Music Classification By Parsing Main Melody

Cheng-Che Lu and Vincent Shin-Mu Tseng
Department of Computer Science and Information Engineering
National Cheng Kung University
Tainan, Taiwan, ROC
E-mail{ cclu, vincent }@idb.csie.ncku.edu.tw

ABSTRACT

Music classification is an important problem in multimedia databases due to its widespread applications. In this paper we investigated the extraction of main melody in a music based on music theory and what music features are most suitable to represent a melody and then use decision tree to classify music. We tested various kinds of music like classic, folk, kid and pop music collected from finalemusic [5] and discovered the interesting diversities between different music styles, such as the tempo, grace, slur, etc. Experimental results show that the classification accuracy was achieved as good as 90% by the most suitable features for 2-way classification.

1. INTRODUCTIONS

In recent years, the large digital music has been extensively used due to its convention for human, like MIDI [8], mp3, etc. Basically MIDI is a main base of the digital music, because most composers use MIDI style to compose pieces in the initial stage and transfer MIDI into other digital music style for users' needs. MIDI can be translated into XML [18] style with small storage, and it is popular, due to compatibility with MusicXML software [12]. On the other hand, symbolic music representations [10] is becoming more important in the near future for the development of the MPEG standard. In practice, the works in [7][14] already translated MIDI music into XML style with described music characters, and they also proposed some methods for music retrieval or music classification.

In this paper we investigated the extraction of main melody in a music based on music theory. Based on the music features extracted from the main melody such as meter, slur, etc., we calculated the melodic variations that were also regarded as one of the important music feature. After extracting the main music features, we used decision tree to classify music. To evaluate the most suitable features, we tested various kinds of music like classic, folk, kid and pop music collected from finalemusic [5]. The experimental results show that specific music features obtained from the main melody are very helpful for music classification, and it was shown that the classification accuracy was achieved as good as 90% by the most suitable features for 2-way classification. Hence, these discovered music features can be applied in wide applications like music recommendation systems [2][3][4].

The rest of the paper is organized as follows. A brief introduction to previous work is given in Section 2. In Section 3, the method is described in details. The performance of the approach was evaluated in Section 4. A conclusion is given in Section 5.

2. RELATED WORK

For music information retrieval, it is usually based on comparison between melodies. For example, Tseng *et al.* [15] considered the problem of extracting melody and used the N-index technique to reduce the retrieval errors. Liu *et al.* [6] proposed that the repeated and longest patterns was usually the main melody, and Uitdenbogerd *et al.* [17] proposed that the successive highest pitch in staves can be regarded as the main melody, which can be converted into a sequence of intervals and they were more appropriate for comparisons between music. In addition to melody extraction, in order to improve accuracy of the music retrieval, McKay *et al.* [7] and Weyde *et al.* [19] segmented melody into motifs and phrases, and then they were encoded according to variations. Besides, Miller *et al.* [9] thought that people easily hear some specific variations in listening the melodies. In other words, melody is one of the most important music features that can be used to recognize music style.

In current MIDI music classification, Chai *et al.* [1] collected mono music that contains Irish, German and Austrian folk songs, and translated melody into four sequences — absolute pitches, absolute pitches with duration, intervals and contours, and then employed hidden Markov model to classify the folk songs. In addition, Shan *et al.* [13] collected Enya, Beatles, and Chinese folk songs, and then translated songs into mono music by Uitdenbogerd *et al.* [16] proposed. Shan *et al.* used the triad and 7^{th} chords to annotate a melody, so a melody can be showed a sequence of patterns. Finally, Shan *et al.* used heuristic method to classify music. These researches focused on the sequence of chords or notes respectively and they ignored other music features in composing a piece such as dynamics, staccato, etc. We investigated the extraction of main melody in a music based on music theory and what music features are most suitable to represent a melody.

3. METHOD

In this section, we describe the extraction of non-repeated main melody and music features in details. The flow chart of method is shown Figure 1. The main idea is to extract the non-repeated main melody from a piece and discover the melodic variations in music features. We state how a non-repeated main melody is extracted from a piece in Section 3.1. In Section 3.2, we illustrate how a melody is translated into a sequence of representations and what representations are suitable for data mining. The melodic features are interpreted and weighted in Section 3.3.

