# DISAMBIGUATING MUSIC EMOTION USING SOFTWARE AGENTS

*Dan Yang*
Systems Science
University of Ottawa
Ottawa Canada K1N 6N5
dyang006@uottawa.ca

*WonSook Lee*
School of Information Technology and Engineering
University of Ottawa
Ottawa Canada K1N 6N5
wslee@uottawa.ca

## ABSTRACT

Annotating music poses a cognitive load on listeners and this potentially interferes with the emotions being reported. One solution is to let software agents learn to make the annotator's task easier and more efficient. Emo is a music annotation prototype that combines inputs from both human and software agents to better study human listening. A compositional theory of musical meaning provides the overall heuristics for the annotation process, with the listener drawing upon different influences such as acoustics, lyrics and cultural metadata to focus on a specific musical mood. Software agents track the way these choices are made from the influences available. A functional theory of human emotion provides the basis for introducing necessary bias into the machine learning agents. Conflicting positive and negative emotions can be separated on the basis of their different function (reward-approach and threat-avoidance) or dysfunction (psychotic). Negative emotions have strong ambiguity and these are the focus of the experiment. The results of mining psychological features of lyrics are promising, recognisable in terms of common sense ideas of emotion and in terms of accuracy. Further ideas for deploying agents in this model of music annotation are presented.

Keywords: annotation, emotion, ambiguity, agents

## 1. INTRODUCTION

Music retrieval systems typically lack query-by-emotion, leaving it up to the user to know which artist, album, and genre names correlate with the desired musical emotion. Progress in this area is hindered by a "serious lack of annotated databases that allow the development of bottom-up data-driven tools for musical content extraction" [8]. In this paper we show that machine learning techniques can be embedded in annotation tools using software agents, in support of high-level design goals.

One high-level design choice is to model musical meaning as compositional, and Emo allows different users to focus on different aspects of the stimulus material to specify its musical meaning and emotion. A diverse range of stimulus material is offered in this environment, so that subjects are very willing to elaborate further about their mental state, and the underlying influences they are focusing on can be traced. Such an integrated environment overcomes the problem of cognitive overload of other environments where the user is burdened with dissimilar tasks of emoting and reporting. The use of multiple stimulus materials (audio, graphics, text lyrics, text reviews) assists user need prompting to better articulate and externalise their feelings.

The problem of the reliability of emotion ratings is one of the main motivations for the design of Emo. Consistency of response is understood relative to the combination of stimuli leading up to the response, and emotion is not taken as an absolute property of any single media file by itself. While variations in response are allowed, there are also unwanted variations such as drift due to the influence of extreme songs, and deterioration in the quality of response due to fatigue. Outliers such as these can be detected by software agents that check for consistency between some combination of input stimuli and past ratings.

## 2. RELATED WORK

### 2.1. Psychological perspectives

The structure of emotion studied by psychologists includes core affect [14], emotion, mood, attitude [13], and temperament [21]. Psychological perspectives on music and emotion have tended to focus on people's verbal reports of feeling states and on whether these emotion words can be structured according to how many and which dimensions [14][15][19].

Dimensional ratings are quick single-item test scales for eliciting emotions [15], suitable for repeated use in applications such as annotating short musical segments. The aim is to not confuse the user with the need to discriminate between similar emotions, so related emotions are placed close together on the dimensional scale. This makes the single-item test rating a useful entry point for eliciting emotion. Some studies directly relate dimensional scale ratings to musical features such as tempo, or perceived energy [12]. Others have depicted the relationship between dimensional scales and musical features with greater complexity [9]. Music emotion is generally seen as irreducible to simply one or two dimension ratings. For example, the online All Music Guide [4] uses over 160 different discrete emotion categories (e.g. trippy, quirky) to describe artists, by using up to 20 emotion words to describe the tone of their career's work. Baumann's Beagle system [1] mines text documents to collect all the emotion-related words (e.g. ooh, marry, love, wait, dance, fault), used in lyrics or online reviews

