Lyric Oriented Rhythmic Framework Generating for Automatic Chinese Song Composition

Hao Wu [a], Lu Xie[b] and Xihong Wu[c]

Speech and Hearing Research center, Peking University, Peking, China

[a]wuhao@cis.pku.edu.cn, [b]xlu@cis.pku.edu.cn, [c]wxh@cis.pku.edu.cn

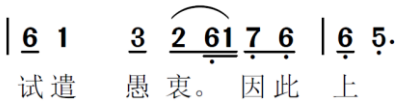
Abstract. This paper proposes a lyric oriented rhythm framework generating (LORFG) method for automatic Chinese song composition. For a given lyric, this method constructs a rhythmic framework, in which each linguistic prosodic unit (such as sentence, prosodic phrase, prosodic word, and Chinese characters) in the lyric is properly assigned with certain musical time span in the meter of musical beat, and thus build a melody rhythmic pattern for composing a song. The framework is constructed in a hierarchical way: For the topmost paragraph level, we used passage analysis and paragraph grouping to generating basic music form. For lower levels such as sentences and phrases, tree matching based pattern searching is performed to find the best matching rhythm patterns from predefined rhythm tree bank. For the lowest level (words and characters), an aligning method is performed to assign each Chinese character with musical notes so as to optimize and refine the final framework.

Keywords: Automatic Song Composition, Rhythmic Framework,

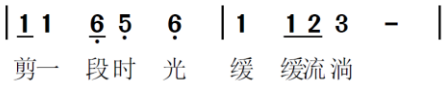
Introduction

Automatic Song Composition (ASC) is an extensive task for Automatic Composition (also known as Algorithm Composition). Deriving the issues from general automatic composition task [1][2][3], ASC raises its extra challenge of giving consideration on the given lyric; and one major issue of this challenge is problem of keeping the rhythmic consistence between the melody and lyric.

The term ‘rhythm’ means a "movement marked by the regulated succession of strong and weak elements [1]." While rhythm is regulated by prosodic structure of the sentence in spoken language, it is regulated by the music form and beat structure in music [5][6]. When composing a song, it is important that the prosodic segmentation in lyric should be consistent with the music form and metric level beats in generated melody. Without such constraint, the lyric could be incorrectly received and by listeners [7]. This problem could be illustrated with the following example as in Fig .1 (a) , which is drawn from a classical Chinese song “红楼梦.引子” (the introduction from the Stone Story). The lyric was made by Xueqin Cao, and the melody was composed by Chinese composer Liping Wang.



(a)



(b)

Fig.1 lyric in wrong melody rhythmic framework

In this example, while the phrase “试遣愚衷” and “因此上” belong to different sentences, they are incorrectly grouped into one melody sentence. This introduces significant confusion on the lyric.

Another example drawn from pop song “荷塘月色”(moonlight in lotus pond) in Fig.1 (b) gives the illustration of this problem when it is occurring in a shorter scale rhythmic span. In this example, the lyric piece has linguistic structure like [剪[一段]][时光][缓缓][流淌]. However, in the song, the lyric is modulated by the melody rhythmic structure and segmented with a prosodic structure of [[剪一]段][时光][[缓[缓流]淌], so people can hardly identifies the correct word [一段] and [缓缓]. Instead , people tend to join the syllable “jian” from “剪” and “yi “ from “一” together and group them into a wrong word such as “建议”. The syllable “huan” from second “缓” and “liu “ from “流” are also tend to be joined together and grouped into a wrong word such as “环流”.

Since both lyric prosody and melody form are organized in a hierarchical way, the above mentioned rhythmic consistence should be further kept in every level. For example, if the given lyric contains three paragraphs, it is necessary to derive a three-section musical form structure, such as AAA, ABA, or ABB etc., so that each lyric paragraph can be assigned with correspondent musical section; also, for each less scale prosodic units (e.g. sentence, phrase,…) in the lyric paragraph, it is necessary to be assigned with proper rhythmic segmentation with notable boundaries (e.g. ending with fairly long and stable note or chord) so that it can be perceptional separated from neighboring segmentations. This is defined as lyric oriented rhythmic framework generating (LORFG) in our research.

