

# MOOD EXPRESSION IN REAL-TIME COMPUTER GENERATED MUSIC USING PURE DATA

Marco Scirea, Yun-Gyung Cheong, Byung Chull Bae

IT University of Copenhagen

## ABSTRACT

This paper presents an empirical study that investigated if procedurally generated music based on a set of musical features can elicit a target mood in the music listener. Drawn from the two-dimensional affect model proposed by Russell, the musical features that we have chosen to express moods are *intensity*, *timbre*, *rhythm*, and *dissonances*. The eight types of mood investigated in this study are being *bored*, *content*, *happy*, *miserable*, *tired*, *fearful*, *peaceful*, and *alarmed*. We created 8 short music clips using PD (Pure Data) programming language, each of them represents a particular mood. We carried out a pilot study and present a preliminary result.

## 1. INTRODUCTION

In this paper we present our empirical study to explore how the manipulation of a set of musical features can express different moods in music. As an affect model, we have employed the two-dimensional affect model proposed by Russell [16]: pleasant-unpleasant and arousal-sleepy. In addition to Russell's bipolar dimensional model, we also considered Thayer's explication for the arousal - including both energetic arousal and tense arousal [19]. The musical features that we employ to express moods are *intensity* (general volume), *timbre* (brightness), *rhythm* (strength, regularity, and tempo), and *dissonances*. The first three features (i.e., intensity, timbre, rhythm) were inspired by Liu's study [12] about mood information extraction from classical music pieces. We applied the same principles to generate music instead.

Moods are different from emotions; as defined by Beedie *et al.* [3], emotions are short and intense states, while moods are less strong and may last for much longer. Although music can evoke emotions in the listeners, we concentrated on moods in this paper, with the purpose to use this work for digital games. In games, we believe that gamers are likely to listen to the background music (e.g., background music) longer than music that would evoke a particular emotion.

The research question we aim to answer in this study was “can the music generated based on the combination of the features arouse different set of moods with a fine-granularity?” We selected eight types of mood for the study: *bored*, *content*, *happy*, *miserable*, *tired*, *fearful*, *peaceful*, and *alarmed*. To this end, we first located, based on Russell's study, the coordinates of the emotion words on the two-dimensional affect space. Then, we created eight clips of mood music according to the pleasantness-arousal coordinates obtained. To generate music for our study we employed a real-time procedural music generator that we had developed using PD (Pure Data) programming language [17]. Our music generation approach

does not take into account chord sequences, leitmotifs, or improvisation. Instead, we aim to create a very minimalistic ambient music created by simple random number generators. This allows us to test our hypothesis of being able to display moods through only the manipulation of the mood defining features we consider. Despite a variety of studies towards procedural music generation [4], [6], [18], there has been little research on investigating the relationship between music and affect.

We carried out a pilot study that ten undergraduate/graduate students participated in. While listening to the generated 30 seconds music clip, each study participant specified the degree of pleasantness and arousal that they felt in the music, and then chose at most two mood words that could best represent the music. According to our early study results so far, only three affect words (*peaceful*, *alarmed*, and *bored*) were correctly perceived with reasonably good accuracies. Overall, however, the pleasantness/arousal answers from the participants were closer to our expected values.

## 2. MUSIC MOOD TAXONOMY

The set of adjectives that describe music mood and emotional response is immense and there is no accepted standard; for example Katayose *et al.* [7] use a set of adjectives including *Gloomy*, *Serious*, *Pathetic* and *Urbane*.

Russell [16] proposed a model of affect based on two bipolar dimensions: *pleasant-unpleasant* and *arousal-sleepy*, theorizing that each affect word can be mapped in this bi-dimensional space by a combination of these two components.

Thayer [19] applied Russell's model to music using as dimensions *stress* and *arousal* (see Figure 1). Although the name of the dimensions is slightly different from Russell's, their semantic meaning is the same. Since valence and arousal are commonly used terms, we will use these terms in this paper.

