Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Music/Lyrics Composition System Considering User's Image and Music Genre

Chisa Nakamura
Onisawa Lab., Graduate School of Systems and Information Engineering, University of Tsukuba
Tsukuba, Japan
nakamura@fhuman.esys.tsukuba.ac.jp

Takehisa Onisawa
Graduate School of Systems and Information Engineering, University of Tsukuba
Tsukuba, Japan
onisawa@iit.tsukuba.ac.jp

***Abstract*— This paper proposes a music/lyrics composition system consisting of two sections, a lyric composing section and a music composing section, which considers user's image of a song and music genre. First of all, a user has an image of music/lyrics to compose. The lyric composing section presents initial lyrics chosen at random from database that is constructed using Markov Chain and existent lyrics classified by music genres. If presented lyrics do not fit user's image, a part of lyrics not fitting user's image is changed by some other words. When satisfied lyrics are obtained, the music composition section starts. This section composes music fitting lyrics generated by the lyric composing section using the music composition system. And this section generates melody and backing patterns according to music genres. The section presents combinations of lyrics and 16 measures music. A subject evaluates each combination of lyrics and music whether it fits his/her image of a song. According to subject's evaluation music melody is changed by Genetic Algorithms and a part of lyrics are changed. These procedures are repeated until satisfied combination of lyrics and music melody is generated. In order to verify the validity of the presented system, subject experiments are performed.**

***Keywords*—music, lyrics, genre, genetic algorithm, markov chain**

I. INTRODUCTION

A lot of musical works in many genres such as classical music, popular music, modern music are heard everywhere at the present day, and music is indispensable to our daily lives. Jean-Jacques Rousseau says *There is a sound, the soul grasps for it, and there it has a ringing word* in Essay on the Origin of Language [1]. That is, music and words are inseparably related to each other. Music has a wealth of expression in itself. However, a song, i.e., music with words has more wealth of expression. Therefore, a song sometimes stirs one's heart harder than only words or only music.

There are many studies on automatic music composition and automatic lyrics composition in order to realize intelligent music information processing of human. For instance, *LYRICA* [2] composes lyrics automatically to existing melodies. *Tra-ra-Lyrics* [3] aims to create a computer program capable of generating lyrics to given melodies, and *VOCALOID* [4] is known as a singing software. There are also some other studies on music composition [5, 6].

As for music/lyrics composition, the following two procedures are usually considered. The one is to compose lyrics first and after that melody composition is followed. The other is to compose music first and lyrics composition is followed. In this way music and lyrics are usually composed separately because different technical knowledge is needed for music composition and lyrics one. Although composer's impressions/image should be usually reflected to composed songs, it is difficult to reflect composer's impressions/image to music/lyrics if they are composed separately by other composers. By the way, some artistes compose music/lyrics by themselves. They usually compose lyrics/music checking composed lyrics and/or music repeatedly so that their individual impressions/image can be reflected to songs well. It is necessary to consider user's impressions/image for the construction of a music/lyrics composition system. Moreover, there are various genres in lyrics, e.g., lyrics of *popular music*, lyrics of *nursery rhyme*, lyrics of *Japanese Enka* (peculiar to Japanese songs), and also various genres in music, e.g., *popular music*, *ballade*, *nursery rhyme*, *Japanese Enka*. Therefore, it is also necessary to consider genres for the construction of a lyrics/music composition system. However, few existing systems consider user's impressions/image and genres.

This paper aims to develop a lyrics/music composition system considering user's impressions/image and music genres. Furthermore, even if a user has little knowledge of lyrics and/or music composition, he/she can compose lyrics/music using the presented system. This paper has following organization. Chapter 2 shows the structure of a music/lyrics composition system. Chapter 3 describes experiments using the presented system and shows experimental results with discussion. The final chapter concludes this study.

