Music Segmentation by Rhythmic Features and Melodic Shapes

Hung-Chen Chen, Chih-Hsiang Lin, and Arbee L.P. Chen*
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan 300, R.O.C.
alpchen@cs.nthu.edu.tw

ABSTRACT

According to the musicology, musical content and musical structure are both major components of a music work. Most approaches of music retrieval, classification, and analysis use the information of the musical content, but not the information of the musical structure. The main reason is that the musical structure usually needs to be analyzed manually by experts, which is time-consuming and impractical. In this paper, we propose an approach for automatic music segmentation to extract the phrases and sentences of the musical structure. In addition to the rhythmic features, the melodic shape is also used to improve the effectiveness of the music segmentation. The experiments are performed to show that our approach is practical.

Keywords

Musical Features, Music Segmentation, Musical Structure, Music Retrieval

1. INTRODUCTION

According to the musicology, musical content and musical structure are both major components of a music work. The relationship of these two components is especially close from the viewpoints of composers. The musical structure is used to express the musical content while the musical content is used to determine the musical structure. However, most approaches of music retrieval, classification, or analysis use the information of the musical content, but not the information of the musical structure.

Many researchers indicate that the performance of the approaches for music retrieval, classification, and analysis could be improved by combining the information of the musical structure. For example, [12] shows that the learning rules of melodic expression are significantly improved when considering the musical structure. Even though, most researchers only focus on musical content without considering musical structure. The main reason is that the musical structure usually needs to be analyzed manually by experts, which is time-consuming and impractical. Therefore, a practical approach for structure analysis is urgently needed.

In this paper, an approach of automatic music segmentation is proposed to extract the phrases and sentences of the musical structure. Both of the rhythmic features and the concept *melodic shape* are used for music segmentation in our approach.

The rest of this paper is organized as follows. In Section 2, we present related work and define key terms of this paper. In Section 3, we present our segmentation approach. In Section 4, we use an example music object for music segmentation and perform experiments to show that our approach is practical. Finally, we conclude this paper in Section 5.

2. RELATED WORK AND DEFINITIONS

Many researchers in musicology and music psychology fields agree that repetition is a universal characteristic in musical structure modeling [7] [8]. Accordingly, various algorithms are developed to find the exact or approximate repeating patterns in a music object [2] [5] [10] [11]. However, these algorithms for finding repeating patterns do not perform structure analysis since the discovered repeating patterns may overlap or not satisfy the requirements of musicology.

Several researchers have proposed approaches for music segmentation in phrase-level and sentence-level [3] [14]. The shortcoming of these approaches is that only the rests or the lengthened notes are considered for segmentation. [1] introduces the Local Boundary Detection Model (LBDM) to calculate boundary strength values for each interval of a melodic surface, *i.e.*, pitch, duration, and rest, according to the strength of local discontinuities. However, to choose an appropriate weight for each interval is very difficult.

[6] provides a concept of melodic shape for phrase observation. In the case of Essen Folksong Collection, the statistics shows musical phrases tend to exhibit an arch-shaped pitch contour. Of the 36075 phrases in the database, three-quarters of all phrases are between six and ten notes in length. Table 1 shows the statistics for different shape types in the database. The top four shape types are *convex*, *descending*, *ascending*, and *concave*. Using the prior knowledge showed in Table 1, we design a heuristic approach for phrase-based segmentation. Then, repeating sentences can be obtained by merging these extracted phrases.

Here, we define the key terms of this paper.

Definition 1: Terminative Note

The note, that satisfies one of the two conditions described below, is marked as a terminative note.

(1) The note preceding a *rest*. (A rest is an interval of silence in a music work, marked by a sign indicating the length of the pause [13].)

(2) The note with a sudden change of boundary strength in terms of the duration intervals [1]. (The way to compute the boundary strength and detect a sudden change is described in Subsection 3.1.1.)

Definition 2: Music Piece

A music piece is a piece of music, which starts either in the beginning of a music object or right after a terminative note and ends with a terminative note.

Definition 3: Phrase

A phrase is a fragment of a music piece or a complete music piece, which satisfies the length constraint and exposes a certain melodic shape.

Definition 4: Sentence

A sentence is a basic unit of a music structure. Each sentence consists of one or more successive phrases, which exactly or

* To whom all correspondence should be sent.

approximately repeats in a music object.

Table 1. The statistics and definitions of all melodic shape types in Essen Folksong Collection.