Thus the music is divided in four clusters as in Figure 1: **Anxious/Frantic** (Low Valence, High Arousal), **Depression** (Low Valence, Low Arousal), **Contentment** (High Valence, Low Arousal) and **Exuberance** (High Valence, High Arousal). These four clusters have the advantage of being explicit and discriminable. Also they are the basic music-induced moods (even if with different names) as discovered by Kreutz [9] (*Happiness*, *Sadness*, *Desire* and *Unrest*) and Lindstrom [11] (*Joy*, *Sadness*, *Anxiety* and *Calm*).

## 3. MUSICAL MOOD FEATURES

In order to generate mood-based music, we used four musical features – intensity, timbre, rhythm, and dissonances, which are mainly inspired by Liu *et al.* [12]. While Liu *et al.*'s research

focused on mood information extraction, we applied their approaches to generate music instead. This section extends our previous approach [17], introducing a new feature called dissonances.

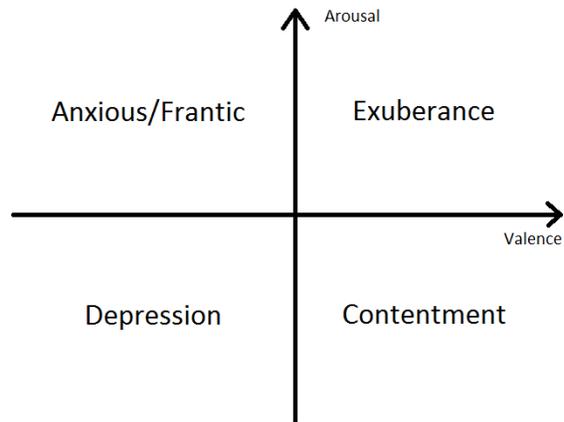


Figure 1: Thayer's bi-dimensional model [19].

### 3.1 Intensity

Intensity is defined by how strong the volume of the music is. It is an *arousal*-dependent feature: high arousal corresponds to high intensity; low arousal to low intensity.

Intuitively the more stress is present in the music, the more it will have a high volume. Calm pieces of music, in a similar manner, have a lower one.

### 3.2 Timbre

Timbre is what we could call the brightness of the music, that is, how much of the audio signal is composed by bass frequencies. It is often associated with "how pleasing to listeners"[1]. In previous literature audio features such as MFCC (Mel-Frequency Cepstral Coefficients) and spectral shape features have been used to analyze this timbral feature.

We associated this timbral feature with *valence*: the more positive the valence, the higher will the timbre be. The brightness of *Exuberance* music, for example, is generally higher than that of music in *Depression*, which will result in greater spectral energy in the high sub bands for *Exuberance*.

Generally, timbre is a factor that is very dependent on the instrumentation choice. In our case we acted on the synthesizers, our instruments, to generate brighter and darker sounds. In our generator we had three different sets of "instruments" (they were actually the same synthesizers with different settings to make them sound different) for high, low and neutral valence.

### 3.3 Rhythm

We included three features related to rhythm: strength, regularity and tempo [12].

- Rhythm strength: how prominent the rhythmic section is (drums and bass). This feature is *arousal* dependent.
- Regularity: how steady the rhythm is. This feature is *valence* dependent.
- Tempo: how fast the rhythm is. This feature is *arousal* dependent.

In a high valence/high arousal piece of music, for instance, we can observe that the rhythm is strong and steady. In a low valence/low arousal, on the other hand, the tempo is slow and the rhythm cannot be as easily recognized.

We acted on these features in different ways. To influence rhythm strength, we changed how much the drums are prominent in the music. Having the instruments play notes on the beat or the upbeat created different feeling of regularity and irregularity. For example, in *Contentment* music, we favored a steady rhythm with notes falling on the beats of the measure. In *Depression* music, on the other hand, we gave more space to upbeat notes. Finally, to influence the tempo we just acted on the BPMs (Beats Per Minute) of the music.

### 3.4 Dissonances

What we mean by dissonance is the juxtaposition of two notes very close to each other: for example C and C#. The distance between these two is just a semitone, which gives the listener a generally unpleasant sensation (like when you hear someone singing out of tune).

