An Automatic Metadata Generation Method for Culture-Based Music Impressions

Kamarul Arif bin Yusoff and Kiyoki Yasushi

Abstract—Communication plays an important role in this edge. Music is one of the successful medium for people to communicate because of its universal characteristic. It is necessary to generate metadata for music impressions in order to communicate between cultures. Our system provides method to obtain culture-based musical impression. The method consists of 5 steps: 1) Music features extractions. 2) User impression collections. 3) User impression table creation. 4) Semantic correlation calculation and 5) Obtaining the highest ranking of semantic correlation. This method is able to realize the music impression generation according to Malaysian/Indonesian, Indian and Japanese Gagaku music as sample music. By this approach, an automatic generation for intercultural music or other media impression can be realized in order to bridge the culture divide.

Index Terms—Music impression, culture-based music, metadata generation, intercultural.

I. INTRODUCTION

In general, communication can be defined as information transferring between two parties. The process of transferring the information can be in various ways such as speech, discussion, music, sign language, writing and touch. When the internet becomes more popular in early 1980s, people get an easier tool to communicate. Music is one of the best way to communicate between people even though the communication dealing with infants [1]. It is because some groups of people the might not have ability to speak using language like children or people who cannot speak in common languages.

"Music is, by its very nature, essentially powerless to express anything at all, whether a feeling, an attitude of mind, a psychological mood, a phenomenon of nature...."(Stravinsky,1936)

By using music, people might be able to express their feeling and sharing their thought by giving impression towards the music.[2]

This research focusing the method to extract the music impression and by doing that, we realizes the music impression retrieval by music query. Since the music databases are one of the important technologies which widely used to access music, this technology also could help the process of extracting the music impression.

Mainly, there are two types of method in music data

Kamarul Arif bin Yusoff is with the Malaysian Public Service Department, Cyberjaya 63000 Malaysia (e-mail: kamarul.yusoff@jpa.gov.my).

Kiyoki Yasushi is with the Department of Environmental Information at Keio University, Fujisawa, 252-0882 Japan (e-mail: kiyoki@sfc.keio.ac.jp).

retrieval. First is text based which is by providing the text annotation, the occurrence of some events will be presented according to the given text. This kind of query might consume of time if it deal with big amount of data. Another method is using content-based. This type of approach considers multimedia content as a query even though the content might be dynamic and have sequence of contents. This kind of retrieval system such as humming and voice recognition is able to become key or context for music retrieval.

Our system will use music data as a query and calculate the correlation among retrieval candidate to extract the impression which is similar among music data inside the database. We try to find the highest correlation between the query music data and the candidate music data. We propose the simple, yet effective in order to extract an impression new music data (query).

The key technology for this system is using music sample to represent various types of music. Since every people and individual define each of music differently, it is very hard to make transformation matrix to represent each individual perception towards certain type of music. In order to overcome this problem, we collect sample music and each of the song will assigned the suitable impression according to each person.

Our basic system concept is offer the generation of metadata for music impression. This system is able to receive new music or a query and do certain process. Basically, the process contains the music features extraction and correlation calculation. Our method in this system did musical feature extractions into six elements namely key, tempo, pitch, rhythm, harmony and melody. After that, system calculates the semantic correlation between query and among sample music and generates the impression for the query.

II. RELATED WORKS

There are several related studies about music impression and emotions. The role of individuality in expressing music is discovered through the paper [3][4]. Individual expression plays very important part to provide music emotion recognition. Personal impression was shown as a good indicator and improved the accuracy of emotion recognition. The individual or personal expression is very important because each individual is exposed with different personality, culture, demographic and music experience. Because of these differences, our study is proposing the best method to extract the music impression.

To detect the music impression, it is necessary to choose

the music elements. The musical elements chosen can be a modulator in conveying emotion or mood. Some studies analysis tempo as major elements to detect emotions [4]. However, most research shows, more musical elements involve to be analysed to represent particular music and to gain the accuracy of impression. [5], [6], [7].

The proliferation of musical data in the network is unavoidable and lead to rapid growth in music retrieval research. Besides retrieving the musical data, the access to musical impression gain interest to researchers. More method introduced to deal with music and impression [8], [9]. The idea is to generate metadata for music and it can be access by dealing with impressions. Some engine utilizes the text based query such as impression words and from the words, the associated music are accessible. However, the query also can be done by multimedia [10] such as audio files and picture files.