Psychological research has explored ways of unifying the dimensional and discrete approaches to emotion ratings. Sloboda and Juslin [15] note that dimensional and discrete models can be complementary to each other. One accessible approach is the PANAS-X test scale [22] which has two dimensional ratings called Positive Affect (PA) and Negative Affect (NA). The dimensional ratings function as entry points to more detailed ratings of discrete emotions under each axis (e.g. Fear under NA). The two PANAS-X dimensions can be mathematically related to Russell's circumplex model[14]. Russell's Arousal is the sum of PA and NA, while Russell's Valence is the difference (PA – NA). Tellegen, Watson and Clark [19] use the Valence dimension (pleasant-unpleasant) as the top-level entry point of a 3-layer model. This unified model offers the benefits of dimensional ratings, plus a theoretical basis that links the entry-point of the hierarchy to the discrete emotion categories at the base (as shown in Figure 1).

### 2.2. Systems for music annotation

A number of online systems exist for annotating popular music emotion. The All Music Guide collects user responses from the web in dimensional form (e.g. exciting/relaxing, dynamic/calm). Moodlogic.com collects user emotion ratings from the web in bipolar dimensional form (general mood positive or negative), in multivalent dimensional form (e.g. brooding, quirky) as well as discrete terms (e.g. love, longing). Moodlogic.com allows query-by-emotion using 6 discrete emotion categories (aggressive, upbeat, happy, romantic, mellow and sad). Songs are regularly labelled by two or more emotions. A query for two emotions together, 'both happy and upbeat', retrieved about half the songs in the database. There was no way to disambiguate the results into happy, upbeat as separate emotions.

Microsoft MSN.com has 115 Mood/Theme discrete category names. No theory of emotion is used in this taxonomy, which is based on a mix of artist names, emotion names, country names, and names of parts of the daily routine such as workout, dinner, etc. Musicat [3] also used names of everyday contexts such as bedtime as well as moods (happy, romantic) to label listening habits. The system learns which songs are played in which listening habit.

The above taxonomies tended to be ad hoc lists mixing together words for feelings, thoughts and everyday activities instead of systematically examining these affective, cognitive and behavioural aspects of emotion.

### 2.3. Systems for music data mining

Human listening is very effective at organizing the stream of auditory impulses into a coherent auditory image. If digital signal processing primitives can be used to discern features of interest to a human listener, then these are useful to add to the music emotion annotation environment. The evaluation of the best features is hindered by a lack of standardized databases [5]. Current feature extraction tools are very low-level, such as MPEG-7 Low Level Descriptors [6].

Recently, wavelet techniques have been developed that tile the acoustic landscape into smaller features that correspond to musical elements such as octaves [20]. Another innovation is the automatic discovery of feature extractors. Sony's Extractor Discovery System (EDS) [12], uses genetic programming to construct trees of DSP operators that are highly correlated to human-perceived qualities of music.

Baumann's Beagle system [1] demonstrates the relevance of mining music reviews and lyrics for emotion words co-occurring with artist names and song names. State-of-the-art performance on extracting individual names from text is about 90%, but accuracy falls below 70% when compound relations such as (artist, song, emotion) because errors multiply.

## 3. EMO SYSTEM

### 3.1. Motivation

Initially we were looking for online datasets of music already annotated by discrete emotion labels, and what we found was an ad hoc mix of discrete terms, including cognitive, behavioural and affective words. Single-item dimensional ratings did not separate like emotions, such as happy/upbeat or anger/fear. Hierarchical models of music emotion recognition have been reported by Liu [10], and we decided to extend the hierarchical approach further to the level of discrete emotion categories. The key problem we found annotating music in greater detail was the cognitive load on the annotator in both listening and reporting in detail. The solution we are developing uses software agents that learn to make the annotation task more efficient.

### 3.2. Emotion model

In this paper we focus on Negative Affect as these emotions are less distinguishable from each other than positive emotions are from each other [21]. This finding is shown in the way emotions such as fear and anger are highly correlated in the dimensional model. In real terms, this can be seen in the way that related episodes of negative emotion such as hostility, paranoia and sadness occur in depressive illness.

