Automatic Song Composition from Japanese Lyrics with Singing Voice Synthesizer

Satoru Fukayama\textsuperscript{1}, Kei Nakatsuna\textsuperscript{1}, Shinji Sako\textsuperscript{2}, Takuya Nishimoto\textsuperscript{1}, Nobutaka Ono\textsuperscript{1} and Shigeki Sagayama\textsuperscript{1}

\textsuperscript{1}Graduate School of Information Science and Technology, the University of Tokyo, Japan
\textsuperscript{2}Graduate School of Engineering, Nagoya Institute of Technology, Japan

\{fukayama, nishi, onono, sagayama\}@hil.t.u-tokyo.ac.jp, tsuma@alab.t.u-tokyo.ac.jp, s.sako@nitech.ac.jp

\begin{abstract}
Automatic composition techniques are important in sense of upgrading musical applications for amateur musicians such as composition support systems. In this paper, we present an algorithm that can automatically generate songs from Japanese lyrics. The algorithm is designed by considering composition as an optimal-solution search problem under constraints given by the prosody of the lyrics. To verify the algorithm, we launched Orpheus which composes with the visitor’s lyrics on the website, and 87,000 songs were produced within the 2 years of operation. Evaluation results on the generated songs are also reported, indicating that Orpheus can help users to compose their own original Japanese songs.

\textbf{Index Terms}: prosody, automatic composition, dynamic programming
\end{abstract}

\section{1. Introduction}

By looking around what kind of music are surrounding us, we find song is one of the familiar style in music. However, since composition requires some techniques, not everyone can create their songs with their own lyrics. Therefore, implementing a system that can take over this difficulty may enlarge the possibility of musical activities. For instance, we will be able to compose songs for gifts, or generate songs for a commercial use with the lyrics including the name of the products. With this motivation, we set our goal to invent a method to compose a song automatically from arbitrary input of Japanese lyrics. We also aim to implement a system that can assist users to compose their original songs with their lyrics.

It can be said that generating music automatically itself is a well discussed problem. Automatic composition with computers have been one of the interesting issue to be challenged. The first automatically generated musical piece by computer was the “ILLIAC suite” \cite{1}. Since then, various methods as stochastic process \cite{2}, and approaches based on artificial intelligence \cite{3} were exploited to generate musical pieces. Although considerable amount of research has been done on automatic composition, much less is done on composing melody of songs from the lyrics. One possible reason for this can be the difficulty in reflecting the meaning of the lyrics to the generated melody. Besides considering the semantic analysis of the lyrics in composing songs, there are still information left can be used for composition such as syntactic information and prosodic information. Syntactic information analyzed from the lyrics with NLP techniques were exploited in song composition from Japanese lyrics \cite{4}. However no system that uses prosody has yet been attempted.

Thus, our approach would be developing a method to generate a melody automatically which satisfies the constraints given by the prosody of Japanese lyrics. We define melody composition here as generating a melody given the lyrics, patterns of rhythms (by “rhythm tree” which is described later in the next section), and harmony sequence with specifications of tonality and scale.

\section{2. Melody Composition Algorithm Exploiting Prosody of Japanese Lyrics}

We firstly review the properties of Japanese prosody and argue the importance of considering prosodic information when composing Japanese songs. Secondly, we describe a method to generate rhythm of the melody. Then, we discuss how to obtain a proper pitch given the rhythm, harmony sequence, and the lyrics. Finally, we introduce our automatic song composition system Orpheus, which is based on our proposed algorithm and used in the evaluations.

\subsection{2.1. Japanese Prosody and its Role in Composition}

Japanese is said to “have a fixed shape consisting of a sharp decline around the accented syllable, a decline that is usually analysed as a drop from a H\textsuperscript{1} tone to a L\textsuperscript{2}” \cite{5}. Furthermore, “the place of the accent is lexically contrastive, as in ka’mi ‘god’ vs. kami ‘paper’” \cite{5}. (Fig. 1) A melody attached to the lyrics cause an effect similar to the accent. Therefore we can assume that the prosody of Japanese lyrics imposes constrains on pitch motions of the melody.

