A Japanese Lyrics Writing Support System for Amateur Songwriters

Chihiro Abe* and Akinori Ito*
* Graduate School of Engineering, Tohoku University, Sendai, Japan
E-mail: abechi@spcom.ecei.tohoku.ac.jp, aito@spcom.ecei.tohoku.ac.jp Tel: +81-022-717-7085

***Abstract*—In this paper, we propose a lyrics writing support system focused on the number of syllables, rhyme and word accent. The system generates candidate sentences that satisfy user-specified conditions based on N-gram, and presents them. Users can use the system like a dictionary, and write lyrics be choosing presented sentences. In our subjective evaluations, we have investigated how the system is utilized for writing lyrics actually.The log of using the system and the questionnaires showed that users want the system to present words suitable for their images, and they used the presented words as keywords of a lyrics rather than as they are.**

Index Terms: Songwriting, Lyrics, Mora, Rhyme, N-gram

I. Introduction

Recent boom of writing, playing and sharing music by amateur musicians is exploding worldwide. This phenomena is supported by rapid development of digital technology such as desktop music (DTM) tools and video sharing service like YouTube. User-created music will play more and more important role in the contents in the Internet.

Creating a music clip with song needs several processes, such as composing, writing lyrics, arranging, singing the song, playing instruments, recording and creating the video. Among them, the central parts of songwriting are composition and lyrics writing. Composition has long history and has its established theory and methodology. There also have been a huge number of works on using computer for composition [1] such as composition assistance [2] or automatic composition [3], [4]. However, lyrics writing is less popular compared with composition. There are few universities and colleges that have lyrics writing courses. Researches of lyrics writing and information technology are rare, either. A few exceptions are automatic lyrics generators for Portuguese [5] and Tamil [6]. However, automatic lyrics writing does not seem to be required by the most of amateur musicians because lyrics is main part to convey ideas and messages of the musician to the listeners. Rather generating all of the lyrics automatically, assistance of lyrics writing seems to be more meaningful for most musicians.

In this paper, we propose a system that helps a songwriter to write lyrics by proposing words that matches given rhythm and rhyme.

II. Constraints in Japanese Lyrics

A. Melody-first Approach and Lyrics of Japanese Song

On writing a song, there are two choices: the "lyrics-first" approach and the "melody-first" approach. We assume the melody-first approach that is employed in most popular songs, where a composer first composes melody and a lyrics writer writes lyrics that match the melody. When the melody-first approach is employed, we need to find lyrics with suitable rhythm and rhyme.

The basic unit of Japanese song is a *mora* [4], [7], which is different from a syllable that is a unit for most of other languages. A mora in Japanese is basically a vowel or a consonant followed by a vowel. Because one mora is basically corresponds to one note, we need to find a word or a phrase that has a specified number of morae according to the number of notes in the melody. In addition, as Japanese language has pitch accent system [8], the accent of words need to be consistent with pitch change of the melody, or the words in the lyrics could be misunderstood. Therefore, our songwriting support system aims to present words or phrases that have the following properties:

- Having suitable number of morae to the given melody
- Having rhyme specified by the user
- Having accent consistent with the given melody

To search phrases that match these properties, we use a N-gram language model, trained from large number of text. We predict words that are likely to appear in a given context using the N-gram, and then we choose phrases that match the conditions of morae, rhyme and accent specified by the user. Finally the chosen phrases are presented as candidates of lyrics.

B. Length, Rhyme and Accent

We introduce the following three constraints when searching phrases: mora length, vowel condition and accent.

As mentioned before, the mora length is number of morae in a phrase, which should coincide with the number of notes. For example, a word *konnichiwa* (hello) have five morae (*ko-n-ni-chi-wa*), and thus this word can be used as lyrics for five notes. Note that Japanese word or phrase is written using *kanji* or *kana* characters, each of which has ambiguity of reading. Therefore, we need dictionary-based processing (morphological analysis [9]) for determining the pronunciation of words and number of morae.

The vowel condition is needed to rhyme the lyrics. In this work, we only consider the vowel as a factor of rhyme. Therefore, "*tabeta*" (ate) and "*nonda*" (drank) rhyme because the both words end with /a/.

