J Supercomput (2013) 65:274–286
DOI 10.1007/s11227-010-0447-6

Bridging the semantic gap in multimedia emotion/mood recognition for ubiquitous computing environment

Seungmin Rho · Sang-Soo Yeo

Published online: 4 June 2010

Abstract With the advent of the ubiquitous era, multimedia emotion/mood could be used as an important clue in multimedia understanding, retrieval, recommendation, and some other multimedia applications. Many issues for multimedia emotion recognition have been addressed by different disciplines such as physiology, psychology, cognitive science, and musicology. Recently, many researchers have tried to uncover the relationship between multimedia contents such as image or music and emotion in many applications. In this paper, we introduce the existing emotion models and acoustic features. We also present a comparison of different emotion/mood recognition methods.

Keywords Emotion/Mood recognition · Multimedia features · Semantic analysis · Ubiquitous computing

1 Introduction

Recently, with advances in the field of multimedia information retrieval, we face a new possibility that multimedia can be automatically analyzed and understandable by the computer to some semantic level. Due to the diversity and richness of music content, many researchers have been pursuing a multitude of research topics in this field, ranging from computer science, digital signal processing, mathematics, and

S. Rho
School of Electrical Engineering, Korea University, Anam-Dong, Seongbuk-Gu, Seoul, 136-713, Korea
e-mail: smrho@korea.ac.kr

S.-S. Yeo (✉)
Division of Computer Engineering, Mokwon University, Do-An-dong, Seo-gu, Daejeon 302-729, Korea
e-mail: ssyeo@msn.com

statistics applied to musicology. Most traditional content-based multimedia retrieval (CBMR) techniques [1, 25] have focused on low-level features such as energy, zero crossing rate, audio spectrum, etc. However, these features were not enough to give semantic information of multimedia contents (especially in music) and gave serious limitation in retrieving and recommending appropriate music in various situations within the same time but different location. For example, a person wants to listen to soft music when he/she wakes up in the morning and he/she prefers fast beat music when he/she exercises at the gym. For this, more semantic information such as mood and emotion should be recognized from low level features such as beat, pitch, rhythm, and tempo. Here is another example, which shows a person is very sad for some reason. Depending on his/her personality, he/she may want to listen to some music that may cheer him/her up, or easy music that can make him/her calm down.

Due to the abovementioned limitations of low-level feature-based approaches, some researchers have tried to bridge the semantic difference, which is also known as semantic gap, between the low-level features and high-level concepts. With low-level feature analysis only, we might experience many difficulties in identifying the semantics of musical content. Similarly, it is difficult to correlate high-level features and the semantics of music. For instance, a user's profile, which includes educational background, age, gender, and musical taste, is one possible high-level feature. Due to this semantic gap, Semantic Web technology is considered as one of promising methods to bridge it. Recently with the development of Semantic Web, ontology has been widely used to make easy knowledge sharing and reusing.

In the following section, we describe a brief overview on the most widely used multimedia emotion/mood models. In Sect. 3, we illustrate various multimedia features. Section 4 describes diverse emotion/mood recognition methods and compares among them. In Sect. 5, we introduce the music emotion recognition system using semantic web technology. In the last section, we conclude the paper.

2 Related work

Many researchers have explored models of emotions and factors that give rise to the perception of emotion in multimedia. Many other researchers investigate the problem of automatically recognizing emotion in multimedia.

2.1 Definition of emotion and mood

In many cases, the terms emotion and mood have been used interchangeably. It is noted that, in most psychology related books and papers [21], "emotion" usually has a short duration (seconds to minutes) while "mood" has a longer duration (hours or days). The idea that the semantic meaning attached to the word "emotion" inherently carries an ephemeral connotation with it, whereas the word "mood" carries with it an extended time frame that begins with listening and ends at some indiscriminate point after listening is complete.

