

Research in the field of Music Information Retrieval (MIR) has concentrated on extraction of more detailed information from music files by means of audio signal analysis (Tzanetakis and Cook 2002, Pampalk 2001). Extraction of such features allows detailed description of spectral content (e.g. brightness, intensity), tempo, rhythm etc. and has been used as a basis for categorisation of music by genre, artist, or similarity to other music items. They can therefore be used to accurately (and automatically) identify and recommend 'similar' music files to the experimenter.

More recently, effort has been expended toward using such signal parameters in the classification of music files in terms of the emotion they express (Lu et al 2006, Yang et al 2008). The aim is to extract the 'mood' of the music. A related study, just started by the authors of this paper, aims to improve the robustness of signal analysis and mood classification of music

by the inclusion of measures of musical structure into these analysis and classification schemes (Knox et al 2007).

The experiment described in this paper examines listener health, performance and experience while listening to music during game play. Such experiments provide an ideal opportunity to study the effects of music mood (as identified by nascent emotion classification systems) during this everyday activity.

Section II of this paper provides an overview of the experiment, which studied participant health, performance and experience during entertainment console game play, including the effect of subject-selected (preferred) music on these measures. Section III introduces the concept of emotion classification schemes based on audio signal analysis, and Section IV presents the results of applying these schemes to the music used in the game play experiment. Section V proposes a means by which these emotion classification algorithms may be made more robust, and presents some initial experimental results toward this goal.

II. GAME PLAY EXPERIMENT

Measures obtained from participants completing entertainment console games included behavioural (pedometer rate, completion time and performance), physiological (heart-rate), and experience measures (enjoyment). There were three sound conditions during the test: The game soundtrack (game), a track selected by the experimenter (E-S), and music chosen by the participant (preferred). Mean and standard deviation performance and estimate measures in each sound condition are presented in Table 1.

Table 1. The mean (M) and standard deviation (SD) performance and estimate measures in each sound condition during the game play experiment.

		Heart Rate	Ped Rate	Score	Time	Enjoy
Game	M	95.3	22.1	1.4	100.1	6.1
	SD	12.1	13.3	1.0	14.8	2.2
Pref	M	99.1	26.3	2.3	94.3	7.4
	SD	14.0	14.9	0.7	14.2	2.0
E-S	M	97.4	21.2	1.7	93.5	5.6
	SD	13.4	11.7	1.0	16.3	2.3

A 3 x 2 x 1 mixed design ANOVA was carried out on the data. The within subjects factor was the sound condition, with 3 levels; (i) Soundtrack (ii) Preferred (iii) Experimenter-Selected. The two between-subjects factors were gender and age. There were 5 dependent measures of performance; (i) Heart Rate (ii) Pedometer Rate (iii) Score (iv) Time and (v) Enjoyment. There was a significant effect of sound on the combined dependent variables [$F(10,34)=7.609$, $p<.001$; Wilks Lambda=.088; partial eta squared=.692]. The main effect of gender, main effect of age, and the interaction between sound and gender, and sound and age, did not reach significance.

- There was a main effect of sound on Heart Rate [$F(2,20)=19.891$, $p<.001$; partial eta squared=.665]. A series of paired t-tests revealed that Heart Rate was significantly highest when exposed to Preferred music in comparison to Experimenter-Selected ($t(24)=-3.761$, $p<0.001$) and Soundtrack ($t(24)=-3.029$, $p<0.01$) the other sound conditions, and significantly lowest in Soundtrack ($t(24)=-2.561$, $p<0.05$, in comparison to Experimenter-Selected).
- There was a main effect of sound on Pedometer Rate [$F(2,20)=6.412$, $p<.01$; partial eta squared=.391]. A series of paired t-test revealed that Pedometer Rate was highest when exposed to Preferred music in comparison to Experimenter-Selected ($t(24)=-3.712$, $p<0.001$) and Soundtrack ($t(24)=-3.428$, $p<0.01$).
- There was a main effect of sound on Time [$F(2,20)=4.526$, $p<.05$; partial eta squared=.312]. A series of paired t-tests revealed that Time was fastest when exposed to Preferred and Experimenter-Selected music ($t(24)=-0.601$, $p=.554$) in comparison to Soundtrack ($t(24)=2.725$, $p<0.05$ for Experimenter-Selected, and ($t(24)=3.285$, $p<0.01$ for Preferred).
- There was a main effect of sound on Enjoyment [$F(2,20)=14.561$, $p<.001$; partial eta squared=.593]. A series of paired t-tests revealed that enjoyment was greatest in Preferred music in comparison to Experimenter-Selected music ($t(24)=-4.191$, $p<0.001$) and Soundtrack ($t(24)=-4.571$, $p<0.001$).
- Although the main effect of sound did not reach significance, a series of paired t-tests revealed that Performance was significantly better when exposed to Preferred music in comparison to Experimenter-Selected ($t(24)=-3.412$, $p<.01$) and Soundtrack ($t(24)=-4.530$, $p<0.001$) for Experimenter-Selected.

