Study of soundscape emotions alteration by a blend of music signals

Stone Cheng¹; Shi-Shiang Niu²; Cheng-Kai Hsu³
¹, ², ³ National Chiao Tung University, Taiwan

ABSTRACT
This study presents an approach for analyzing the ingredient of emotions aroused by the music signals, and applied to the soundscape emotions analysis. The proposed system integrated variety of emotion models, including two-dimensional emotion space. Training process for emotion recognition is preceded in a variety of features by 192 music clips to build emotional classification model between each other in order to construct two dimensional analysis of the emotive states. Eleven features are extracted into music and audio categories. Each feature used different length of frame for analysis. The proposed system classified the category of four emotions in the emotion plane by support vector machine (SVM), and draws the variation of emotion ingredients evoked by musical signals. A Gaussian mixture model (GMM) is used to demarcate the boundaries of “Exuberance”, “Contentment”, “Anxious”, and “Depression” on the emotion plane. A graphic interface of emotion arousal locus on two-dimensional model of mood is established to represent the tracking of emotional transition. The soundscape survey procedure is carried out by studying the soundscape emotion locus tracking on selected soundscape sets to evaluate the effectiveness of emotions alteration by a blended music signals. Preliminary evaluations indicate that the proposed algorithms produce results agreed well.

Keywords: Soundscape, Music emotion, Support vector machine (SVM), Gaussian mixture model (GMM)

1. INTRODUCTION
Rapidly expanding research on music emotion classification or mood prediction assumed the music is always in a constant emotion [1,2,3]. Zhang [4] suggested that music is composed of three major elements, “intensity,” “timbre,” and “rhythm,” respectively. Intensity refers to the energy in each sub-band, timbre refers to spectral shape features, and rhythm includes rhythm strength, rhythm regularity, and tempo. Zentner et al [5] contributed to an understanding of music’s universal appeal by identifying emotions that are most frequently induced by music and by deriving and replicating a structural model of music-induced emotions using rigorous analytic techniques. As for the system interface designed for music emotion expression, Laurier and Herrer [6] developed “Mood cloud” system which can instantaneously display the ratio presented by five major kinds of emotion when playing music. The five major kinds of emotion include: “happy”, “sad”, “aggressive”, “relax”, and “party”.

In this study, Thayer’s model of mood [7] is adopted as the emotion plane of classification and detection, as shown in Figure 1. The model is consisted of four quadrants: (i)Contentment, (ii)Depression, (iii)Anxious, (iv)Exuberance. The horizontal axis is the stress/valance, and the vertical axis is the energy/arousal. Although such classification has few emotional adjectives, the model can roughly distinguish the common emotional response. All moods can get a reasonable explanation and a suitable position by the two-axis coordinate in this model. The horizontal axis is defined as the emotional feeling of pressure; the vertical axis is defined as the emotional feelings of the energy. The compositions of music have many common characteristics such as volume, tempo, mode, timbre, etc. These characteristics own some obvious relationship with the model’s horizontal axis and vertical axis. For example, fast tempo and high volume also represent high energy music, a minor music and low frequency timbre lead to depression atmosphere. However, music performances
might have different human response even in the same section and does not lead the listener to a single emotion from the start to the end. Several studies use multiple emotion labels to determine what emotion this music is [8]. Two main approaches to music information retrieval (MIR) are metadata retrieval and content-based retrieval. Existing MIR systems select music files from large databases according to the music content, such as artist, title, genre, acoustic feature, melody, and rhythm. To find the music files that are similar in styles or emotions is a perception manner to create listening situation. Therefore, how to browse music clips based on emotion effectively and efficiently has become an important line of research. Higher volume intensity in classical music would lead to a dramatic emotional change is set as an assumption on the course of music emotion tracking [9]. Based on this empirical rule, music emotion identification can be done by locating the section of possible emotional changes, and assuming that each section is a stable emotion. But another point is that the music’s mood involves a lot of subjective cognition of anthropogenic factors and personal background or experience. During the period of music playing, will the listener’s emotion remain on the same state? It can be presumed that people may have different emotional responses when appreciating different music genres, and different music genres are characterized by different emotions likewise. The mechanisms of emotions evoked by music depend on the subjective nature of the individual’s perception and emotional experience. What are the ingredients of the emotions induced by musical section structures? The soundscape survey procedure may be carried out both using traditional noise mapping procedures and studying people assessments on psychoacoustic parameters collected through a structured sound emotions analysis. In particular, with reference to the soundscape perceived in a specific environment, the following emotion-related questions are usually on the evaluation sheet: “loudness”, “pleasant” to “unpleasant”, “stressful” to “relaxing”, “boring” to “vibrant”. Can the time-varying ingredients of soundscape emotions analysis provide the qualitative data about the subjective acoustic perceptions? Which emotive states are most (or least) frequently induced by music? Instead of indicating what specific mood the whole music is and make musical emotion group classification, this study addresses above questions by adopting an algorithm on evaluating real-time time-varying emotion ingredients on Thayer’s emotion model to trace the emotions progressively aroused by the flow of music signals, which might further close to the individual listener perception.