II. STRUCTURE OF MUSIC/LYRICS COMPOSITION SYSTEM

A. Outline of Music/Lyrics Composition System

Fig. 1 shows the outline of a music/lyrics composition system consisting of two sections, a lyric composing section and a music composing section. Sixteen measures songs are generated by this system, where a musical work including music and lyrics is called a song in this paper.

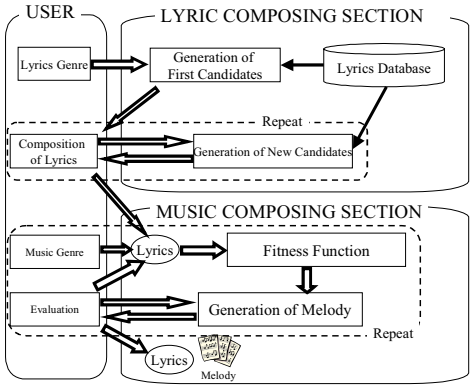

Figure 1. Outline of music/lyric composition system

In the lyric composing section, at the first step, a user inputs a lyrics genre, i.e., *popular music lyrics*, *nursery rhyme lyrics*, or *Japanese Enka lyrics*. Then, the section presents the first candidates of lyrics chosen from lyrics database at random, where lyrics database are constructed using Markov Chain [7] and have lyrics of three genres as mentioned above. If presented lyrics don't fit user's image, a part of lyrics not fitting user's image is changed by some other words, where presented lyrics as the next candidates are chosen according to probabilities in Markov Chain. After repeating the presentation of lyrics and evaluation, if satisfied lyrics are obtained, the music composing section starts.

In the music composing section, at the first step, a user inputs a music genre, i.e., *popular music*, *ballade, nursery rhyme*, or *Japanese Enka*. Based on generated lyrics and the inputted genre, the music composing section using the music composing system [8, 9] generates melodies using Interactive Genetic Algorithms [10, 11], where the accompaniment part is generated according to an inputted music genre. In addition to the fitness function evaluating the difference between the number of words in lyrics composed by the lyrics composing section and the number of notes in melodies composed by the music composing section, user's subjective evaluations are considered in the music composing section. A user evaluates each combination of lyrics and music whether it fits user's image/impressions of a song.

B. Lyric Composing Section

1) Structure of lyric composing section: Fig. 2 shows the structure of the lyric composing section. The section chooses the first candidates of lyrics from lyrics database at random according to the lyrics genre inputted by a user. If presented lyrics don't fit user's image, a part of lyrics not fitting user's image is modified by some other words, where presented words following modified words as candidates are chosen according to probabilities in Markov Chain as mentioned after. A user can also modify lyrics by himself/herself freely.

2) Database construction: In this paper three kinds of lyrics genres are considered, *popular music rylics*, *nursery rhyme lyrics*, and *Japanese Enka lyrics*. Many lyrics are collected using lyrics web search sites [12, 13]; 834 songs of *popular music*, 250 songs of *nursery rhyme*, and 228 songs of *Japanese Enka*. Database of each lyrics genre are constructed using Markov Chain [7] as follows. Collected lyrics are divided into morphemes using Morphological Analysis [14]. Then, a word set of the combination of morpheme 1(written as word 1) and morpheme 2 (written as word 2) is considered, and the number of the sets is counted for all collected songs in each genre. For example, let the combination of word 1 and word 2, and the number of their combinations be obtained as shown in Table 1 as the results of searching web sites. Then, probabilities of the combination of *your face*, *your look*, *your feeling* or *your song* are k_1/k , k_2/k , k_3/k , and k_4/k , respectively, as shown in Table 1, where $k = k_1 + k_2 + k_3 + k_4$.