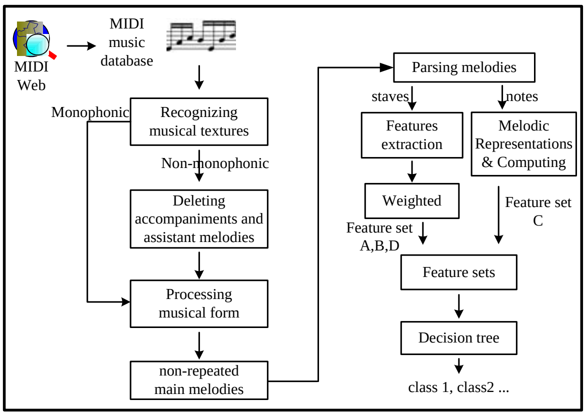

Fig. 1 System architecture.

3.1. Non-repeated Main Melody Extraction

Melody [9] is composed of a sequence of the pitches and durations of the notes. Composers are to make the music sound more harmonious so extra music textures are often added in a piece. However, a texture of the western music consists of melody (horizontal) and harmony (vertical). In the following, textures can be divided into three categories: 1) Monophonic texture: It is single melody without accompaniment. 2) Homophonic texture: A texture of one main melody with accompaniment. 3) Polyphonic texture: Two or more melodic lines of equal complexity and they are sounding simultaneously. Besides, there are repeated and similar phrases in a piece that are most memorable. In this paper, the repeated and similar phrases are called main melody. Miller *et al.* [9] argued that main melody is enough to represent a song. The extraction of main melody is as shown in the following step:

Stage 1: Each score can be translated into the XML style and then we parse it to acquire the monophonic texture. Fig.2 (a) shows that the tonic of a homophonic texture is recorded in the treble clef.

Stage 2: The various music forms must be considered for extracting a main melody such as A、AB and ABA etc. Most pieces belong to one of them and we only need to process ABA form because of the back segment A is repeated performance according to music theory. We can delete the redundancy by extraction from its XML score. For example, Fig.2 (b) shows that the back redundant parts A must be deleted in ABA form. The remained part AB is a main melody we need as shown in Fig.2(c).

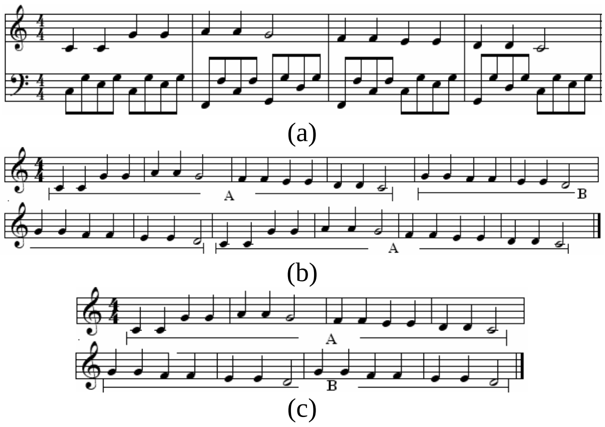

Fig. 2 An example illustrating how to extract main melody by parsing a homophonic texture piece.

3.2. Representations of Melody

Integrating the three points below, we can use a sequence of symbols to represent a main melody after considering the pitch of a note, a note with duration and a chord.

Fig. 3 A same song with pitches in different octaves.

(A) The pitch of a note can be represented as relative or absolute pitch separately. Absolute pitch means that the notes with same chroma in different octaves are regarded as same pitch. Relative pitch means that the notes with same chroma in different octaves are regarded as different pitches. We employed two representations because of composers may compose a same piece with different octaves. A same song with different pitch is composed as shown in Fig. 3.