2.2 Multimedia and emotion/mood

During the last decade, many researchers have investigated the influence of music factors like loudness and tonality on the perceived emotional expression [16]. They analyzed those factors using diverse techniques, some of which are involved in measuring psychological and physiological correlation between the state of particular musical factor and emotion evocation. Juslin and Sloboda [9] investigated the utilization of acoustic cues in the communication of music emotions by performers and listeners and measured the correlation between emotional expressions (such as anger, sadness, and happiness) and acoustic cues (such as tempo, spectrum, and articulation).

2.3 Multimedia emotion/mood model

Traditional mood and emotion research in multimedia has focused on finding psychological and physiological factors that influence emotion recognition and classification. During the 1980s, several emotion models were proposed, which were largely based on the dimensional approach for emotion rating.

Ortony, Clore, and Collins [18] developed their theoretical approach under the assumption that emotions develop as a consequence of certain cognitions and interpretations. Therefore, their theory exclusively concentrates on the cognitive elicitors of emotions. The authors claimed that these cognitions are determined by three aspects: events, agents, and objects. These events, agents or objects are appraised according to an individual's goals, standards, and attitudes. Emotions represent valenced reactions to these perceptions of the world. One can be pleased about the consequences of an event or not (pleased/displeased); one can endorse or reject the actions of an agent (approve/disapprove) or one can like or not like aspects of an object (like/dislike). On the other hand, the dimensional approach focused on identifying emotions based on their location on a small number of dimensions such as valence and activity.

The dimensional approach focuses on identifying emotions based on their location on a small number of dimensions such as valence and activity. Russell's [27] circumflex model has had a significant effect on emotion research. This model defines a two-dimensional, circular structure involving the dimensions of activation and valence. Within this structure, emotions that are across the circle from one another, such as sadness and happiness, correlate inversely. Thayer [29] suggested a two-dimensional emotion model that is simple but powerful in organizing different emotion responses: stress and energy. The dimension of stress is called *valence* while the dimension of energy is called *arousal*.

Figure 1 shows the dimensional emotion model and Fig. 2 illustrates emotional model modified from Thayer's two-dimensional model. As shown in Fig. 2, the two-dimensional emotion plane can be divided into four quadrants with various emotion adjectives placed over them.

2.4 Multimedia emotion recognition

Automatic emotion detection and recognition in speech and music is growing rapidly with the technological advances of digital signal processing and various effective

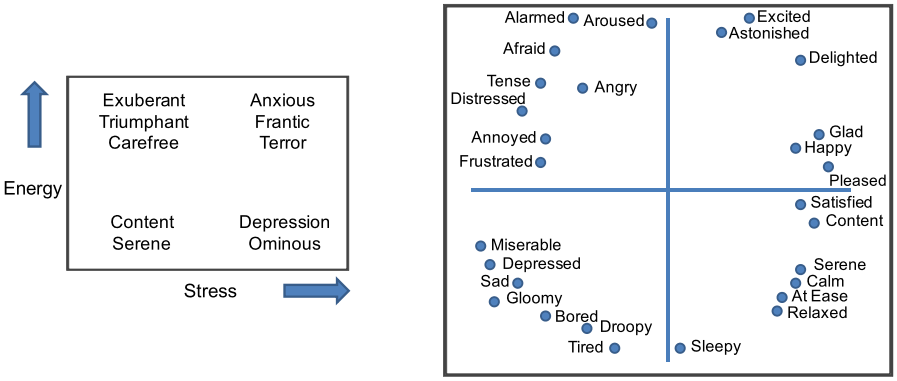

Fig. 1 Thayer's two-dimensional emotion model [29] (*Left*) and Russell's circumplex model [27] (*Right*)

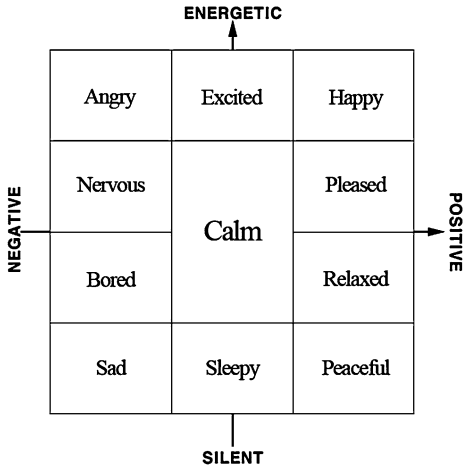

Fig. 2 Modified Thayer's two-dimensional emotion model [26]

feature extraction methods. Emotion recognition can play an important role in many other potential applications such as music entertainment and human-computer interaction systems.