Figure 1 – Four-factorial model of music-induced emotions. (a) Emotion Model of various studies (b) Thayer’s Emotion Model

2. SYSTEM ARCHITECTURE

The most common method for emotion-based music information retrieval classification in the literatures has been to extract audio characteristics from music which are standalone values taken as an average over the entire piece or clip. Figure 2 illustrates the proposed sequential framework system for real-time emotion ingredients tracking of music signals. The framework progressively characterizes music-induced emotions. Both testing data and training data are WAVE file of music. We use 192 clips of music in four emotion groups to train the system. Each mood classification category contained 48 emotional pre-defined music data. Eleven features (Figure 3) are extracted in two...
categories, music and audio genre. Music related features are extracted (i) density of music event, (ii) sound volume, (iii) timbre, (iv) mode, (v) dissonance. These features directly related to emotions.[10,11]

Figure 2 – Block Diagram of the System.

<table>
<thead>
<tr>
<th>Features</th>
<th>Feature Genre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key</td>
<td>Music</td>
</tr>
<tr>
<td>Key clarity</td>
<td></td>
</tr>
<tr>
<td>Mode</td>
<td></td>
</tr>
<tr>
<td>Tonal centroid</td>
<td>Audio</td>
</tr>
<tr>
<td>Spread</td>
<td></td>
</tr>
<tr>
<td>Flatness</td>
<td></td>
</tr>
<tr>
<td>Centroid</td>
<td></td>
</tr>
<tr>
<td>Audio Spectrum Envelope</td>
<td></td>
</tr>
<tr>
<td>Audio Spectrum Centroid</td>
<td></td>
</tr>
<tr>
<td>Audio Spectrum Spread</td>
<td></td>
</tr>
<tr>
<td>Audio Spectrum Flatness</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3 – Features Extraction items

3. EMOTIONS CLASSIFICATION AND SCORE COUNTING

Scoring method is used to simulate the process of music listening. Its concept is that the listeners’ emotion is influenced by the feeling at the previous time period, and as the time goes on, the emotional feeling will also release gradually. The present emotion position $P_t$ of listeners is defined on the emotion plane, and $P_t$ is the accumulation of the past listening process. The intensities of previous three features and the specific weight $w$ determine the present emotion score $P_t$. As shown in Eq. (1), $f$ is the index of feature, representing the three different features, $S_f(t)$ is the intensity of $f$-th feature at the time $t$, and $w_x(f)$, $w_y(f)$ are the weights of each feature on the x axis and y axis on the emotion plane. The whole scoring process is shown in Figure 4. In the accumulating process of $P_t$, the accumulated score $p_t$ at every moment multiplies the decay function, representing the decaying of time, i.e. the release or decaying of emotion. The decay function limits the range of emotions scoring.

$$p_t = \sum_{f=1}^{n} \left[ w_x(f) \cdot S_f(t) \cdot x + w_y(f) \cdot S_f(t) \cdot y \right]$$

In the emotion tracking diagram for the training music data, the final emotion scores of stress and energy are marked on the emotion plane to displays the results of emotion score counting. Through all training data displacement of coordinates of the emotion track, with the manually mood classification by listener, the SVM classifier is employed to compute the distribution and boundary of each type of the emotion model. The details of music emotion trajectory and emotion boundaries training are discussed in next section. To classify the music clip complex moods, audio characteristics are processed with a support vector machine (SVM) classifier. An SVM works as a binary classifier by taking a set of input data and predicts which of two possible classes the input is a member of. Figure 5 shows emotion ingredients of
each music time frame sequence. The first row shows 1/11 “Contentment”, 3/11 “Depression”, 6/11 “Anxious”, and 1/11 “Exuberance”, which indicated the emotion ingredients of each music frame.

\[ P_1 = p_1 \times \sigma(0) \]
\[ P_2 = p_1 \times \sigma(0) + p_2 \times \sigma(0) \]
\[ P_3 = p_1 \times \sigma(2) + p_2 \times \sigma(1) + p_3 \times \sigma(0) \]
\[ P_4 = p_1 \times \sigma(3) + p_2 \times \sigma(2) + p_3 \times \sigma(1) + p_4 \times \sigma(0) \]

Figure 4 – Score counting sequence

Figure 5 – Emotion ingredients of music time frame sequence. The first row shows 1/11 “Contentment”, 3/11 “Depression”, 6/11 “Anxious”, and 1/11 “Exuberance”.

4. BOUNDARIES OF THE EMOTION PLANE

4.1 A. Emotional Boundaries and Graphical Interface

As the music signals proceed, the listener will produce a general emotional response after the end of music fragments. In this study, 192 music samples which had been marked by listener will be processed through audio frame down-sampling; feature extraction; emotion score counting as mentioned above.

