

# Automatic Composition System based on Genetic Algorithm and $N$ -gram Model Considering Chord Progression

Naoki SUGIYAMA, Masanobu YOSHIDA, Katsuya ASANUMA and Yuko OSANA

School of Computer Science, Tokyo University of Technology  
 1404-1 Katakura, Hachioji, Tokyo, 192-0982, Japan Email: osana@cs.teu.ac.jp

**Abstract**—In this paper, we propose an automatic composition system based on genetic algorithm and  $N$ -gram model considering chord progression. In the proposed system, the tone and the length of sounds are expressed in a form of gene, and some melodies are generated by genetic algorithm. In this system, the fitness of the gene is calculated based on (a) transition of rhythm, (b) transition of tone, (c) transition of tone and length, (d) transition of the number of sounds per bar, (e) transition of chord progression, (f) similarity of rhythms between bars, (g) rate of unique sounds of scales and (h) difference between two sounds. We carried out a series of computer experiments and confirmed the effectiveness of the proposed system.

## 1. Introduction

Since the automatic composition was realized based on the Markov process in 1957, much research on the automatic composition has been carried out[1]. As such methods, the methods based on various techniques such as the Markov process, the Hidden Markov Model(HMM), the neural networks, the fuzzy inference, the genetic algorithms[2] and the genetic programming.

For example, in ref.[3], the automatic composition is realized by the genetic algorithm[2] and  $N$ -gram model[4]. Although this model can realize the automatic composition which reflects the features of melodies used in the learning, they cannot always generate natural melodies.

In this paper, we propose the automatic composition system based on genetic algorithm and  $N$ -gram model considering chord progression. In the proposed system, the tone and the length of sounds are expressed in a form of gene, and some melodies are generated by genetic algorithm.

## 2. Automatic Composition System based on Genetic Algorithm and $N$ -gram Model considering Chord Progression

### 2.1. Outline

In this paper, we propose the automatic composition system based on genetic algorithm and  $N$ -gram model considering chord progression. In the proposed system, the tone and the length of sounds are expressed in a form of gene, and some melodies are generated by genetic algorithm. In

this system, the fitness of the gene is calculated based on (a) transition of rhythm, (b) transition of tone, (c) transition of tone and length, (d) transition of the number of sounds per bar, (e) transition of chord progression, (f) similarity of rhythms between bars, (g) rate of unique sounds of scales and (h) difference between two sounds.

### 2.2. Learning of Feature in $N$ -gram Model

#### (1) $N$ -gram Model for Rhythm

In the proposed system, two  $N$ -gram models for rhythm are used.

#### (a) $N$ -gram Model for Rhythm per Two Bars $M_{R1}$

The sequence of rhythm per two bars in the block  $j$  for the melody  $m$  is given by

$$R_1(m_j)_{N_1^{s(m_j)+2}} = \{R_1(m_j)_1, R_1(m_j)_2, \dots, R_1(m_j)_i, \dots, R_1(m_j)_{N_1^{s(m_j)+2}}\} \quad (R_1(m_j)_i \in C_{R1}) \quad (1)$$

where  $N_1^{s(m_j)}$  is the number of sounds and rests in the block  $j$  of the melody  $m$ . In the melody of four-four time,  $C_{R1}$  is given by

$$C_{R1} = \{b_1, b_2, e_1, e_2, s_1, \dots, s_{96}, r_1, \dots, r_{96}\} \quad (2)$$

where  $b_1$  and  $b_2$  show the beginning of the block.  $b_1$  shows the case where the first sound begins at the beginning of the block and  $b_2$  shows the case where the first sound begins at the end of the previous block.  $e_1$  and  $e_2$  show the end of the block.  $e_1$  shows the case where the last sound ends in the block, and  $e_2$  shows the case where the last sound continue to the next block.  $s_1 \sim s_{96}$  show the beginning of the sound,  $r_1 \sim r_{96}$  show the beginning of the rest, and  $1 \sim 96$  show the position in the block. In the proposed system, the minimum length of the tone (unit length) is set to  $1/3$  length of the demiquaver.