The Watson model [19][21] explains negative emotion as the threat-avoidance function in the structure of emotions. Taking fear and anger as an example, these categories are both high in negative affect on a dimensional scale (high Negative Affect in Figure 1), but they can be distinguished apart based on their pattern of response to the threat. Fear anticipates a threat and triggers flight, while anger can involve a fight against the threat. By finding more information about the pattern of response to a threat, using information from both the lyrics and the music, negative emotions can be distinguished from each other into discrete classes.

Taking anger and guilt as an example, these are both high in negative affect on a dimensional scale (high Negative Affect in Figure 1) and practically uncorrelated with Positive Affect. Guilt is related to feelings that persist a little time after some event, while Hostility is directly involved with some threat event.

Sadness and guilt show some separation at the middle level of the emotion model, because sadness is slightly correlated with Positive Affect while guilt is not. This implies that sadness and guilt can be better differentiated by assessing the Positive Affect component, given lack of separability in terms of Negative Affect.

The Tellegen-Watson-Clark model discussed in Section 2.1 is useful in linking the dimensional and discrete levels of emotion. The experiments and results in section 4 are based on this model shown in Figure 1.

## 4. EXPERIMENTS

### 4.1. Music emotion intensity prediction

This first experiment was designed to implement a classifier for music emotion intensity, understood in terms of the psychological models of Russell [14] and Tellegen-Watson-Clark [19], where intensity represents the sum of the PA and NA dimensions (in Figure 1).

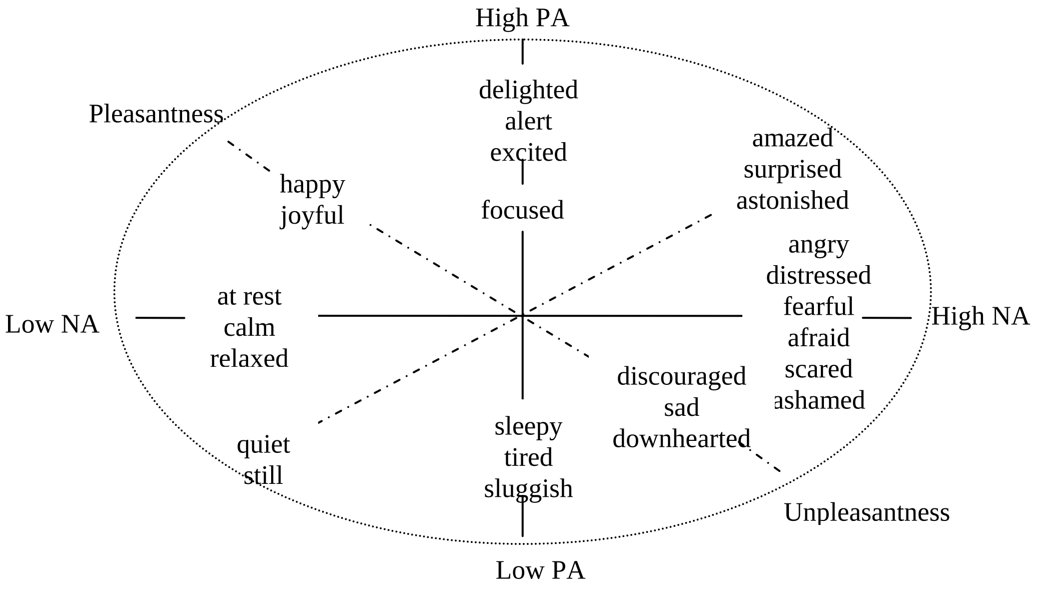

Figure 1. Elements of the Tellegen-Watson-Clark emotion model [19] [21]. Dotted lines are top-level dimensions. The Positive Affect (PA) and Negative Affect (NA) dimensions shown as solid lines form the middle of the hierarchy, and provide heuristics needed to discern the specific discrete emotion words based on function. Discrete emotions that are close to an axis are highly correlated with that dimension, e.g. sad is slightly correlated with positive affect.

The emotional intensity rating scale was calibrated to Microsoft’s emotion annotation patent [17] , using the same method to train a volunteer. The initial database consisted of 500 randomly-chosen rock song segments of 20 seconds each taken beginning a third of the way into the song.