In our system, this process is integrated into a three-stage algorithm: first we use passage analysis based on structural similarity to group the lyric paragraphs to determine the basic music form, such as AAA, ABA, etc. In the second stage, for each paragraph, a corresponding prosodic tree is generated with automatic prosodic analysis based on conditional random field (CRF) [8]. A pattern searching process is then performed in a prebuilt prosodic tree bank to find the song of which the lyric prosodic tree best matches the given prosodic tree. Due to the data sparse problem in the prosodic tree bank, there is significant possibility to directly find the exact framework. So at last stage, a character level alignment is performed to fill the rhythmic tree with timespan-assigned characters, and thus get the final rhythm framework. The form of resulted sentence framework can be illustrated with Fig. 2.

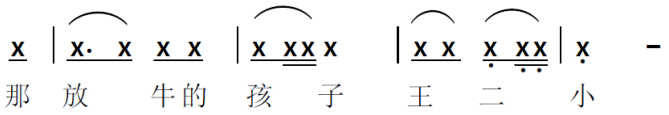


Fig.2 the melody rhythmic framework for given lyric sentence
“那放牛的孩子王二小”

The rest parts of this paper are organized as follows: in the second and third section, we gives brief introduction on the hierarchical rhythmic structure for lyric and music. In the fourth section, we present the algorithm in detail. Then we illustrate the effectiveness with examples and draw the conclusion.

Hierarchical rhythmic framework for melody and lyric

Hierarchical Music Form of Song Melody and Representation. In basic principles of music form, melody is defined as an organized succession of musical tones [9]. This succession involves of different kinds of structural units, the principal of which is the phrase—a complete musical utterance. A melody, then, ordinarily consists of a succession of phrases. Regulated by the relation between successive ones, the phrases are grouped into longer melodic segmentations. Borrowing concepts from language, these groups are categorized according to their scale as: periods, sentences, and paragraphs. In this way, musical forms are not only additive but also hierarchical: phrases are conjoined to produce a period, and periods are grouped into sentence, which in turn part of a larger paragraph. Also, in our research, there is necessity of giving structure analysis finer than phase level; so we down extend such hierarchy system with an extra sub-phrase level of “rhythmic element”, which represents the basic elementary regular figure of melody segmentation, and generally lasts for duration of less than two beats. With such hierarchy, the entire decomposition of a song can be illustrated with a tree structure. For example, for the first section of very popular Chinese folk song “lullaby” (摇篮曲), the decomposition of melody can be shown as figure 2

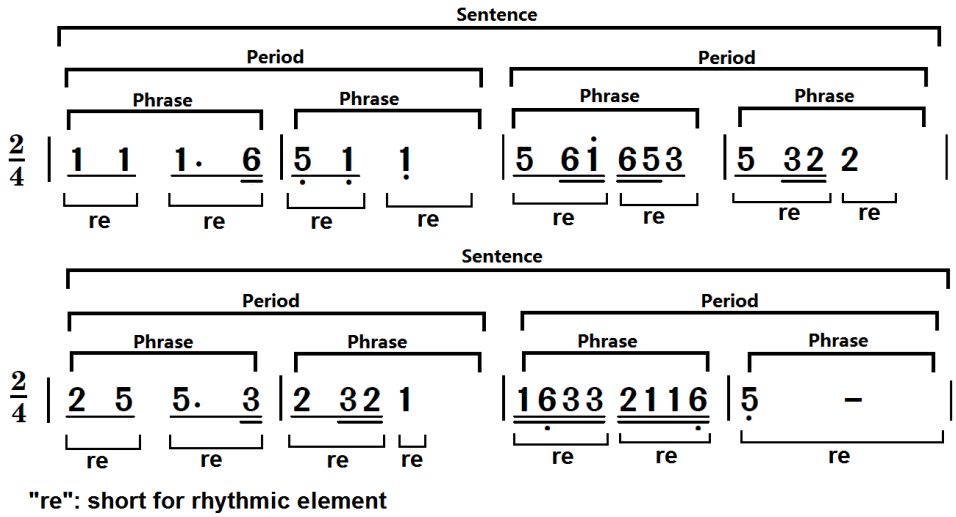


Fig.3 the rhythmic structure for Chinese folk song “lullaby” (摇篮曲)

In this paper, the rhythmic structure for a song is represented by labeling the position of each notes in the melody with a tag that indicates what is the position of this note (at beginning, ending, or middle) in the sub-section unit. The tags and its meaning are list in table 1.