These results indicate that health and experience measures were optimal when listening to self-selected music: players experienced highest heart rate, performance, steps, enjoyment and most positive mood change.

III. MOOD CLASSIFICATION

It has been shown by several researchers that emotional aspects of music are important search criteria when seeking/browsing for music (Huron 2000, Huron and Aarden 2002, Li and Mitsunori 2003, Yang and Lee 2004, Vignoli

2004, Lu et al 2006). These studies suggest that incorporating qualitative music features, such as emotion (Chai and Vercoe 2000) make music information retrieval better suited to representing the needs and preferences of the user. If emotional information is to be extracted from large databases efficiently, the process must be automated – i.e. extracted using signal analysis methods (Huron and Aarden 2002).

Automatic emotion detection in music is a relatively new field of study, and the lack of research in this area has been attributed to the subjective nature of emotion and cultural differences in experiencing emotion (classification of a particular song as communicating a certain mood may not apply to all listeners). However human beings display a common emotional response to a wide range of stimuli (Ekman and Friesen 1998), and there is marked agreement in the emotion being expressed by music between different listeners (Krumhansl 2002). This suggests cultural differences may not be so strong as to make emotional categorisation culture-specific. We should emphasise a key point here: Our aim is not to identify music's emotional effect on the listener – rather it is to identify the perceived 'mood' of the music. Indeed, whether music elicits an emotional response in the listener is still a matter of some debate. However some approaches to mood classification concentrate on the feelings of the listener, as opposed to the perceived emotion being expressed by the music (e.g. Yang et al 2008).

Studies examining automatic feature extraction for emotion in music typically aim to label music in terms of the two-dimensional energy-stress model of Thayer (1989). This model is based upon Russell's circumplex model of affect (Russell 1980), and adapted for music. In Thayer's model, the axes are energy (arousal) and stress (valence), with four quadrants equating to contentment, depression, exuberance and anxious/frantic. Valence is the horizontal axis (positive to negative from left to right), and arousal is the vertical axis (low to high from bottom to top). See Figure 1.

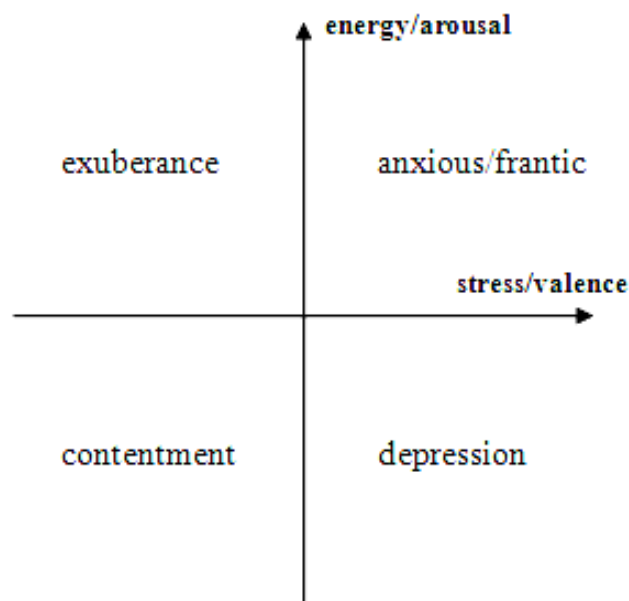


Figure 1. Thayer's stress/energy model, showing the four mood clusters; contentment, depression, exuberance and anxious/frantic.

