The Sixteenth International Symposium on Artificial Life and Robotics 2011 (AROB 16th '11),
B-Con Plaza, Beppu,Oita, Japan, January 27-29, 2011

Music Recommendation System Using the Time-series Discrete Wavelet Transform and the FastICA

K. Horiike[1], Y. Yoshitomi[2], T. Tokuyama[3] and M. Tabuse[2]

1: Taka Dance Fashion Co., Ltd., 1-13 Ikutamamae-machi, Tennouji-ku, Osaka 543-0072, Japan
2: Graduate School of Life and Environmental Sciences, Kyoto Prefectural University,
1-5 Nakaragi-cho, Shimogamo, Sakyo-ku, Kyoto 606-8522, Japan,
E-mail: yoshitomi@kpu.ac.jp
3: Shofu Inc., 11 Kamitakamatsu-cho, Fukuine, Higashiyama-ku, Kyoto 605-0983, Japan

Abstract: We previously proposed a content-based music recommendation system that uses several strategies for music recommendation. Each strategy, based on the user's music evaluation history up to the present time, is composed of a feature vector and a decision rule for music recommendation. Vocal signals are an important factor in music recommendation. Therefore, we use FastICA to separate vocal components from music. In an experiment to evaluate the proposed system, 10 users rated 100 music files in a music database. Of these 100 music files, 49 contained vocal components. For all 100 music files, the mean recommendation accuracy by the proposed system was 67.2%, and the mean recommendation accuracy of random recommendation was 32.1%. For songs containing vocal signals, the mean recommendation accuracy by the proposed system was 45.3% with FastICA and 35.8% without FastICA.

Keywords: Content-based music recommendation, Vocal characteristics, Independent component analysis, FastICA.

I. INTRODUCTION

Currently, many computer users enjoy a large number of music files stored on their computer or using the Internet. The increasing amounts of available audio data require the development of a music recommendation system. There are two kinds of music recommendation systems: content-based recommendation and collaborative filtering. We previously proposed a content-based music recommendation system [1].

In this study, in our music recommendation system, we add the time series of Mel-frequency cepstral coefficients (MFCCs) [2], [3] as the feature parameters. Moreover, because vocal signals are an important factor in a listener's musical tastes, we add a module using FastICA [4] for separating vocal signals from music.

II. FEATURE PARAMETERS

In the content-based music recommendation system, we use three kinds of feature parameters for characterizing music: time series of wavelet transform coefficients, time series based on MFCCs, and rhythmic content [5]. We explain these feature parameters in the following subsections.

1. Wavelet transform coefficients

Original audio data $s_k^{(0)}$, where k denotes the element number in the data, are used as the level-0 wavelet decomposition coefficient sequence. The $s_k^{(0)}$ data are decomposed into the multi-resolution representation (MRR) and the coarsest approximation by repeatedly applying the discrete wavelet transform (DWT). The wavelet decomposition coefficient sequence $s_k^{(j)}$ at level j is decomposed into two wavelet decomposition coefficient sequences at level $j+1$, as shown below in (1) and (2):

$$s_k^{(j+1)} = \sum_n \overline{p_{n-2k}} s_n^{(j)}, \tag{1}$$

$$w_k^{(j+1)} = \sum_n \overline{q_{n-2k}} s_n^{(j)}, \tag{2}$$

where p_k and q_k denote the scaling and wavelet sequences, respectively, and $w_k^{(j+1)}$ denotes the development coefficient at level $j+1$. The development coefficients at level J are obtained by using (1) and (2) iteratively from $j=0$ to $j=J-1$.

In the present study, we use Daubechies wavelet for the DWT. As a result, we obtain the following relation between p_k and q_k:

$$q_k = (-1)^k p_{1-k}. \tag{3}$$

It is known that the histogram of the wavelet coefficients of each domain of MRR sequences has a distribution centered at approximately 0 when the DWT is performed on audio data [6]. We use the time series of wavelet coefficients of each domain of the MRR sequences as elements of the feature vector.