Tabel 1. The Tags for Melody Structure Labeling

The position of note / the level	beginning	middle	ending	single
Sentence	b4	m	e	s4
Period	b3	m	e	s3
Phrase	b2	m	e	s2
rhythmic element	b1	m	e	s1

Considering that a beginning for higher level units also indicate a beginning for lower level unit (e.g. the beginning of a sentence also indicates a beginning of a phrase and a word), only the topmost level beginning are labeled for each note. The tags for notes that span a whole single unit are treated in this way as well. Also, since the level for each unit ending can be determined by the level of following unit beginning, the level tag for each ending note can be ignored, and thus the ending tags can be merged to one tag “e”. The tags for notes at middle of the units are treated in this way as well and merged to one tag of “m”. For example, for a sentence in Chinese folk song “歌唱二小放牛郎 ”， the derived melody rhythmic framework and corresponding note level tagging can be illustrated by Fig.4:

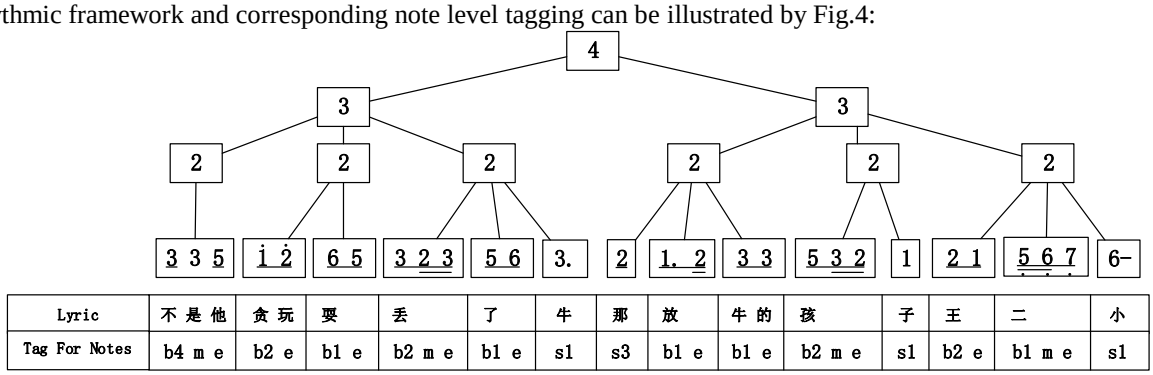


Fig.4 hierarchical melody structure and corresponding tags for a sentence in “歌唱二小放牛郎 ”

Hierarchical Lyric Structure. The lyric for a song can also be analyzed in similar way as in melody analysis. Though being one kind of poetic literary form, Chinese lyrics still follow the basic language rules and structure. Lyric has to convey a complete thought and follows general narrating process [7]. The basic structure of lyric has to be in consistent with the general narrating passages, and follow similar

hierarchical structure. Based on such fact, in this paper, we derive the lyrics analysis based on general syntactic analysis for Chinese text.

This analysis is performed in hierarchical manner. Similar to general passage, lyric can be decomposed into multiple scale units such as paragraphs, sentences, phrases, words and Chinese characters. However, considering the difficulty in long passage analysis, paragraph level analysis is separated alone as a preliminary procedure. Also, for those levels lower than sentence, it is more reasonable that we use rhythmic segmentation rather than syntactic segmentation in the context of oral representations such as talking and singing. In our research, the sub-sentence lyric structure is constructed with 3 levels as intonational phrase, phonological phase, and prosodic word [10]. Given that, the entire decomposition of lyric can also be illustrated as a tree structure just like in melody analysis.