It is noted that the labeling of these axes varies depending on the study the model is used in, and the adjectives describing the four quadrants, or clusters, can differ but retain a similar meaning. Further, it is possible to overlay specific mood adjectives relating to varying points in the four quadrants - see Picard (1997), Juslin (2000), and Hevner (1935).

Generally, the circumplex model is preferred to identification of individual adjectives in automatic mood classification, due to the increased ease with which the two axes (energy and valence) may be managed computationally (Lu et al 2006). However other researchers point to the ambiguity of attempting to classify music in terms of only four mood quadrants, and that emotion states may vary widely within each quadrant. This has led to attempts to label any point in the Thayer axes as an individual ‘emotional state’, and categorising music according to exactly which point it lies on the axes (Yang et al 2008).

A. Feature extraction

Extraction of acoustical features from digital music recordings has been used for various audio classification applications such as genre or music similarity. Likewise, emotion classification schemes are based on audio signal analysis parameters. There are many studies which indicate the significance of acoustic and musical features to emotion in music. A useful overview of the various features in question can be found in Juslin (2001). Many of these features can be extracted by audio analysis techniques, and have formed the basis for emotion classification studies. Lu et al (2006) base their classification upon measures of signal intensity, timbre and rhythm. These include spectral intensity, spectral shape (brightness, bandwidth, and how the spectrum changes over time – measured over spectral sub-bands), rhythm, tempo and rhythm regularity. Yang et al (2008) extract further features, including those utilized by Tzanetakis and Cook (2002), and those produced by the PsySound programme (Cabrera 1999), which also produces measures of loudness, dissonance and pitch.

B. Mood Classification

Once extracted, the acoustical features of music files are used in a subsequent classification process. Typically these processes seek to assign a ‘weighting’ to various acoustical features extracted from the music files, and then consider these when attempting to classify the music in terms of the emotion it expresses. That is, the relative importance of each feature is different for varying emotional expressions (Juslin 2000, Lu et al 2006).

Different approaches have been taken by various researchers concerned with this task, although typically a corpus of training data is studied and acoustical features are extracted. The perceived emotion of each music file in the corpus is already known (through subjective listening tests, or annotation by the experimenters or music ‘experts’). This training set is then used as the basis for classifying new music inputs, the acoustic features of the new sample being compared in some way to the training corpus.

Lu et al (2006) consider intensity, timbre and rhythm features separately. Acoustical features are extracted from the training

set, and these are normalised across the corpus (by z-scores). The Karhunen-Loeve transform is applied to remove correlation between the extracted parameters, and thus the values considered in subsequent classification stages actually represent statistical measures (mean and standard deviation) of the extracted acoustical parameters. These features are then considered in a hierarchical classification scheme based on Gaussian Mixture Models (GMM). In this scheme, music clips are first classified as belonging to one of two mood groups. Group 1 is made up of the mood clusters contentment and depression, and Group 2 is made up of the clusters exuberance and anxious/frantic (as seen in Figure 1). The next stage groups the clips into the individual clusters (e.g. contentment). It is at this stage that the relative importance of rhythm and timbre features in the expression of emotion in music are taken into account. See Figure 2.

The authors state the example of music in the contentment cluster being ‘brighter’ than in the depression cluster, and that therefore timbre features are important when discriminating between these two clusters. However rhythm is deemed to be more important when discriminating between exuberance and anxious/frantic clusters. The λ_1 and λ_2 values in Figure 2 (relative weighting) are manipulated at this stage to fine-tune the classifier accuracy.

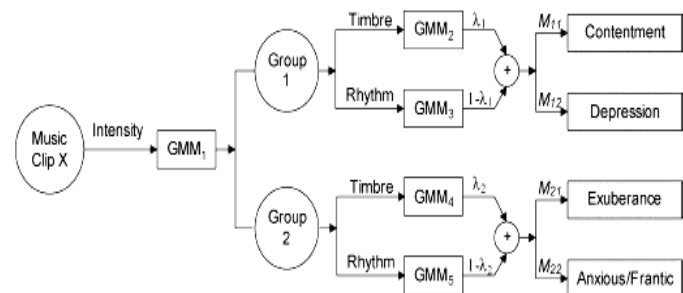


Figure 2. The hierarchical mood classification scheme of Lu et al.