One of the first studies of emotion detection in music is presented by Feng et al. [6]. Their work, based on Computational Media Aesthetics (CMA), analyzes two dimensions of tempo and articulation which are mapped into four categories of moods: happiness, anger, sadness, and fear. Lu et al. [16] developed a hierarchical framework for extracting music emotion automatically from acoustic music data. They used music intensity to represent the energy dimension of the Thayer model, and timbre and rhythm for the stress dimension.

FEELTRACE [3] is software that is designed to let observers track the emotional content of stimuli (such as words, faces, music, and video) as they perceive it and taking full account of gradation and variation over time. Yang et al. [33] developed a music emotion recognition (MER) system from a continuous perspective and represented

each song as a point in the emotion plane. They also proposed a novel arousal/valence computation method based on regression theory.

Dunker et al. [4] present a multimodal mood classification framework by using various audio features such as MFCC, Spectral Centroid, Spectral Crest Factor, Audio Spectrum Flatness, and visual features Color Histogram, Haar Wavelet, and Color Temperature Histogram, respectively. They also present multiple classifiers and models such as the Gaussian Mixture Model (GMM) and the Support Vector Machine (SVM).

3 Multimedia features

There are many low-level and high-level features that can be computed from acoustic musical signals. These following conceptual features are originally classified by Juslin and Slobada [9]. To automatically recognize emotions from multimedia contents, the researchers used both spectral features (including Spectral Centroid, Spectral Flatness, Spectral Spread, Spectral Flux, and Mel-Frequency Cepstral Coefficients) and temporal (including Scale, Intensity, Rhythm, and Harmonicity), respectively [23, 26].

3.1 Conceptual emotion/mood features

3.1.1 Tempo

Tempo is usually considered the most important factor when affecting emotional expression in music. For example, BPM (Beats per Minute) = 100 or designation such as presto, allegro, moderato, andante, and adagio. Fast tempo may be associated with various expressions of activity/excitement, happiness/joy/pleasantness, and fear. Slow tempo may be associated with various expressions of calmness/serenity, sadness, tenderness, and disgust.

3.1.2 Loudness

Loud music may be associated with various expressions of intensity/power, tension, anger, and joy, and soft music with softness, tenderness, sadness, solemnity, fear, and activity.

3.1.3 Pitch

High pitch may be associated with various expressions of happy, graceful, serene, dreamy, and exciting. Low pitch may suggest dignity/solemnity, sadness, and excitement, as well as boredom.

3.1.4 Melody

Wide melodic range may be associated with joy, whimsicality, and uneasiness, and a narrow range with expressions such as sad, dignified, sentimental, tranquil, delicate, and triumphant.

3.1.5 Harmony

Consonant harmony may be associated with expressions such as happy/gay, relaxed, graceful, serene, dreamy, and majestic; and dissonant harmony with excitement, tension, and sadness.

3.1.6 Rhythm

Regular and smooth rhythm may be perceived as expressing happiness, dignity, majesty, and peace; irregular and rough rhythm, amusement, uneasiness, and anger.

3.1.7 Timbre

Timbre with many harmonics may suggest anger, disgust, fear, activity, or surprise. Timbre with few and low harmonics may be associated with pleasantness, boredom, happiness, or sadness.