III. METHOD AND IMPREMENTATION

A. Basic Method

We introduced the pillar of this research which is the method or our approach to obtain the music impressions. Since we found the human being will be different in their impression towards certain things [2], [4] based on their experience and demographics, we proposed this method to achieve musical impression. We used one selection of new music to enter the system as query and from that, the system will be able to produce the impression for the music according to particular user.

Inside the system, all sample music will be calculated with the query or the new music to get the highest correlation. By using the highest correlation the system will assume as the musical impression for the query. This approach is depicted in Fig. 1.
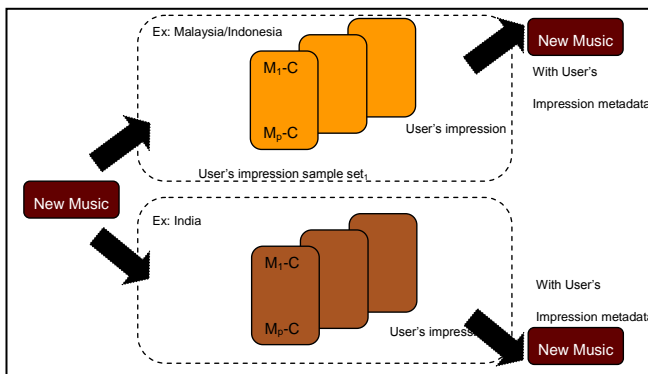

Fig. 1. Musical impression extraction process.

B. The Method

1) STEP1: Music selection and feature extraction.

In music selection, we proposed the main idea or the uniqueness of this research. Since we want to entertain the difference impression according to difference people, we found that using sample of music is the best way to get the musical impression.

In this study we start with the music selection according to particular country. We using MIDI (Music Instrument Digital Interface) file as the medium for our analysis. MIDI file is used because of it was designed to be a general standard for the music industry in terms of hooking up computer technology with varying musical instruments [11], [12]. Since the MIDI file is a performance translator or it is control the related event such as playing keyboard or varying the modulation wheel, so, it is very comprehend to be used with computer to do some calculation. The calculation will produce some characteristic of the MIDI or the music. This characteristic we will extract them as a musical features.

Our feature extraction concept is based on [5], [6], [7], [13]. We extracted six musical elements from the MIDI file. The musical elements extracted are key, tempo, pitch, rhythm, harmony and melody. Basically the idea of processing flow for getting the feature extraction and impression is shown in Fig. 2.

As for the Key musical elements extraction, we changed the value of tonality for each key based on Krumhansl value [14]. These changes made to make the key value become non-zero value. In the proposed tonality value described in Chapter 3, key value for D♭ (major code) and a (minor code) are "0.000000" or Zero. By using Krumhansl approach [14], the value for D♭ (major code) and a (minor code) will be non-zero value.
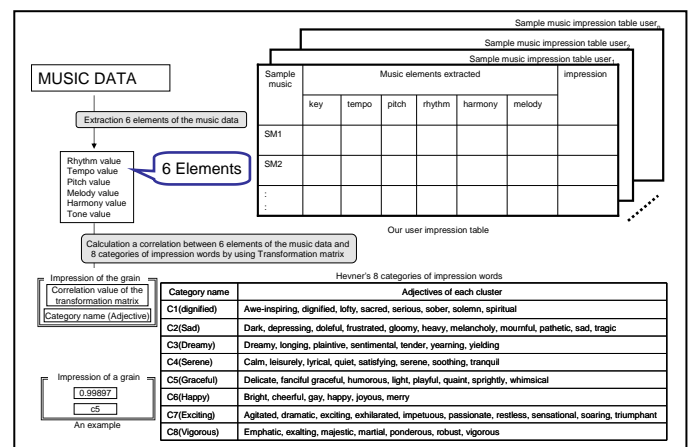

Fig. 2. Basically the idea of processing flow.

2) STEP 2: User's impression collection extraction.

The user impression towards ones music is different from others. In order to acquire user impression from particular music, we have to get it from the user themselves. However, it is very difficult to ask user to listen and perceive all music. We used sample of music to do musical features extraction and at the same time it was able to ask user to give their perception within reliable amount of music files.

