Evaluation of Music Classification Method based on Lyrics of English Songs

Mizuki Furuya, Hung-Hsuan HUANG, and Kyoji Kawagoe,

***Abstract*—Music is used for people practicing sports, for elderly individuals, and to help train the mind. Recently in music information science, studies have been conducted on music therapy and on music classification from a therapeutic point of view. However, most of these studies have classified music based on melody and tempo. No classification method that is based on lyrics has been performed for music therapy support. The authors previously propose a music classification method using emotional words included in lyrics toward music therapy support. As this method was developed for Japanese lyrics, it is necessary to evaluate the method for English lyrics. In this paper, the results of such evaluation is described. We also describe an improved method, appropriate for English lyrics.**

***Index Terms*—lyrics, music, recommendation, emotion.**

I. INTRODUCTION

MUSIC is effective for improving concentration and reducing stress, especially for people practicing sports, for elderlyindividuals, and to help train the mind. Recently in music information science, music therapy and a music classification method such purposes have been sufficiently proposed [1].

Although their study established associations between music and feelings, especially associations between melody and tempo on people's feelings, a music classification method that is only based on lyrics was not considered. Lyrics are an important element of music in addition to melody and tempo. Therefore, it is worthwhile to use lyrics to classify music for the purpose of music therapy.

We previously proposed a music classification method that uses emotional words found in Japanese lyrics toward music therapy [5]. In this method, after calculating a characteristic vector from the occurrence number of emotional words, music is classified from this feelings vector using a clustering method. By this method, we can realize the music classification based on the Japanese lyrics.

In this paper, we first evaluate our method to show that the method can be applied for English songs. We also describe our modified classification method with similar process as the original method. We then evaluate the modified method to show that the result of the modified method can be improved.

II. RELATED WORK

In this section, we describe a study about the influence given to feelings and study the feelings contained in music. First, we describe a study that uses lyrics. Then, we describe the elements that constitute music. Subsequently, we describe the influence of music on people's feeling.

M. Furuya is with Graduate School of of Information Science and Engineering, Ritsumeikan University, Japan, e-mail: is0086hk@ed.ritsumei.ac.jp.
H.H. HUANG and K. Kawagoe are with College of Information Science and Engineering, Ritsumeikan University.

A. Study using lyrics

Dan Yang analyzed feelings on both music and lyrics using a machine learning method [2]. He discovered associations between songs and their lyrics using feelings classification method. McVicar et al. studied the associations between music and lyrics for music recommendation. They discovered the common characteristics between music and lyrics [3].

B. Element of music and influence

Balkwill and Thompson discussed melody and tempo with regard to their influence on people's feelings [4]. They found associations between listener impressions and music melody and tempo. Their study showed that music with a fast tempo caused people to feel more active, whereas music with a slow tempo caused people to feel calmer.

III. EXISTING MUSIC CLASSIFICATION METHOD BASED ON LYRICS WRITTEN IN JAPANESE [5]

In our previous paper [5], we proposed a music classification method that uses emotional words found in Japanese lyrics. Figure 1 shows the classification procedure for our proposed method.
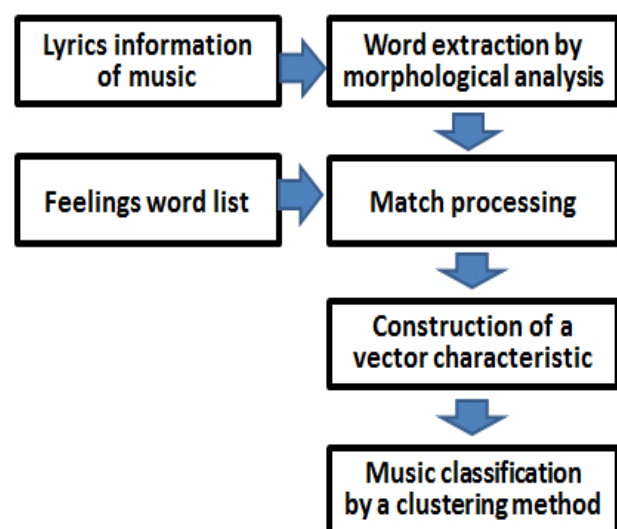

Fig. 1.   Classification procedure

As Figure 1 shows, our method is performed through four processes; word extraction by morphological analysis, match processing, construction of a vector characteristic that represents the feelings found in the music and music classification by a clustering method.

1) First, we create the feelings word list to express each feeling. The categories of emotional words are joy, comfort, sorrow, excitement, goodwill.

2) Next, we perform morphological analysis to extract information about each word from the lyrics of a song.

