Automatic Song Composition from Japanese Lyrics with Singing Voice Synthesizer

Satoru Fukayama, Kei Nakatsuma, Shinji Sako,
Takuya Nishimoto, Nobutaka Ono and Shigeki Sagayama

1 Background

Support systems for song composition are important in sense of upgrading musical abilities of amateur musicians. If we suppose a user who cannot compose nor follow the score, the system should consist of automatic composition module to generate a song and singing voice synthesizer to send feedbacks of the result to the users. One possible architecture of the the system which users can easily try is a compostion system from the lyrics. Although considerable research has been done on automatic composition[1][2][3], much less is done on composing songs from the lyrics, and a question remains that what information in the lyrics should be exploited for generating songs. Guido d'Arezzo invented a method to compose a melody by choosing the notes whichcorrespond to the vowels in the lyrics[4]. Hayakawa used syntactic information of Japanese lyrics[5].

2 Aims

Our objective is to design a composition support system that generate a song automatically from Japanese lyrics and send feedbacks of the results with singing voice synthesizer. We define melody composition here as generating a melody given the lyrics, patterns of rhythms, and harmony sequence with specifications of tonality and scale.

3 Method

In this section, we firstly discuss how prosody of the lyrics can be exploited when compositing Japanese songs. Secondly, we argue that composition of pitch sequence can be formalized as a search problem of optimal combination of pitches which maximize the probability of the melody. Finally, we show how we can implement the composition support system with singing voice synthesizer.

3.1 Japanese Prosody and Melody 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[1] tone to a L[2]"[6]. Furthermore, "the place of the accent is lexically contrastive, as in ka'mi 'god' vs. kami' 'paper"'[6]. A melody attached to the lyrics cause an effect similar to the accent. Therefore we can assume that the prosody of Japanese lyrics imposes constraits on pitch motions of the melody.

3.2 Composition Algorithm Considering the Prosody of the Lyrics

We can say that there are certain tendency in melodies. 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. 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 occurence and transition of pitches on the basis of écuriture of composition, such as harmony and counterpoint. Althogh exploiting these écuritures are not always indispensable to discuss how can we generate a cutting edge contemporary music automatically, still we can assume that these écuritures would secure the quality of generated songs with our algorithm for the purpose of composition support system for amateur musicians.

If a certain melody were obtained, the melody would satisfy these constraits as we disucussed 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 $X_1^N = x_1 x_2 \cdots x_N$, and the sequence of conditions on pitch sequence be $Y_1^N = y_1 y_2 \cdots y_N$, where each y_n involves chord label with annotations of scale and tonality(c_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 $P\left(X_1^N | Y_1^N\right)$ as conditional probability for X_1^N given Y_1^N which represent the tendency of pitch sequences X_1^N under condition Y_1^N. The composition of pitch for melody can be considered as finding an optimal sequence X_1^{N*} given Y_1^N which maximize $P\left(X_1^N | Y_1^N\right)$:

$$X_1^{N*} = \underset{X_1^N}{\operatorname{argmax}} P\left(X_1^N | C_1^N\right). \tag{1}$$

By assuming

$$P\left(x_n | X_1^{n-1}, Y_1^N\right) \simeq P\left(x_n | x_{n-1}, Y_1^N\right), \tag{2}$$