In the  $N$ -gram model of rhythm per two bars  $M_{R1}$ , the probability is estimated by

$$P(R_1(m_j)_i^{i+1} | R_1(m_j)_{i-N+1}^i) = \frac{N(R_1(m_j)_{i-N+1}^{i+1})}{N(R_1(m_j)_{i-N+1}^i)} \quad (3)$$

where  $N(R_1(m_j)_{i-N+1}^{i+1})$  is the number of the rhythm sequence  $R_1(m_j)_{i-N+1}^{i+1}$ , and  $N(R_1(m_j)_{i-N+1}^i)$  is the number of the rhythm sequence  $R_1(m_j)_{i-N+1}^i$ .

#### (b) $N$ -gram Model for Rhythm per Four Bar $M_{R2}$

The sequence of rhythm per four bars in the block  $j$  for the melody  $m$  is given by

$$R_2(m_j)_{N_2^{(m_j)+2}} = \{R_2(m_j)_1, R_2(m_j)_2, \dots, R_2(m_j)_i, \dots, R_2(m_j)_{N_2^{(m_j)+2}}\}$$

$$R_2(m_j)_{N_2^s(m_j)+2} \quad (R_2(m_j)_i \in C_{R2}) \quad (4)$$

where  $N_2^s(m_j)$  is the number of sounds and rests in the block  $j$  of the melody  $m$ . In the melody of four-four time,  $C_{R2}$  is given by

$$C_{R2} = \{b_1, b_2, e_1, e_2, s_1, \dots, s_{192}, r_1, \dots, r_{192}\}. \quad (5)$$

In the  $N$ -gram model of rhythm per four bars  $M_{R2}$ , the probability is estimated by

$$P(R_2(m_j)_i^{i+1} | R_2(m_j)_{i-N+1}^i) = \frac{N(R_2(m_j)_{i-N+1}^{i+1})}{N(R_2(m_j)_{i-N+1}^i)} \quad (6)$$

where  $N(R_2(m_j)_{i-N+1}^{i+1})$  is the number of the rhythm sequence  $R_2(m_j)_{i-N+1}^{i+1}$ , and  $N(R_2(m_j)_{i-N+1}^i)$  is the number of the rhythm sequence  $R_2(m_j)_{i-N+1}^i$ .

## (2) $N$ -gram Model for Tone $M_T$

In the  $N$ -gram model for tone  $M_T$ , the probability is estimated by

$$P(T(m)_i | T(m)_{i-N+1}^{i-1}) = \frac{N(T(m)_{i-N+1}^i)}{N(T(m)_{i-N+1}^{i-1})} \quad (7)$$

where  $T(m)_i$  is the tone of the  $i$ th sound in the melody  $m$ ,  $N(T(m)_{i-N+1}^i)$  is the number of tone sequence  $T(m)_{i-N+1}^i$ , and  $N(T(m)_{i-N+1}^{i-1})$  is the number of tone sequence  $T(m)_{i-N+1}^{i-1}$ .

## (3) $N$ -gram Model for Tone and Length of Sound

In the  $N$ -gram model for tone and length  $M_S$ , the probability is estimated by

$$P(S(m)_i | S(m)_{i-N+1}^{i-1}) = \frac{N(S(m)_{i-N+1}^i)}{N(S(m)_{i-N+1}^{i-1})} \quad (8)$$

where  $S(m)_i$  is the tone and length of the  $i$ th sound in the melody  $m$ ,  $N(S(m)_{i-N+1}^i)$  is the number of the sound sequence  $S(m)_{i-N+1}^i$ , and  $N(S(m)_{i-N+1}^{i-1})$  is the number of the sound sequence  $S(m)_{i-N+1}^{i-1}$ .

## (4) $N$ -gram Model for Number of Sounds per Bar

In the  $N$ -gram model  $M_B$ , the probability is estimated by

$$P(B(m)_i | B(m)_{i-N+1}^{i-1}) = \frac{N(B(m)_{i-N+1}^i)}{N(B(m)_{i-N+1}^{i-1})} \quad (9)$$

where  $B(m)_i$  is the number of sounds in the  $i$ th bar in the melody  $m$ ,  $N(B(m)_{i-N+1}^i)$  is the number of the sequence  $B(m)_{i-N+1}^i$ , and  $N(B(m)_{i-N+1}^{i-1})$  is the number of the sequence  $B(m)_{i-N+1}^{i-1}$ .