A rather different approach is taken by Yang et al (2008), who aim to link music to specific points on Thayer’s axes, rather than solely the four mood clusters. Their training set is labeled in a series of subjective listening tests. The listeners input their emotional response to the music (as opposed to the perceived emotion of the music) on a set of Arousal/Valence axes. This data is then considered along with the extracted acoustical parameters and used to train a regressor. Therefore the scheme is similar to the more standard classification approach, but the axes are viewed as a continuous space. The authors state that the ambiguity of mood cluster classification is avoided - the aim is to classify new music samples as existing at (or near to) a specific point on Thayer’s axes.

IV. GAME PLAY MUSIC ANALYSIS

The game play experiment provides an excellent opportunity to apply such emotion classification schemes to music selected (preferred) by the subjects taking part in the test. As the GMM classification scheme of Lu et al can be seen as being fairly ‘typical’ of emotion classification by signal analysis, and that high prediction accuracy is stated, these

methods were applied to the music files selected by the test subjects. In addition, this method has only been applied previously to western classical music. In this study, there is the opportunity to observe the performance of the algorithm upon collections of contemporary western popular music, such as those commonly held on portable music players.

At the time of writing, the preferred music of only 23 test subjects was available for analysis. The emotion classification algorithm training set was also fairly small – based on 80 music clips, of which 20 clips were representative of each of the four mood clusters. The training clips were extracted from a corpus of western popular music, and tagged for the appropriate mood cluster manually by the authors. Following the original implementation of Lu et al, the training and test corpus consists of 20 second clips of the music files in question. The files are then downsampled to 16kHz, 16 bit mono, and further segmented into a series of frames 32ms in length, spanning the duration of each clip. The intensity, timbre and rhythm parameters are then extracted and used as a basis for the GMM classification stage.

Of the 23 test subjects studied, 7 chose ‘content’ music, 4 chose ‘depressed’, 9 chose ‘exuberant’ and 3 chose ‘anxious/frantic’ music as defined by the mood classification algorithm. The majority of test subjects (16) chose music that was positively valenced, and roughly equal numbers chose low arousal (11) and high arousal music (12). The small numbers involved (and also the need to collapse the data in order to compare mood classifications to the measures for all the preferred music), make formal statistical analysis of the data difficult. Only one mood cluster showed significant effect on the test variables – ‘depressed’ music resulted in pedometer rates significantly lower than the preferred music mean ($t(22) = -3.567, p < 0.05$). However some interesting trends are noted:

- Those choosing depressed music had lower heart rate, pedometer rate and took longer to complete the test than those choosing music from the other mood clusters. However they also scored highest for enjoyment.
- Anxious/frantic music coincided with the highest pedometer rates.
- Subjects who chose content music had the fastest overall time of completion, and the lowest enjoyment.
- Those choosing music of high arousal had the highest heart rate and pedometer rate, but had the lowest scores and took the longest time.
- Those choosing positively valenced music had the quickest time, but the lowest enjoyment.

It is also interesting to note that the experimenter selected music is classified as anxious/frantic (high arousal, negative valence). From the above observations, we might expect (bearing in mind the lack of statistical significance) that such a selection is likely to produce, for example, high pedometer rates and lowest scores. This is, at the very least, useful information for the experimenter, who might wish to control for the likely effects of music according to its determined mood. In addition, mood classification algorithms can aid the

experimenter in automatically accessing music according to mood criteria for whatever purposes they may require.

V. TOWARD IMPROVED MOOD CLASSIFICATION

The aforementioned mood classification schemes are based solely on extracted acoustical parameters, and have been cited as gaining a good degree of classification accuracy. Indeed the performance of the algorithm in classification of the game play test music was impressive – only two music clips from 23 were not classified correctly by the algorithm, and had to be manually tagged for mood. However, the authors of such schemes themselves recognise that extraction of ‘more powerful’ acoustic features is required to better represent music primitives in mood detection (Lu et al 2006).