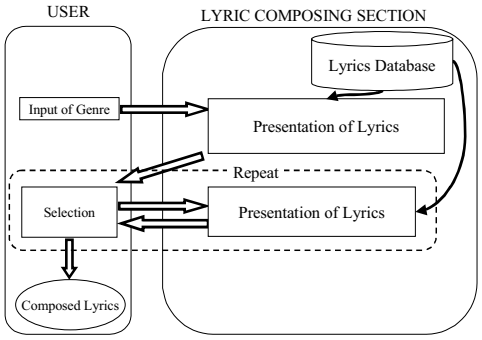

Figure 2. Structure of lyric composing section

TABLE 1. COMBINATION OF WORDS AND PROBABILITIES

Set of Words		The Number of Sets	Probability
Word 1	*Word 2*		
your	Face	k_1	k_1/k
your	Look	k_2	k_2/k
your	feeling	k_3	k_3/k
your	Song	k_4	k_4/k

3) Lyrics composing using databse: The lyric composing section uses words sets shown by expression (1). That is, the contents of Table 1 are expressed by expression (1).

$$\begin{pmatrix} a_1{:}X_1 \\ a_2{:}X_2 \\ \vdots \\ a_n{:}X_n \end{pmatrix} = \begin{pmatrix} a_1{:} & b_{11}(P_{11}) & b_{12}(P_{12}) & \cdots & b_{1m}(P_{1m}) \\ a_2{:} & b_{21}(P_{21}) & b_{22}(P_{22}) & \cdots & b_{1m}(P_{2m}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_n{:} & b_{n1}(P_{n1}) & b_{n2}(P_{n2}) & \cdots & b_{nm}(P_{nm}) \end{pmatrix}, \quad (1)$$

where $X_i (i = 1,2,...,n)$ consist of elements b_{ij} and probabilities $P_{ij} (i = 1,2,...,n, j = 1,2,...,m)$. Elements a_i are morphemes obtained by Morphological Analysis of existent lyrics, elements b_{ij} are morphemes following a_i , P_{ij} are probabilities of elements b_{ij} following elements a_i , $\sum_{j=1}^{m} P_{ij} = 1$, n is the total number of morphemes, and m is the maximum number of b_{ij} following a_i . If $n < m$, then some b_{ij} has blank, i.e., no element. If b_{ij}

has no element for some i and j, its probability is assumed to be 0. Examples of words sets generated by Markov Chain are shown in Table 2. Using this table, the following sentences are generated. At the first step, an element a_i is chosen at random as the initial element. Here, let a_1 be chosen as shown in Table 2 and be *there*. Element b_{ij} following a_1 is *are* or *is* as shown in Table 2. Therefore, the next element is chosen from b_{ij}, i.e., *are* or *is* in this example, according to each probability and user's evaluation. Here, *are* is assumed to be chosen. Next, a_i is chosen, of which element is *are*. Here, let a_2 be *are*. Elements b_{ij} following a_2 are *white*, *gray*, or *blue* as shown in Table 2. The next element is chosen from b_{ij}, i.e., form *white*, *gray*, or *blue* in this example, according to each probability and user's evaluation. Repeating these procedures, an example of generated lyrics is: *There are white clouds in the sky.*

TABLE 2. EXAMPLE OF WORDS SETS GENERATED BY MARKOV CHAIN

a_i		X_i	b_{ij}	P_{ij}
a_1	There	X_1	are	0.7
			is	0.3
a_2	are	X_2	white	0.5
			gray	0.25
			blue	0.25
a_3	white	X_3	clouds	0.5
			snow	0.25
			dog	0.25
a_4	clouds	X_4	in	0.75
			are	0.25
a_5	in	X_5	the	0.5
			that	0.5
a_6	the	X_6	sky	0.4
			book	0.3
			cat	0.3
a_7	sky	X_7	.	0.5
			is	0.5

The number of collected words is 6,000 through 10,000 depending on lyrics genres and the number of words following word 1 is 1 through 20 also depending on lyrics genre. If a user chooses another word at user's option rather than the word chosen by Markov Chain, lyrics composing procedures are continued based on the chosen word.