Japanese accent system is based on pitch. A word or an intonational phrase of standard Japanese have an accent type,

Accent type	Word	Meaning
0	o-mu-re-tsu	an omelet
1	sa-n-ka-ku	a triangle
2	sa-ka-mi-chi	a sloping road
3	na-ki-go-e	a cry

Fig. 1. Japanese accent type

Input	*kyou wa* (today is)
Length	4
Vowel	a * * *
Accent	1

Fig. 2. An example of input conditions

ta i to ru	(title)
ta i i ku	(physical training)
ta i yo u	(sun)
sa i da a	(soda pop)
ja zu fe su	(jazz festival)
ma mi i to	(with mom)

Fig. 3. Example of generated candidates

$C \leftarrow \{h(s_0)\}$
$C_a \leftarrow \emptyset$
while $C \neq \emptyset$ **do**
  for all $s \in C$ **do**
    Remove s from C
    for all $w \in V$
      $S(sw) \leftarrow S(s) + \log P(w|h(s)) + \eta(w)$
    end for
    Let $\hat{w}_1, \ldots, \hat{w}_{|V|}$ be all words in V such that $S(s\hat{w}_1) \geq \ldots \geq S(s\hat{w}_{|V|})$
    for $i \leftarrow 1, \ldots, B$
      if length$(s\hat{w}_i) = n$ **then**
        Add $s\hat{w}_i$ to C_a
      else if length$(s\hat{w}_i) < n$ **then**
        Add $s\hat{w}_i$ to C
      end if
    end for
  end for
end while
Output $s \in C_a$ in descending order of $S(s)$

Fig. 4. The sentence generation algorithm

which is specified by a number. Fig. 1 shows examples of accent types, words and perceived pitch patterns. Word of accent type 0 starts with low pitch, then the pitch raises at the second mora, then the pitch stays high until the end of the word. The type 1 word starts with high pitch, then the pitch falls at the second mora. Words of the other type n ($n > 1$) start with low pitch, then the pitch raises at the second mora, then the pitch falls again after the n-th mora. The falling position is word-specific, and called "the accent nucleus." On singing a word with melody, the most important position is the accent nucleus. If the melody raises just after the accent nucleus, the accent of that word sounds weird, or can be misheard as a different word.

III. The Lyrics Writing Support System

A. Overview of the system

The purpose of this system is to present the user (a songwriter) several candidates of lyrics that can appear after a certain context. The user first specifies several words (we call them *the input sentence*), and then specifies the conditions of the phrase after the input sentence. The condition includes number of morae that should appear, rhyme, and the accent position. The system then generate candidate phrases, and give them scores considering N-gram probabilities and the specified conditions. Finally, the phrases with highest scores are presented as the candidates. Fig. 2 and 3 show examples of the input and output of the system. In Fig. 2, conditions "four-morae phrases," "vowel of the first mora is /a/" and "accent type 1" are specified. From Figure 3, we can see that the generated phrases have four morae and the first vowel is /a/.

B. Sentence Generation

We exploited a 3-gram language model to generate a phrase after the given context. We used the blog articles as a corpus. Let w_i be an i-th word, and c_i be the part of speech of that word. Then we combined probabilities given by word 3-gram and part-of-speech(POS) 3-gram, as follows.

$$P'(w_i|w_{i-2}w_{i-1}) = \lambda P(w_i|w_{i-2}w_{i-1}) + (1-\lambda)P(w_i|c_i)P(c_i|c_{i-2}c_{i-1}) \quad (1)$$

Here, $0 \leq \lambda \leq 1$ is a combination coefficient. According to the preliminary experiment, we chose $\lambda = 0.8$.

Using the 3-gram probability, we calculate the score of a candidate $w_1 \ldots w_n$ as

$$S(w_1 \ldots w_n) = \log P'(w_1 \ldots w_n) + \eta(w_1 \ldots w_n) \quad (2)$$

$$\log P'(w_1 \ldots w_n) = \log P(w_1) + \log P(w_2|w_1) + \sum_{i=3}^{n} \log P(w_i|w_{i-2}w_{i-1}) \quad (3)$$

Here, $\eta(w_1 \ldots w_n)$ is a penalty function that penalizes a candidate that does not satisfy the specified conditions such as rhyme and accent. In the implementation, we set the penalty of the accent as 1 and the penalty of rhyme can be tuned using the user interface.

Using this score, the phrase candidates are generated. The generation is based on the beam search. First, let the mora length to be generated be n and the input sentence be s_0. Let $h(s)$ be a function that gives the last two words of word sequence s. Let V be the vocabulary. Then the candidates are generated by the algorithm as shown in Fig. 4.