## (5) $N$ -gram Model for Chord Progression

In the  $N$ -gram model  $M_C$ , the probability is estimated by

$$P(C(m)_i | C(m)_{i-N+1}^{i-1}) = \frac{N(C(m)_{i-N+1}^i)}{N(C(m)_{i-N+1}^{i-1})} \quad (10)$$

where  $C(m)_i$  is the chord in the  $i$ th bar in the melody  $m$ ,  $N(C(m)_{i-N+1}^i)$  is the number of the sequence  $C(m)_{i-N+1}^i$ , and  $N(C(m)_{i-N+1}^{i-1})$  is the number of the sequence  $C(m)_{i-N+1}^{i-1}$ .

## 2.3. Automatic Composition System based on Genetic Algorithm

### (1) Gene Expression for Melody

In the proposed system, the gene is expressed by rhythm and tone.

The rhythm information is expressed by 0 (rest), 1 (the beginning of the sound), 2 (the state that the sound continue) per unit of length. In this expression, the sequence "02" are not permitted, only the sequence "01" are permitted. If there is the sequence "02" in the gene, the following sequence including "02"  $R(g)_a^b$

$$R(g)_j \in \begin{cases} \{1, 2\} & (j = a - 1) \\ \{0\} & (a \leq j < i) \\ \{2\} & (i \leq j \leq b) \\ \{0, 1\} & (j = b + 1) \end{cases} \quad (11)$$

is changed into

$$R(g)_j = \{0 \text{ or } 2\} \quad (a \leq j \leq b). \quad (12)$$

The tone information is expressed by the tone and octave.

### (2) Generation of Initial Gene

In the proposed system, initial genes are generated based on the feature of training melodies. The feature of training melodies are expressed in the Markov model  $M_{R3}$  for the state transition sequence and the occurrence probability of the sound  $o$  at the position  $i$   $P_i(o)$ .

In the Markov model  $M_{R3}$ , the probability  $P(R_1(m_j)_i | R_1(m_j)_{i-1})$  is given by

$$P(R_1(m_j)_i | R_1(m_j)_{i-1}) = \frac{N(R_1(m_j)_{i-1}^i)}{N(R_1(m_j)_{i-1})} \quad (13)$$

where  $R_1(m_j)_i$  is the rhythm state at the position  $i$  in the block  $j$  of the melody  $m$ ,  $N(R_1(m_j)_{i-1}^i)$  is the number of the sequence  $R_1(m_j)_{i-1}^i$  and  $N(R_1(m_j)_{i-1})$  is the number of the sequence  $R_1(m_j)_{i-1}$ .

The occurrence probability of the first sound  $P(R_1(m_1)_2)$  is given by

$$P(R_1(m_1)_2) = N(R_1(m_1)_2) / N^m \quad (14)$$

where  $N(R_1(m_1)_2)$  is the number of state  $R_1(m_1)_2$ , and  $N^m$  is the number of training melodies.

The first state after the state  $b_1$  is generated based on the Eq.(14), and the remain rhythm state sequences are generated randomly based on the Markov model  $M_{R3}$ . The tone of each sound is determined based on the probability  $P_i(o)$ .

The probability of the sound  $o$  at the position  $i$   $P_i(o)$  is given by

$$P_i(o) = N_i^o / N_i \quad (15)$$

where  $N_i^o$  is the number of sounds whose tone is  $o$  and begin at the position  $i$ , and  $N_i$  is the number of the sounds which begin at the position  $i$ .

### (3) Evaluation

#### (a) Fitness on Transition of Rhythm

The fitness on transition of rhythm is calculated based on the  $N$ -gram models  $M_{R1}$  and  $M_{R2}$ .