Acoustic feature extraction used a number of tools to give a broad mix from which to select the best features.
Wavelet tools [20] were used to subdivide the signal into bands approximating octave boundaries, and then energy extraction and autocorrelation were used to estimate Beats per Minute (BPM). Other acoustic attributes included low-level standard descriptors from the MPEG-7 audio standard (12 attributes). Timbral features included spectral centroid, spectral rolloff, spectral flux, and spectral kurtosis. Another 12 attributes were generated by a genetic algorithm using the Sony Extractor Discovery System (EDS) [12] with simple regression as the population fitness criteria. Labels of intensity from 0 to 9 were applied to instances by a human listener reporting the subjective emotional intensity, following exactly the human listening training method in the Microsoft patent [17].

The WEKA package [23] was used for machine learning. The results below were calculated using Support Vector Machine (SVM) regression.

#### *4.1.1. Results and analysis for emotion intensity*

This experiment confirmed the results of Liu [10] which found that emotional intensity was highly correlated with rhythm and timbre features. We achieved almost 0.90 correlation (mean absolute error 0.09), the best features being BPM, Sum of Absolute Values of Normed Fast Fourier Transform (FFT), and Spectral Kurtosis [7].

### 4.2. Disambiguation of emotion using text mining

In our compositional model of musical emotion, non-acoustic features such as lyrics words or social contextual content do play a role of focusing specific emotions and help account for the range of emotional responses to the same song. One problem with this approach is the size of the vocabulary used in expressive lyrics, which could be over 40,000 different words for songs in English. Various feature-reduction strategies are used in classic machine learning, but it is not certain how well these apply to emotion detection. We chose an established approach, the General Inquirer[18], to begin to explore the available techniques for verbal emotion identification. This system was chosen for its good coverage of most English words, and its compactness of representation, with 182 psychological features.

Of 152 30-second clips of Alternative Rock songs labelled with emotion categories by a volunteer, only 145 songs had lyrics. The emotion categories of the PANAS-X schedule [22] were used. Lyrics text files were transformed into 182-feature vectors using the General Inquirer package, and the WEKA machine learning package was used. The original text files were also examined using the Rainbow text mining package[11].

#### *4.2.1. Results and analysis for text disambiguation*

This experiment tested the idea that general psychological features driving emotions, such as seeking and attaining goals or reacting to threats, can be conveyed specifically in text and can add focus to the way music is interpreted.

| | |
|---|---|
| Hostility | expletives, not, get, got, want, never, don’t, go, no, my, oh, fight, burn, show, had, you |
| Sadness | love, life, time, say, slowly, hold, feel, said, say, go, if |
| Guilt | one, lost, heart, face, alone, sleep, mistake, memory, lies, eyes, die, silence, remember |

Table 1. Lyric words that distinguish lyrics by negative emotion i.e. have high information gain.

| | |
|---|---|
| Hostility | Words about no gain from love or friendship,<br>Not being a participant in love or friendship, Words about not understanding,<br>Words expressing a need or intent |
| Sadness | Loss of well-being,<br>Words about a gain of well-being without color or relationship |
| Guilt | Saying-type words,<br>Talking about gains from love or friendship, Passive-type words |

Table 2. General Inquirer[18] psychological features of lyrics text that most distinguish lyrics by negative emotion in the WEKA C4.5 decision tree.

***<u>Hostility, Sadness and Guilt</u>***

The negative affect behaviours are related to threat avoidance, so words strongly related to distinguishing each negative emotion from other negative emotions were ranked using their information gain [23]. The data shown in Table 1 includes forms of hostile display shows in the form of threatening expletives, and sounds showing lack of constraint such as ah, oh. Other words tended to connote commands or threats, such as no, don’t etc. There were references to ‘weapons’ and destruction such as fire, burning etc. Words that favoured guilt over hostility are related to waking/sleeping, mistakes, and reflection. The references to low energy activities in guilt are interesting, considering that the music is as arousing as hostility.