3.2 Spectral features

3.2.1 Mel-frequency cepstral coefficients (MFCC)

MFCC is commonly used in speech recognition system. Because of its good discriminating ability, it is also used in audio classification system [7, 13]. These are computed from FFT (Fast Fourier Transform). The log spectral coefficients are perceptually weighted by a nonlinear map of the frequency scale, which is called Mel-scaling, using a triangular band-pass filter bank. Then the Mel-weighed spectrum is transformed into MFCC with the COS transformation.

$$C_n = \sqrt{\frac{2}{K}} \sum_{k=1}^{K} (\log S_k) \cos\left[\frac{n(k-0.5)\pi}{K}\right] \quad n = 1, 2, \ldots, L \tag{1}$$

Here, K is the number of band-pass filters, S_k is the Mel-weighted spectrum after passing kth triangular band-pass filter, and L is the order of the cepstrum.

3.2.2 Spectrum flux (SF)

Spectrum Flux is defined as the average variation of spectrum between the adjacent two frames in an audio segment,

$$SF = \frac{1}{(N-1)(K-1)} \times \sum_{n=1}^{N-1} \sum_{k=1}^{K-1} \left[\log(A(n,k)+\delta) - \log(A(n-1,k)+\delta)\right]^2 \tag{2}$$

where

$$A(n,k) = \left| \sum x(m) w(nL - m) e^{-j\frac{2\pi}{L}km} \right| \tag{3}$$

and $x(m)$ is the input discrete audio signal, *w(m)* the window function; L is the window length; K is the order of DFT (Discrete Fourier Transform), a very small value to avoid calculation overflow, and N is the total frame number in one audio segment.

3.2.3 Spectrum centroid (SC)

The spectral centroid is a measure used for characterizing a spectrum. It indicates where the "center of mass" of the spectrum is. Perceptually, it has a robust connection with the impression of "brightness" of a sound [28]. It is calculated as the weighted mean of the frequencies present in the signal, with their magnitudes as the weights. Equation for computing spectral centroid is as follows:

$$C_t = \frac{\sum_{n=1}^{N} M_t[n] \times n}{\sum_{n=1}^{N} M_t[n]} \tag{4}$$

where $M_t[n]$ is the magnitude of the Fourier transform at frame t and frequency bin n.

3.2.4 Spectrum rolloff (SR)

The rolloff is another measure of spectral shape. The spectral rolloff is defined as the frequency R_t below which 85% of the magnitude distribution is concentrated. The following equations show the estimation of spectral rolloff.

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 \times \sum_{n=1}^{N} M_t[n] \tag{5}$$

3.2.5 Short-time average energy function

Short-time average energy is one of the important features in audio processing. The average energy indicates the loudness of the audio signal. It is easy to separate the voice and noise signals using the average energy function. The short-time energy function is defined by

$$E_m = \frac{1}{N} \sum_{n=0}^{N-1} x(n)^2 \times h(m-n) \tag{6}$$

where m is the number of samples, $x(n)$ is the input signal and $h(m)$ is a hamming window function reduce the spectral leakage. For a square window of size N, $h(m)$ is expressed as

$$h(m) = \begin{cases} 0.54 - 0.46\cos(\frac{2\pi m}{N-1}) & \text{for } 0 \le m \le N-1 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

3.2.6 Average zero crossing rate

The zero crossing rate (ZCR) indicates the frequency of signal amplitude sign change. To some extent, it indicates the average signal frequency. The average zero crossing rate is calculated as follows:

$$\mathrm{ZCR} = \frac{\sum_{n=1}^{N} |\operatorname{sgn} x(n) - \operatorname{sgn} x(n-1)|}{2N} \tag{8}$$

where $\operatorname{sgn} x(n)$ is the sign of $x(n)$ and is 1 if $x(n)$ is positive and -1 if $x(n)$ is negative.

3.2.7 Color histogram

The color histogram measures the distance between adjacent pixels based on the distribution of luminance levels. To construct the color histogram, Rho and Park [24] compared each pixel of an image with all 24 moods, which are defined by Yang and Lee [32], to find the closest mood, then attribute it to that color mood's bin in the color histogram. A reasonable assumption about the luminance of a pixel L_{p} is that is modeled as

$$L_{\mathrm{p}} = W_R \times (R) + W_G \times (G) + W_B \times (B) \tag{9}$$

that is a weighted linear combination of 8-bit red (R), green (G), and blue (B) components, W_k being suitable weights [20].