#### (a-1) Fitness on Transition of Rhythm per Two Bars

The fitness on transition of rhythm per two bars of the gene  $g$ ,  $V_{R1}(g)$  is given by

$$V_{R1}(g) = \frac{2}{N^B} \sum_{j=1}^{N^B/2} \left( \frac{1}{N_1^S(g_j)} \sum_{i=1}^{N_1^S(g_j)} P(R_1(g_j)_i^{i+1} | R_1(g_j)_{i-N+1}^i) \right)$$

(16)

where  $N^B$  is the number of bars in the melody, and  $N^B$  was set to 16.  $N_1^S(g_j)$  is the number of sounds and rests in the block  $j$ . In Eq.(16),  $P(R_1(g_j)^{i+1} | R_1(g_j)_{i-N+1}^i)$  is the probability calculated by the  $N$ -gram model  $M_{R1}$ .  $R_1(g_j)^{i+1}$  is the  $i$ th state in the block  $j$  of the melody expressed in the gene  $g$ , and  $R_1(g_j)_{i-N+1}^i$  is the  $i-N+1 \sim i-1$ th rhythm sequence in the block  $j$  of the melody expressed in the gene  $g$ .

#### (a-2) Fitness on Transition of Rhythm per Four Bars

The fitness on transition of rhythm per four bars of the gene  $g$ ,  $V_{R2}(g)$  is given by

$$V_{R2}(g) = \frac{4}{N^B} \sum_{j=1}^{N^B/4} \left( \frac{1}{N_2^S(g_j)} \sum_{i=1}^{N_2^S(g_j)} P(R_2(g_j)_i^{i+1} | R_2(g_j)_{i-N+1}^i) \right) \quad (17)$$

where  $N_2^S(g_j)$  is the number of sounds and rests in the block  $j$ . In Eq.(17),  $P(R_2(g_j)_i^{i+1} | R_2(g_j)_{i-N+1}^i)$  is the probability calculated by the  $N$ -gram model  $M_{R2}$ .  $R_2(g_j)_i^{i+1}$  is the  $i$ th state in the block  $j$  of the melody expressed in the gene  $g$ , and  $R_2(g_j)_{i-N+1}^i$  is the  $i-N+1 \sim i-1$  rhythm sequence in the block  $j$  of the melody expressed in the gene  $g$ .

#### (b) Fitness on Transition of Tone

The fitness on transition of tone is calculated based on the  $N$ -gram model  $M_T$ .

The fitness on transition of tone of the gene  $g$ ,  $V_T(g)$  is given by

$$V_T(g) = \frac{1}{N^S(g)} \sum_{i=1}^{N^S(g)} P(T(g)_i | T(g)_{i-N+1}^{i-1}) \quad (18)$$

where  $N^S(g)$  is the number of sounds in the melody expressed by the gene  $g$ .  $P(T(g)_i | T(g)_{i-N+1}^{i-1})$  is the probability calculated by the  $N$ -gram model  $M_T$ ,  $T(g)_i$  is the number of  $i$ th tone in the melody expressed by the gene  $g$ , and  $T(g)_{i-N+1}^{i-1}$  is the  $i-N+1 \sim i-1$ th tones in the melody expressed by the gene  $g$ .

#### (c) Fitness on Transition of Tone and Length

The fitness on transition of tone and length is calculated based on the  $N$ -gram model  $M_S$ .

The fitness on transition of tone and length of the gene  $g$ ,  $V_S(g)$  is given by

$$V_S(g) = \frac{1}{N^S(g)} \sum_{i=1}^{N^S(g)} P(S(g)_i | S(g)_{i-N+1}^{i-1}) \quad (19)$$

where  $P(S(g)_i | S(g)_{i-N+1}^{i-1})$  is the probability calculated by the  $N$ -gram model  $M_S$ .  $S(g)_i$  is the  $i$ th tone and length in the melody expressed by the gene  $g$ , and  $S(g)_{i-N+1}^{i-1}$  is the  $i-N+1 \sim i-1$ th tone and length.

#### (d) Fitness on Transition of the Number of Sounds per Bar

The fitness on transition of the number of sounds per bar  $V_B(g)$  is calculated by the  $N$ -gram model  $M_B$ , and is given

by

$$V_B(g) = \frac{1}{N^B} \sum_{i=1}^{N^B} P(B(g)_i | B(g)_{i-N+1}^{i-1}) \quad (20)$$

where  $P(B(g)_i | B(g)_{i-N+1}^{i-1})$  is the probability calculated by the  $N$ -gram model  $M_B$ .  $B(g)_i$  is the number of sounds of the  $i$ th bar in the melody expressed by the gene  $g$ , and  $B(g)_{i-N+1}^{i-1}$  is the numbers of sounds in the  $i-N+1 \sim i-1$ th bars in the melody expressed by the gene  $g$ . Here, the number of sounds does not include the number of rests.