Sadness was also interesting in strongly referring to positive words such as love, life, feel etc. This higher correlation of sadness to positive affect is predicted in the emotion model.

Table 2 shows the psychological features mined from lyrics using WEKA's implementation of C4.5. Informally, these features appear to make sense, and the machine learning has mined verbal patterns that one would expect to correspond with these negative emotions, such as needing-type words associated with anger.

***Love, Excitement, Pride***

| | |
|---|---|
| Love | Not knowing-type words,<br>Not political,<br>Not loss of well-being,<br>Not negative,<br>Not failing,<br>Gains from love and friendship,<br>Passive-type word,<br>Not saying-type word |
| Excitement | Gains of well-being from relationship,<br>Animal-type words |
| Pride | Political words,<br>Respect,<br>Initiate change,<br>Knowing-type words |
| Attentive | Knowing-type words,<br>Color words |
| Reflective | Passive type words |
| Calm | Completion of a goal-type words |

Table 3. General Inquirer[18] features that most distinguished lyrics by positive emotion in the WEKA C4.5 decision tree.

Table 3 shows the psychological features that WEKA mined from song lyrics associated with positive emotions. Informally, these results are recognizable in terms our common-sense understanding of emotions.

### 4.3. Disambiguation of emotion using data fusion

This experiment fused together both acoustic and text features to maximise the classification accuracy.

There was an increase in accuracy of successful classification from 80.7% to 82.8%, a decrease in mean error from 0.033 to 0.0252. The decrease in root relative squared error was from 30.62% to 25.04%.

These results do not distinguish very much between the two procedures on such a small training set without testing, but the numbers do not contradict the informal discussion in the preceding section. A larger study would have more scope for examining the different types of errors in classification from the acoustic and text features.

## 5. CONCLUSION AND FUTURE WORK

This paper evaluated a structured emotion rating model for embodiment in software agents to assist human annotators in the music annotation system Emo. In experiments we found the structured emotion model useful in the context of a compositional model of musical meaning and emotion, where text features focused attention on more specific music emotions. Experiments were designed to explore this model, and we focused on negative emotions where there is the greatest ambiguity.

Results were given for a single-attribute test to rate emotion intensity (the sum of positive and negative energy in the model), based on 500 songs. About 90% accuracy was achieved using both timbral and rhythmic features. For learning to distinguish like-valenced emotions, a sample of 145 full-text lyrics showed promising results. Informally, the verbal emotion features based on General Inquirer appeared to correlate with significant emotion experiences reported by listeners. The small sample size precluded robust testing at this exploratory stage. The allmusic.com[4] song browser, with 1000s of songs classified by mood, could be one way to increase the sample size significantly.

Future work will investigate the way in which a compositional model of musical meaning and emotion can be deployed using graphical user interface devices for the user. The system tracks the focus of attention as each emotion is experienced by the user, and the resulting annotation trees can be mined to help confirm the theory of music as compositional. Subtle shifts in cognitive focus can correspond with shifts in musical meaning and emotion. Different methods of verbal emotion identification will be investigated, as this is a new and rapidly growing area of machine learning research. The existing bootstrap database appears adequate, and with further use there will be more songs added to the database, complete with appropriate stimulus material such as lyrics and cultural data. The emerging Semantic Web also provides further opportunities to find musical stimulus material by means of a shared music ontology [1]. Graphical media types are also relevant as stimulus material for popular music such as music videos. Visual features could be extracted from music videos as MPEG-7 video descriptors, and related to the function of each emotion. Some researchers believe that a more promising approach to rating emotion is to use direct physiological means [2]. This type of input could be added to the resource hierarchy of Emo.

## 6. REFERENCES

[1] Baumann, S, & Klüter, A., “Super-convenience for Non-musicians: Querying MP3 and the Semantic Web”, *Proceedings of the International Symposium on Music Information Retrieval,* Paris, France, 2002.

[2] Cacioppo, J.T., Gardner, W.L. & Berntson, G.G., “The Affect System Has Parallel and Integrative Processing Components: Form Follows Function”, *Journal of Personal and Social Psychology,* Vol. 26, No. 5, 1999, 839-55.