They computed 64-bin gray-scale histograms over each image. H is a 64-bin histogram computed by counting the number of pixels in each bin of 4 gray levels, so

$$H[k] = \#\text{ of pixels} \quad (k = L_{\mathrm{p}}/4,\ 0 \le k \le 63) \tag{10}$$

3.3 Temporal features

3.3.1 Scale

Scale is an overall rule of the tonic formation of music. In [26], Rho et al. defined scale as a set of key, mode, and tonality. For accurate scale features, they first computed chromagram from the frequency representation of musical signal. After that, they applied the key profile matrix proposed by Krumhansl [11]. The following equations show the process of combining the chromagram and key characterization:

$$\mathit{Tonality} = \mathbf{C} \cdot \mathbf{KeyProfileMatrix} \tag{11}$$

$$\mathit{Key} = \max_{\mathit{KeyIndex}} (\mathit{Tonality}(\mathit{Idx})) \tag{12}$$

Here, vector **C** has 12 elements and represents the summed chromagram of each acoustic frame. **KeyProfileMatrix** is a key profile matrix composed of 12-by-24 elements. *KeyIndex* indexes **KeyProfileMatrix**, where $\mathit{KeyIndex} = 1, 2, \ldots, 24$. From the inner product of **C** and **KeyProfileMatrix** in (11), they obtained a tonality score for each key. Finally, they picked the key having the maximum tonality in (12) as the most appropriate one.

3.3.2 Intensity

The average energy (AE) of the overall wave sequence is widely adopted to represent the loudness of music and its standard deviation (σ) can be used to present the regularity of the loudness. AE is defined by:

$$\mathrm{AE}(x) = \frac{1}{N}\sum_{t=0}^{N} x(t)^2, \qquad \sigma(\mathrm{AE}(x)) = \sqrt{\frac{1}{N}\sum_{t=0}^{N}\big(\mathrm{AE}(x) - x(t)\big)^2} \tag{13}$$

where x is an input discrete signal, t is the time in sampling units, and N is the length of x in the sample.

3.3.3 Rhythm

Rhythm, which is composed of rhythmic features such as tempo and beat, is one of the most important elements in music. Beat is a fundamental rhythmic element of music. Tempo is usually defined as the beats per minute (BPM) and is used to represent the global rhythmic feature of music. The tempo and regularity of the beats can be measured in various ways. For beat tracking and tempo analysis, Rho et al. [26] used the algorithm proposed by Ellis et al. [5]. As the rhythm feature, they used the overall tempo (in beats per minute) and the standard deviation of the beat intervals.

3.3.4 Harmonicity

Harmonics can be observed in musical tones. In monophonic music, harmonics are easily observed in the spectrogram. However, it is difficult to find harmonics in polyphony, because many instruments and voices are performed at the same time. This problem can be solved by computing the harmonicity as follows:

$$H(f) = \frac{\max_{f=1}^{N}\left(\sum_{k=1}^{M}\min(\|X(f)\|, \|X(kf)\|)\right)}{\frac{1}{N}\sum_{f=1}^{N}\|X(f)\|} \tag{14}$$

Here, M denotes the maximum numbers of overtones considered, f is the fundamental frequency, and X is the short-time Fourier transform (STFT) of the source signal. In the equation, the *min* function is used in such a way that only the strong fundamental and strong harmonics result in a large value for H. In [26], Rho et al. measured average of each frequency using (14).