[1]H: high
[2]L: low

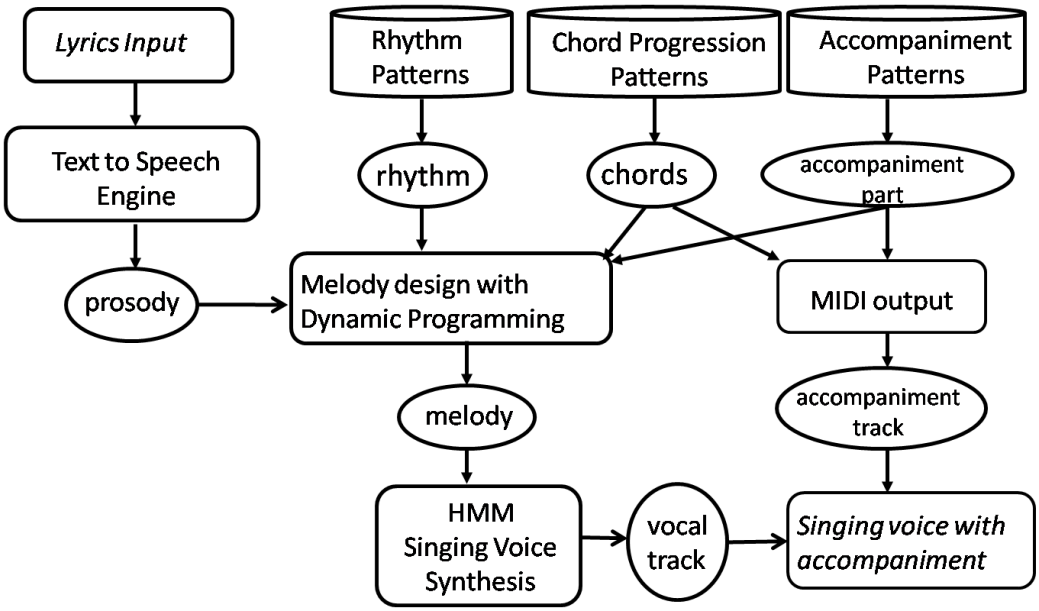

Figure 1: System generates songs and the singing voices with the lyrics input and the choices of patterns.

equation (1) will be as follows:

$$X_1^{N*} = \underset{X_1^N}{\operatorname{argmax}} \prod_{n=1}^{N} P\left(x_n | x_{n-1}, Y_1^N\right), \tag{3}$$

where $P\left(x_1|x_0, Y_1^N\right) = P\left(x_1|Y_1^N\right)$. Since there are 128^N possible sequence of pitches, it is computationally unfeasible to search the optimal sequence by calculating probabilities for all of the possible sequences. However, we can obtain the optimal pitch sequence in order $O(N)$ by using dynamic programming[7].

3.3 *Orpheus*: Implementation with Singing Voice Synthesizer

Orpheus is an automatic composition system that we implemented using melody composition algorithm based on prosody. This system computes melody from the lyrics input with choices of chord progressions, rhythm patterns, and accompaniment instruments. We used Galatea-Talk[8] text-to-speech engine to analyze the prosody of Japanese lyrics, and HMM singing voice synthesizer[9] to generate the vocal part. We also implemented the system as a web-based application.

4 Evaluation Results

We did two experiments to evaluate the system. Firstly, we asked a classical music composer to evaluate generated songs in five-grade evaluation. The 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. Secondly, we uploaded our system to get comments from a large number of users on the internet. During a year of operation, 56,000 songs were generated by the users and 1378 people answered the questions about *Orpheus* and the generated songs. 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.

5 Conclusion

This research attempted to design a system which generates a song automatically from the lyrics using prosody information and sends a feedback to the users with singing voice synthesizer, which enables users to make their original songs easily. The results indicate that our method and implemented system was an enjoyable solution for amateur musicians. We plan to extend the composition algorithm to handle "stress accent" languages, such as English, by putting constraints on metric structure of the melody.

References

[1] L. Hiller and L. Isaacson.: Experimental Music. McGraw-Hill, New York (1959)

[2] I. Xenakis: Formalized Music. Revised edition. Pendragon Press, New York (1992)

[3] D. Cope: Computers and Musical Style. A-R Editions. Madison, Wisconsin (1991)

[4] D. G. Loy: Composing with computers — a survey of some compositional formalisms and music programming languages. Current Directions in Computer Music Research. pp. 291–396. The MIT Press, Cambridge, Massachusetts (1989)

[5] K. Hayakawa et. al.: Automatic song composer from phrase structure of lyrics., 57th IPSJ Annual Convention, pp.11–12. (1998) (in Japanese)

[6] M. E. Beckman and J. B. Pierrehumbert: Intonational structure in Japanese and English, Phonology Yearbook 3, pp. 255–309. (1986)

[7] R. E. Bellman: Dynamic Programming. Princeton University Press, Princeton, New Jersey (1957)

[8] S. Kawamoto et al.: Galatea: Open-Source Software for Developing Anthropomorphic Spoken Dialog Agents, Life-Like Characters, pp. 187–212, Springer-Verlag. (2004)

[9] S. Sako et. al.: A Singing Voice Synthesis System Based on Hidden Markov Model., Transactions of IPSJ, pp.719–727. (2004) (in Japanese)

Keywords(up to 5) Automatic Composition, Prosody, Dynamic Programming

Topic areas Production, Synthesis, Assistance