Dissonance doesn't mean that it always sounds poorly. In fact most music pieces contain dissonances, as they can be used as clues that express something's wrong. The listener's ear can also be trained to accept dissonances through repetition. In general, the bigger the interval between the two dissonant notes, the easier it is on the listener's ear: a C and a C# are always dissonant, but the dissonance is more evident if the notes are played from the same octave and not on two different ones.

Already in the first study we noticed that these features, originally devised to extract mood information, were enough to generate different moods. But we also realized that we could strengthen the impression by introducing dissonances in the music: for *Exuberance* and *Contentment* we use a diatonic scale, while for *Anxious* and *Depression* an altered one. We believe this is an important feature that cannot be ignored when wanting to show more precise moods in music.

Dissonance feature is *valence* depending. In our study we just used two scales: a C major scale (C D E F G A B) for positive and a Eb Harmonic Minor scale minus the third grade (Eb F [Gb] Ab Bb B D) for negative valence. Music built on a minor scales is generally

considered more somber than when made in a major key. This is not technically correct in our system because it would require a grade of organization and harmony that would make plain which is the root note. The notes of the harmonic minor scale are the same as the natural minor except that the seventh degree is raised by one semitone, making an augmented second between the sixth and seventh degrees. For our unstructured music this means that we have a whole-and-a-half interval between B and D and two half intervals ( *D-Eb* and *Bb-B*). The removal of the third grade (*Gb*) makes even more difficult to the listener’s ear to identify the key, effectively making the dissonances sound as such.

#### 4. AN EXPERIMENTAL PILOT STUDY

We conducted a pilot study to check whether our system could possibly represent higher definition moods and the users could recognize the differences in music. Ten students from IT University of Copenhagen, Denmark volunteered to participate in our pilot study.

As seen in Figure 2, we used Russell’s two-dimensional valence/arousal space to locate various types of mood. It is worth noting how these appear in a circular orbit around the origin, this means that the closer we get to the center the more indistinct the mood would result.

We tried to express some of the adjectives through our music generator. After noticing that some are so close in the space that they were very difficult to differentiate (e.g., *astonished* and *aroused*), we finally decided to make eight clips. This decision was also brought forth from the desire of keeping the length of the experiment below ten minutes so that the tester wouldn’t get tired and so affect the quality of the data. The final emotions that we selected are bored, content, happy, miserable, tired, fearful, peaceful and alarmed. It should be noted that we have two moods that don’t appear in Russell’s study: *fearful* and *peaceful*. With these we wanted to express some feelings that are more commonly found in music, as Russell’s study was only focused on emotions and not on music.

We defined fearful to be a mood with *medium-low valence* and *medium-high arousal*, which would put it very close to the *frustrated-annoyed-angry* cluster as in Figure 2. Peaceful, on the other hand, has *medium-high valence* and *medium-low arousal*, so it would be part of the *content-satisfied-calm* cluster. With this and Content mood, we could explore if people could see a difference between such closely located moods.

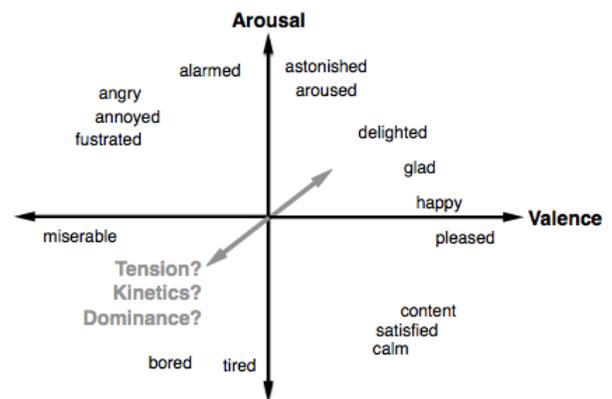
When we defined the moods, we asked the tester to place the mood they feel in the valence/arousal space. We believed this was important as people might have different definitions of the mood adjectives. It should also note that, for the majority of the participants, English was not their first language.

As the valence/arousal space is not something that most people use in ordinary life, we employed the SAM (Self Assessment Manikin) pictures with two sliders (representing a Likert value from one to five), with the texts describing the meaning of the dimensions [13].