3) Then, we perform match-processing to determine the number of times that emotional words appear in the lyrics. Through this match processing, we count the number of times that lyrics information matches an emotional word in the feelings word list and create the characteristic vector.

4) The characteristic vector is composed of the ratio between the lyrics word count and the number of times that a particular word is included in each feeling counted by matching processing. The expression of the characteristic vector calculation is shown in Equation 1.

$$X = (\frac{A_1}{B}, \frac{A_2}{B} \ldots\ldots\ldots \frac{A_{10}}{B}) \qquad (1)$$

A_i :Number of times word is included in feelings

B : Lyrics word count

5) After matching process, we classify music by clustering the set of music with similar impressions for the characteristic music vector.

With this procedure, a music classification based on lyrics according to feelings can be realized.

We conducted the evaluation experiment to observe the classification accuracy. Ten subjects composed of eight men and two women between the age of 20 and 29 participated in the experiment. We used 50 musical pieces which the subjects usually listen. The evaluation of the experimental result was performed by finding the classification precision between the set of classified musical pieces and the set of correct answers. We performed clustering based on lyrics and realized music classification according to feelings.

As stated in [5], it was indicated that our method can classify music only from music lyrics enough. Table 1 shows the evaluation results described in [5]. It is because it was reported [11] that the maximum accuracy of music classification obtained by manual classification was 44% of F-measure. The F-measure for Sorrow feeling is the same result using our automated classification method, although the F-measures for the two Excitement and Goodwill feelings was inferior to those of the manual classification. Therefore, we showed the effectiveness of this classification method by achieving target classification precision in the feeling category of sorrow in the evaluation experiment.

As evaluation of our classification method were conducted using Japanese lyrics only, it is important to observe whether our method can be applied for music classification using non-Japanese lyrics. Thus, we conducted the same evaluation of our music classification method using English lyrics.

IV. MODIFIED MUSIC CLASSIFICATION BASED ON LYRICS

In this section, we describe the changes from the feeling classification method based on Japanese lyrics. As changes from the feeling classification in Japanese lyrics, we change the following:

TABLE I
EMOTION CLASSIFICATION PRECISION OF JAPANESE SONG LYRICS [5]

Feeling	Joy	Sorrow	Excitement	Goodwill
Recall	27.3 %	36.7 %	20.0 %	22.2 %
Precision	37.5 %	55.0 %	37.5 %	26.1 %
F-measure	31.6 %	44.0 %	26.1 %	24.0 %

- the way of creating the feeling word list in English
- the way of collecting the music dataset for evaluation experiment.

For word extraction through a morphological analysis, match processing, creating a vector characteristic, and a clustering method, we use the same processing as that used in the original method.

To create an emotional word list to express all feeling of English lyrics, we use Wordnet-Affect. Wordnet-Affect is a dictionary that grants the emotional information to words of the English vocabulary database Wordnet. The categories of emotional words used in our method are joy, sadness, calmness, excitement, fear, anger, depression and peace. For these emotions words, a feeling word list was created by extracting words represented using only single-word.

V. EVALUATION EXPERIMENT

In this section, we describe the evaluation experiment. We conducted two evaluation experiments to obtain the accuracy of the classification method.

The first experiment (EXP-1) is an evaluation obtaining classification accuracies for each of the same emotions that were used in the classification of Japanese lyrics. The other experiment (EXP-2) is an evaluation using a two-dimensional feeling space of positive-negative and activities-inactive.

A. Test Music Dataset

The music dataset used for these evaluation experiments is composed of 60 pop-songs. We collected 15 songs in each of four categories of emotions, i.e. joy, sadness, calmness, and excitement from a large music database, All Music [7]. The feelings that are defined in All Music are used as the correct emotions in the evaluation experiment.

B. Evaluation result of emotion classification (EXP-1)

In this section, we describe the result of the EXP-1 evaluation for each emotion. The emotion categories used in this experiment are joy, calmness, sadness, and excitement. These emotional categories are the same as those used in the previous classification experiment with Japanese lyrics.

An evaluation of this experiment was conducted by calculating the precision of the classification between a set of classified musical pieces and a set of correct answers. We used 60 songs, i.e. 15 songs for each category defined in the All Music.

The result is shown in Table 2. From this table, it is clear that the precision values are lower than those for our method using the previous Japanese lyrics shown in Table 1. Moreover, the maximum F-measure was only 18.8%, which is below the random guess value of 25%.