Existing taxonomies do not take into account many features identified as being key to expressing emotion in the composition/performance of music, such as:

- Mode - the primary set of pitch intervals.
- Tonality - the relationship existing between tones or tonal spheres within the context of a particular style system (Meyer 1956).
- Pitch - perceived frequency.
- Tonic, or the tonal centre - a stable point often emphasized melodically or rhythmically. The tonic largely defines the pitch of a section of music (Krumhansl 1990).
- Key: Described by the tonic and the mode.
- Interval (harmonic and melodic).

Inclusion of such measures of musical structure, along with acoustical features already extracted in peer studies, is likely to provide a fruitful approach towards finding a more accurate means of automatically classifying the emotional content of music. An EPSRC-funded research project in its early stages (and for which the author of this paper is principal investigator) aims to investigate this possibility, and the following section discusses initial investigations toward this goal.

A. Implication-realisation

It has long been established that expectation plays a pivotal role in a listener's experience of music (Meyer 1956). This theory is based on the assumption that while listening to music an individual will form expectancies about its continuation. If these presuppositions are violated it evokes a corresponding emotional reaction (Huron 2006). From an evolutionary perspective such responses serve a biological purpose, however some are more clearly defined than others. A high intensity noise may evoke a ‘fight or flight’ response whereas the variations of a melodic line can generate a more subtle and often complex response, typically calling on high order cognitive functions.

We seek to investigate this aspect of melodic perception and cognition by examining the implication-realisation (I-R)

model (Narmour 1990) with respect to the classification process proposed by Lu et al (2006).

The investigation into musical expectation is being undertaken using a method proposed by Eerola and Toivainen (2004) in their MIDI toolbox for the MATLAB environment. By adapting this method we may generate results which are meaningful in the context of contemporary western music. The framework judges melodic expectancy based on the Gestalt-based principles of proximity, similarity and good continuation. More formally these principles are drawn from the I-R model and revisions suggested by others (Krumhansl, 1982 and 1995, von Hippel 2000) and have been found to predict melodic expectancies with reasonable accuracy.

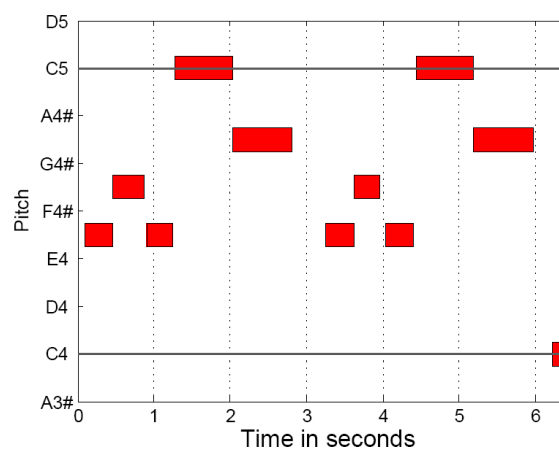
The I-R model is based on a three note archetype containing an implied and realised interval. The former creates implications for the melody's continuation and the next interval carries out its implication (Eerola and Toivainen 2004). In general terms the principles which lie behind the framework focus on the distance and direction relationships between these two intervals. Of the five principles which constitute this theory, two originate from the original I-R model: proximity and registral return. In addition there are also measures of consonance (Krumhansl 1995), tonality (Krumhansl 1982), tessitura and mobility (von Hippel 2000).

An example of the analysis is shown on a fragment of music in Figure 3. The first step of the process is to determine the key, based on analysis of a MIDI representation of the piece as extracted from the audio (shown in piano roll format in Figure3(a)).

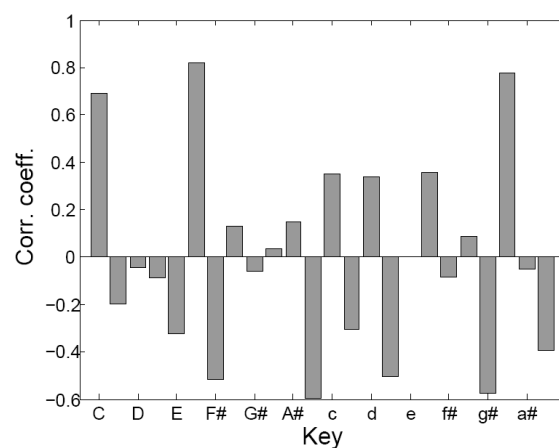
This is achieved by comparing the distribution of pitch classes throughout the piece with the 24 major/minor key profiles, a process developed by Krumhansl (1990). Figure 3(b) shows a graphical representation of this correlation. From this statistical analysis we can see the correlation of all of the pitch classes with the largest similarity to the actual key signature (in this case F major). Figure 3(c) shows an alternative projection of this data using self-organising maps (SOM) trained with the 24 key profiles.