Similar to the tagging system in melody structure labeling, the tags used for labeling the sub-paragraph level prosodic structure is defined in table 2:

Table 2. Structural tags for lyric analysis

Character position in the prosodic unit / The level of prosodic unit	Beginning	middle	ending	Single
Sentence	B4	M	E	S4
Intonational phrase	B3	M	E	S3
Prosodic phrase	B2	M	E	S2
Prosodic word	B1	M	E	S1

For example, for the above mentioned sentence in Chinese folk song “歌唱二小放牛郎 ”，the derived lyric tagging can be illustrated by Fig.5.

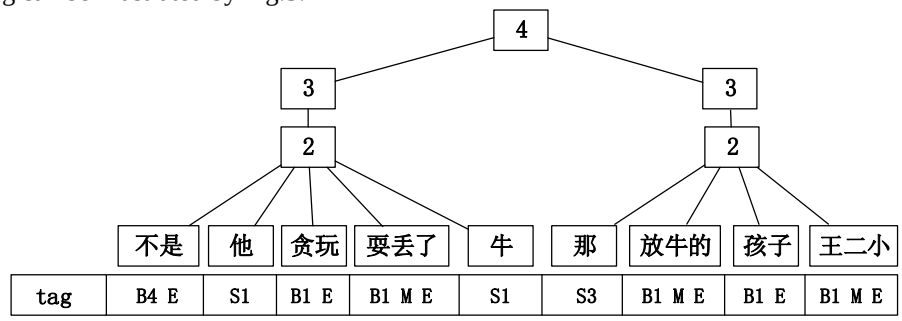


Fig.5 lyric structure and corresponding tags for a sentence in “歌唱二小放牛郎 ”

Rhythmic Alignment between Melody and Lyric. As been presented before, to make song understandable, the rhythmic segmentation of lyric should be in consistence with that of the melody.

One issue has to be concerned is that the evaluation of the hierarchical music structure relies significantly on a series of complicated melodic regulations, which are related to the long span grouping and collocation for rhythmic units [11]. Unfortunately, the hierarchical sematic segmentations for lyrics seldom satisfy the regulations. So it’s not operable to directly derive the rhythmic framework from the semantic or prosodic structure of lyric (e.g. using the semantic tree as the tree structure of the rhythmic framework).

For this reason, instead of directly mapping, we proposed an indirect strategy that selecting and composing the rhythmic framework from a pre-established song set. For each song in the set, the melodic rhythmic framework and the prosodic segmentation are extracted and labeled with above mentioned tags, and stored as the melodic rhythmic pattern and prosodic pattern for making new songs. When inputting a lyric, a search in the database is performed to find patterns that locally or globally best matches the given lyric. These patterns are then used to construct the rhythmic framework. Since these patterns are drawn from the songs that satisfy the long span rhythmic segment regulations, the result framework can meet the rules as well.

Automatic Rhythmic Framework Generating

Principle of the Algorithm. The basic principle of our algorithm is illustrated in Fig. 6

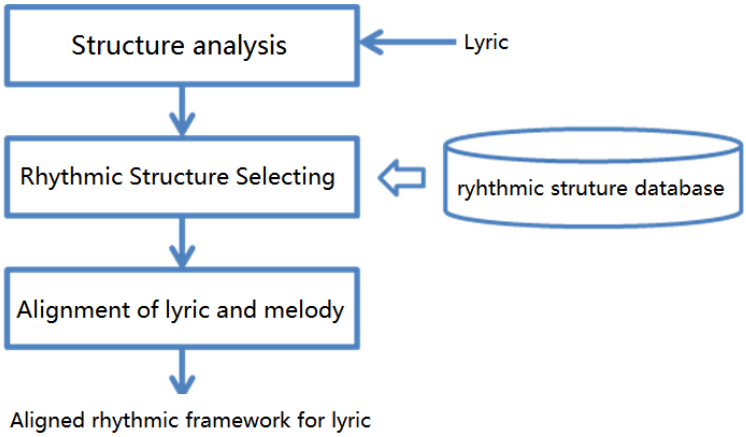


Fig.6 The basic flowchart of LORFG

The whole algorithm concerns a rhythmic structure database and three procedures including lyric structure analysis, rhythmic structure selecting, and alignment for lyric and melody.

