Modeling of Music Recommendation Methods to Promote the User's Singing Motivation – For Next-Generation Japanese Karaoke Systems

Satoshi Isogai* and Miwa Nakanishi

Keio University, Yokohama, Japan
isogaisatoshi@a3.keio.jp, miwa_nakanishi@ae.keio.ac.jp

Abstract. This study attempted to build a model that recommends music choices to encourage karaoke-system users to sing by using data about the music preferences and inner characteristics of each user. First, we conducted an auditory experiment in two phases. Additionally, we analysed the acoustics and lyrics of music pieces. Using these data, we built a map of the music based on user impressions, and used this map to reveal the relationship between the user's most favourite music piece and the music piece that a user was highly motivated to sing. Thus, we were able to establish a basic model of the system that recommends the music piece a user would be highly motivated to sing.

Keywords: music recommendation, singing motivation, karaoke system.

1 Introduction

Recently, with the spread of the Internet, it is possible to compile the characteristics of users and the characteristics of the goods and services in a large database. The providers of goods and services want to find a method of associating the users' characteristics with the characteristics of their offerings. For example, in a typical shopping site, the system recommends goods for each user based on the user's age, gender and purchase history data. Many recommendation methods have been developed with a focus on recommending 'as similar as possible' products and services. However, the purpose of the recommendation system is to increase the motivation of the user to purchase goods and services. Presenting 'as similar as possible' products and services as recommendations do not necessarily correspond to this goal.

In psychological theory, users' motivation for products or services is categorised as a type of intrinsic motivation. Hunt [1] discussed the psychological theory that intrinsic motivation is evoked when human beings perceive an adequate gap between their own characteristics and those of an object, from the viewpoints of emotion, cognition, and ability. This theory is interesting in the study of the recommendation method that motivates users to make a purchase.

* Corresponding author.

S. Yamamoto (Ed.): HIMI/HCII 2013, Part I, LNCS 8016, pp. 439–448, 2013.

This study attempts to develop a recommendation model that applies psychological theory to engineering. In particular, this study aims to build a model that compares the characteristics of music selections to the characteristics of each user in order to make recommendations that would motivate the user to sing a selection.

2 Experiments

2.1 Experiment 1

The participants were 20 students (10 males and 10 females whose average age was 22 years) who had singing experience using a karaoke machine. After asking them to listen to several music pieces, we asked them to answer 36 questions, including questions on their impression of, experience with and singing motivation for each music piece. The music piece choices consisted of 80 tracks from four genres. For each genre, there were 20 tracks of Japanese popular music sung by a male (J-Pop male), Japanese popular music sung by a female (J-Pop female), music pieces sung in a foreign language (non-Japanese) and music pieces closely tied with movies or anime films.

In addition, we asked participants to use a free scale from 0 to 6 in order to assess 26 adjectives, such as ‘novelty’ or ‘quiet’, when answering questions about their impressions of the music pieces, with 0 meaning ‘strongly disagree’ and 6 meaning ‘strongly agree’. They also used a free scale from 0 to 6 to indicate how motivated they would be to sing a music piece (singing motivation) and to listen to a music piece (listening motivation). They used a multiple-choice method to answer questions on how know they were about each music piece and their impression of the listening experience.

2.2 Experiment 2

A different set of students (10 males and 10 females whose average age was 22 years) participated in a similar experiment. In this case, the participants listened to 55 music pieces. Of the 55 music pieces, 51 were commonly known and 4 were each participant’s favourite music pieces and music pieces which he or she tended to sing when using a karaoke machine. These preferences were determined during preliminary research study.

The music pieces were edited to be approximately 50 s long, usually ending at the climax of the second verse. Participants listened to each music piece three times. The order of the music pieces was randomised across participants. To help participants separate their impression of one music piece from the next, a 5 s long mechanical sound was inserted between each music piece. Participants listened to the music with headphones that included a noise-cancelling function, and they entered their answers on a computer. We explained the content of the experiment and obtained the informed consent of the individuals who agreed to participate in this experiment.

3 Appointing the Impression Map of Music

If it is possible to map the impressions of music, it seems that it is possible to capture the impressions of listeners for different features of the music. Therefore, instead of mapping listeners' impressions of the acoustic aspects of the music, we mapped their impressions of the emotional aspects.