4 Emotion/mood recognition methods

Automatic emotion recognition and extraction in multimedia systems is growing rapidly with the advancement of digital signal processing, audio analysis and feature extraction tools [31]. Emotion/Mood recognition in music is beginning to be seen as a relevant field of music information retrieval and promises to be an effective means of classifying songs [17]. Recent research works by Carvalho and Chao [2], Li and

Table 1 Comparison of current emotion/mood recognition methods

	Emotion model	Features	Emotions	Dataset (Size)	Precision
[2]	–	Musical Surface*, Spectral Flatness Measure, Spectral Crest Factor, MFCC	Happiness, Spooky, Fear, Passionate	200	63.45∼67%
[14]	–	Musical Surface, MFCC, DWCH	Cheerful, Depressing, Relaxing, Exciting, Comforting, Disturbing	235	70∼83%
[26]	–	Tempo, Articulation	Happiness, Sadness, Anger, Fear	353	75 ∼86%
[15]	Thayer's model	Centroid, Roll off, Spectral Flux	Depression, Contentment, Exuberance, Anxious	250	76.6 ∼94.5%
[32]	TWC model	12 features from Sony EDS System [20]	Hostility, Sadness, Pride, Guilt and Love, Excitement	152	82.8%
[8]	Modified Thayer's model	Pitch, Tempo, Loudness, Tonality, Key, Rhythm, Harmonics	Angry, Bored, Calm, Excited, Happy, Nervous, Peaceful, Pleased, Relaxed, Sad, Sleepy	165	91.52 ∼94.55%

*Musical Surface = {Centroid, Roll off, Spectral Flux, Zero Crossings, Average Silence Ratio}

Ogihara [14], Yang and Lee [32], and Han et al. [8], among others, will be discussed below. Table 1 shows the comparison of emotion/mood recognition methods.

One of the pioneer research works on emotion detection in music is presented by Feng et al. [6]. They presented mood detection on the viewpoint of Computational Media Aesthetics by analyzing two music dimensions, tempo and articulation. In the procedure of making music, they derived four categories of mood: happiness, anger, sadness, and fear. This categorization is based on both Thayer's two-dimensional model [29] and Juslin's theory [9].

Liu et al. [15] classified various features into three categories: intensity, timber, and rhythm. All the moods were mapped into the Thayer's two-dimensional space [29]. Also, they used the Gaussian mixture model (GMM) as a classifier. To track music mood, they considered musical mood variation and proposed a mood boundary detection method based on threshold adaptation.

Another emotion recognition research is done by Li and Ogihara [14]. Their system extracts relevant audio descriptors such as timbral features using both Daubechies Wavelet Coefficient Histograms and MARSYAS system of Tzanetakis and Cook [30]. They considered emotion recognition as a multiclass classification problem and trained on the extracted the features using the Support Vector Machine.

Yang and Lee [32] implemented a system called *Emo*, a music annotation prototype system, which combines inputs from both human and software agents to better study human listening. 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.

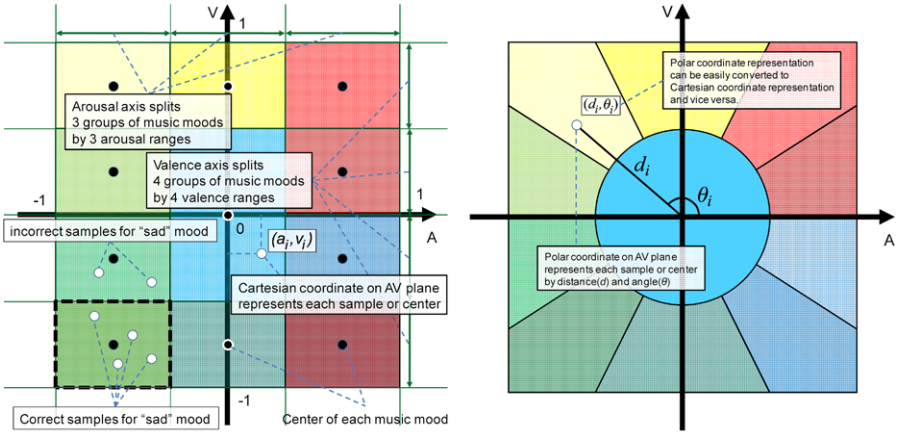

Fig. 3 Music emotion separation policy in AV plane in *Cartesian* (**a**) and *Polar* (**b**) representation [8]

Rho et al. [26] developed a music recommendation system and emotion recognition system. The recognition process consists of three steps: (i) several distinct features were extracted from music; (ii) those features were mapped into eleven emotion categories on Thayer's two-dimensional emotion model (illustrated in Fig. 2); (iii) two regression functions were trained using SVR and then arousal and valence values were predicted. For emotion recognition, they reformulated it into a regression problem based on support vector regression (SVR). Through the use of the SVR-based emotion classifier, they achieved 87.8% accuracy.

