A LOOP SEQUENCER THAT SELECTS MUSIC LOOPS BASED ON THE DEGREE OF EXCITEMENT

Tetsuro Kitahara, Kosuke Iijima, Misaki Okada, Yuji Yamashita, and Ayaka Tsuruoka
College of Humanities and Sciences, Nihon University, Japan
{kitahara,iijima,misaki,yuji,tsuruoka}@kthrlab.jp

ABSTRACT

In this paper, we propose a new loop sequencer that automatically selects music loops according to the *degree of excitement* entered by the user. A loop sequencer is expected to be a good tool for non-musicians to compose music because it does not require expert musical knowledge. However, it is not easy to appropriately select music loops because a loop sequencer usually has a large scale of loop collection (e.g., more than 3000 loops). It is therefore necessary to automatically select music loops based on the user's simple and easy input. In this paper, we focus on the degree of excitement. In typical techno music, the temporal evolution of excitement is an important feature. Our system allows the user to enter the temporal evolution of excitement by drawing a curve, then the system automatically selects music loops according to the entered excitement. Experimental results show that our system is easy to understand and generates satisfying musical pieces for non-experts of music.

1. INTRODUCTION

A loop sequencer such as ACID PRO 7 or GarageBand is a popular tool for composing musical pieces. It enables the user to compose musical pieces by concatenating short musical materials called *music loops*. Because music loops are usually audio data, the user can easily compose high-quality pieces (e.g., without expert knowledge) as compared to inputting musical notes, one by one, in a MIDI sequencer.

A loop sequencer requires a huge number of music loops in order to enable users to compose a wide variety of music. For example, ACID PRO has more than 3,000 music loops. However, it is not easy for most users to listen to all 3,000 music loops, consider which loops match their inspiration, and then select the appropriate loops. It is therefore not common to enjoy composing music with a loop sequencer, even though expert musical knowledge is not required to use this tool.

There are many studies that investigate automatic music composition and computer-aided music composition. Fukayama et al. proposed a web-based automatic music composition system based on a probabilistic model [1].

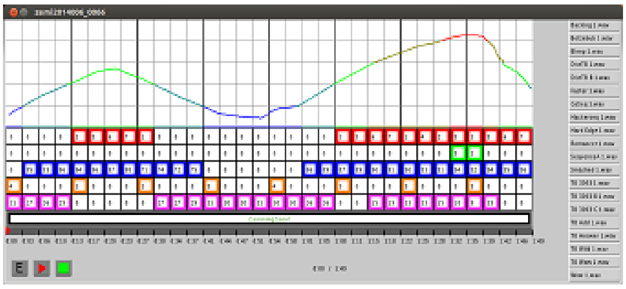


Figure 1. Screenshot of our system

Ando et al. proposed the use of interactive evolutional computation to automatically generating melodies [2]. In addition, many researchers have explored new automatic and/or computer-aided music composition methods [3]. However, all of these studies were aimed at automating and/or supporting note-level music composition, and thus they did not focus on audio-based music composition tools like a loop sequencer.

In this paper, we narrow the target to techno music, and propose a loop sequencer that automatically selects music loops according to the user's input of the desired *degree of excitement*. The temporal evolution of excitement is one of the most important features in techno music, and is expressed by what kinds of music loops are selected and how many. A typical piece of techno music may start with a simple repetition of the bass drum. Then, the excitement may increase as new loops (i.e., those of other drums, bass lines, and melodic lines) are added. This system enables users to input the temporal evolution of the excitement by drawing a curve, and then the system generates a musical piece that most matches the desired degree of excitement.

2. SYSTEM OVERVIEW

The screenshot of this system is shown in Figure 1. A list of music loops is shown in the right side of the screen. The panel for drawing an excitement curve is placed in the upper part of the screen, and the panel for displaying the selected music loops is placed in the lower part. The user first draws an excitement curve. Then, the system automatically determines how many loops will be placed and which loops will be placed at each measure. The user can remove, add, and change these loops at a later point.

3. METHOD FOR SELECTING MUSIC LOOPS

3.1 Formulation with hidden Markov model

The automatic selection of music loops is formulated using a hidden Markov model (HMM). For simplicity, we assume that all loops have the same length (one measure). A musical piece consists of several instrumental parts (e.g., drums, bass, and synth), and all music loops have been classified into the instrumental parts in advance. Let $\mathcal{M}_i$ be a set of music loops for the part i.

The problem of selecting music loops is to find the most likely $\boldsymbol{s}_n = (s_{n,1}, \cdots, s_{n,I})$ $(s_{n,i} \in \mathcal{M}_i \cup \{0\})$ for each measure n, where $s_{n,i}$ represents the music loop for the n-th measure of the part i. When we do not select any loops for the n-th measure of the part i, we denote this by $s_{n,i} = 0$. Our system supposes that a sequence of the excitement, $\boldsymbol{x} = [x_1, \cdots, x_N]$, is given. Therefore, the selection of music loops is formulated as:

$$\hat{\boldsymbol{S}} = \underset{\boldsymbol{S}}{\operatorname{argmax}}\, P(\boldsymbol{S}|\boldsymbol{x}),$$

where $\boldsymbol{S} = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_N]$. The excitement curve is freely drawn, so its value may vary within each measure. Therefore, we use the mean value of the excitement within the n-th measure for x_n.