C. Music Composing Section

1) Structure of music composing section: Fig. 3 shows the structure of the music composing section that uses the music composition system [8, 9] applying Interactive Genetic Algorithms [10, 11]. Human evaluation and the fitness function are used in Interactive Genetic Algorithms. The music composing section presents combinations of lyrics and 16 measures music. In this paper a melody with four measures is called one phrase.

2) Musicg genre, backing pattern and chord progression: In this paper, 4 kinds of music genres, *popular music*, *ballade*, *nursery rhyme* and *Japanese Enka*, are considered. Musical works have the following part structures; one melody part and three or four backing patterns. Table 3 shows kinds of musical instruments played in backing patterns in each genre and tempos expressing how many quarter notes are played in a minute. An example of backing patterns in *popular music* is shown in Fig. 4. Since musical works in these 4 genres have usually quadruple time, this section aims at the composition of quadruple time musical works.

The chord progressions are used in order to have a wealth of expression and to harmonize a melody part with backing patterns. Chord progressions are considered according to music genres.

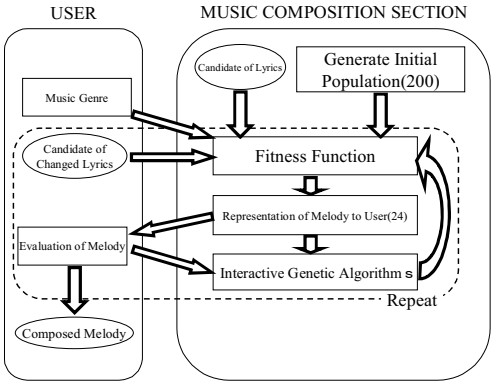

Figure 3. Structure of music composing section

TABLE 3. KINDS OF MUSICAL INSTRUMENTS PLAYED IN BACKING PATTERN AND TEMPOS

Genres	Instruments	Tempos
Popular Music	Piano, Bass, Drums, Wind Instrument	140
Ballade	Bass, Guitar, Wind Instrument	100
Nursery Rhyme	Bass, Agogo, Wind Instrument	120
Japanese Enka	Piano, Drums, Violin	76

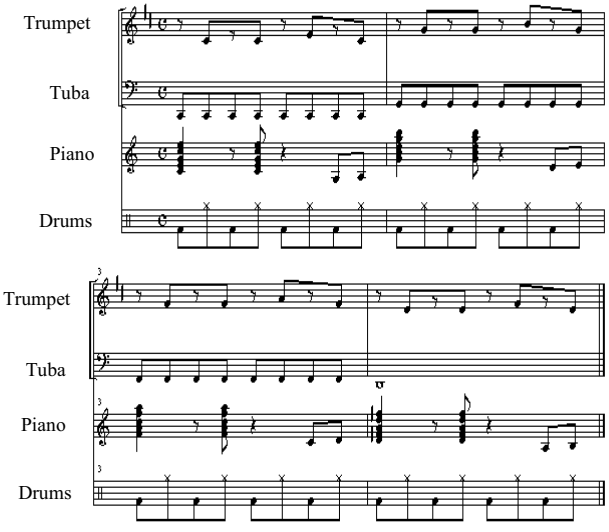

Figure.4. Example of backing patterns (Popular music)