(B) Note with duration means that the note sounded for a specific time such as whole note, half note, etc. Composers use notes with duration to enrich the rhythm. In this paper we treats thirty-second note as a base unit. For example, a quarter note equals eight thirty-second notes. A melody can be converted into a sequence of pitches with duration as shown in Fig. 4.

(C) Chord [9] is a combination of two or more notes which sound simultaneously. A chord can be represented as the highest, average and lowest pitches if a chord consists of three or more notes. A chord with two notes can be represented as the highest and lowest pitches. In Fig. 4, the first chord with two eighth notes is represented as a relative highest pitch C3 and a relative lowest pitch C4. The C3_4 can be denoted as a relative pitch C3 with four units. The C3_4&C4_4 denotes that a chord with two notes. The representations are convenient for analyzing the variations of the pitch of a melody.

Relative pitch：<C3_4&C4_4, G3_4, E3_4&C4_4, G3_4, C3_4&G4_4, G3_4, E3_4&G4_4, G3_4, F2_4&A4_4…..>
Absolute pitch：<C_4&C_4, G_4, E_4&C_4, G_4, C_4&G_4, G_4, E_4&G_4, G_4, F_4&A_4..>

Fig. 4 Chords are represented as a symbolic sequence.

3.3. Features Interpretation and Weighting

In music theory, Miller [9] stated that the music involves other characters except for the notes. In other words, we can utilize music characters to analyze the variations between melodies according to music theory. We described the specific thirty features that are useful for music classification as shown in Table 1. First, the music features in set A are about format of the music score and usually implied the music style the composers want to express. Then, both sets of B and D are related to format of the performing notes, which can be sounded interesting. Hence, we adopted both sets of B and D to analyze the variations between melodies. To discover the melodic variations, the related music features are standardized by formula (1) as shown below.

$$w_i = \frac{\sum_{i=1}^{n} x_i}{M} \tag{1}$$

where w_i denotes the time ratio of the feature in a melody; x_i denotes the time duration of the feature in a melody; M denotes the whole time of a melody.

Feature set	Term
A	key signature, tempo, time signature
B	Chord, slur, tie, staccato, accent and fermata
C	maximum pitch, minimum pitch, average relative pitch, average absolute pitch, average positive interval and average negative interval
D	appoggiatura, mordent, trill, arpeggio, turn, crescendo, diminuendo, pianisissimo, pianissimo, piano, mezzo piano, mezzo forte, forte, fortissimo and fortisissimo

Table 1 Music features used in this paper.

Besides, the music feature in set C is about melodic gamut. We thought that the gamut is useful for interpreting a melody and encode all pitches of a melody into the figures (value: 1~88) according to a piano keyboard. Hence, we compute the important features, including maximum pitch, minimum pitch, average relative pitch, average absolute pitch, average positive interval and average negative interval. Finally, we put the music features (as Table 1) of all melodies above described into the features pool for music classifier.

Number of training samples	Feature set	Class set						*Mean*
		C-F	C-K	C-P	F-K	F-P	K-P	
60	A	50%	75%	75%	70%	70%	55%	65.8%
	B	60%	65%	55%	70%	60%	50%	60.0%
	C	45%	70%	45%	60%	50%	50%	53.3%
	D	45%	55%	65%	65%	60%	50%	56.7%
	A&B	50%	65%	60%	70%	75%	50%	61.7%
	A&B&C	50%	70%	65%	75%	70%	50%	63.3%
	A&B&C&D	65%	85%	65%	80%	65%	50%	68.3%
	Mean	52.1%	69.3%	61.4%	70.0%	64.3%	50.7%	61.3%
80	A	45%	75%	75%	65%	70%	65%	65.8%
	B	50%	65%	60%	70%	65%	65%	62.5%
	C	55%	50%	55%	70%	50%	60%	56.7%
	D	45%	55%	65%	60%	50%	50%	54.2%
	A&B	55%	75%	60%	75%	70%	60%	65.8%
	A&B&C	55%	80%	65%	75%	70%	60%	67.5%
	A&B&C&D	65%	85%	65%	80%	70%	55%	70.0%
	Mean	52.9%	69.3%	63.6%	70.7%	63.6%	59.3%	63.2%
100	A	55%	70%	***75%***	70%	70%	***65%***	67.5%
	B	50%	70%	60%	70%	60%	***65%***	62.5%
	C	45%	65%	50%	65%	50%	***65%***	56.7%
	D	55%	55%	70%	60%	65%	50%	59.2%
	A&B	65%	80%	70%	75%	70%	***65%***	70.8%
	A&B&C	***70%***	85%	65%	75%	***75%***	***65%***	*72.5%*
	A&B&C&D	55%	***90%***	***75%***	***80%***	70%	55%	70.8%
	Mean	56.4%	*73.6%*	66.4%	70.7%	65.7%	61.4%	65.7%