Han et al. [8] also developed an emotion recognition system. They have tested SVR-based emotion classifier in both *Cartesian* and *Polar* coordinate system empirically. They observed that the result indicates the SVR classifier in the *Polar* representation produces satisfactory result which reaches 94.55% accuracy superior to the SVR (in Cartesian) and other machine learning classification algorithms such as SVM and GMM. The motivation for *Polar* coordinates came from observing misclassifications of Calm (refer to Fig. 2), and the results suggest that the Calm/not-Calm dimension (radius) is somehow more natural for classification than Arousal and Valence. Their proposed emotion separation policy in the Arousal–Valence (AV) plane is shown in Fig. 3.

5 Emotion recognition using semantic technology in ubiquitous computing environment

In [26], Rho et al. developed extended music ontologies to enable mood and situation reasoning in a music recommendation system. Those ontologies are described in Web Ontology Language (OWL) [19] language using Protégé editor [22]. Due to its rich expressiveness and decidability, they chose DL language for representing their ontology. A DL-based reasoning engine can answer semantic queries with regard to

various types of mood and situation. For a music recommendation, they reason about the user's mood and situation using both collaborative filtering and ontology technology.

Ko et al. [10] proposed an emotion recognition system for human brain signal by measuring EEG (Electroencephalogram) signals with relative power values and a Bayesian network. They measured EEG signals related to emotion and decomposed them into 5 frequency ranges from 0 to 50 Hz. They used audio and visual contents to induce emotions, and the brainwaves were transformed into power spectrum values using a FFT (Fast Fourier Transform). Due to the difficulty of measuring human emotions, they used probability inference and a Bayesian network.

In [12], Leon et al. proposed an emotion detection mechanism that recognizes or classifies the positive and negative emotional changes in real-time using physiological signals for the pervasive computing environment. They used pre-trained AANN (Auto Associative Neural Network) in conjunction with a SPRT (Statistical probability ratio test)-based decision method to identify the emotional states. They observed that their proposed recognition system along with portable sensing equipment can integrate emotional states into decision systems in ubiquitous computing.

6 Conclusion

There are various methods for emotion/mood recognition in music via musical feature analysis; however, there is no standard method for emotion recognition. Nowadays, many researchers from different fields are paying attention for solving these problems. Therefore, we reviewed emotion/mood recognition methods and compared them. We are extending our preliminary surveys on the use of psychological model analysis and semantic technologies for the future work.

Acknowledgements This work was supported by the Korea Research Foundation Grant funded by the Korean Government [KRF-2008-357-D00223]. We also very thank Byeong-jun Han for his effort on the artwork.

References

1. Birmingham W, Dannenberg R, Pardo B (2006) An introduction to query by humming with the vocal search system. Commun ACM 49(8):49–52
2. Carvalho V, Chao C (2005) Sentiment retrieval in popular music based on sequential learning. In: SIGIR
3. Cowie R, Douglas-Cowie E, Savvidou S, McMahon E, Sawey M, Schröder M (2000) 'FEELTRACE': an instrument for recording perceived emotion in real time. In: ISCA Workshop on speech and emotion, Northern Ireland, pp 19–24
4. Dunker P, Nowak S, Begau A, Lanz C (2008) Content-based mood classification for photos and music—a generic multi-modal classification framework and evaluation approach. In: ACM international conference on multimedia information retrieval, Vancouver, Canada, pp 97–104
5. Ellis D, Poliner PW, Graham E (2007) Identifying 'cover songs' with chroma features and dynamic programming beat tracking. In: IEEE international conference on acoustics, speech, and signal processing (ICASSP), vol 4, pp 1429–1432
6. Feng Y, Zhuang Y, Pan Y (2003) Music retrieval by detecting mood via computational media aesthetics. In: Proc of the IEEE/WIC international conference on web intelligence