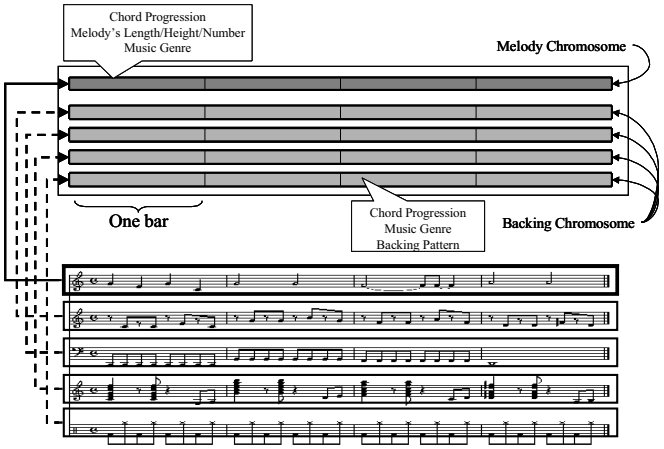

Figure 5. Relation between chromosome and bar

3) Correspondence between Music Features and Chromosome: In this section, one phrase is expressed by one individual, i.e., one chromosome. Fig. 5 shows correspondence between music features and melody/backing chromosomes. The melody chromosome has music features such as chord progression, melody length/height, the number of notes, and a music genre. On the other hand the backing chromosome has music features such as chord progression, a music genre and backing patterns. Because a melody plays an important role to reflect user's image/impressions, GA operations such as crossover and mutation are performed to only information on melody length/height and the number of notes in the melody chromosome. Furthermore, in order to harmonize a melody and backing patterns, information on the chord progressions used in the melody chromosome is copied to the one used in the backing chromosome.

4) Melody and Backing Pattern Composing: The music composing section composes music by the following procedures. A user chooses a music genre from 4 kinds of genres. Therefore, inputs to this section are a chosen music genre and composed lyrics. This section generates initial 200 melody individuals and backing individuals (4 measures music individuals), where backing ones are dependent on a chosen music genre. In order to fit composed melodies to lyrics the difference between the number of notes and that of words is evaluated by fitness function (2). The fitness value becomes low as the difference becomes large.

$$Fitness1 = \begin{cases} 100(M = F) \\ \dfrac{0.01}{M - F}(M > F) \\ -100(M < F) \end{cases} \qquad (2)$$

$$M : The \quad Number \quad of \quad Notes \quad of \quad Melodies$$
$$F : The \quad Number \quad of \quad Lyrics$$

Next, this section presents combinations of lyrics and 4 measures music with top 6 fitness values to a user. A user evaluates each combination of lyrics and music whether it fits

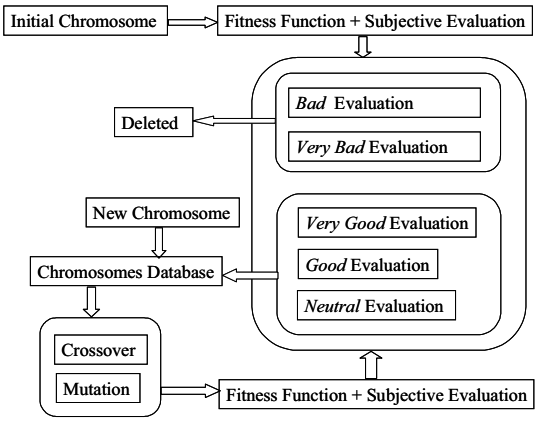

Figure 6. Procedures of GA operations of melody chromosome

user's image/impressions of a song with 5-points scale as follows: 5-*very good*, 4-*good*, 3-*neutral*, 2-*bad*, 1-*very bad*. According to user's evaluations, music melody individuals are evolved by GA operations, crossover and mutation, in the procedures shown in Fig. 6. Individuals with *very good*, *good*, or *neutral* evaluation are saved in the individual population. On the other hand, individuals with *bad* or *very bad* evaluation are deleted. New individuals are generated by copying, where the number of copied individuals is determined according to evaluation, *very good*, *good*, or *neutral*. If the number of copied individuals is less than 200, the shortage is supplemented copying individuals in the individual population at the previous generation. When a user evaluates a melody individual as *very good*, this individual is saved as an elite one in the current generation.

III. Experiments and Remarks of Results

A. Outline

In order to verify the presented system 4 kinds of subject experiments are performed. The first experiment (Experiment A) is performed to verify the validity of lyrics database based on lyrics genres. The second one (Experiment B) is for the verification of validity of backing patterns based on music genres. The third one (Experiment C) is for subjective evaluation, that is, user's satisfaction degree of composed music/lyrics. The last one (Experiment D) is for objective evaluation, that is, how someone else different from composer himself/herself evaluates composed music/lyrics.