Figures 4 and 5 illustrate the scoring system based on the above principles relating to melodic expectation.

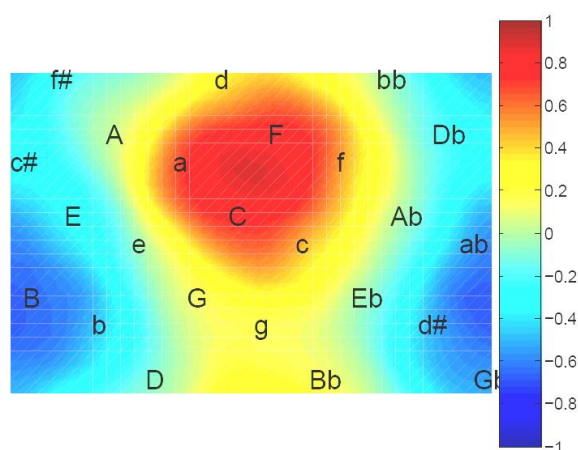
In this example a 12 note fragment was terminated on different notes in order to judge the quality of the melody. The original ending provided the highest mean score (Figure 4) with alternatives deviating 1 and 9 semitones from the original respectively. The lower mean scores obtained by these intervals demonstrates their decreasing suitability for melodic continuation in this context – see Figure 5.



(a)



(b)



(c)

Figure 3 (a) extracted MIDI piano roll representation of a music clip in F major, ending in C. (b) pitch class distribution throughout the piece. (c) SOM visualisation of the extracted key signature.

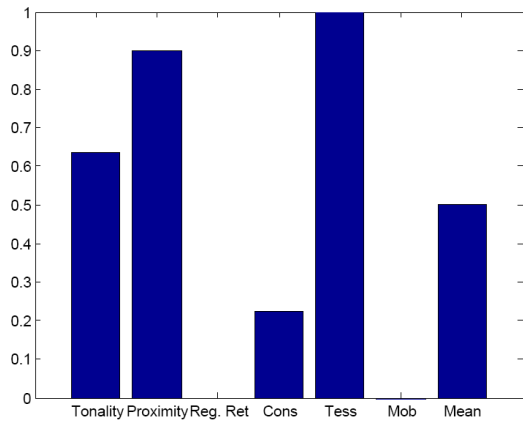
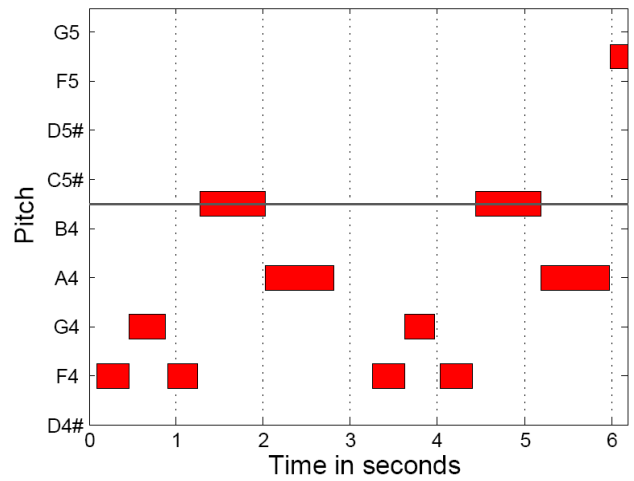
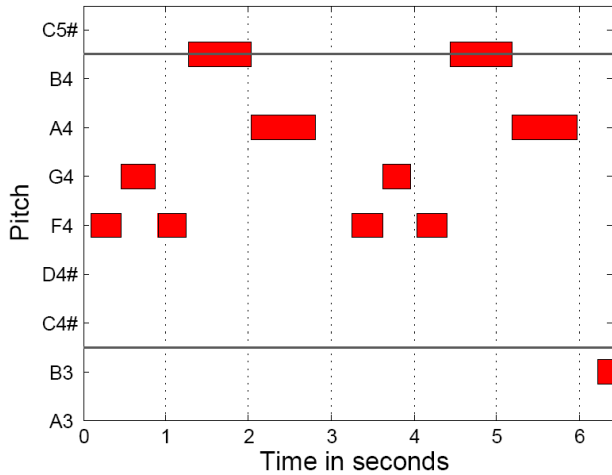