[3] Chai, W., “Using User Models in Music Information Retrieval Systems”, *Proceedings of the International Symposium on Music Information Retrieval,* Plymouth, MA, USA, 2000.

[4] Datta, D., “Managing metadata”, *Proceedings of the International Symposium on Music Information Retrieval,* Paris, France, 2002.

[5] Downie, J.S., “Towards the Scientific Evaluation of Music Information Retrieval Systems”, *Proceedings of the International Symposium on Music Information Retrieval,* Baltimore, MD, USA, 2003.

[6] ISO/IEC TC1/SC29/WG11, “ISO/IEC 15938-6 Information Technology – Multimedia Content Description Interface (MPEG-7) – Part 6: Reference Software”, Geneva, Switzerland, 2000.

[7] Kenney, J. F., Keeping, E. S., Section 7.12 in “Mathematics of Statistics” Pt.1, 3rd ed. Princeton, NJ; Van Nostrand, pp.102-103, 1962.

[8] Leman, M., “GOASEMA – Semantic description of musical audio”. Retrieved from http://www.ipem.ugent.be/.

[9] Leman, M., Vermeulen, V., De Voogdt, L., Taelman, J., Moelants, D. & Lesaffre, M., “Correlation of Gestural Music Audio Cues and Perceived Expressive Qualities”, *Lecture Notes in Artificial Intelligence,* Vol. 2915, 40-54, Springer Verlag, Heidelberg, Germany, 2004.

[10] Liu, D., Lu, L. & Zhang, H.J., “Automatic Mood Detection from Acoustic Music Data”, *Proceedings of the International Symposium on Music Information Retrieval,* Baltimore, MD, USA, 2003.

[11] McCallum, A., “Bow: A Toolkit for Statistical Language Modelling, Text Retrieval, Classification and Clustering”. From www-2.cs.cmu.edu/ ~mccallum/bow.

[12] Pachet, F. & Zils, A., “Evolving Automatically High-Level Music Descriptors from Acoustic Signals”, *Lecture Notes In Computer Science*, Vol. 2771, 42-53 Springer Verlag, Heidelberg Germany, 2004.

[13] Russell, J.A., Weiss, A. & Mendelsohn, G.A., “Affect Grid: A Single-Item Scale of Pleasure and Arousal”, *Journal of Personality and Social Psychology*, 57, 3, 495-502, 1989.

[14] Russell, J.A., “Core affect and the psychological construction of emotion”, *Psychological Review* Vol. 110, No. 1, 145-172, Jan 2003.

[15] Scherer, K.R., “Toward a dynamic theory of emotion”, *Geneva Studies in Emotion,* No. 1, 1-96 Geneva Switzerland, 1987.

[16] Sloboda, J.A. and Juslin, P.N. “Psychological perspective on emotion”, in Juslin, P.N. and Sloboda, J.A.(eds.), *Music and Emotion,* Oxford University Press, New York, NY, USA, 2001.

[17] Stanfield, G.R., “System and methods for training a trainee to classify fundamental properties of media entities”, US Patent Application No. 20030041066, Feb 27, 2003.

[18] Stone, P. J., *The general inquirer a computer approach to content analysis.* MIT Press, Cambridge MA USA, 1966.

[19] Tellegen, A., Watson, D. & Clark, L.A., “On the dimensional and hierarchical structure of affect”, *Psychological Science,* Vol. 10, No. 4, July 1999.

[20] Tzanetakis, G., “Manipulation, Analysis and Retrieval Systems for Audio Signals”, PhD Thesis, Princeton University, Princeton, NJ, USA, 2002.

[21] Watson, D., *Mood and Temperament,* Guilford Press, New York, NY, USA, 2000.

[22] Watson, D. & Clark, L.A, “The PANAS-X Manual for the Positive and Negative Affect Schedule – Expanded Form”. Retrieved from http://www.psychology.uiowa.edu/

[23] Witten, I.H., & Frank, E., *Data Mining: Practical machine learning tools and techniques with Java implementations,* Morgan Kaufmann, San Francisco, CA, USA, 2000.