TABLE II
EMOTION CLASSIFICATION PRECISION RESULT (EXP-1)

Feeling	Joy	Calmness	Sad	Total
Recall	6.7 %	13.3 %	20.0 %	13.3 %
Precision	16.7 %	20.0 %	17.6 %	18.2 %
F-measure	9.5 %	16.0 %	18.8 %	15.4 %
The number of classified music	6	10	17	33

C. Discussions of EXP-1

In Table 2, evaluation results that used the same feeling as category classification in Japanese showed low values. We could not extract emotions from 27 songs, which is about one-half of the 60 songs because fewer words in the lyrics matched the words on the emotional word list and there are many songs in which an emotion cannot be extracted. Concretely, there are only 296 emotional words on the list, which is much fewer than 694, the number of Japanese emotional words. Therefore, we need to improve the method by increasing the number of basic feeling words as well as the number of feeling words. The following EXP-2 was conducted to check the effectiveness for of the improved method from the first viewpoint.

D. Evaluation using two dimensional space (EXP-2)

To increase the basic words for each emotion, we increased the number of emotion categories from multiple dimensions to two-dimensions. This increase can affect more matches of a word in the lyrics as well as emotional words in the list on of emotional word dictionary to use. Two-dimensional spaces are composed of two independent axes, a positive-negative axis and an activities-inactive axis. Through this classification method using a two-dimensional space, the number of words emotion words list can increase, which is considered to lead to an increase in the amount of music emotion extraction. In the classification of emotions, we used the emotion classification model of Russell [8], which was also used in the study by Hu et al [9]. Figure 2 shows Russell's emotion classification model.

There are eight types of emotion categories used in this experiment, i.e., anger, fear, depression, peace, joy, excitement, calmness, and sadness, which were also used in the experiment EXP-1. The number of feeling words in the list was increased to 694 words through an increase in the number of emotions. We classified these feelings according to the emotion model, and classified these eight emotions according to Russell's emotion model. Table 3 shows the classification map in a two-dimensional space.

TABLE III
EMOTION CLASSIFICATION

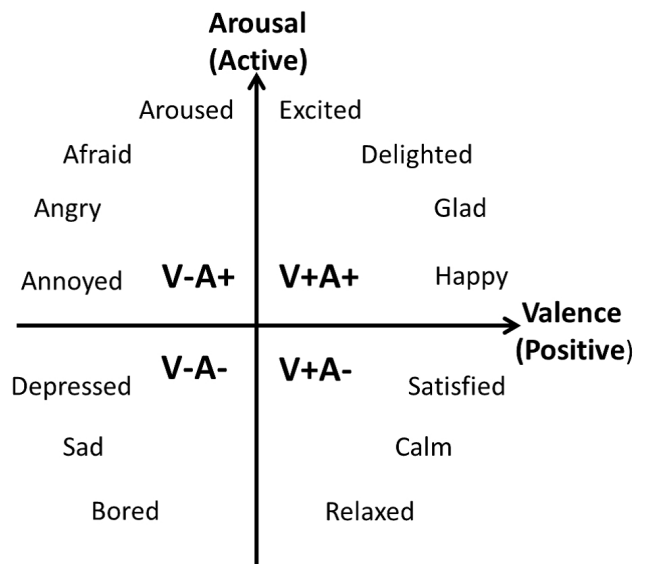| Category	Emotion
V+A+	Joy,Excitement
V-A-	Calm,Peace
V-A+	Afraid,Anger
V-A-	Sadness,Depression

Fig. 2.   Russell's model of mood [8]

For each song, we create a characteristic vector representing the related feeling. The expression of the characteristic vector calculation is shown in Equation 2.

$$(Valence, Arousal) = \frac{|V+| - |V-|}{N}, \frac{|A+| - |A-|}{N} \quad (2)$$

V+ : Number of occurrences of positive emotion
V - : Number of occurrences of negative emotion
A+ : Number of occurrences of active emotion
A - : Number of occurrences of not active emotion
N : The total number of words of the lyrics.

Music is classified from this characteristic vector using a clustering method, called a ward method. In the experiment, we used 60 songs i.e. 15 songs for each category of emotion defined. Table 4 lists the results of our experiment.

TABLE IV
EMOTION CLASSIFICATION PRECISION RESULT (EXP-2)

Feeling	V+A+	V+A-	V-A+	V-A-	Total
Recall	6.7 %	13.3 %	60.0 %	20.0 %	25.0 %
Precision	16.7 %	20.0 %	64.3 %	33.3 %	38.5 %
F-measure	9.5 %	16.0 %	62.1 %	25.0 %	30.3 %
The number of classified music	6	10	14	9	39

E. Discussions of EXP-2

As Table 4 indicates, the feeling category of V+A- indicating anger and fear, was shown to have the highest classification precision among the four emotion categories. This result, indicates that direct emotional words are used frequently to express the emotions of anger and fear, compared to the other feeling categories. Therefore, emotion extraction through emotional words using an emotion dictionary is effective and attains the high precision value. Although we

obtained a good result for one feeling category, the emotion extractions for the other categories was worse and remains to be improved. In particular, we were unable to categorize 21 songs, one-third of the 60 original songs considered.