In this case, we asked users to listen twenty three MIDI from the beginning of the music until the end and pick any impression word that suit their feeling or perception towards the music. By then we gathered the highest amount of the impression words for particular cluster [5][6][7][13]. Cluster which has highest number of impression words chosen by the user for particular music will assigned as impression for that

music. Fig. 2 shows user's impression collection process. To make it simpler, user's impression collection follows these steps:

1) User listens to one sample music. User listens from beginning to the end of the music.
2) User chooses impression words as many as they feel suit to the music based on their feeling.
3) We calculate number of impression word for each cluster.
4) The highest number of impression word for particular cluster will be assigned as the impression for the music.

3) STEP 3: User's impression table creation.

Each individual has their perception, thinking and emotion towards songs. Hence, each individual is unique and each of their table is different. After collecting user's impression for each of music, we create one table for each of them.

The user table for musical impression is basically two dimensional tables. It consists of sample music axis and musical impression and musical features axis. We place the sample music in the row and musical features and impression in the column.

1) One table will be represented for one user.
2) For each table, row represents sample music. Insert all sample music in the rows.
3) Columns represent musical features and user impression cluster.
4) Insert all the six musical elements in the column. There will be six columns named key, tempo, pitch, rhythm, harmony and melody and each of the will be filled by musical element respectively.
5) Last but not least, one column will be dedicated for impression word cluster for each of the sample music.

4) STEP 4: Semantic correlation calculation.

Our research is to determine ones new music which is a query to the system and determine its impression based on the sample music. In order to determine the impression, we do semantic calculation with query (new music) and each of sample music which already assigned with impression from the user. There are many semantic calculation can be done. However in our research, we did cosine similarity for semantic calculation:

$$\cos ine\theta = \frac{(x_1 x_2) + (y_1 y_2)}{\left(\sqrt{x_1^2 + x_2^2}\right) \cdot \left(\sqrt{y_1^2 + y_2^2}\right)} \qquad (1)$$

where:

- x_1, y_1: musical elements for new music (query)
- x_2, y_2: musical elements for one sample music in the user impression table

On the cosine similarity, the distance value for each calculation between query and the sample music shows the degree of similarity between them. In this equation, the distance between two points is obtain from inner product value $(x_1 x_2) + (y_1 y_2)$ and normalized by the Euclidean distance $\left(\sqrt{(x_1^2 + x_2^2)}\right)\left(\sqrt{(y_1^2 + y_2^2)}\right)$. This result is the ratio or *cosine* angle between to point/vector. Fig. 3 illustrate the cosine.
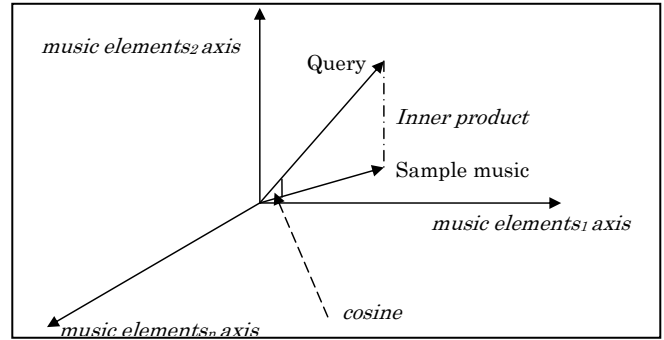

Fig. 3. Cosine illustration.

5) STEP5: Obtain highest ranking among sample music.

After semantic calculation was done, for each of the table are produced distance result between new music (query) and each of sample music. Distance value shows how close and the similarity between query and sample music.

We sort the similarity calculation value in each table. From the sorting, we obtain the highest rank or the top value. In cosine similarity, from the cosine value calculated between two vector (query and sample music), we sort the cosine value. From the sorting, we get the highest value. The highest value means the close similarity between query and sample music. So, the sample music which assigned with impression by particular user is representing the similar impression with the query. For example, if the highest cosine value is obtained from sample music A, while sample music A is assigned by user with impression cluster C2 or group of Sad impression, we infer that the music sample A represent the new music (query) impression.