#### (e) Fitness on Transition of Chord Progression

The fitness on transition of chord  $V_C(g)$  is calculated by the  $N$ -gram model  $M_C$ , and is given by

$$V_C(g) = \frac{1}{N^B} \sum_{i=1}^{N^B} P(C(g)_i | C(g)_{i-N+1}^{i-1}) \quad (21)$$

where  $P(C(g)_i | C(g)_{i-N+1}^{i-1})$  is the probability calculated by the  $N$ -gram model  $M_C$ .  $C(g)_i$  is the chord of the  $i$ th bar in the melody expressed by the gene  $g$ , and  $C(g)_{i-N+1}^{i-1}$  is the  $i-N+1 \sim i-1$ th chord sequence in the melody expressed in the gene  $g$ .

#### (f) Fitness on Similarity of Rhythms between Bars

The fitness on similarity of rhythms between bars  $V_{RS}(g)$  is given by

$$V_{RS}(g) = f_{RS} \left( \frac{1}{N^B/4 C_2} \sum_{i=1}^{N^B/4} \sum_{j=i+1}^{N^B/4} \left( L_{ij}^g + a \cdot \left( \frac{n_{ij}^g}{N_i^S} + \frac{n_{ji}^g}{N_j^S} \right) \right) \right) \quad (22)$$

where  $L_{ij}^g$  is the rate of sounds and rests whose length and position are same in the blocks  $i$  and  $j$ .  $N_i^S$  is the number of sounds and rests in the block  $i$ ,  $N_j^S$  is the number of sounds and rests in the block  $j$ ,  $n_{ij}^g$  is the number of sounds and rests in the block  $i$  which begin at the same time to the block  $j$ ,  $n_{ji}^g$  is the number of sounds and rests in the block  $j$  which begin at the same time to the block  $i$ , and  $a$  is the constant.

#### (g) Fitness on Rate of Unique Sounds of Scale

The most of the melody in C major consists of C, D, E, F, G, A, B. These sounds are unique sounds of scale of C major.

The fitness on rate of unique sounds of scale in the gene  $g$ ,  $V_{US}(g)$  is given by

$$V_{US}(g) = f_{US} (N^{US}(g)/N^S(g)) \quad (23)$$

where  $N^S(g)$  is the number of sounds in the melody expressed by the gene  $g$ , and  $N^{US}(g)$  is the number of the unique sounds of scale in the melody expressed by the gene  $g$ .  $f_{US}(\cdot)$  is given by

$$f_{US}(u) = \begin{cases} 1 & (u > \theta_{US}) \\ u/\theta_{US} & (\text{otherwise}) \end{cases} \quad (24)$$

where  $\theta_{US}$  is the threshold for the rate of unique sound, and is given by

$$\theta_{US} = \min_m \{N^{US}(m)/N^S(m)\} \quad (25)$$

where  $N^S(m)$  is the number of sounds in the melody  $m$ , and  $N^{US}(m)$  is the number of unique sounds of scale in the

melody  $m$ .

### (h) Fitness on Difference between Two Sounds

The fitness on difference between two sounds of the gene  $g$ ,  $V_D(g)$  is given by

$$V_D(g) = \frac{1}{2D_{max} + 1} \sum_{i=-D_{max}}^{D_{max}} f_{D_i} \left( \frac{N_i^D(g)}{N^D(g)} \right) \quad (26)$$

where  $D_{max}$  is the maximum difference between two sounds,  $N^D(g)$  is the number of differences between two sounds in the melody expressed in the gene  $g$ , and  $N_i^D(g)$  is the number of differences between two sounds whose value is  $i$ .  $f_{D_i}(\cdot)$  is given by

$$f_{D_i}(u) = \begin{cases} (1-u)/(1-\theta_{D2}^i) & (\theta_{D2}^i < u) \\ 1 & (\theta_{D1}^i \leq u \leq \theta_{D2}^i) \\ u/\theta_{D1}^i & (\theta_{D1}^i) \end{cases} \quad (27)$$