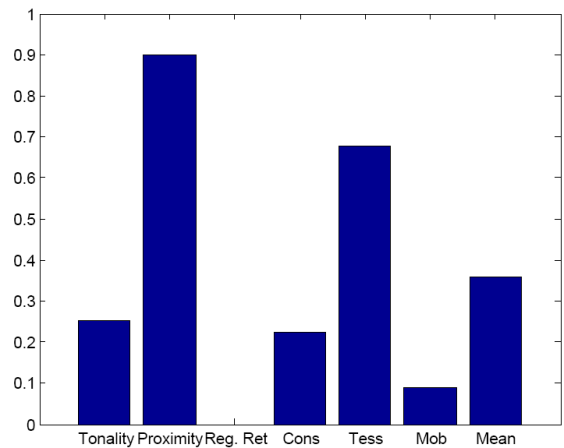
Figure 4. Melodic expectation means for the music clip shown in Figure 3(a).



(c)

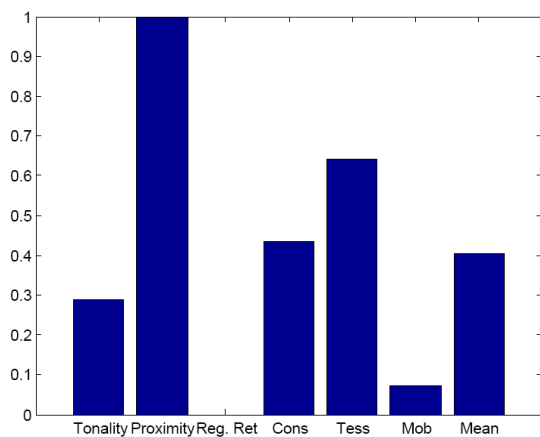


(a)



(d)

Figure 5. Piano roll and melodic expectation means for the music clip. (a) - (b) modified to end in B , and (c) - (d) modified to end in F#.



(b)

VI. CONCLUSION

The results of the mood classification of preferred music choices from the game play experiment are interesting. Music classified as being ‘depressed’ resulted in significantly lower pedometer rates than music from other mood clusters. Some interesting trends are observed (although statistical significance is not reached) for music from other mood clusters, and also when classified by valence or arousal. It is anticipated that further tests (and access to more of the music files used in the experiment) may glean more significant results as regards performance measures for varying music mood. It is also anticipated that uncovering such effects, and the ability to automatically access music of a specific mood, will prove valuable tools for the experimenter.

The performance of the mood classification scheme applied to the game play results is promising. However the accuracy of the scheme is dependent upon the size and quality of the training data used. It is also very dependent upon judicious weighting of the influence of intensity, timbre and rhythm acoustical features. It is recognised that more powerful

analyses are required to make these schemes more robust. By taking the factors discussed in section V into consideration we hope to judge the fitness of the melodic continuation - which is key to assessing the type of emotion expressed by the music. Throughout the course of the research we seek to include more aspects of Narmour's model with the aim of further improving the emotion classification process.