Rhythmic Structure Database. As being shown in the graph, a musical material database is used to contain a set of musical structural materials. These materials include the rhythmic structure tree, the motive level rhythmic patterns and the 2-gram statistic of these patterns. This database is setup by tagging the structural information for songs in a training song set.

In preliminary researching stage, we need the rhythmic pattern and lyric text to be as simple as possible. Also, when building database, the selected songs should have good rhythmic synchronization. Since most nursery rhymes meet these requirements well, we focus our task on this kind of songs in this paper.

The songs are obtained from an attached score collection of the popular Chinese scoring software "composer master"[12]. In original "composer master" set, there are 326 songs which have good rhythmic synchronization. All the songs are scored with digitalized numbered musical notation. After removing 37 songs which have no corresponding lyrics, we set up a training score data set including 289 songs. Each song in database was converted into Music XML format. Furthermore, we extended the Music XML format so that we can incorporate with the structural and the lyric information.

We use this data set to generate the material data set. For each song, we applied the manual tagging and automatic tagging to extract the different level of signatures and others structural tags and lyric rhythmic tags.

Melody Structure Tagging. In this research, we derive the framework by manual-automatic hybrid method. For each song, we first spotted all the motives with automatic motives clustering algorithm proposed by David Cope [14][15][16]. Then, each rhythmic elements (which are referred to as "signature" and "unification" in Cope's books) were manually tagged with structural labels defined in Table 1.

CRF based Lyric Analysis and prosodic tagging. The hierarchical structure of lyric is derived with a CRF based prosodic tagging method [8]. The basic labeling rules are borrowed and from prosodic transcription system described in [10], but with necessary simplification. For each Chinese character in lyric, the corresponding prosodic level and the character position of the level is marked with designed tags as illustrated in table 2.

Hierarchical Matching for Lyric structure and Melody Rhythm.

The generation of rhythmic framework involves three stages: the first stage is the passage analysis and section grouping, in which the paragraphs of input lyric are grouped into sections; in the second stage of hierarchical prosodic tree matching, a search is performed to gather a collection of candidate framework patterns; in the third stage of melody-lyric alignment, an dynamic programming based alignment is performed that assign each character in the lyric with beat spans in the candidate framework, and finally the best alignment result is selected as the final rhythmic framework.

Passage Analysis and Section Grouping for Input Lyric. The input lyric has already segmented into paragraphs. In passage analysis, these paragraphs are grouped in to sections based on the similarities between each pair of them. The first paragraph is defined as the "A" section, then each of the following paragraph are compared with the section "A", those paragraph similar to section A is categorized as section A, while other paragraph are categorized as section B in our current implementation.

The similarity between paragraph is measured based on the numbers of sub-sentences in the paragraphs and the number of Chinese characters in corresponding sub-sentence of two paragraphs (The

term sub-sentences mentioned here refers to the segment that separated by comma). In our study, when satisfying the following conditions, two paragraphs are considered to be similar:

1. Exact the same: two paragraphs contain same number of sub-sentences, and have same number of Chinese characters in each pair of corresponding sub-sentences.

2. Matchable lyric paragraphs: the two paragraphs contain same number of sub-sentences, but the numbers of Chinese characters in each pair of corresponding sub-sentences are slightly different (less than two characters).

3. Matchable lyric paragraphs with sub-sentences merging: the two paragraphs contain different number of sub-sentences, but when merging two or more successive sub-sentences, the resulted combination can be matchable to the sub-sentence at the corresponding in the other paragraph.

Hierarchical Matching for Lyric structure. In this stage, a series of candidate rhythmic frameworks are selected for each of grouped lyric sections from the database.

For each lyric section, the CRF based prosodic analysis is also performed. The resulted prosodic tree is then used to compare with the prosodic trees in the database. A rapid tree matching method is used for selecting the best matching tree from the rhythmic material database for a given lyric.

Having the same number of prosodic words in two trees, the similarity of the trees is measured with the F1 score as defined in (1)[17].

$$F_1 = \frac{2PR}{P+R} \qquad (1)$$

In which the precision P and recall R are defined as:

$$P = \frac{S}{T}, R = \frac{S}{G} \qquad (2)$$

Where S represents the number of tree nodes that have matchable positions, T represents the numbers of the tree nodes in input prosodic tree, and G represents the numbers of tree nodes in reference prosodic tree.