3.1 Method

We created an impression map of the music for each experiment and compared the two. Thus, we confirmed the universality of using an impression map. The map was made using the average rating score for each adjective, analyzing the principal components of each experiment.

3.2 Results

For both results, we adopted the second principal component as a guide to establish the 70% cumulative contribution ratio. The loadings of both Experiment1 and Experiment 2 are very similar. The first principal component can be interpreted as indicating 'energetic', because the parameters 'flashy' and 'dynamic' make a positive contribution and the parameter 'quiet' makes a negative contribution. The second principal component can be interpreted as indicating 'familiar', because the parameters 'healthy' and 'flesh' make a positive contribution and the parameters 'alien' and 'stubborn' make a negative contribution. For this reason, impressions of the music for each user can be explained by the spatial relationship formed by the two axes of 'energetic' and 'familiarity'. This can be considered universal irrespective of the type of music and the participant.

4 Establishment of Impression Map Mapping Based on the Features of Music

When considering building a real system, in order to know the position in space of each music piece impression, it is not practical to ask the user to rate all music pieces as described above. Therefore, we built a model that positions any music on an impression map based on 'energetic' and 'familiarity'. We built the model using the data from Experiment 2 and we verified the model using the data from Experiment 1.

4.1 Quantitative Estimation of 'energetic'

From past studies [2, 3], it can be said that the use of acoustic feature to estimate the impression of music is effective. This study attempts to build a model of estimating 'energetic' using acoustic features.

4.1.1 Method

From the results of the second experiment, we obtained the principal component 'score of energetic' and built a model that had 'score of energetic' as its objective

variable and acoustic features as its explanatory variable. In particular, we used five acoustic features: MFCC-13 [4], the spectral centroid, spectral roll off, brightness [5] and chroma vector [6]. We used these features to generate 80 variables in a size 16 vector-quantisation codebook. We adopted six explanatory variables. There is a high correlation with 'scores of energetic'.

The model was built from six variables that are high correlation coefficients combined of 80 variables and 'scores of energetic'. The estimate parameters were derived using the least-square method.

$$E = -0.43 \times \sqrt{x1} - 0.19 \times x2 - 0.37 \times \log(x3) + 1.08 \times \log(x4) - 0.52 \times \sqrt{x5} + 1.60 \times \log(x6) + 2.65 \quad (1)$$

E: Scores of energetic

There was a high correlation between the 'scores of energetic' that were derived by analysing music features from Experiment 2 and the 'scores of energetic' that were derived by analysing the principal components as clarified in Section 3 (R^2=0.82).

4.1.2 Verification

The correlation is very high between the 'scores of energetic' and the 'scores of energetic' (R2=0.66). Therefore, Equation 1 is accurate as a model to estimate the 'energetic' music characteristics for each music piece. In other words, using Equation 1, we can estimate the 'energetic' characteristics of any piece of music without listening to it.

4.2 Quantitative Estimation of 'familiar'

Other than acoustic features, there are lyrics and degrees of recognition of music that can be obtained as data. By using these facts, we could build a model that substantively estimates 'familiar'.

4.2.1 Method

From the results of the second experiment, the second principal component of 'score of familiar' was obtained and a model was built that used 'score of familiar' as its objective variable and a degree of recognition of music as well as its lyric features as its explanatory variables. We used 0 for 'nothing' responses and 1 for high-awareness responses. When 70% or more of the responses was 1, then the level of awareness was set at 1 and when 70% or less of the responses was 1, then the level of awareness was set at 0.

On the other hand, after the lyrics were divided into words, they were analysed by focusing on the adjectives. This analysis used 'Hevner's adjective circle' [7]. Using this method, we divided the lyrics into eight groups of 66 words representing the feelings suitable to represent the music. Collecting about 50–70 synonyms for each group, we created a list of related words. Each group was counted up when a word in the lyrics matched the word list for each music piece. In addition, the total number of the counted words for each group was divided by the total number of the counted words for all groups (Table 1).