REFERENCES

- Cabrera, D. (1999). PsySound: A Computer Program for Psychoacoustical Analysis. *Proceedings of the Australian Acoustic Society Conference*. Pp. 47-54.
- Cassidy, G.G. & MacDonald, R.A.R (2008). The Effects of Self-Selected vs. Experimenter-Selected Music on Driving Game Performance and Evaluation. (Submitted to *Musicae Scientiae*).
- Chai, W and B. Vercoe. (2000). Using user models in music information retrieval systems, *Proceedings of International Symposium on Music Information Retrieval*, Oct.
- Eerola, T. and P. Toivainen., (2004). MIDI Toolbox: MATLAB Tools for Music Research. Jyväskylä, Finland: University of Jyväskylä.
- Ekman, P. and W.V. Friesen. (1998). Constants Across Culture in the Face and Emotion. In Jenkins, Oatley and Stein (eds.), *Human Emotions: A Reader*. Malden, MA: Blackwell.
- Hervner, K. (1935). Expression in music: a discussion of experimental studies and theories. *Psychology Review*, vol. 42, pp. 186–204.
- von Hippel, P. (2000). Redefining pitch proximity: Tessitura and mobility as constraints on melodic intervals, *Music Perception*, vol. 17, no. 3, p. 315-327.
- Huron, D. (2000). Perceptual and Cognitive Applications in Music Information Retrieval. *Proceedings of International Symposium on Music Information Retrieval*.
- Huron, D. and B. Aarden. (2002). Cognitive Issues and Approaches in Music Information Retrieval, chapter in unpublished book '*Music Information Retrieval*', S. Downie and D. Byrd (eds.).
- Huron, D. (2006). *Sweet anticipation: music and the psychology of expectation*. MIT Press, Cambridge, Mass.
- Juslin, P. (2000). Cue utilization in communication of emotion in music performance: relating performance to perception. *Journal of Experimental Psychology: Human Percept. Perf.*, vol. 16, no. 6, pp. 1797–1813.
- Juslin, P. (2001). Communicating Emotion in Music Performance. In P. N. Juslin and J. A. Sloboda, *Music and Emotion: Theory and Research*. Oxford, U.K.: Oxford Univ. Press.
- Knox, D., S. Beveridge and R. MacDonald. (2007). Emotion Classification in Contemporary Music. DMRN+2: Digital Music Research Network One-day Workshop 2007, Queen Mary, University of London, December.
- Krumhansl, C.L. and E. J. Kessler. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys, *Psychological Review*, vol. 89, no. 4, pp. 334-368.
- Krumhansl, C. (1990). *Cognitive Foundations of Musical Pitch*. Oxford University Press, USA.
- Krumhansl, C.L. (1995). Music Psychology and Music Theory: Problems and Prospects. *Music Theory Spectrum*, vol. 17, no. 1, pp 53-80..
- Krumhansl, C. (2002). Music: A Link Between Cognition and Emotion. *Current Directions in Psychological Science*, Volume 11, Number 2, April, pp. 45-50.
- Li, T., and O. Mitsunori. (2003). Detecting emotion in music. In *Proceedings of International Symposium on Music Information Retrieval: ISMIR*.
- Lu, L., D. Liu. and H. Zhang. (2006). Automatic Mood Detection and Tracking of Music Audio Signals, *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 14, No. 1, January.
- Meyer, L.B. (1956). *Emotion and meaning in music*. The University of Chicago Press.
- Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. University of Chicago Press.
- Pampalk, E. (2001). Islands of Music: Analysis, Organization and Visualization of Music Archives. Master's thesis, Vienna University of Technology, December.
- Picard, R.W. (1997). *Affective Computing*. Cambridge, MA:MIT Press.
- Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality Social Psychology*, vol. 39, pp. 1161–1178.
- Thayer, R.E. (1989). *The Biopsychology of Mood and Arousal*. Oxford, U.K.: Oxford University Press.
- Tzanetakis, G. and P. Cook. (2002). Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing*, Vol. 10, No. 5.
- Vignoli, F. (2004). Digital Music Interaction Concepts: A User Study. In *Proceedings of International Symposium on Music Information Retrieval: ISMIR*.
- Yang, D. and W. Lee. (2004). Disambiguating Music Emotion Using Software Agents. In *Proceedings of International Symposium on Music Information Retrieval: ISMIR*.
- Yang, Y.H., Lin, Y.C., Su, Y.F., Chen, H.H., Yang, Y.H., Lin, Y.C., Su, Y.F., & Chen, H.H. (2008). A regression approach to music emotion recognition. *IEEE Trans. Audio, Speech and Language Processing (TASL)*: 16 (2), 448–457.