In Experiment A, subjects are 17 boys/girls and all are twenties. They evaluate which genre 8 lyrics composed by the presented system come under, i.e., *popular music lyrics*, *nursery rhyme lyrics*, or *Japanese Enka lyrics*, where presented lyrics are composed using database of each lyrics genre.

In Experiment B, 17 boys/girls, the same subjects as the ones in Experiment A, evaluate which genre of backing pattern the presented backing pattern comes under, where presented backing patterns are composed using backing patters in each music genre.

In experiment C, subjects are 22 boys/girls and are all twenties. One of 22 subjects has experience in music/lyrics

composition. Three subjects have experience in music composition. Eighteen subjects have experience in neither music composition nor lyrics one. Eleven subjects out of 22 choose *popular music lyrics*. Six out of them choose the *popular music genre* and 5 of them choose the *ballade genre*. Six subjects out of 22 choose the *nursery rhyme genre*, and 5 choose *Japanese Enka genre* as lyrics and music genres. As for music composition, each subject repeats music composition ten times because according to [10], the desirable number of repetition is from 10 to 20 for user's tiredness in Interactive Genetic Algorithms. After composing music/lyrics, each subject answers the following questionnaires which are about user's satisfaction degree for composed music/lyrics.

Questionnaire:

a) Do you feel that the system supports composition of music/lyrics?

b) Do you feel that composed lyrics fit your image?

c) Do you feel that composed song fits your image?

Each subject answers questionnaires with 5-points scale as follows.

5: I feel so. 4: I feel so a little. 3: I don't know. 2: I don't much feel so. 1: I don't feel so.

After this experiment, each subject evaluates the ranking of his/her composed songs at the first, the fourth, the seventh and the tenth generations, where the songs are presented to each subject at random. The composed song with the first rank, with the second rank, with the third rank and with the last rank are given 4, 3, 2 and 1 points, respectively.

In Experiment D, subjects are 17 boys/girls and all are twenties. Each subject evaluates lyrics and songs composed by other subjects using the presented system. Each subject answers the following questionnaires.

Questionnaire:

a) Do you think that the lyrics are good?

b) Do you think that the song is good?

Each subject answers questionnaires with 5-points scale as follows.

5: I think so. 4: I think so a little. 3: I don't know. 2: I don't think so much. 1: I don't think so.

B. Experimental Results and Discussion

1) Results of Experiment A: Results of Experiment A are shown in Fig. 6. It is found that lyrics in the *popular music genre*, the ones in the *nursery rhyme genre* and the ones in the *Japanese Enka genre* all composed by the presented system fit user's image of lyrics of *popular music* at 88.7%, that of *nursery rhyme* at 93.5% and that of *Japanese Enka* at 73.2%, respectively, and that these rates are the highest in each genre. From these results it can be said that lyrics composed using lyrics database can reflect user's image well.

2) Results of Experiment B: Results of Experiment B are shown in Fig. 7. It is found that backing patterns in the *popular music genre*, in the *ballade genre*, in the *nursery rhyme genre*, and in the *Japanese Enka genre* all generated by the presented system fit user's image of *popular music* at 64.0%, that of *ballade* at 65.0%, that of *nursery rhyme* at 85.0%, and that of *Japanese Enka* at 89.0%, respectively, and that these rates are the highest in each genre.