2. MFCCs

MFCCs are obtained for each frame of a sound signal by using reported techniques [3]. The following are used as elements of the feature vector: the time-series of the mean values and the standard deviations of 12-dimensional MFCCs, the logarithmic power of 12-dimensional MFCCs, the 12-dimensional MFCC difference between frames, and the 12-dimensional MFCC logarithmic power difference between frames.

3. Rhythm content

Rhythmic content feature parameters are obtained by using reported techniques [5]. A set of feature parameters based on a beat histogram are calculated. These are as follows:

- *A0*, *A1*: the relative amplitude (divided by the sum of the amplitudes) of the first and second histogram peaks, respectively [5];
- *RA*: the ratio of the amplitude of the second peak divided by the amplitude of the first peak [5];
- *P1*, *P2*: the period of the first and second peaks, respectively, in beats per minute (bpm) [5];
- *SUM1*, *SUM2*, *SUM3*: sum of beat strength in the histogram in the range of 40–90, 90–140, 140–250 bpm, respectively.

Each of the three kinds of feature parameters, which are *a*: [*SUM1*], *b*: [*A0*, *A1*, *SUM1*, *SUM2*, *SUM3*], c: [*A0*, *A1*, *P1*, *P2*, *RA*, *SUM1*, *SUM2*, *SUM3*], is used as a feature parameter in the proposed system described in Section IV.

III. FastICA

Independent component analysis (ICA) [7], which is a statistical and computational technique for separating hidden factors in a signal, is a promising technique for separating vocal signals from songs [8]. In this study, we use FastICA [4], for performing ICA.

IV. PROPOSED SYSTEM

To explain the decision rules for music recommendation, we describe a set of music indices as $M = \{m \mid 1, \cdots, N_M\}$, an evaluation by a user as a score $s\ (1 \le s \le 5)$, where 1, 2, 3, 4, and 5 mean "dislike," "slightly dislike," "neutral," "slightly favorite," and "favorite," respectively, and a set of evaluated music indices as $M_s = \{m_s \mid 1, \cdots, N_{Ms}\}$.

1. Elements of decision rules for the recommendation

A. Decision rules using one kind of feature parameter

After principal component analysis (PCA) on the feature vectors obtained from unevaluated music m' file and the number N_{Ms} of evaluated music files of a user, the principal components up to the l th component are selected under the condition that the accumulated contribution ratio first exceeds 80% at the l th component. The score s for music m^* file having the maximum value of similarity to music m' file among the number N_{Ms} of evaluated music files is assigned to the score of music m' file. The similarity is calculated as the inverse value of the Euclidean distance in the l-dimensional feature vector space obtained by the above PCA. When the assigned score is 4 or 5 ("slightly favorite" or "favorite"), the unevaluated music m' file is recommended for the user.

B. Decision rules using two or three kinds of feature vectors

Table 1 shows the decision rules for music recommendation using two or three kinds of feature parameters.

Table 1. Decision rules for music recommendation using two or three kinds of feature parameters

A	Score (1 kind of feature parameters) = 4 or 5 Score (another) = 3〜5
B	Score (1 kind of feature parameters) = 4 or 5 Similarity order (another) : within top 50 %
C	Score (1 kind of feature parameters) = 4 or 5 Similarity (another) : average or higher than average

When using 3 kinds of feature parameters, 'another' is replaced by 'other two', and then naming replacement (A=>D, B=>E, C=>F) is performed.

2. Recommendation system

The system finds a suitable strategy for music recommendation for each user among 74 combinations of decision rule and feature parameter(s), as shown in Table 2, based on the user's music evaluation history up to the present time. Assuming that we have evaluated $K-1$ music files, one unevaluated music m file is judged recommendable or not recommendable according to the following two criteria.

Criterion 1:

One music m^{**} file among $K-1$ evaluated music files is considered an unevaluated music file, and each strategy of the 74 combinations of decision rule and feature parameter(s) is used to judge whether music m^{**} file is recommendable by using $K-2$ evaluated music files as training samples. Then, each recommendation is checked using the score of music m^{**} file to determine whether each judgment is correct. The recommendation accuracy for each strategy is obtained by selecting one music m^{**} file in order among $K-1$ evaluated music files. The recommendation strategy having the best accuracy among

Table 2. Strategies for music recommendation in our system.