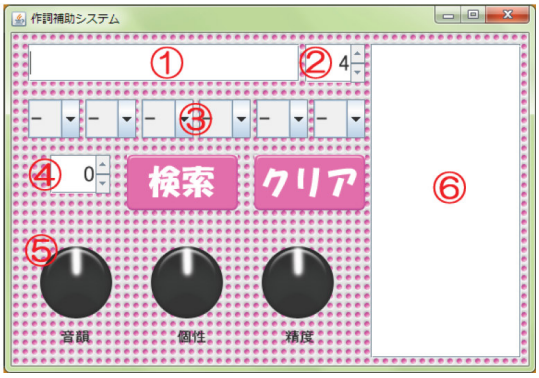

Fig. 5. The user interface

TABLE I
NUMBER OF SEARCH USING THE SYSTEM

	Part. 1	Part. 2	Part. 3	Part. 4
# of system use	48	13	15	47

TABLE II
AN EXAMPLE OF LYRICS WRITTEN BY A PARTICIPANT

	Original lyrics	Lyrics by participant 1
1	*Sakura sakura*	*Sakura sakura*
2	*Yayoi no sora wa*	*Hanayagu komichi wa*
3	*Miwatasu kagiri*	*Sumireni irodori*
4	*Kasumi ka kumo ka*	*Mejiro ka hibari ka*
5	*Nioi zo izuru*	*Shirabe zo chikazuku*
6	*Iza ya iza ya*	*Iza ya Iza ya*
7	*Mi ni yukan*	*Fure ni yukan*

C. The User Interface

Fig. 5 shows the user interface of the implemented system. First, the user types the input sentence in ①, then specify the mora length by ②. Vowel conditions and accent type can be specified by ③ and ④, respectively. After specifying the conditions, the candidate phrases are generated by clicking the [search] button. The parameters such as the penalty value for rhyme, the combination coefficient λ and the beam width B can be tuned using the knobs ⑤. Finally, the generated candidates are displayed at ⑥.

IV. EVALUATION EXPERIMENT

We conducted an evaluation experiment to investigate how the proposed system is actually used for writing lyrics. We asked the participants to accomplish two lyrics writing tasks, and recorded logs of how the system was used. After that, we asked the participants to fill a descriptive questionnaire. Use of the lyrics candidates presented by the system was left to the participants. The participants were four graduate students (three males and one female). Two of them had experience of composition, but none of them had experience of lyric writing.

A. Task I: writing the second verse

In the first task, we asked the participants to write lyrics on the melody of Japanese folk song *Sakura Sakura* [10], as the next verse of the original song. The purpose of this experiment is to investigate the usage of the system when writing lyrics that have similar theme and rhyme to the original lyrics. The procedure of the experiment is as follows:

1) We presented the participant a card of the original lyrics, and the participant listened to the melody of the song.
2) The participant was asked to write lyrics using the system. He/she could listen to the melody of the song when needed.
3) After writing the lyrics, the participant was asked to write the lyrics on the score.
4) Finally, the participant was asked to fill in the questionnaire.

Table I shows how many times each participant used the system. As the table shows, frequency of using the system differed from participant to participant. All of the participants made a behavior to search candidates multiple times using the same input sentence and only changing the condition. This result suggests that the condition settings such as rhyme and accent were not used as a method of finding the most suitable phrases but used as a way of obtaining "different" candidates.

Next, we analyzed the questionnaire to find relationship from the system usage and the parts of the lyrics where the system's results were used. TABLE II shows the example for the participant 1. We traced usage of the system from the questionnaire of the participant 1.

- First, he imagined *hibari* (a skylark) from the word *kumo* (cloud) in the original lyrics (line 4), because characters of *hibari* in Japanese literally means "cloud-sparrow."
- Then he used the system using "*hibari ka*" as an input sentence, and find "*ongaku*" (music) among the candidates. From it, he decided the theme of the lyrics to be "bird's tweet."
- Next, he recalled *mejiro* (a white-eye) that is a white bird singing in Spring, inspired by the word *kasumi* (mist) in the original lyrics (line 4).
- Then he chose a word *shirabe* (tune) denoting the birds' tweet, and used the system using "*shirabe ga*" as an input sentence. From the candidates, he chose "*chikazuku*" (get near, line 5), which ended with the same vowel as the same line of the original lyrics (*izuru* and *chikazuku*).

The analysis of this participant revealed that the system was used not only as a tool that outputs phrases that satisfy certain conditions but also a tool for creativity support or brainstorming. The participant did not necessarily used the presented candidates as is, but the candidates were used as seeds of the imagination for the lyrics.

B. Task II: freestyle lyrics writing

The next task is to write lyrics for a musical piece without lyrics. In this task, we did not instruct the participants a specific target like the previous experiment. We chose "The Trout Quintet" by F. Schubert as a music for writing lyrics. Fig. 6 shows the score of the piece.

Procedure of the experiment is as follows.