Table 1. Classification results of eight groups of adjectives (The upper row is the number; the lower row is the percentage)

	Group of related words1	Group of related words2	Group of related words3	Group of related words4	Group of related words5	Group of related words6	Group of related words7	Group of related words8	sum
music1	0	1	1	0	0	4	1	3	10
music1 (percent)	0	0.1	0.1	0	0	0.4	0.1	0.3	1

The model was built from 'awareness' and the percentage of each group 1–7. The estimate parameters were derived using the least-square method.

$$F = -2.23 \times x1 - 1.95 \times x2 + 0.04 \times x3 + 0.37 \times x4 + 0.23 \times x5 + 0.23 \times x6 - 0.70 \times x7 + 0.84 \times x8 - 0.07 \quad (2)$$

F: Scores of familiar

The correlation is very high between 'scores of familiar', which was derived by analysing lyrics and awareness from Experiment 2, and 'scores of familiar', which was derived by analysing the principal components from Section 3(R^2=0.61).

4.2.2 Verification

We verified the accuracy of the model using 40 music pieces from Experiment 1, not including foreign language music pieces. When we applied these 40 music pieces to Equation 2, the coefficient of determination was 0.53. Therefore, Equation 2 is accurate as a model to estimate the 'familiar' characteristics of music by using lyrics and the awareness for each music piece. In other words, we can estimate the 'familiar' characteristics of any piece of music without listening to it by using Equation 2.

5 Trend Analysis of the User's Music Preference

To build a system that recommend music pieces that karaoke users would be highly motivated to sing based on each users' most favourite music piece, this study attempts to vectorise the positional relationship to a music piece that a user is highly motivated to sing from the user's most favourite music piece on the impression map that we built in Section 3.

5.1 Common Map and Each User's Map

The impression map built in Section 3 represents the average of all participants' music impressions, that is, a common user's map. On the other hand, each user's impression of a music piece can vary based on the user's characteristics (e.g. preference of music and movies or other hobbies). Therefore, instead of using a common map, we will examine in each user's map the positional relationship between the user's most favourite music piece and the music piece that the user is highly motivated to sing. Each user's map is made from the rating scores of each user by multiplying the loadings of experiment.

5.2 Trend Analysis

We defined a piece of music that has a singing motivation score above 5 as the music piece that a user is highly motivated to sing. We defined music pieces that are sung well during karaoke as the user's most favourite music pieces. We examined the positional relationship between the two.

Based on where users' most favourite music pieces appear on the map, the trend was divided to show on a separate map of each participant the relationship between the different relative positions of the users' most favourite music pieces and the music pieces that they were highly motivated to sing. Figure 1 shows that groups 1, 2 and 3 define the user's favourite music piece as 'familiar', 'energetic', and 'lethargic', respectively. Unfamiliar music pieces were excluded, because there are very few cases in which an unfamiliar music piece is the user's most favourite music piece.

Next, we used the following process to examine the positional relationship to the music piece that a user is highly motivated to sing from the user's most favourite music piece (Figure 2). At first, for each participant, we used the vectorisation of the relationship between user's most favourite music piece and the music piece that a user was highly motivated to sing in order to obtain the distance and orientation. After dividing the distance by one and the orientation into increments of $\pi/6$, we found out in which area a music piece that a user was highly motivated to sing was placed when seen from the viewpoint of the user's most favourite music piece.

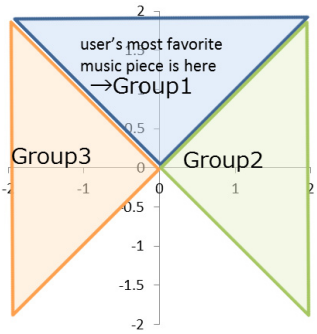

Fig. 1. Grouping of participants by the position of their most favourite music pieces

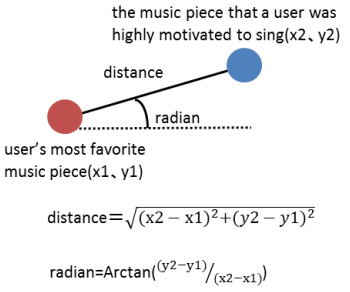

Fig. 2. Relationship of the position of the music piece that a user was highly motivated to sing when compared with the user's most favourite music piece

Figures 3, 4 and 5 show the probability of the presence of a music piece that a user is highly motivated to sing when the user's most favourite music piece is placed in the centre. In group 1 (Figure 3), a music piece that a user is highly motivated to sing tends to be placed in the unfamiliar area, which is slightly lower when viewed from the user's most favourite music piece. In group 2 (Figure 4), a music piece that a user is highly motivated to sing tends to be placed in the upper left area (inferior to 'energetic' but more than 'familiar') when viewed from the user's most favourite

music piece. In group 3 (Figure 5), a music piece that a user is highly motivated to sing tends to be placed in the lower right area (inferior to 'familiar' but more than 'energetic') when viewed from the user's most favourite music piece.