where  $\theta_{D1}^i$  and  $\theta_{D2}^i$  ( $\theta_{D1}^i \leq \theta_{D2}^i$ ) are the thresholds on the difference between two sounds  $i$ , and are given by

$$\theta_{D1}^i = \min \{ N_i^D(m) / N^D(m) \} \quad (28)$$

$$\theta_{D2}^i = \max \{ N_i^D(m) / N^D(m) \} \quad (29)$$

where  $N^D(m)$  is the number of differences between two sounds in the melody  $m$ , and  $N_i^D(m)$  is the number of differences whose value is  $i$  in the melody  $m$ .

### (i) Fitness of Gene $g$

The fitness of the gene  $g$  is given as the weighted sum of the witnesses calculated in (a)~(h).

$$V(g) = w_{R1}V_{R1}(g) + w_{R2}V_{R2}(g) + w_T V_T(g) + w_S V_S(g) + w_B V_B(g) + w_C V_C(g) + w_{RS} V_{RS}(g) + w_{US} V_{US}(g) + w_D V_D(g) \quad (30)$$

where  $w_{R1}$ ,  $w_{R2}$ ,  $w_T$ ,  $w_S$ ,  $w_B$ ,  $w_C$ ,  $w_{RS}$ ,  $w_{US}$  and  $w_D$  are the coefficients.

## 2.4. Estimation of Chord

In the proposed system, the chord per bar is estimated by the 3-layered hierarchical neural network which is trained by the Back Propagation (BP) algorithm[5]. In the neural network, the rate of tone in three bars is given to the network, and the chord of the center bar is estimated.

## 3. Computer Experiment Results

Here, we show the computer experiment results to demonstrate of the effectiveness of the proposed system.

### 3.1. Generation of Melodies

Figure 1 show the melodies generated by the proposed method.

### 3.2. Features of Generated Melodies

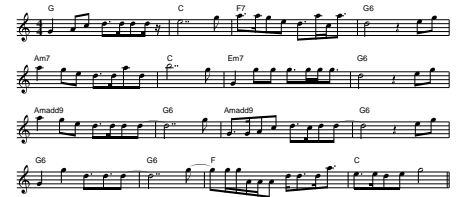
Figure 2 show the distribution of two sounds of trained melody and 30 generated melodies. As shown in Fig.2, the distribution of difference of tone between two sounds of the generated melodies is similar to that of the training data.



(a) from Japanese Children Song

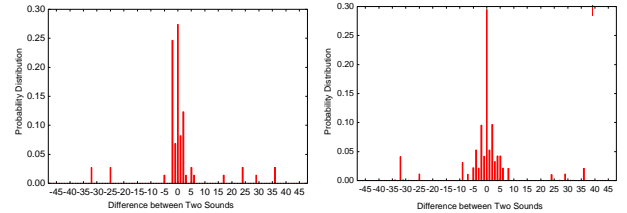


(b) from Game Music



(c) from Animation Song

Figure 1: Generated Melodies.



(a) Training Data

(b) Generated Melodies

Figure 2: Distribution of Difference of Two Sounds.

## 4. Conclusions

In this paper, we have proposed the automatic composition system based on genetic algorithm and  $N$ -gram model. We carried out a series of computer experiments and confirmed the proposed model can generate natural melodies which have the features similar to the melodies used in the learning.

## References

- [1] K. H. Burns: "The history and development of algorithms in music composition, 1957–1993," Ball State University, 1994.
- [2] D. E. Goldberg: Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Longman Publishing, 1989.
- [3] M. Tomari, M. Sato and Y. Osana: "Automatic composition based on genetic algorithm and  $N$ -gram model," Proceedings of IEEE Conference on System, Man and Cybernetics, Singapore, 2008.
- [4] C. D. Manning, H. Schütze: Foundations of Statistical Natural Language Processing, MIT Press, 1999.
- [5] D. E. Rumelhart, J. L. McClelland and the PDP Research Group: Parallel Distributed Processing, — Explorations in the Microstructure of Cognition, Vol.1 : Foundations, The MIT Press, 1986.