IV. EXPERIMENTAL STUDIES

The implementation of this method we divided into two type of experiment. First, Conjugate Test and second is Simple Test. In Conjugate Test, the songs selections we collect are twenty three songs and the user participate in answering questionnaires are twelve people. For Simple Test, the songs selections only involve seven songs and the people who participate to answer the questionnaire imply to seven people. This two test trying to show the result of number of song and people involve will not effect the result produced the musical impression extraction.
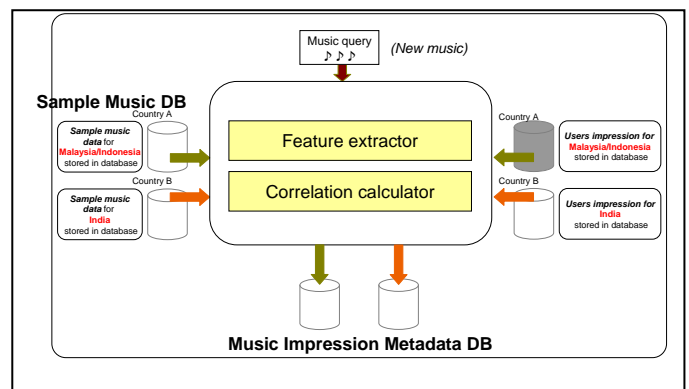

Fig. 4. Basic system concept.

In this test we separate two types of songs which we divide to two cultures. First are Malaysian/Indonesian music and the other one is from Indian music. We ask twelve people to

answer our questionnaire. We provided all 62 adjectives based on Hevner adjective cyclic [5], [6], [7], [13]. Users listened to each of the music and freely tick any adjectives they perceived to the music. We conduct online questionnaire so that user can easily access and listen at anytime. From the questionnaire we gather the adjective according to each of the music and count amount of highest adjectives. The highest ranking represents the impression cluster for the each music. As for the simple test, we select seven music and we ask about seven peoples to response to our online questionnaire.

The new music is a query to our system. The query is our objective to be assigned its impression automatically based on the retrieval candidate impression. For our experimental study, we used four queries to be assigned its musical impression in all three tests. This experimental study trying to fulfill the system concept based on the system architecture shown in Fig. 4.

V. RESULT

TABLE I: RESULT FROM CORRELATION CALCULATION

Query	Retrieval Candidate		
	Song name	*Correlation*	*Impression*
Query1	Kangen	0.900987	C3
Query2	Hati Yang Terluka	0.926454	C2
Query3	Mimpi Laila	0.985139	C6
Query4	Aku Tetap Aku	0.994764	C2
User 2			
Query1	Kangen	0.900987	C4
Query2	Hati Yang Terluka	0.926454	C3
Query3	Mimpi Laila	0.985139	C4
Query4	Aku Tetap Aku	0.994764	C4
User 3			
Query1	Kangen	0.900987	C3
Query2	Hati Yang Terluka	0.926454	C2
Query3	Mimpi Laila	0.985139	C8
Query4	Aku Tetap Aku	0.994764	C2
User 4			
Query1	Kangen	0.900987	C3
Query2	Hati Yang Terluka	0.926454	C2
Query3	Mimpi Laila	0.985139	C6
Query4	Aku Tetap Aku	0.994764	C2
User 5			
Query1	Kangen	0.900987	C2
Query2	Hati Yang Terluka	0.926454	C2
Query3	Mimpi Laila	0.985139	C7
Query4	Aku Tetap Aku	0.994764	C2

Table I shows our research result. In this table, we show the impression we achieved automatically from the retrieval sample music which found as highest correlation with the query. From this table we got the impression for each of query song for each user automatically.

The table shows result among 5 users for retrieving impression from Malaysian/Indonesian music. We also successfully achieved the result for Indian music and Japanese Gagaku music.

VI. CONCLUSION

We believed that communication plays and important role nowadays because of the rapid enhancement of technologies. One of the best medium to communicate is music and from it we can express thought and impression. Our objective in this research is to generate the musical impression based on different culture so that different culture can communicate each other through music media.

The idea of music communication not only for social aspect but it can be in economic side. In music industries for example, the communication using music makes the industries easily to exchange market of musical outcome or entertainment product.