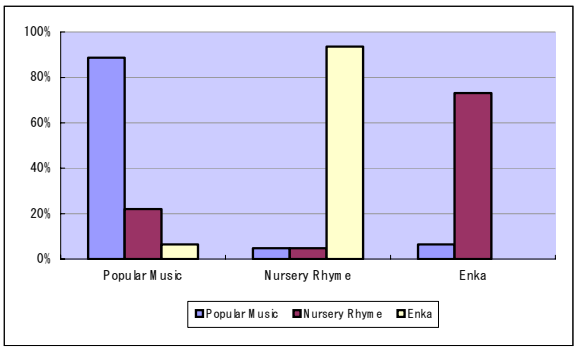

Figure 6. Results of Experiment A

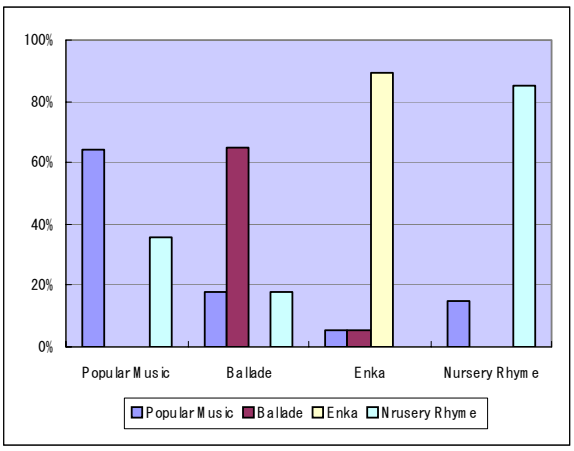

Figure 7. Results of Experiment B

From these results it can be said that backing patterns in the *nursery rhyme genre* and in the *Japanese Enka genre* can reflect user's image well. However, the rate in the *popular music genre* and that in the *ballade genre* are lower than that in *nursery rhyme genre* and that in the *Japanese Enka genre*. The followings are considered as probable cause of these results: The sound source is made by MIDI file and MIDI file is apt to give mechanical or childish impressions to listeners. As a matter of fact some subjects have comments that all backing patterns sound like nursery rhyme because MIDI file is used. The following is considered in order to deal with this problem: Each subject chooses backing patterns fitting user's image by himself/herself.

TABLE 4. POPULATION MEAN AT A 95% CONFIDENCE OF SATISFACTION DEGREE (EXPERIMENT C)

	Question a)	**Question b)**	**Question c)**
Lower Limit	4.3	4.0	3.9
Average	4.5	4.3	4.2
Upper Limit	4.7	4.5	4.5

TABLE 5. POPULATION MEAN AT A 95% CONFIDENCE (EACH GENERATION)

	First	Fourth	Seventh	Tenth
Lower Limit	1.8	1.8	1.9	2.7
Average	2.1	2.3	2.4	3.2
Upper Limit	2.5	3.0	3.0	3.7

3) Results of Experiment C: Subjects take 10 through 30 minutes for the lyrics composition. Results of Experiment C are shown in Table 4 and Table 5. Table 4 shows population mean at a 95% confidence of user's satisfaction degree in Experiment C. It is found that lower limits on questionnaires a), b) and c) are larger than 3.9. These results show that subjects feel that the presented system supports composition of music/lyrics and that composed songs fit their image well.

Table 5 shows population mean at a 95% confidence of user's satisfaction at the first, the fourth, the seventh and the tenth generations. The average seems to become high as generations progress. By Wilcoxon signed-rank test between the seventh and the tenth generations, statistics z is 1.68, and probability p is 0.046 at one-sided test. If the level of significance is 0.05, then $p < 0.05$. Then, it can be said that there is a difference between subjects' evaluation of the seventh songs and that of the tenth songs, and that the presented system can compose songs fitting subjects' images more as generations progress.

From these results of Experiment C, it is found that the presented system can compose music/lyrics fitting user's image and that the system can support inexperienced users to compose music/lyrics.