\subsection{2.2. Composition of Rhythm}

In order to design a rhythm on the given lyrics, two problems have to be considered. The first problem is the allocation of rhythms (by “rhythm tree” which is described later in the next section), and harmony sequence with specifications of tonality and scale.

\subsection{2.2.1. Allocation of Lyrics on Melody}

In order to solve the first problem, we assumed that melody consists of segments and the lyrics should be divided into these segments. For instance, 2 bars can be handled as a segment for
Figure 1: Pitch contour of ‘ka’mi’ and ‘kami’

a song with a length of 8 bars. This is because the lyrics input by the user are not always “formatted” as poem like the usual

texts are.

Furthermore, in most classical Japanese songs, one syllable
(mora) corresponds to a note in a melody. Thus the number of
notes in each segment is determined by allocating the syllables.
When we consider the constraints on dividing the lyrics, the
following 3 criteria can be assumed:

1. A similar number of syllables in each segments is pref-
ered.
2. The border of the segments should not be crossed over
within a word.
3. Too short lyrics should be iterated prior to allocation.

Under these constraints, we can solve the syllable allocation
problem by using dynamic programming.

2.2.2. Keeping Unity of Rhythm in Melody

Even though the numbers of notes in each of the segments are
decided, there still exist a large degree of freedom in rhythm.
One possible way to put constraints on rhythm is to arrange that
the generated rhythm belongs to a same “family” of rhythm. It is
reasonable to assume that the “rhythm family” does not change
in a relatively short song.

Here the second problem, that is, the requirement to keep
the unity of rhythm in melody arises. To cope with this problem,
we introduce a “rhythm tree” that is one rhythm has a similar
feature when one can be derived by uniting or dividing the note
just one on the another. In practice, a tree structured templates
of rhythm as shown in Fig. 2 are prepared by hand beforehand.
Since the number of notes in each segment corresponds to
the number of notes, this tree structured template determine the
rhythm in each segment.(Fig. 2)

2.3. Composition of Pitches

2.3.1. Composition with Probabilistic Inference

In this section, we discuss on how to obtain a pitch se-
quence. Although there are some discussions on the definition
of melody, still it is likely to say there are certain tendency in
melody. For example, in case of song, pitches of the melody
would be constrained by the usual voice range of the singer. The
prosody of the lyrics also impose constraints on pitch motions
of the melody. As we reviewed at section 2.1, pitch motions of
Japanese songs largely follow the up-ward and down-ward motions based on the prosody of the lyrics. Furthermore, chord
progression, bass line of the accompaniment part and durations
of each notes impose constraints on occurrence and transition of
pitches on the basis of écuretiré de composition, such as har-
mony and counterpoint. Although exploiting these écuretiré

Figure 2: By using the “rhythm tree”(above), rhythm corre-
sponding to the number of syllables are generated with con-
sideration of keeping the unity of rhythm feature in the same
song.

are not always indispensable to discuss how can we generate
a cutting edge contemporary music automatically, still we can
assume that these écuretiré would secure the quality of gener-
ated songs with our algorithms for the purpose of composition
support system for amateur musicians.

If a certain melody were obtained, the melody would sat-
sify these constraints as we discussed above. Conversely, we
can compose a song by finding the melody which optimally
meets the condition. Let the pitch sequence as a sequence of
MIDI note number be $X^N = x_1 x_2 \cdots x_N$, and the sequence
of conditions on pitch sequence be $Y^N = y_1 y_2 \cdots y_N$, where
each $y_n$ involves chord label with annotations of scale and
tonality($r_n$), duration of the note($d_n$), MIDI note number of
the accompaniment bass ($b_n$), and pitch accent information, i.e.