A: the pitch value of the first note in target phrase
B: the pitch value of the last note in target phrase
C: the average pitch value of the remaining notes in target phrase

Melodic Shape Type	Number of Phrases	Percentage	Arch Shape	Definition
Convex	13926	38.6%		A<C ∧ B<C
Descending	10376	28.8%		A>C>B
Ascending	6983	19.4%		B>C>A
Concave	3496	9.7%		A>C ∧ B>C
Others	1294	3.5%		

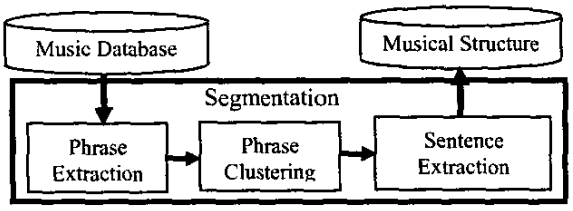

Figure 1. The process of our music segmentation approach.

3. MUSIC SEGMENTATION

The process of our music segmentation approach is extended from that of [14], which is described in Figure 1. There are three main steps of the process, *i.e.*, phrase extraction, phrase clustering, and sentence extraction. Note that our approach is designed for monophonic music objects which are symbolic score-like representation, such as MIDI format.

3.1 Phrase Extraction

3.1.1 Decision of Terminative Note

According to Definition 1, there are two kinds of terminative notes in a music object. The first kind of terminative note is the note preceding a rest and the second kind is the note with a sudden change of boundary strength in terms of duration intervals. We can detect whether there exists a rest which follows the target note by calculating the interval between the current offset and the next onset. On the other hand, the sudden change of boundary strength in terms of duration intervals can be detected by using Equation (1) and (2) presented in [1]. After the calculations, the sequence $[S_1, S_2, \ldots, S_n]$ can be obtained from a music object with the length n. Any local peak in this sequence indicates that the corresponding note is a terminative note.

$$DEG_{i,i+1} = |D_i - D_{i+1}| / (D_i + D_{i+1}) \quad (1)$$

where D_i *is the duration interval of the target note* N_i
D_{i+1} *is the duration interval of the following note* N_{i+1}
$DEG_{i,i+1}$ *is the degree of duration changing between* N_i *and* N_{i+1}

$$S_i = D_i \times (DEG_{i-1,i} + DEG_{i,i+1}) \quad (2)$$

where $DEG_{i-1,i}$ *is the degree of duration changing between* N_{i-1} *and* N_i
$DEG_{i,i+1}$ *is the degree of duration changing between* N_i *and* N_{i+1}
S_i *is the strength of the note* N_i

3.1.2 Heuristic Approach for Phrase-based Segmentation

After identifying the positions of all the terminative notes, the music pieces of a music object are extracted according to the terminative notes. However, a music piece is not always a phrase according to Definition 3. It may happen that there is no terminative note between two successive phrases. Therefore, we need to decompose some music pieces into phrases. The statistical information of the phrase length in [6] can help us to choose the candidate music pieces for decomposition. For phrase-based segmentation, we make an assumption that if a music piece is too long (> k), there may be more than one phrase composes it. Based on the statistical information, we set the k to be 12 (twice of 6) for our approach and experiments. If the length of a music piece is less then or equal to 12, the music piece is directly marked as a phrase. Otherwise, we try to decompose the music piece into phrases by using the concept of melodic shapes. The minimum length of a phrase, which is decomposed from a music piece, is fixed to 6 to avoid over-segmenting. Besides the original definitions of the four melodic shapes in Table 1, we add additional constrains for the convex and the concave. That is, the pitch difference between the first note and last note of a convex or a concave cannot exceed two semitones. By using the statistic information of each melodic shape for decomposition, the priority assigned to each melodic shape is based on the global possibility. When trying to decompose a music piece, we first check whether a prefix music fragment with certain length of the music piece exposes a convex melodic shape. If all the prefix music fragments with lengths from 6 to 12 do not expose a convex melodic shape, the descending melodic shape is then used to check the possible phrase, and so on. If a prefix music fragment is marked as a phrase with certain melodic shape, the prefix music fragment is then removed from the corresponding music piece. Again, the length of the remaining music piece will be checked to see whether it is over 12. If so, the remaining music piece will require further decomposition.