7. Foote JT (1997) Content-based retrieval of music and audio. In: Multimedia storage and archiving systems II. Proceedings of SPIE. SPIE Press, Bellingham, pp 138–147
8. Han B, Rho S, Dannenberg R, Hwang E (2009) SMERS: music emotion recognition using support vector regression. In: Proceedings of international society for music information retrieval, pp 651–656
9. Juslin PN, Sloboda JA (2001) Music and emotion: theory and research. Oxford University Press, Oxford
10. Ko K-E, Yang H-C, Sim K-B (2009) Emotion recognition using EEG signals with relative power values and Bayesian network. Int J Control Autom Syst 7(5):865–870
11. Krumhansl C (1990) Cognitive foundations of musical pitch. Oxford University Press, Oxford
12. Leon E, Clarke G, Callaghan V, Sepulveda F (2007) A user-independent real-time emotion recognition system for software agents in domestic environments. Eng Appl Artif Intell 20(3):337–345
13. Li SZ (2000) Content-based classification and retrieval audio using the nearest feature line method. IEEE Trans Speech Audio Process 8(5):618–625
14. Li T, Ogihara M (2004) Content-based music similarity search and emotion detection. In: ICASSP, pp 705–708
15. Liu D, Lu L, Zhang HJ (2003) Automatic mood detection from acoustic music data. In: International symposium on music information retrieval, Baltimore, Maryland, USA
16. Lu L, Liu D, Zhang HJ (2006) Automatic mood detection and tracking of music audio signals. IEEE Trans Audio, Speech Audio Process 14(1):5–18
17. Meyers O (2007) A mood-based music classification. PhD thesis, MIT
18. Ortony A, Clore GL, Collins L (1998) The cognitive structure of emotions. Cambridge University Press, Cambridge
19. OWL web ontology language (2010) Available at: http://www.w3.org/TR/owl-ref/
20. Pachet F, Zils A (2004) Evolving automatically high-level music descriptors from acoustic signals. In: LNCS. Springer, Berlin
21. Paulo N et al (2006) Emotions on agent based simulators for group formation. In: Proceedings of the European simulation and modeling conference, pp 5–18
22. Protégé Editor (2010) Available at: http://protege.stanford.edu
23. Rho S, Hwang E (2009) Content-based scene segmentation scheme for efficient multimedia information retrieval. Int J Wirel Mob Comput (IJWMC) 3(4):299–311
24. Rho S, Park J (2010) Intelligent multimedia services using semantic web technologies in internet computing environments. J Internet Technol (JIT) 11(3):353–360
25. Rho S, Han B, Hwang E, Kim M (2008) MUSEMBLE: a novel music retrieval system with automatic voice query transcription and reformulation. J Syst Softw 81(7):1065–1080
26. Rho S, Han B, Hwang E (2009) SVR-based music mood classification and context-based music recommendation. ACM Multimedia, Beijing, pp 713–716
27. Russell JA (1980) A circumplex model of affect. J Personal Soc Psychol 39
28. Schubert E, Wolfe J, Tarnopolsky A (2004) Spectral centroid and timbre in complex, multiple instrumental textures. In: Proceedings of the international conference on music perception and cognition, North Western University, Illinois
29. Thayer RE (1989) The biopsychology of mood and arousal. Oxford University Press, New York
30. Tzanetakis G, Cook P (2000) MARSYAS: a framework for audio analysis. Organised Sound 4(30)
31. Van de Laar B (2006) Emotion detection in music, a survey. In: 20th Student conference on IT
32. Yang D, Lee W (2004) Disambiguating music emotion using software agents. In: Proc int conf music information retrieval, pp 52–58
33. Yang Y-H, Lin Y-C, Su Y-F, Chen H-H (2008) A regression approach to music emotion recognition. IEEE Trans Audio Speech Lang Process (TASLP) 16(2):448–457