Table 2 Classification between categories.

4. EXPERIMENTAL EVALUATION

In this section, we evaluate proposed approaches about extraction of main melody and weighted music features by testing various kinds of music namely classic, folk, kid and pop music that are collected from finalemusic [5]. In order to verify that was also suitable for classification between sub-categories, we further divided classical music into three sub-categories namely baroque, classic and contemporary periods. For the classifier, we employed the decision tree for two-way classification. All the experimental datasets were split randomly into training set (80%) and test set (20%). In the training phase, different number of samples (each class was provided with 30, 40 and 50 training samples) were used to build the decision tree. After the training, each built classifier was tested by other cases (each class was provided with 10 cases) not contained in the training set for further evaluation.

4.1 Performance Analysis of Feature Sets

First, we want to investigate the effect of music feature sets so we grouped four categories into six clusters that each cluster consists of two categories for two-way classification. Feature sets are divided into single feature set or combinatorial feature sets.

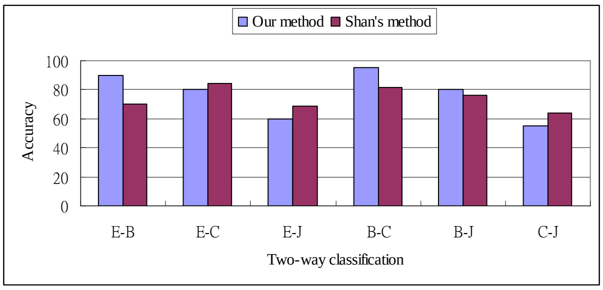

Fig. 5 Comparisons of the accuracy with Shan [13].

Table 2 shows the result where C, F, K, P denotes the categories of classic, folk, kid and pop respectively. The result indicates that the music features sets are able to improve the accuracy from 45% to 90%. We discover that the combinatorial feature sets are worth for music classification especially for the specific category such as classic-kid (C-K), folk-kid (F-K), etc. We review folk and kid songs and then discover that most folk songs have more chords and dynamics than kid songs. Most kid songs are C Major. In 100 training samples, we obtain the average accuracy from 56.4% to 73.6% if we do not consider single or combinatorial feature sets. The average 72.5% accuracy can be obtained using combinatorial feature sets (A&B&C) if we consider an optimal combination. In addition, the single feature set A was better than other single feature set in accuracy. We assume that each experiment can be based on feature set A and then they can be combined other feature sets to improve the accuracy. In order to prove that proposed approaches about extraction of main melody and weighted music features are reliable so we classified the music Shan [13] used. They are performed by two-way classification using 60 training samples. Fig. 5 shows that our some results are more than Shan's such as E-B, B-C and B-J by using combinatorial feature sets A&B and A&B&C. The Enya and Beatles' songs can be distinguished after combining feature set (C) because of they have the diverse pitches in most melodies. The Beatles, Chinese and Japan folk songs can also be distinguished after combining feature set (B) due to Beatles' songs using more chords or ties.

4.2 Performance Analysis of Sub-categories

Furthermore, we tested that proposed approaches about extraction of main melody and weighted music features were effective for sub-categories so we divided classical music into three different periods. Table 3 shows the results where C1, C2, C3 denotes the three sub-categories of classical music, namely baroque, classic and contemporary periods respectively and the suitable music feature sets are able to improve the accuracy from 45% to 80%.