In addition to the demographic data of the study participants, such as age and gender, data relating to their music preference (such as genre) and average time for listening to music were gathered in an open way, by having as answers a four point scale: “*always*” “*often*”, “*seldom*”, “*never*”.

Overall the experiment survey presented: a personal data section, a mood recognition questionnaire (where the participants listened to the clip and then specified which valence/arousal level they felt in the music) and finally a section in which they could select up to two emotion adjectives to describe the piece of music.

As an effort to influence the participants as little as possible with the emotion words, we divided their decision of music recognition in two parts. So while listening to the music, the participants could already set valence and arousal values without seeing the emotion adjectives. Once the clip ended, the last part of the form appeared.



**Figure 2:** The Valence-Arousal space, labeled by Russell’s direct circular projection of adjectives [16]. Includes semantic of projected third affect dimensions: “tension”[5], “kinetics” [15], “dominance” [14]. In our study we haven’t considered this third dimension as it’s still not very defined.

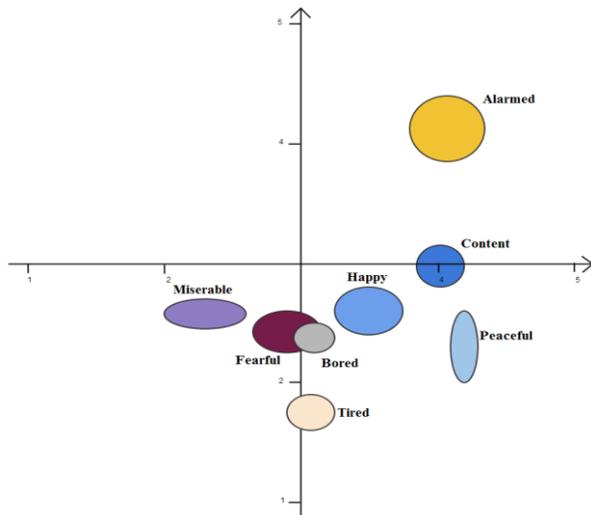
#### 5. RESULTS

Figure 3 and Table 1 shows the study results including percentage of correct answers by the study participants. The least correct answers were for *happy* music, as low as 30%. The mood types that had highest recognition were **peaceful** and **alarmed** (with an 80% of correct guesses) and **bored** (with 60%).

The mean ratings of arousal and valence for each type of mood music are summarized in Table 2, with a five point scale from 1 (most negative for valence and most calm for arousal) to 5 (most positive for valence and most active/stressful for arousal). For example, we can see how **peaceful** is perceived as a high valence (4.2) and medium/low arousal (2.3) mood, which is very close to what we expected.

Some interesting results are yielded by **fearful** and **alarmed**:

- *Fearful* mood appears to be perceived as a slightly low arousal emotion (2.4), while our expectation was for it to be medium/high.
- *Alarmed* mood is expected to be an almost pure arousal mood with just a very small negative valence (or almost neutral), but it turned out high valence (4.1) with high arousal (4.1). It seems that the participants had no issue in recognizing the arousal component, but found the music to have a positive valence.



**Figure 3:** Distribution of the mood types recognized by the participants. (Vertical axis represents Arousal and horizontal axis represents Valence.)

Bored	Content	Happy	Miserable	Tired	Fearful	Peaceful	Alarmed
60%	40%	30%	40%	40%	40%	80%	80%

**Table 1:** Percentage of correct answers by participants

	Valence	Valence STD	Arousal	Arousal STD
Peaceful	4.2	0.422	2.3	1.252
Fearful	2.9	0.994	2.4	0.699
Bored	3.1	0.568	2.4	0.516
Content	4	0.667	3	0.667
Alarmed	4.1	1.135	4.1	1.101
Happy	3.5	0.972	2.6	0.843
Miserable	2.3	1.252	2.6	0.516
Tired	3.1	0.738	1.8	0.632

**Table 2:** Average ratings of valence and arousal for various moods (scale from one to five)

## 6. CONCLUSIONS

We had some interesting results regarding emotional adjectives. There seems to be a consensus on the semantic meaning of these words is lacking and, moreover, correlations between different emotion words seem to emerge (for example *content*, *peaceful* and *happy*). This made our early analysis seem to have pretty negative results for most moods (apart from *peaceful*, *alarmed* and *bored*). By examining attentively the data, however, we could see how the results were much closer to what we expected. We should also note that the emotional meaning associated to chord sequences differs between individuals (even though some chord sequences have a more shared emotional perception) as found out by Yasuda and Abe [20].