Thus, we built a model to estimate the location of a music piece that a user is highly motivated to sing in order to make a recommendation.

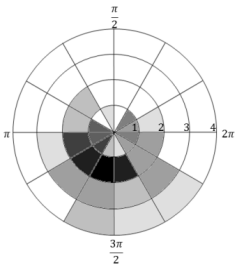

Fig. 3. Probability that there is a music piece that the user is highly motivated to sing (group 1)

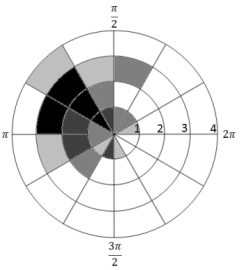

Fig. 4. Probability that there is a music piece that a user is highly motivated to sing (group 2)

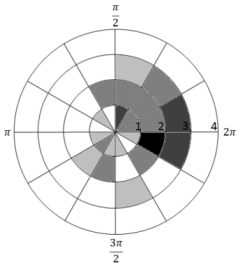

Fig. 5. Probability that there is a music piece that a user is highly motivated to sing (group 3)

6 Establishing a Personalised Music Mapping Method Based on the User's Profile

We attempted to construct a model to convert a common map to an individualized map for each user by using the users' profile to estimate positions without asking the users to listen to music. We built a model using the data from Experiment 2. We verified the model using the data from Experiment 1.

6.1 Analysis of the Relevance between a Personalised Music Map and a User's Profile

The correlation between the 'energetic' scores of all participants obtained from the experiment and the 'energetic' scores derived using Equation 1 in Section 4 is very high(R^2=0.69). The correlation between the 'familiar' scores of all participants obtained from the experiment and the 'familiar' scores derived using Equation 2 in Section 4 is relatively low (R^2=0.36). The results show that 'familiar' is a more sensitive index to individuality than 'energetic'.

Next, we focused on the analysis of the personality of the scores of 'familiar'. We focused on the mere exposure effect [8], which is well-known in the field of psychology in analyzing. The effect is that both the degree of courtesy and impressions of the target increase with repeated contact. According to this theory, a person feels that the music they have heard personally on several occasions is

'familiar'. With this in mind, we analysed for each user the relationship between the variability of 'familiar', 'degree of singing experience' and 'degree of viewing experience'. For each music piece, we defined it as '+' if the position on each user's map moved positively from the common position to the 'familiar' one, and defined it as '−' if the position moved negatively. We examined the trend of the direction of movement, the degree of singing experience and the degree of viewing experience. In addition, we determined 'awareness' by using the results of the answers to questions on awareness of music pieces. First, we defined music pieces that had a high degree of recognition as those music pieces that 70% of subjects responded with a choice other than 'nothing'. We defined other music as low-awareness music. Figure 6 shows the trend of the direction of movement by each degree of awareness of viewing experience. Focusing on high-awareness music, Figure 7 shows the trend of the direction of movement by the level of singing experience, separated by gender.

Figure 6 shows that participants who have heard a piece of music, even though their awareness is low, tended to feel the music as more familiar'. In addition, participants who have never heard a music piece that has a high 'awareness' and for which a lot of people feel 'familiar', have a tendency to feel the music piece as 'unfamiliar'. As seen in Figure 7, male subjects tend to feel music pieces that they recently sang during a karaoke session as more 'familiar'.