Number of training samples	Feature set	Class set		
		C_1-C_2	C_1-C_3	C_2-C_3
60	A	65%	50%	50%
	A&B	70%	55%	45%
	A&B&C	70%	55%	55%
	A&B&C&D	60%	70%	55%
80	A	65%	60%	50%
	A&B	75%	70%	45%
	A&B&C	70%	80%	55%
	A&B&C&D	60%	70%	60%
100	A	65%	60%	50%
	A&B	75%	70%	50%
	A&B&C	75%	80%	55%
	A&B&C&D	60%	65%	55%

Table 3 Classification between sub-categories.

5. CONCLUSIONS

In this paper, we investigated the extraction of main melody in a music based on music theory and employee the suitable feature sets to classify music. Experimental results show that the classification accuracy was achieved as good as 90% by the most suitable features for 2-way classification. We discovered the interesting diversities between different music styles, such as the tempo, grace, slur, chords, staccato, etc. Through the experimental results, we confirm that the combinatorial feature sets are important for music classification. Moreover, the proposed most suitable features can quickly raise classification accuracy by using only 60 to 100 training samples. For future work, it is helpful for retrieving users' needs rapidly from music database. A combinatorial approach [11] has been proposed to solve. We may use these proposed music features to predict user's music hobby and further develop music recommendation systems. Besides, we will also test more categories of music in the future so as to extend the applications.

REFERENCES

[1] Chai, W. and Vercoe, B. Folk Music Classification Using Hidden Markov Models. In *Proc. of International Conference on Artificial Intelligence*, 2001.

[2] Chai, W. and Vercoe, B. Using User Models in Music Information Retrieval Systems. In *Proc. of Intl. Symposium on Music Information Retrieval*, 2000.

[3] Chen, H. C. and Chen, A. L. P. A Music Recommendation System Based on Music Data Grouping and User Interests. In *Proc. of ACM Intl. Conference on Information and Knowledge Management*, 2001.

[4] Fang-Fei Kuo, Meng-Fen Chiang, Man-Kwan Shan, Suh-Yin Lee: Emotion-based music recommendation by association discovery from film music. *ACM Multimedia*, pp.507-510, 2005.

[5] Finale Showcase. http://www.finalemusic.com/showcase/

[6] Liu, C. C., Hsu, J. L. and Chen, A. L. P. Discovering Nontrivial Repeating Patterns in Music Data. *IEEE Transactions on Multimedia*, 3(3), 2001.

[7] McKay, C., Fiebrink, R., McEnnis, D., Li, B. and Fujinaga, I. Ace: A Framework For Optimizing Music Classification. *6th International Conference on Music Information Retrieval*, 2005.

[8] MIDI Manufacturers Association. http://www.midi.org/

[9] Miller, H. M. and Williams, E. Introduction to Music, 1991.

[10] Musicnetwork. http://www.interactivemusicnetwork.org/wg_standards/index.html

[11] Pachet, F., Roy, P. and Cazaly, D. A combinatorial approach to content-based music selection. *IEEE Multimedia*, 7(1): 44-51, 2000.

[12] Recordare. http://www.recordare.com/software.html

[13] Shan, M. K. and Kuo, F. F. Music Style Mining and Classification by Melody. *IEICE Transactions on Information and Systems*, E86-D(4), 2003.

[14] Themefinder. http://www.themefinder.org/

[15] Tseng, Y. H. Content-Based Retrieval for Music Collections. *SIGIR*, pp. 176-182, 1999.

[16] Uitdenbogerd, A. L. and Zobel, J. Manipulation of Music for Melody Matching. In *Proc. of the ACM Multimedia*, pp. 235-240, 1998.

[17] Uitdenbogerd, A. L. and Zobel, J. Melodic Matching Techniques for Large Music Databases. In *Proc.of the ACM Multimedia*, pp. 57-66, 1999.

[18] W3Schools. http://www.w3schools.com/

[19] Weyde, T. and Datzko, C. Efficient Melody Retrieval With Motif Contour Classes. *ISMIR*, pp. 686-689, 2005.