While just early results of our pilot study with a small number of participants seem to indicate some positive results, it also shows us the problem of using emotional adjectives in this particular type of testing. As future work, we plan to conduct our study with more participants, revising the usage of adjectives to express emotions by either eliminating them completely or giving the subjects a definition of them to resolve a possible problem of ambiguousness. Another direction we are looking into currently is the use of digital games to collect ground truth data as games are known to serve as an effective platform to crowdsource data from the user [10][8][2].

## 7. REFERENCES

1. Aucouturier, J.-J., Pachet, F. and Sandler M. (2005). "The Way It Sounds": Timbre Models for Analysis and Retrieval of Music Signals
2. Barrington, L., O'Malley, D., Turnbull, D., & Lanckriet, G. (2009, June). User-centered design of a social game to tag music. In *Proceedings of the ACM SIGKDD Workshop on Human Computation* (pp. 7-10). ACM.
3. Beedie, C., Terry, P., & Lane, A. (2005). Distinctions between emotion and mood. *Cognition & Emotion*, 19(6), 847-878.
4. Brown, D. (2012, July). Mezzo: An Adaptive, Real-Time Composition Program for Game Soundtracks. In *Eighth Artificial Intelligence and Interactive Digital Entertainment Conference*.
5. Eerola, T., Lartillot, O., & Toivainen, P. (2009, October). Prediction of Multidimensional Emotional Ratings in Music from Audio Using Multivariate Regression Models. In ISMIR (pp. 621-626).
6. Houge, B. (2012) Cell-based music organization in tom clancy's endwar. In *Eighth Artificial Intelligence and Interactive Digital Entertainment Conference*.
7. Katayose, H., Imai, M., & Inokuchi, S. (1988, November). Sentiment extraction in music. In *Pattern Recognition*,

1988., *9th International Conference on* (pp. 1083-1087).  
IEEE.

8. Kim, Y. E., Schmidt, E. M., & Emelle, L. (2008, February). MoodSwings: A Collaborative Game for Music Mood Label Collection. In *ISMIR* (Vol. 8, pp. 231-236).
9. Kreutz, G., Ott, U., Teichmann, D., Osawa, P., & Vaitl, D. (2008). Using music to induce emotions: Influences of musical preference and absorption. *Psychology of music*, 36(1), 101-126.
10. Law, E. L., Von Ahn, L., Dannenberg, R. B., & Crawford, M. (2007, September). TagATune: A Game for Music and Sound Annotation. In *ISMIR* (Vol. 3, p. 2).
11. Lindström, E., Juslin, P. N., Bresin, R., & Williamon, A. (2003). "Expressivity comes from within your soul": A questionnaire study of music students' perspectives on expressivity. *Research Studies in Music Education*, 20(1), 23-47.
12. Liu, D., Lu, L., & Zhang, H. (2003, October). Automatic mood detection from acoustic music data. In *ISMIR*.
13. Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4), 261-292.
14. Mehrabian, A., & Russell, J. A. (1974). An approach to environmental psychology. the MIT Press.
15. Mion, L., & De Poli, G. (2008). Score-independent audio features for description of music expression. *Audio, Speech, and Language Processing*, IEEE Transactions on, 16(2), 458-466.
16. Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6), 1161.
17. Scirea, M. (2013). Mood Dependent Music Generator. In *Advances in Computer Entertainment* (pp. 626-629). Springer International Publishing.
18. Smith, B. D., & Garnett, G. E. (2012, July). Improvising musical structure with hierarchical neural nets. In *Proc. AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*.
19. Thayer, R. E. (1989). *The biopsychology of mood and arousal*. Oxford University Press.
20. Yasuda T., & Abe J. (2003) Are emotions perceived differently among individuals? *Music Perception*, 21(1), 156-157.