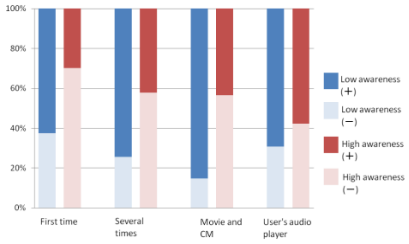

Fig. 6. Moving direction of the degree of viewing experience by awareness

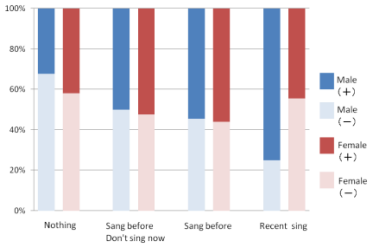

Fig. 7. Moving direction of the degree of singing experience by gender

When we analysed the results in more detail, we found that among the music pieces that have a low 'awareness' and especially the music pieces that have a negative value on the 'scores of familiar' on the common map, 'friendliness' on each user's map moves positively by a higher degree of viewing experience (average 0.4). In addition, among the music pieces that have a high 'awareness', especially the music pieces that have a value under −0.5 for 'scores of familiar' on the common map, 'friendliness' on each user's map moves negatively by a lower degree of viewing experience (average −0.5). Furthermore, only among male participants, the music pieces that have a value under 0.5 for 'scores of familiar' on the common map, 'friendliness' on each user's map moves positively by a higher degree of singing experience (average 0.6).

6.2 Modeling the Conversion from the Common Map to Each User's Map

Table 2 shows the rules to convert from the common map to each user's map, established by the above analysis.

Table 2. Movement pattern by acoustic feature and user characteristic

	Music characteristic	User characteristic	Pattern
Rule1	Lower awareness "Score of familiar" is under 0	Contained in the playback device	0.4
Rule2	Lower awareness "Score of familiar" is under 0	heard repeated to movies and CM	0.65
Rule3	Higher awareness "Score of familiar" is under −0.5	listened for the first time	−0.5
Rule4	"Score of familiar" is under −0.5	Male participant Recently sang	0.6

The correlation coefficient between 'score of familiar' on the common map and 'score of familiar' on each user's map was 0.562 in Experiment 1 and 0.604 in Experiment 2. When we applied the above rule to convert from the common map to each user's map, the correlation coefficient was 0.574 in Experiment 1 and 0.626 in the Experiment 2. In this way, the impression of personal "familiarity" for each user could be estimated more accurately. Using this method, we can also estimate the position of any piece of music on each user's map without the process of listening.

7 Conclusion

This study attempted to develop a method of recommending music on the basis of user's singing motivation with the aim of applying the method to next-generation karaoke services.

First, to locate music at the emotional side, we built an impression of space. As a result, we found 'energetic' and 'familiar' as the two axes that made up the impression of space. Next, in the impression map, we examined the relationship of the position between the music piece for which users are highly motivated to sing and the most favourite music piece of each user.

When considering the application of the method to a real system, we built a model that can find the position in space of each music piece impression based on acoustic features, lyrics and user awareness. Furthermore, to build a system with high accuracy, we established a method to reconstruct the spatial impression of each user using the user's profile.

After constructing the above results, we proposed a method of recommending music that is expected to highly motivate the user to sing by using the music characteristics and the user characteristics as data.

References

1. Hunt, J.M.V.: Motivation inherent in information processing and action. In: Harvey, O.J. (ed.) Motivation and Social Interaction, Cognitive Determinants, pp. 35–94. Ronald, New York (1963)
2. Kim, Y.E., et al.: Music Emotion Recognition: A State of the Art Review. In: ISMIR 2010, pp. 255–266 (2010)
3. Nishikawa, et al.: Design and Evaluation of a Musical Mood Trajectory Estimation Method Using Lyrics and Acoustic Features. IPSJ-SIGMUS 2011-MUS 91(7), 1–8 (2011)
4. Nishimura, et al.: Noise-robust speech recognition using band-dependent weighted likelihood IPSJ SIG 2003(124), pp. 19-24 (2003)
5. Juslin, P.N.: Cue utilization in communication of emotion in music performance: relating performance to perception. Journal of Experimental Psychology: Human Perception and Performance 26(6), 1707–1813 (2000)
6. Bartsch, M.A., et al.: To Catch a Chorus: Using Chroma-based Representations for Audio Thumbnailing. In: WASPAA 2001, pp. 15–18 (2001)
7. Hevner, K.: Experimental studies of the elements of expression in music. Amer. J. Psychol. 48, 246–268 (1936)
8. Zajonc, R.: Attitudinal effects of mere exposure. Journal of Personality and Social Psychology 9(2 Pt. 2), 1–27 (1968)