Figure 6 – The trajectory distribution of 192 training samples on the music emotion plane.
All 192 training music's trajectories are marked as shown in Figure 6, green mark trajectories group represents comfortable mood music tracks, blue mark for sad mood music tracks, red mark for anxiety, yellow mark for uplifting. Nevertheless, when using different training music samples, the result trajectories will have slightly different distribution.

Figure 6 shows that different mood trajectories indeed have different distributions. A GMM is used to calculate the distribution and the PDF (Probability Density Function) for each mood trajectories. For the summative emotion class of the emotion trajectory coordinate, the example with GMM as the classifier is as follows: the original PDF of each class is deemed as the distribution of several Gauss function and after a large amount of data training; the process is shown as Figure 7. Eventually, the music emotional distribution and boundaries of four emotions are constructed on the emotion plane. This graphical interface is used to track the emotion transition caused by complete test music.

In testing mode, the proposed approach is evaluated with our testing database. We first present the performance of music emotion tracking locus on selected music clip. Figure 8 illustrates an example mood tracking result of a testing music piece, “Tchaikovsky 1812 Overture”, drawn the locus of arousal emotions. It can be seen from Figure 8 that most all of the emotions are in “Contentment” using the proposed algorithm. In the mood tracking experiments, we also find that it is possible that soundscape clips may be applied by the proposed framework, as shown in Figure 9. In such a case, an individual mood of soundscape may not provide enough information, and it would be better to provide several emotions with the corresponding confidences.
5. SYSTEM EXPERIMENTS

Soundscapes vary over time (minutes/hours) and this can affect people’s perception of them. The soundscape survey procedure is carried out by conducting a soundscape emotions evaluation and alteration by a blend of music signals to compare the results from proposed system analysis the actual listener experience. In testing mode, the proposed approach is evaluated with our testing database. In this framework it is possible to design for particular and predictable psychological effects by tracking the emotions of soundscape. The annoyance sound may be attributable to the sonic environment to understanding soundscape quality, allows one to engage in soundscape design for quality of life. We first present the performance of soundscape emotion tracking locus on selected soundscape sets. Figure 10 shows the proceeding of test soundscape drawn the locus of arousal emotions. The test soundscape is rivulet sound with cicadas chirping. The locus of soundscape’s emotion are in “Anxious” mostly (56 secs), which can be connected to results regarding noise sensitivity. Distal situational awareness is predominantly determined by the loudest (foreground) sound events and proximal situational awareness by the subtle (background) sounds.[12] We choose a music clip with its emotion locus is in the range of “Contentment”, as shown in Figure 11. The emotion locus of blended audio signals is shown in Figure 12. Our research indicates that an annoying sound is the main reason of its annoyance emotion. In particular people complained because of its constant, frequent, or unpredictable presence or because it had particular source properties. The blended soundscape experiment suggests that more subtle interventions that address the attention attracting properties might often be possible. For example, it may make singing sounds more audible and in doing so shift the appraisal of the sonic environment towards the “Contentment” emotion. The loudness effect of this intervention may be small, while its well-being effect may be important.

This study was aimed at developing a new assessment technique for soundscapes. These results highlight the complexity involved in defining a soundscape affectional effect, and have implications for measuring and designing soundscape within multi-purpose environments. However, this study has successfully altered the emotions locus tracking by blending a music clip.
Figure 10 – Emotion locus of soundscape sample: Rivulet Stream-Nature Sounds.

Figure 11 – Emotion locus of singing song sample

Figure 12 – Analysis on ratings of emotional responses to a blend soundscape emotion
6. CONCLUSIONS

This work presents an approach for analyzing the ingredients of emotions evoked by the musical signals. Eleven feature sets, including “key”, “key clarity”, “mode”, “tonal centroid”, “audio spectrum envelope”, “audio spectrum centroid”, “audio spectrum spread”, “Spread”, “Centroid”, and “Flatness” are extracted from WAV file of music signal to represent the characteristics of a music clip. 192 clips of emotion-predefined music are used to train the system and mark the trace on the Thayer’s model of emotion plane, which is composed of four music emotion categories, “Contentment”, “Depression”, “Exuberance”, and “Anxious”. GMM algorithm is used to calculate the distribution of emotion group and demarcate the emotions classification based on the training data. A graphic interface of emotions arousal locus on emotion plane is established to represent the real-time tracking of dynamic emotional ingredients caused by music under testing. We present the performance of soundscape emotion tracking locus on selected soundscape sets (rivulet sound with cicadas chirping). The locus of test soundscape’s emotion is in “Anxious” mostly, which may be considered as an annoying sound. By choosing music clip with its emotion locus is in the range of “Contentment” and blended with soundscape sample, the emotion locus of audio signals is switched to “Contentment” emotion. Preliminary evaluations indicate that the proposed algorithms produce soundscape emotions alternation is achieved in this case.

REFERENCES