4) Results of Experiment D: Results of Experiment D are shown in Table 6, where L1 and M1 mean Lyrics 1 and Music 1, respectively. As for lyrics evaluations, lower limits are 3.7, 3.9 3.6 and 3.8, and averages are 4.4, 4.3 4.1 and 4.4. Then, it is found that subject evaluations are affirmative about lyrics composed by other persons using the presented system. On the other hand, as for song evaluations, lower limits are 3.2, 3.0 3.2 and 3.5, and averages are 3.8, 3.6, 3.8 and 4.2. Although evaluations of songs are lower than those of lyrics, subject evaluations are rather affirmative. However, comparing evaluations in Experiment D with those in Experiment C, it is found that even if composers evaluations of composed music/lyrics are high, someone else but composers does not necessarily evaluate them high.

TABLE 6. POPULATION MEAN AT A 95% CONFIDENCE (EXPERIMENT D)

	L1	M1	L2	M2	L3	M3	L4	M4
Lower Limit	3.7	3.2	3.9	3.0	3.6	3.2	3.8	3.5
Average	4.4	3.8	4.3	3.6	4.1	3.8	4.4	4.2
Upper Limit	5.0	4.5	4.7	4.2	4.6	4.4	4.9	4.8

IV. CONCLUSIONS

This paper proposes the music/lyrics composition system considering user's image of a song and music genre. The system consists of two sections, the lyrics composing section and the music composing section. The former section has lyrics database obtained by the analysis of existing lyrics using Markov Chain. This section composes lyrics reflecting user's image by the use of lyrics database based on music genres. The latter section applies Interactive Genetic Algorithms in which the difference between the number of words and that of notes is evaluated by the prepared fitness function. Furthermore, a user evaluates whether composed melodies and lyrics fit user's image. In order to verify the validity of the presented system, subject experiments are performed. Each subject composes music/lyrics with music genres. After the composition of music/lyrics, each subject answers questionnaires from the following points of view: Whether combinations of composed music/lyrics fit subject's image of a song or not, and whether each subject can compose music/lyrics easily or not. Experimental results show that the presented composition system composes music/lyrics fitting user's image and that the system supports inexperienced users to compose music/lyrics. However, the number of samples in the experiments is 22 and it is small for the evaluation of the composition system. The increment of samples should be considered for the further evaluation of the system.

REFERENCES

[1] Jean-Jacques Rousseau, (Translated byYoshihiko Kobayashi), "Essay on the origin of language", Gendaishichosha,1976

[2] Minako Yoshikawa, "Software Creation 2004", Information-Technology Promotion Agency, 2004

[3] H.G. Oliveira, A. Cardoso, F.C. Pereira," Tra-la-Lyrics: An approach to generate text based on rhythm", 4th International Joint Workshop on Computational Creativity,2007, pp.47-54.

[4] YAMAHA Co., "VOCALOID",http://www.vocaloid.com/jp/index.html

[5] S. Imai, and T. Nagao, "Automation Composition Using Genetic Algorithm", Electronic telecommunication society technology research report, No.27, 1998

[6] Hüseyin Göksu , Paul Pigg and Vikas Dixit," Music composition using Genetic Algorithms (GA) and multilayer perceptrons (MLP)", Springer Berlin / Heidelberg, 2005.

[7] M. Kowada, "Markov Chain", Hakujitsu-sha, 1973

[8] M.Unehara, T.Onisawa, "Music composition by interaction between human and computer", New Generation Computing vol.23, no.2, 2005, pp.181-191.

[9] C. Nakamura and T. Onisawa, "Music/lyrics composition system with user's impressions of theme", Proc. of International Symposium on Advanced Intelligent Systems, 2007, pp.460-465.

[10] H. Iba, Japanese Society for Artificial Intelligence ed., "Theory of evolution calculation method", Ohm-sha, 2005.

[11] H. Kitano, "Genetic Algorithm 4", Sangyo-tosho, 2000

[12] Lyrics Master, http://www.kenichimaehashi.com/lyricsmastre/

[13] Uta Map, http//www.utamap.com/

[14] K. Yoshimura, "Base of natural language processing", Science-sha,2000