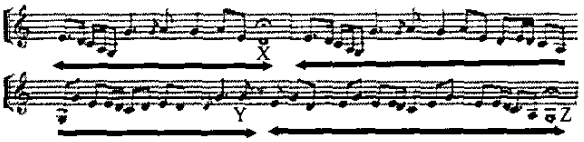

Figure 2. A part of an example music object.

[Example] We take a part of a music object shown in Figure 2 as an example to illustrate the process of phrase extraction. The first step of the process is to identify the positions of all the terminative notes. The note Y is marked as a terminative node because a rest follows it. Similarly, the note X and Z are marked as terminative notes because they both have a sudden change of boundary strength in terms of duration intervals calculating by Equation (1) and (2). Therefore, there are three music pieces in Figure 2.

The next step of the process is to decompose music pieces. Because the length of the first music piece is equal to 12, this music piece is directly marked as a phrase with a descending melodic shape. On the contrary, the length of the second music piece is 28, which implies that this music piece contains more than one phrase. We first detect whether the prefix music fragments of the music piece expose the convex melodic shape. The pitches of the first 12 notes are [64, 62, 60, 57, 55, 67, 67, 69, 67, 69, 64, 62]. Table 2 shows the prefix music fragments and their checking orders for detecting a possible phrase. According to the constraint of the convex, only the sixth and the seventh prefix music fragments could expose a convex melodic shape because their pitches of the last notes fall into the range [64-2, 64+2]. For the sixth prefix music fragment, A, B, and C is 64, 64, and 63.67,

respectively. For the seventh prefix music fragment, A, B, and C, is 64, 62, and 63.7, respectively. Therefore, both of them do not expose a convex melodic shape. Next, we detect whether the prefix music fragments in Table 2 expose the descending melodic shape. Then, we can find that the seventh prefix music fragment satisfies the definition of the descending. As the result, the seventh prefix music fragment is marked as a phrase with a descending melodic shape. Because the length of the remaining music piece is 16, we continuously decompose it into phrases. Finally, three phrases are extracted from the second music piece. The second phrase in the second music piece exposes a descending melodic shape with the length 10. The third phrase exposes an ascending melodic shape with the length 6.

Table 2. The prefix music fragments and their checking orders for detecting melodic shapes.

Order	Length	The pitches of the prefix music fragment
1	6	64, 62, 60, 57, 55, 67
2	7	64,62, 60, 57, 55, 67, 67
3	8	64,62, 60, 57, 55, 67, 67, 69
4	9	64,62, 60, 57, 55, 67, 67, 69, 67
5	10	64,62, 60, 57, 55, 67, 67, 69, 67, 69
6	11	64,62, 60, 57, 55, 67, 67, 69, 67, 69, 64
7	12	64,62, 60, 57, 55, 67, 67, 69, 67, 69, 64, 62

3.2 Phrase Clustering

After phrase extraction, the similar phrases are then grouped for sentence extraction. Each phrase is represented as a music contour [4] with the pitch of the first note. The similarity measure used for phrase clustering is defined in Equation (3) based on the contour and the pitch of the first note. We use the concept of longest common subsequence to measure the similarity between two music contours as shown in Equation (4). The similarity measure for the pitch similarity is shown in Equation (5). Moreover, we dynamically adjust the weights W_c and W_p based on the length of the shorter phrase by Equation (6). By using Equation (3), all phrases are then divided into phrase clusters. Two phrases p_i and p_j are grouped into the same cluster only if $Sim(p_i,p_j)$ is greater than a given threshold.

$$Sim(p_i,p_j) = W_c * Contour\_Sim(p_i,p_j) + W_p * Pitch\_Sim(p_i,p_j) \quad (3)$$

where p_i and p_j are both music phrases
W_c is the weight for contour similarity between the two phrases
W_p is the weight for pitch similarity between the two phrases
$W_c + W_p = 1$

$$Contour\_Sim(p_i,p_j) = |LCS(c_i,c_j)| / MAX(|c_i|,|c_j|) \quad (4)$$

where c_i is the contour of p_i and c_j is the contour of p_j
$|LCS(c_i,c_j)|$ is the length of the longest common subsequence of the contours c_i and c_j
$MAX(|c_i|,|c_j|)$ is the maximum length of the contours c_i and c_j

$$Pitch\_Sim(p_i,p_j) = \begin{cases} 1 - \dfrac{|pit_i - pit_j|}{12}, & \text{if } |pit_i - pit_j| < 12 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where pit_i is the pitch of the first note of p_i
pit_j is the pitch of the first note of p_j
12 is the number of the semitones in a octave