Table 5 shows a comparison of the classification precision of Hu's method [9]. Hu et al. conducted a similar experiment using Chinese song lyrics, which they used to check the classification accuracy of their method. Compared to their results, our method obtained higher classification F-measures in for V+A- and V-A+ but smaller F-measures for V+A+ and V-A-. It is interesting to see a difference in the classification accuracy. V+A+ in Chinese songs was superior to that in English song, while V-A+ in English was superior to that in Chinese. It should be investigate further whether the reason for this difference is due to cultural a characteristics, whether the accuracies of both V+A- and V-A- can be improved by increasing the number of feeling words on the list, and how an accuracy level over 70% can be realized, among other issues.

TABLE V
COMPARISON OF THE CLASSIFICATION PRECISIONS WITH HU'S METHOD[9]

Feeling		Yajie's method	Our method
V+A+	Recall	68.6 %	6.7 %
	Precision	71.0 %	16.7 %
	F-measure	**69.8 %**	9.5 %
V+A-	Recall	75.0 %	13.3 %
	Precision	5.5 %	20.0 %
	F-measure	10.2 %	**16.0 %**
V-A+	Recall	32.7 %	60.0 %
	Precision	65.5 %	64.3 %
	F-measure	43.7 %	**62.1 %**
V-A-	Recall	27.8 %	20.0 %
	Precision	31.2 %	33.3 %
	F-measure	**29.4 %**	25.0 %

VI. CONCLUSION

In this paper, we described the evaluation results of our previously proposed method of music classification using only lyrics. Because our original method was evaluated using Japanese songs, English song lyrics were used in this current evaluation. We also proposed an improved method for increasing the classification accuracy for English songs. In particular, the results obtained show that our classification method can classify English songs in the feeling category of V+A-, which indicates anger and fear. To improve the accuracy of other emotions, we need to expand the emotional word list through the use of non-emotional words using the co-occurrence words and word similarity.

ACKNOWLEDGMENTS

REFERENCES

[1] Dan Yang, Won-Sook Lee. Music Emotion Identification from Lyrics. *11th IEEE International Symposium on Multimedia*. IEEE, San Diego, California, USA, December 14-16, 2009.

[2] Matt McVicar, Tim Freeman, and Tijl De. Bie. Mining the Correlation between Lyrical and Audio Features and the Emergence of Mood. *12th International Society for Music Information Retrieval Conference.* ISMIR, Miami, Florida, USA October 24-28, 2011.

[3] Laura-Lee Balkwill, William Forde Thompson. A cross-cultural investigation of the perception of emotion in music; Psychophysical and cultural cues. Music Perception, 17, 43-64, 1999.

[4] Akira Nakamura. Emotional Display Dictionary, Tokyo Temple Publication, 1979 (in Japanese).

[5] Mizuki Furuya, Hung-Hsuan Huang, Kyoji Kawagoe. Music Classification Method based on Lyrics for Music Therapy Support. *18th International Database Engineering & Applications Symposium.* IDEAS, Porto, Portugal, July 7-9, 2014.

[6] Carlo Strapparava and Alessandro Valitutti. Wordnet-affect: an affective extension of wordnet. In Proceedings of the 4th International Conference on Language Resources and Evaluation, Lisbon, 2004.

[7] "All Music " Available from http://www.allmusic.com. (accessed 2014-10-14)

[8] Yajie Hu, Xiaoou Chen and Deshun Yang. Lyric-Based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method, 10th International Society for Music Information Retrieval Conference ISMIR, Kobe, Japan, October 26-30, 2009.

[9] Russell James. A circumplex model of affect. journal of Personality and Social Psychology, 39(6):1161-1178, 1980.

[10] Rudolf Mayer, Andreas Rauber. Musical Genre Classification by Ensembles of Audio and Lyrics Features, 12th International Society for Music Information Retrieval Conference. ISMIR, Miami, Florida, USA October 24-28, 2011.

[11] Akira Yamada. Suggestion of the feelings extraction method of the song using the lyrics information. *Graduation thesis, Konan University, 2011.* Available from http://www.nadasemi.ii.konan-u.ac.jp/publication/research/2011/final/yamada.pdf. (accessed 2013-11-17) (in Japanese)

[12] Chung-Che Wang, Jyh-Shing Roger Jang, Wennen Wang. An Improved Query by Singing/Humming System Using Melody and Lyrics Information, 11th International Society for Music Information Retrieval Conference (ISMIR 2010), 45-50, 2010