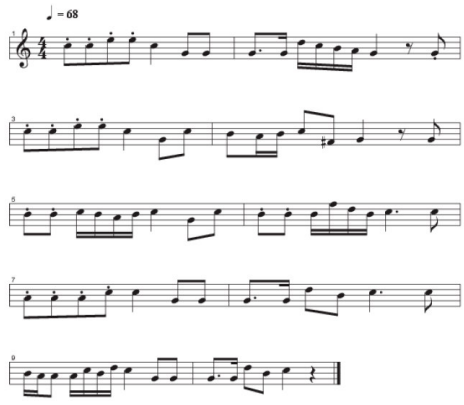

Fig. 6. The score of the piece used in task II (The Trout Quintet)

TABLE III
CHOICES OF THE FIRST PHRASE

Santa	(Santa Claus)
oretachi	(we)
toshiue	(elder)
aitai	(wanna see you)
soretomo	(otherwise)
chikyuugi	(a terrestrial globe)

1) The participant looked at the score (Fig. 6) and listened to the melody.
2) The participant chose the first phrase from the first phrase list shown in Table III.
3) The participant was asked to write lyrics and fill the questionnaire just like the previous experiment.

We analyzed the log of the system to investigate how the system was used. The result is shown in Table IV. Compared with Table I, we can see that number of using the system was fewer than that of the previous experiment, except for the participant 1.

When writing lyrics without explicit constraint such as rhymes in the previous experiment, it seemed to be difficult for the participants to decide when he/she should use the system, which is the reason why the participants did not use the system for lyrics writing. In the questionnaires, we observed an opinion that the output candidates in the task II were unsatisfactory compared with those in task I. The other reason may be that the participants got familiar with writing lyrics compared with task I because task II was conducted after task I.

C. Opinions from the questionnaire

The participants described the following opinions in the questionnaire.

TABLE IV
NUMBER OF SEARCH USING THE SYSTEM

	Part. 1	Part. 2	Part. 3	Part. 4
# of system use	51	5	7	14

- I was inspired from the noun candidates listed by the system
- I used the system to expand my imagination
- This system seems to be useful when theme of the song is clear, such as task I
- It is not good that the presented candidates includes similar words
- This system sometimes makes a mistake in grammar

In addition, the following opinions were gathered as required functions.

- Mood or emotion words as search keys, such as "bright" or "dark"
- Paraphrasing or thesaurus search
- In addition to search phrases after the input sentence, the system should provide a function to search phrases before the input sentence

We are now planning to implement these functions into the system.

V. CONCLUSIONS

In this paper, we created a lyric writing support system for amateur songwriters, and carried out an experiment to investigate how the system was actually used. Requirement to the technical support such as accent and rhyme was different from participant to participant, but it was clear that the participants wanted the system to present candidate phrases that match the mood or image of the song the participant imagined. In addition, we found that the participants used the presented candidate as seeds of inspiration of the lyrics.

In the future work, we explore a method to reflect a user's image of the lyrics to the candidates. In addition, we investigate how to present the candidates so that a user get more inspiration from the candidates.

REFERENCES

[1] R. L. de Mantaras and J. L. Arcos, "AI and Music: From Composition to Expressive Performance", *AI Magazine*, vol. 23, no. 3, pp. 43–57, 2002.

[2] A. Horner and D. E. Goldburg, "Genetic Algorithms and Computer-Assisted Music Composition," In *Proc. Int. Conf. on Genetic Algorithms*, 1991.

[3] J. McCormack, "Grammar based Music Composition," *Complex systems*, vol. 96, pp. 321–336, 1996.

[4] S. Fukayama, K. Nakatsuma, S. Sako, T. Nishimoto and S. Sagayama, "Automatic Song Composition from the Lyrics Exploiting Prosody of Japanese Language," *Proc. Int. Conf. on Sound and Music Computing*, 2010.

[5] H. R. G Oliveira, F. A. Cardoso and F. C. Pereira, "Tra-la-Lyrics: An approach to generate text based on rhythm", *Proc. Int. Joint Workshop on Computational Creativity*, pp. 47–54, 2007.

[6] A. Ramakrishnan A, S. Kuppan and S. L. Devi, "Automatic Generation of Tamil Lyrics for Melodies", *Proc. Workshop on Computational Approaches to Linguistic Creativity*, pp.40-46, 2009.

[7] M. Sadakata, P. Desian, H. Honing, A. D. Patel and J. R. Iversen, "A cross-cultural study of the rhythm in English and Japanese popular music," *Proc. Int. Symp. on Musical Acoustics*, pp. 41–44, 2004.

[8] D. R. Ladd, M. E. Beckmana and J. B. Pierrehumberta, "Intonational structure in Japanese and English," *Phonology*, vol. 3, pp 255–309, 1986.

[9] T. Fuchi and S. Takagi, "Japanese morphological analyzer using word co-occurrence: JTAG," *Proc. Int. Conf. on Computational Linguistics*, pp. 409–413, 1998.

[10] *Sakura Sakura*, Wikipedia. http://en.wikipedia.org/wiki/Sakura_Sakura