$$W_p = -0.1 * MIN(|p_i|,|p_j|) + 1.1 \text{ AND } W_c = 1 - W_p \quad (6)$$

where $MIN(|p_i|,|p_j|)$ is the minimum length of p_i and p_j

3.3 Sentence Extraction

By the result of phrase clustering, a music object can be represented as a sequence of class labels. Simultaneously, we identify the possible starting phrases for sentence extraction by extending the rules presented in [14]. The phrase at the beginning of a music object is marked as a possible starting phrase. In addition, the phrase that follows a note or a rest whose duration is longer than a given threshold δ, *e.g.*, two beats, is marked as a possible starting phrase. Note that a possible starting phrase can only be positioned at the beginning of an extracting sentence. Using the information of possible starting phrases and the sequence of class labels, we can adopt the technique of sequential pattern mining [9] to extract repeating sentences.

4. EXPERIMENTS

An example music object for displaying the result of our approach is shown in Subsection 4.1. Moreover, the experiments for the effectiveness of music segmentation are shown and explained in Subsection 4.2.

4.1 Example Music Object

An indigenous music object named "Uyas Mrmum" [15] is selected to display the result of out approach for music segmentation as shown in Figure 3. After phrase-based segmentation, nine phrases are extracted from the music object. The melodic shapes of the extracted phrased are also shown in Figure 3. After sentence extraction, we can find that the sentence I and the sentence II repeat three times in this music object.

4.2 Effectiveness of the Music Segmentation

To evaluate the effectiveness of the phrase-based segmentation approach, we use 50 indigenous music objects [15] as the testing data set. The phrases of these music objects are extracted by the experts. That is, the end point of each phrase is marked by the experts. Suppose EX is the set of all end points marked by experts and AP is the set of all end points found by our approach. If a point E in EX and a point A in AP are at the same position, the point A is marked as a matching end point. If there is no point in EX at the same position of a point A in AP, the end point A is marked as a redundant end point. On the contrary, if there is no end point in AP at the same position of a point E in EX, the point E is marked as a missing end point. By comparing the end points marked by the experts and the end points found by our phrase-based segmentation approach, the numbers of the matching end points (N_{mat}), the missing end points (N_{mis}) and the redundant end points (N_{red}) are counted to evaluate our approach as shown in Table 3. Here, N_{exp} is the number of end points indicated by the experts. The number of the end points found by our approach is $N_{mat} + N_{red}$.

Similarly, to evaluate the effectiveness of the sentence extraction approach, the repeating sentences are also extracted by the experts from the training data set. The experiment is performed to see how many existing repeating sentences can be found by using our approach. The experimental result for sentence extraction is shown in Table 4. N_{rep} is the number of total repeating sentences selected by the experts and N_{fou} is the number of repeating sentences found by our approach.

According to Table 3, the number of the redundant end points is more than half of the number of the mussing end points. The main reason that causes the redundant end points is that some longer phrases consist of sub-phrases. For example, a sub-phrase with an ascending melodic shape and another sub-phrase with a

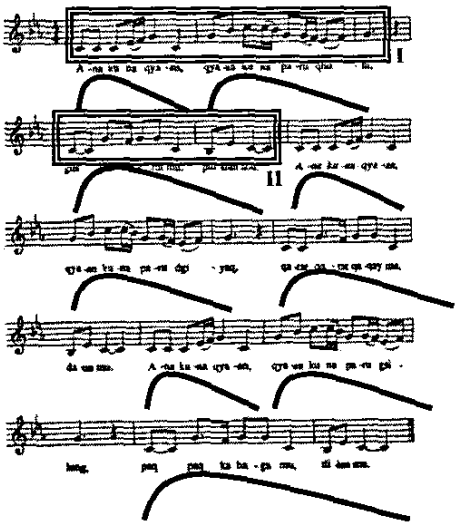

Figure 3. An indigenous music object named "Uyas Mrmum."

descending melodic shape compose of a phrase with a convex melodic shape. This situation is also introduced in [6]. Therefore, our phrase-based segmentation approach may over-segment longer phrases. The problem of phrase over-segmentation doesn't seriously decrease the precision of the sentence extraction, because the step of sentence extraction will merge the successive sub-phrases. However, the precision of the sentence extraction is still lower than that of the phrase extraction. The main reason is that some possible starting phrases are lost, because they don't follow a note or a rest with enough duration. If a possible starting phrase of certain repeating sentence is lost, we will never extract that sentence. Decreasing the threshold δ will cause another problem of sentence over-segmentation. Therefore, we still need to develop a general method to identify all possible starting phrases for improving the effectiveness of the sentence extraction.