Before matching, the input prosodic tree and reference prosodic tree (lyric prosodic tree for candidate rhythmic framework) has to be transformed into another representation that is convenient for comparison. Fig.7 gives the prosodic tree for two lyric sentences: “清晨的阳光清新而又温暖，绿草鲜花展开了笑颜” and “不是他谈玩耍丢了牛，那放牛的孩子王二小”.

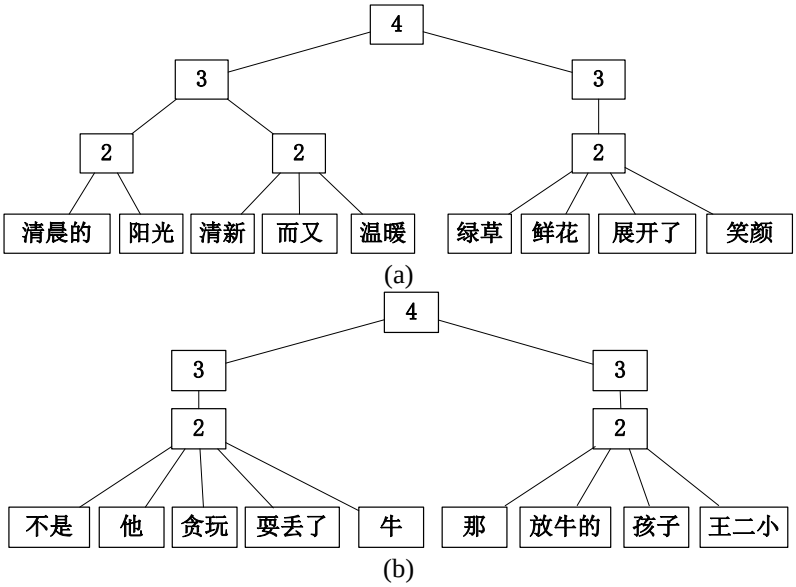


Fig. 7 the prosodic trees for comparision

For prosodic trees given in Fig 7, the corresponding transformed representations are:

(a)：(4(3(2(1清晨的) (1阳光)) (2(1清新) (1而又) (1温暖))) (3(2(1绿草) (1鲜花) (1展开了) (1笑颜))))

(b)：(4(3(2(1不是) (1他) (1贪玩) (1甩丢了) (1牛)))(3(2(1那) (1放牛的) (1孩子) (1王二小))))

Fig. 8 Transformed represation for prosodic trees given in Fig 5

Since this matching is not focused on the exact content but the structure of the given lyric, the content of each node can be replaced with a single symbol Q, thus the representation comes to:

(a)：(4(3(2(1Q) (1Q)) (2(1Q) (1Q) (1Q))) (3(2(1Q) (1Q) (1Q) (1Q))))
(b)：(4(3(2(1Q) (1Q) (1Q) (1Q) (1Q))) (3(2(1Q) (1Q) (1Q) (1Q))).

Fig. 9 Simplized represenation for Fig.6

There are 15 nodes in tree (a) and 14 nodes in tree (b), and the trees contains 13 matchable nodes, which makes G=15，T=14，S=13，P=92.9%，R=86.7%，and F_1=89.7%。

The F1 value is calculated with EVALB program [17]. After filtering out the prosodic trees with different number of prosodic words from the given lyric, the matching is performed on the given lyric and each prosodic tree of the song in database.

Dynamic Programming based Melody-lyric alignment. For each candidate prosodic tree, a character level alignment is then performed on its corresponding rhythmic tree and the given lyric. This alignment is implemented as an instance of the dynamic programming (DP) searching algorithm[18]. The cost of the each pair of input lyric character and the rhythmic element in the tree is given by a matching score, which is trained using the pairs of corresponding characters and rhythmic elements in the songs of the training database.