Strategy no.	Decision rule	Feature vector	
		Main	Other(s)
1	Using only one kinds of feature parameter(s)	Wavelet	
2		MFCC	
3		Rhythm a	
4		Rhythm b	
5		Rhythm c	
6	A	Wavelet	MFCC
7			Rhythm a
8			Rhythm b
9			Rhythm c
10		MFCC	Wavelet
11			Rhythm a
12			Rhythm b
13			Rhythm c
14		Rhythm a	Wavelet
15		Rhythm b	
16		Rhythm c	
17		Rhythm a	MFCC
18		Rhythm b	
19		Rhythm c	
20~33	B	(Same as those described at No.6～19)	
34~47	C	(Same as those described at No.6～19)	
48	D	Wavelet	MFCC, Rhythm a
49			MFCC, Rhythm b
50			MFCC, Rhythm c
51		MFCC	Wavelet, Rhythm a
52			Wavelet, Rhythm b
53			Wavelet, Rhythm c
54		Rhythm a	Wavelet, MFCC
55		Rhythm b	
56		Rhythm c	
57~65	E	(Same as those described at No.48～56)	
66~74	F	(Same as those described at No.48～56)	

all strategies is used to judge whether each unevaluated music m file is recommendable by using $K-1$ music files as training samples.

Criterion 2:

When an unevaluated music m file is judged to be recommendable under criterion 1 and the main feature vector having the second highest similarity to the feature vector obtained from the unevaluated music m file is obtained from the music file having a score of 4 or 5, the unevaluated music m file is judged to be recommendable.

V. PERFORMANCE EVALUATION

1. Conditions

Fig. 1 shows the flowchart for the evaluation of the system. K was set as 100 in using 100 music files in the database, while it was set as 49 in using the 49 music files with vocal.

We used 100 music files in the real world computing (RWC) music database, which is available for research [9]. To evaluate the performance of the proposed system, all music files in the database were assigned scores $s\,(1 \le s \le 5)$ by 10 users. Of the 100 music files, 49 with vocal signals were used to evaluate FastICA as a pre-processing module. The feature parameters from wavelet transform coefficients and those from rhythmic content were calculated under the conditions described in our previous reported paper [1]. To calculate the feature parameters from MFCCs, a sound signal of 10 to 65 seconds from the beginning of a music signal was used. The sound signal was obtained under the following conditions: window length 30 ms, shift pitch 10 ms, Hanning window used as the window function, and 24 filter banks. The sound signal was equally divided into five signals as the time-series. In this experiment, a sound signal of 55 seconds was divided into five sound signals of 11 seconds, and 52

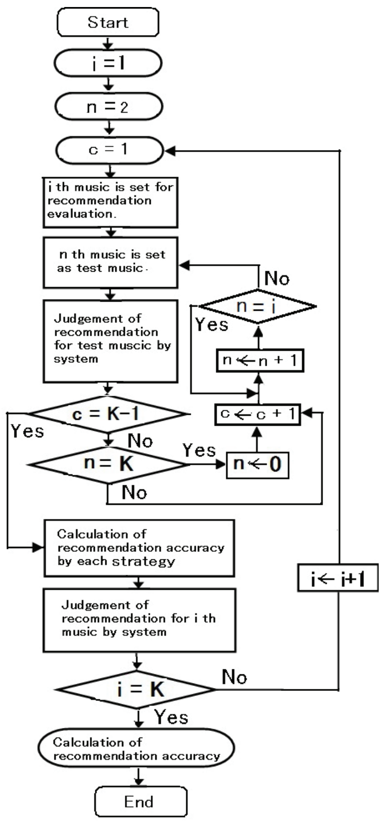

Fig. 1. Flowchart for evaluation of the system

feature parameters were obtained for a sound signal of 11 seconds according to the method described in Section II-2. As a result, we obtained a 260-dimensional feature vector from the sound signal of 55 seconds.