$y_n = (c_n, d_n, b_n, a_n)$. Let us also denote $P(X^N | Y^N)$ as con-
ditional probability for $X^N$ given $Y^N$ which represent the ten-
dency of pitch sequences $X^N$ under condition $Y^N$. The com-
position of pitch for melody can be considered as finding an opt-
imal sequence $X^*_N$ given $Y^N$ which maximize $P(X^N | Y^N)$:

$$X^*_N = \arg \max_{X^N} P(X^N | Y^N).$$  

By assuming

$$P \left( x_n | x_{n-1}, Y^N \right) \approx P \left( x_n | x_{n-1}, Y^1 \right),$$

equation (1) will be as follows:

$$X^*_N = \arg \max_{X^N} \prod_{n=1}^N P \left( x_n | x_{n-1}, Y^1 \right),$$

where $P \left( x_1 | x_0, Y^1 \right) = P \left( x_1 | Y^1 \right)$. Since there are $128^N$ possible sequence of pitches, it is computationally impossible to search the optimal sequence by calculating probabilities for all of the possible sequences. However, we can obtain the optimal pitch sequence in order $O(N)$ using dynamic programming [7].

2.4. Implementation of the Composition System

Orpheus is an automatic composition system that we imple-
mented using proposed algorithm for melody composition. This
system computes melody from the lyrics input with choices of
chord progressions, rhythm pattern, and accompaniment instruments. Flow chart of the processes is shown in Fig. 3. We used Galatea-Talk [8] text-to-speech engine to analyze the prosody of Japanese lyrics, and HMM singing voice synthesizer [6] to generate the vocal part. We also implemented the system as a web-based application.3

3. Evaluation results

We did two experiments to evaluate the algorithm. Firstly, we asked a classical music composer to evaluate generated songs in five-grade evaluation. The results on 59 generated songs are shown in Fig. 5. These results indicate that 83.1% of the generated pieces satisfactorily follow classical music theory, and 91.6% of the songs were voted as attractive aside from musical theory. Example of generated song is shown in Fig. 4.

Secondly, we uploaded our system to get comments from a large number of users on the internet. During the 2 years of operation, about 87,000 songs were generated by the users and 1378 people answered the questions about Orpheus and the generated songs. Summarization of answers in five-grade evaluation is shown in Fig. 6. Judging from the results, about 70.8% commented that the generated songs are attractive, and 84.9% of the users had fun trying this system.

4. Discussions

The evaluations results by a composer indicate that most of the generated songs are able to be called “melody” at least in theory. Songs that are evaluated “very poor” had irregular usages of non-chord tones which rule cannot be described with the relationship between the current note and the previous note.

Evaluation by the users on the internet suggest that our composition system is an enjoyable solution for amateur to compose their original songs. One reason for the result could be the directability on generating songs. Users can generate various melody by typing arbitrary lyrics since the generated song will vary based on the prosody of the lyrics. This may enabled the user not only generating a song automatically but also

3http://orpheus.hil.t.u-tokyo.ac.jp

Figure 3: Flow chart of processes: Orpheus generates songs with the lyrics input and the choices of patterns.

Figure 5: Evaluation results on 59 generated songs by a classical music composer. 83.1% of the generated pieces satisfactorily follow classical music theory, and 91.6% of the songs were voted as attractive aside from musical theory.

Figure 6: Evaluation results on generated songs and the Orpheus by 1378 users. 70.8% commented that the generated songs are attractive, and 84.9% of the users had fun trying this system.
Figure 4: Example of generated song: this song was generated with the lyrics input of weather forecast. The red lines indicates the pitch accents of the Japanese lyrics. The pitch motions of the melody follows the pitch accent of the lyrics.

to generate their original songs with their original lyrics.

5. Conclusion

This research attempted to design an algorithm to compose a song automatically from the lyrics using prosody information, which enables users to make their original songs easily. The results indicate that our method and implemented system Orpheus is an enjoyable solution for amateur musicians.

However, it should be noted that our algorithm can be applied to lyrics written in “pitch accent” languages only. As a next step, we plan to extend the composition algorithm to handle “stress accent” languages, such as English, by putting constraints on metric structure of the melody.

6. Acknowledgements

We would like to thank Hitomi Kaneko for the discussions and the evaluation of our generated songs. This research was funded by the Crest Muse Project.

7. References