Table 3. The effectiveness of the phrase-based segmentation.

N_{exp}	N_{mat}	N_{mis}	N_{red}	Precision	Recall
596	465	131	215	68.38%	78.02%

Table 4. The effectiveness of the repeating sentence extraction.

N_{rep}	N_{fou}	Recall
126	79	62.70%

5. CONCLUSION

In this paper, we propose an approach to automatically segment a music object and extract the phrases and sentences of the music object as its musical structure. Previous research only considers rests and lengthened notes, or variance of pitch and duration for music segmentation. Other musical characteristics, such as smoothness and harmonization, are ignored. In this paper, we adopt the concept of melodic shape to enhance the effectiveness of the music segmentation. The concept of melodic shape can be used to segment a music piece without rests, lengthened notes or a great variance of pitch and duration. As such, our approach extracts phrases which cannot be extracted by other approaches. The first step of our approach is to use both the rhythmic features and the melodic shapes for phrase extraction. After segmenting a music object into phrases, a similarity measure considering the music contour and the pitch of the first note is provided for phrase clustering. Based on the result of the phrase clustering, the repeating sentences are extracted by using a data mining technique. Finally, an example music object is used to show the result of our approach. Experiments are also performed to show the effectiveness of the music segmentation at both the phrase level and sentence level.

Our approach for music segmentation is designed for monophonic music objects. In the future, we will research into automatic segmentation of polyphonic music objects. Moreover, other methods for sentence extraction will also be investigated to improve the effectiveness of the approach.

6. REFERENCES

[1] Cambouropoulos, E., "The Local Boundary Detection Model (LBDM) and its Application in the Study of Expressive Timing," *Proc. of International Computer Music Conference*, 2001.

[2] Dannenberg, R. B. and N. Hu, "Pattern Discovery Techniques for Music Audio," *Proc. of International Symposium on Music Information Retrieval*, 2002.

[3] Friberg, A., R. Bresin, L. Fryden, and J. Sunberg, "Musical Punctuation on the Microlevel: Automatic Identification and Performance of Small Melodic Units," *Journal of New Music Research*, 1998.

[4] Ghias, A., H. Logan, D. Chamberlin, and B. C. Smith, "Query by Humming: Musical Information Retrieval in an Audio Database," *Proc. of ACM International Conference on Multimedia*, 1995.

[5] Hsu, J. L., C. C. Liu, and A. L. P. Chen, "Discovering Nontrivial Repeating Patterns in Music Data," *IEEE Transactions on Multimedia*, Vol. 3, No. 3, 2001.

[6] Huron, D., "The Melodic Arch in Western Folksongs," *Computing in Musicology*, Volume10, 1995.

[7] Krumhansl, C. L., *Cognitive Foundations of Musical Pitch*, Oxford University Press, New York, 1990.

[8] Narmour, E., *The Analysis and Cognition of Basic Melodic Structures*, The University of Chicago Press, Chicago, 1990.

[9] Pei, J., J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M. C. Hsu, "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth," *Proc. of International Conference on Data Engineering*, 2001.

[10] Pienimaki, A., "Indexing Music Databases Using Automatic Extraction of Frequent Phrases," *Proc. of International Symposium on Music Information Retrieval*, 2002.

[11] Rolland, P. Y., "FlExPat: Flexible Extraction of Sequential Patterns," *Proc. of the IEEE International Conference on Mata Mining*, 2001.

[12] Widmer, G., "Learning Expressive Performance: The Structure-Level Approach," *Journal of New Music Research* 25(2), pp. 179-205, 1996.

[13] http://www.wikipedia.org/wiki/Rest.

[14] Yanase, T., "Phrase Based Feature Extraction for Musical Information Retrieval," *Proc. of Communications, Computers and Signal Processing, IEEE Pacific Rim Conference*, 1999.

[15] Yu, C. F., *The new sound from ancestors: a hundred indigenous folk songs in Taiwan*, Taiwan Yuan-Yuan Indigenous Culture Troupe, 1998.

ACKNOWLEDGEMENT

This work was partially supported by National Science Council, Republic of China under Contract Numbers NSC 92-2213-E-007-094 and NSC 92-2213-E-007-091.