This paper produced an easy method to achieve metadata generation of culture-based music impression. The idea to make the method less complex is due to reducing the transmission or congestions of the traffic inside of the communication edge. This research successfully realized the metadata generation method for culture-based music impression automatically

REFERENCES

[1] S. Benjaman, "Music as a Tool in Communications Research," *Nordic Journal of Music Therapy,* vol. 7, no. 1, pp 40-49, 1998.
[2] M. T. Mark and V. G. Flora, "Musical Expression of Moodstates," *Psychology of Music*, vol.19, no. 2, pp.99-109, 1991.
[3] Y. H. Yang, Y. F. Su, Y. C. Lin, H. H. Chen, Y. H. Yang, Y. F. Su, Y. C. Lin, and Chen, "Music emotion recognition: role of individuality," *ACM SIGMM Int. Workshop on Human-centered Multimedia 2007*, pp.13-21,2007.
[4] G. Rosa and A. C. César, *Integrating Emotions and Knowledge in Aesthetics Designs Using Cultural Profiles*, 2007.
[5] K. Hevner, "Expression in music: a discussion of experimental studies and theories," *Psychological Review*, vol. 42, pp. 246-268, 1935.
[6] K. Hevner, "Experimental studies of elements of expression in music," *American Journal of Psychology*, vol. 48, pp.186-204, 1936.
[7] K. Hevner, "The affective value of pitch and tempo in music," *American Journal of Psychology*, vol. 49, pp. 621-630, 1937.
[8] I. Asako and K. Yasushi, "An Automatic Metadata Generation Method for Music Retrieval-by-Impression Dealing with Impression-Transition," in *Proceedings of the 7th IASTED International Conference on Internet and Multimedia System and Applications*. 281-288, 2003.
[9] I. Asako and K. Yasushi, "A Kansei Metadata Generation Method for Music Data Dealing with Dramatic Interpretation," *Information Modeling and Knowledge Bases*, IOS Press, 2004.
[10] I. Sahori, K. Shuichi, I. Asako, and K. Yasushi, "A Music Retrieval System Supporting Intuitive Visualization by the Color Sense of Tonality," in *Proceedings of the 24th IASTED International Conference on Applied Informatics*, pp.153-159. 2006.
[11] B. Edwards. (5 September 2009). Your MIDI recordings are spread over three or more tracks and channels, which makes it difficult to play them on an outboard device or piano that only plays one at a time. Why do you do this?. [Online]. Available: http://www.perfessorbill.com/help/help.htm
[12] D. M. Huber, *The MIDI manual: a practical guide to MIDI in the project studio*, Focal Press, 1999, pp 4.
[13] K. Takahashi and K. Yasushi, "Fundamental framework for media data retrieval systems using media-lexico transformation operator – in the

case of musical MIDI data," *Information Modelling and Knowledge bases XII,* IOS Press, 2001.

[14] C. L. Krumhansl, *Cognitive foundations of musical pitch*, Oxford University Pres*s*, 1990.

Kamarul Arif Bin Yusoff was born in Penang, Malaysia on April 27, 1974 and holds Bachelor Degree in Information Technology (Hons) from Universiti Utara Malaysia (1999) and Master in Media and Governance from Keio University of Japan (2010). Currently, he works as SYSTEM ANALYST at Malaysian Public Service Department in Cyberjaya, Malaysia.

Kamarul Arif is a member of International of Association of Computer Science and Information Technology.

Kiyoki Yasushi received his B.E., M.E. and Ph.D. degrees in electrical engineering from Keio University in 1978, 1980 and 1983, respectively. In 1983 and 1984, he worked at Electrical and Communication Laboratory of NTT. From 1984 to 1996, he was with Institute of Information Sciences and Electronics, University of Tsukuba, as an ASSISTANT PROFESSOR and then an ASSOCIATE PROFESSOR. Since 1996, he has been with Department of Environmental Information at Keio University, where he is currently a PROFESSOR. His research addresses multi database systems, knowledge base systems, semantic associative processing, and multimedia database systems. He serves as the chair on Special Interest Group of Database Systems and as the editor-in-chief on transactions on Databases in Information Processing Society of Japan. He also serves as the program chair for the 7th International Conference on Database Systems for Advanced Applications.