2. Results and discussions

Table 3 shows the recommendation accuracy for 100 music files. The recommendation accuracy of criterion 2 was almost always better than that of criterion 1. The mean recommendation accuracy by the proposed system (criterion 2) was 67.2%, while that of random recommendation was 32.1%. For songs containing vocal signals, the mean recommendation accuracy by the proposed system with FastICA was 45.3%, but that without FastICA was 35.8% under only criterion 1 (Fig. 2). These results confirm that FastICA improves the recommendation accuracy of songs with vocal signals.

Table 3. Recommendation accuracy for 100 music files

User No.	Criterion 1		Criterion 2	
	Recommendation accuracy	Difference to random recommendation	Recommendation accuracy	Difference to random recommendation
1	55.6	15.6	83.3	43.3
2	40.0	11.0	28.6	-0.4
3	66.7	32.7	100.0	66.0
4	59.1	17.1	60.0	18.0
5	44.4	15.4	30.8	0.8
6	50.0	23.0	60.0	33.0
7	84.6	50.6	100.0	66.0
8	68.8	33.8	71.4	36.4
9	35.5	-4.5	37.5	-2.5
10	57.1	47.1	100.0	90.0
Mean	56.2	24.2	67.2	35.1

(%)

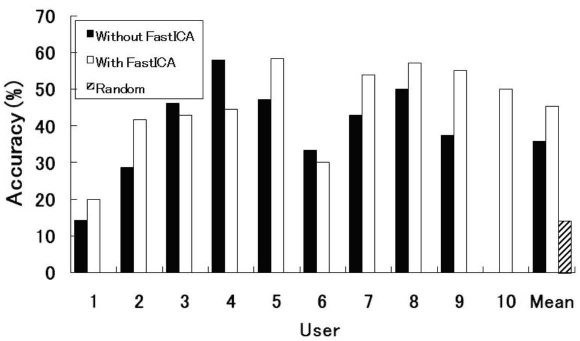

Fig. 2. Effect of FastICA on recommendation accuracy

VI. CONCLUSION

We propose a content-based music recommendation system with several possible strategies. Each strategy is composed of a feature vector and a decision rule for music recommendation. The strategy for each user is based on the user's music evaluation history up to the present time. Since vocal signals are an important factor of a listener's musical tastes, we use FastICA as a pre-processing module to separate vocal signals from songs. In the experiment, 100 music files in the RWC database were used for estimating the performance of the proposed system. The mean recommendation accuracy by the proposed system was 67.2%, and the mean accuracy of random recommendation was 32.1%. For 49 songs containing vocal signals, FastICA improved the mean accuracy of music recommendation.

Acknowledgment

We would like to thank all the participants who cooperated with us in the experiments.

REFERENCES

[1] Tanaka M and Yoshitomi Y (2008), A music recommendation system based on user preferences using the discrete wavelet transform (in Japanese). Proc. of the 7th annual convention of forum on information technology. 2:245-248

[2] Davis S and Mermelstein P (1980), Experiments in syllable-based recognition of continuous speech. IEEE Trans. Acoust., Speech Signal Processing. 28:357-366

[3] Shikano K, Itou K, Kawahara T, Takeda K, and Yamamoto M (2001), Speech recognition system. Ohmsha (in Japanese)

[4] Hyvärinen A (1999), Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. on Neural Networks. 10(3)626-634

[5] Tzanetakis G and Cook P (2002), Musical genre classification of audio signals. IEEE trans. speech audio process. 10(5):293-302

[6] Murata S, Yoshitomi Y and Ishii H (2007), Optimization of embedding position in an audio watermarking method using wavelet transform (in Japanese). Abstracts of autumn research presentation forums of ORSJ 210-211

[7] Hyvärinen A and Oja E (2000), Independent component analysis: algorithms and applications. Neural Networks. 13(4-5)411-430

[8] Feng Y, Zhuang Y and Pan Y (2002), Popular song retrieval based on singing matching. PCM '02 Proc. of the Third IEEE Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing. 639-646

[9] Goto M, Hashiguchi H, Nishimura T, and Oka R (2004), RWC music database: database of copyright-cleared musical pieces and instrument sounds for research purposes (in Japanese). Trans. of IPSJ. 45(3)728-738