One input of this alignment is the rhythmic framework consists of note positions labeled with tags defined in table 1, while the other input is the characters of lyric labeled with tags defined in table2. Treating each Chinese character as a hidden state and the position mark of rhythmic as an observation, the alignment can also be considered as a HMM decoding problem[19]. Ignoring the exact content of the exact Chinese characters, this HMM represents how a series of prosodic boundaries are mapped into a series of musical rhythmic units in a song. The prosodic boundaries transition probability and observation probabilities matrix of this HMM are computed from the songs in the rhythmic structure database and are given by table 3 and table 4.

Table 3 the prosodic boundaries transition probability matrix

	B1	B2	B3	B4	M	E	S1	S2	S3	S4
B1	0.1717	0	0	0	0.2715	0.5568	0	0	0	0
B2	0	0.0896	0	0	0.2985	0.6119	0	0	0	0
B3	0	0	0.1204	0	0.3309	0.5487	0	0	0	0
B4	0	0	0	0.2671	0.3020	0.4309	0	0	0	0
M	0	0	0	0	0.3561	0.6439	0	0	0	0
E	0.5169	0.0192	0.2097	0	0	0.1194	0.0970	0.0090	0.0288	0
S1	0.5931	0.0035	0.2448	0	0	0	0.1379	0	0.0207	0
S2	0.8666	0	0	0	0	0	0.0667	0.0667	0	0
S3	0.7286	0	0.0078	0	0	0	0.0155	0	0.2481	0
S4	0.5909	0	0.0455	0	0	0	0.0114	0	0.0227	0.3295

Table 4 observation probabilities matrix

	b1	b2	b3	b4	m	e	s1	s2	s3	s4
B1	0.2554	0.1822	0	0	0.1040	0.2359	0.0918	0.1307	0	0
B2	0.2089	0.2686	0	0	0.0299	0.2090	0.0896	0.1940	0	0
B3	0	0.0517	0.5992	0	0.0614	0.0590	0	0.0181	0.2106	0
B4	0	0	0	0.5064	0.2172	0.0499	0	0	0	0.2265
M	0.1926	0.0398	0	0	0.1447	0.4379	0.1490	0.0360	0	0
E	0.0879	0.0240	0	0	0.0727	0.4010	0.3835	0.0309	0	0
S1	0.1236	0.1063	0	0	0.0431	0.2443	0.4022	0.0805	0	0
S2	0.1333	0.4333	0	0	0.0334	0.1333	0.1000	0.1667	0	0
S3	0	0.0923	0.4307	0	0.1077	0.1231	0	0.0154	0.2308	0
S4	0	0	0	0.4091	0.2386	0.0909	0	0	0	0.2614

When inputing a lyric section which is representationed in the form like Fig. 9, the corresponding HMM is generated using the probabilaty in table 3 and 4.

In our current implementations, in this HMM, the transition probability matrix is an upper triangular matrix; this indicates that a Chinese character can last for one or more music notes, and one note cannot be assigned with more than one Chinese characters. This is because that any character in the lyric has to be assigned with explicit duration, so if more than one successive Chinese characters are in same pitch, they are assigned with notes each by one.

An alignment is then been performed on the generated HMM and input rhythmic framework, with a WSFT based HMM decoder [20]. After the alignment, each state (the lyric character and its position) in the HMM is assigned with one or more observations (rhythmic markers). This completes the whole rhythmic framework generation.

Examples of Generated Rhythmic Framework. We used the above mentioned method to generate the melody rhythmic framework for given lyric. Since there is no objective and straightforward way to measure to show the goodness of the generated framework, we just give some examples for illustrating the method.

For the given lyric sentence "清晨的阳光清新而又温暖，绿草鲜花展开了笑颜. "(The sunshine in the morning is warm and fresh, little grass and flowers are showing their smiling face)", we derived the rhythmic structure can be shown in Fig.10.

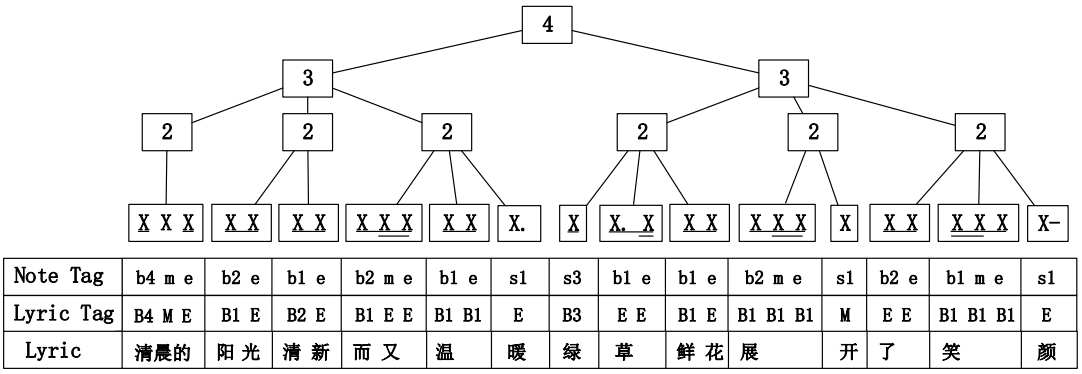


Fig.10 The rhythmic framework for given lyric sentence

Tracing back in the database, this framework is derived from the sentence illustrated in Fig. 3.

Conclusion. In this paper, we presented the problem of automatic lyric oriented rhythmic framework generating (LORFG) in automatic Chinese song composing. We also proposed a melody rhythmic structure generating method for this problem. The method is based on hierarchical tree matching algorithm, and follows a three-stage searching strategy. We also build a nursery rhythm database for model training. The experiment shows that the method can give a reasonable rhythmic framework for given lyric. While current system is still preliminary, further experiments and improvements are going to be made in our future research.

References

[1] Edwards M. "Algorithmic composition: computational thinking in music". Communications of the ACM, vol.54, no.7, pp. 58-67, 2011

[2] Supper M. "A few remarks on algorithmic composition". Computer Music Journal, vol. 25, no.1, pp. 48-53, 2001

[3] C. Ames and M. Domino, "Cybernetic composer: An overview," Understanding Music with AI, pp. 186-205, 1992.

[4] "The Compact Edition of the Oxford English Dictionary **II**". Oxford University Press, p. 2537, 1971

[5] Patel A D, Iversen J R, Rosenberg J C, "Comparing the rhythm and melody of speech and music: The case of British English and French", The Journal of the Acoustical Society of America, p. 3034, 2006

[6] Palmer, C., & Kelly, M. H. "Linguistic prosody and musical meter in song". Journal of Memory & Language, vol. 31, pp. 525–542.

[7] Jinming Zhu, "Interpretation of Linguistics Used in Chinese Lyrics", Master Thesis in Chinese, Tianjin University, 2009

[8] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", in Eighteenth International Conference on Machine Learning, 2001, pp. 282-289.

[9] "Musical Form". Encyclopedia Britannica. Encyclop?dia Britannica Online Academic Edition. Encyclopedia Britannica Inc., 2013

[10] Jianfen Cao, "Prediction of Prosodic Organization based on Grammatical information", Journal of Chinese Information Processing, vol.17,no.3, pp.41-46, 2003

[11] Lerdahl, Fred Autor, and Ray S. Jackendoff. "A generative theory of tonal music". The MIT Press, 1983.

[12] Composer master, http://www.zuoqu.com/html/qybb.htm, Accessed 18 May 2013

[13] Good, Michael. "MusicXML for notation and analysis." The virtual score: representation, retrieval, restoration vol. 12 , pp. 113-124, 2001

[14] D. Cope, "The Algorithmic Composer". WI: A-R Editions, Madison, 2000.

[15] D. Cope, "Computers and musical style". WI: A-R Editions, Madison, 1991.

[16] D. Cope, "Experiments in Musical Intelligence". WI: A-R Editions, Madison, 1996.

[17] M. Emms, "Tree-distance and some other variants of evalb," in International Conference on Language Resources and Evaluation, 2008.

[18] Karlin, Samuel. The structure of dynamic programing models. Naval Research Logistics Quarterly 2.4, pp.285-294, 1955

[19] Juang, Biing - Hwang. "Hidden markov models." Encyclopedia of Telecommunications ,1985.

[20] M. Mohri, etal, "Weighted automata in text and speech processing," in ECAI Workshop, pp. 46-50, 1996