Using Bayes' theorem, this equation is expanded to:

$$\hat{\boldsymbol{S}} = \underset{\boldsymbol{S}}{\operatorname{argmax}}\, P(\boldsymbol{x}|\boldsymbol{S})P(\boldsymbol{S}).$$

When we assume that x_n depends only on $\boldsymbol{s}_n$ and $\boldsymbol{s}_n$ depends only on $\boldsymbol{s}_{n-1}$, the equation becomes the following:

$$\hat{\boldsymbol{S}} = \underset{\boldsymbol{s}_1,\cdots,\boldsymbol{s}_N}{\operatorname{argmax}} \prod_{n=1}^{N} P(x_n|\boldsymbol{s}_n) \cdot P(\boldsymbol{s}_1) \prod_{n=2}^{N} P(\boldsymbol{s}_n|\boldsymbol{s}_{n-1})$$

This equation is equivalent to the use of an HMM in which $\boldsymbol{x}$ and $\boldsymbol{S}$ are regarded as a sequence of observations and a sequence of state transitions, respectively. The most likely $\boldsymbol{S}$ can be determined with the Viterbi algorithm, if $P(x_n|\boldsymbol{s}_n)$, $P(\boldsymbol{s}_1)$, and $P(\boldsymbol{s}_n|\boldsymbol{s}_{n-1})$ are appropriately designed or trained.

3.2 Simplification of formulation

The above-mentioned formulation is difficult to directly apply because it includes a large number of parameters. We therefore simplify the formulation based on the following policies:

Policy 1 We discretize the degree of excitement, in other words, the degree of excitement is any of $1, \cdots, d$ ($d = 5$ in the current implementation).

Policy 2 We separately consider (1) whether a certain loop should be placed for the n-th measure of the part i, and (2) if so, which loop should be placed there. Also, we assume that (1) and (2) are independent.

Policy 3 We assume that which loop should be placed at each measure in each part is determined independently of other parts and/or other measures. However, we select the same loop within every eight measures in order to keep a consistent mood.

Policy 4 For every music loop, the degree of excitement of this loop itself is annotated as any of $1, \cdots, d$.

Based on Policy 2, we reformulate the problem by separately considering (1) whether a certain loop should be placed and (2) which loop should be placed. The random variables representing the former and latter are denoted by $q_{n,i} (\in \{0,1\})$ and $s'_{n,i} (\in \mathcal{M}_i)$, respectively. Thus, $P(s_{n,i})$ can be expanded to:

$$\begin{aligned} P(s_{n,i}) &= P(q_{n,i},\ s'_{n,i}) \\ &= P(q_{n,i})\ P(s'_{n,i}|q_{n,i}) \\ &= P(q_{n,i})\ P(s'_{n,i}). \end{aligned}$$

Therefore, $P(x_n|\boldsymbol{s}_n)$ is expanded as follows:

$$P(x_n|\boldsymbol{s}_n) = \alpha P(x_n|\boldsymbol{q}_n) \prod_{i=1}^{I} P(x_n|s'_{n,i}),$$

where $\boldsymbol{q}_n = (q_{n,1}, \cdots, q_{n,I})$ and α is a constant, because $s'_{n,1}, \cdots, s'_{n,I}$ are independent according to Policy 3.

Here, $P(x_n|\boldsymbol{q}_n)$ models the relationship between the excitement and the number of parts that play music loops. In general, music becomes increasingly exciting as more loops are simultaneously played back. We manually design $P(x_n|\boldsymbol{q}_n)$ based on this idea.

On the other hand, $P(x_n|s'_{n,i})$ models the relationship between the excitement and the music loop itself. According to Policy 4, the music loop $s'_{n,i}$ has an annotation of its excitement $X(s'_{n,i})$. We therefore design $P(x_n|s'_{n,i})$ to assume that $X(s'_{n,i}) - x_n$ follows a normal distribution with zero mean (the variance is experimentally determined).

Similarly, $P(\boldsymbol{s}_1)$ and $P(\boldsymbol{s}_n|\boldsymbol{s}_{n-1})$ are expanded to:

$$\begin{aligned} P(\boldsymbol{s}_1) &= P(\boldsymbol{q}_1) \prod_{i=1}^{I} P(s'_{1,i}), \\ P(\boldsymbol{s}_n|\boldsymbol{s}_{n-1}) &= P(\boldsymbol{q}_n|\boldsymbol{q}_{n-1}) \prod_{i=1}^{I} P(s'_{n,i}|s'_{n-1,i}) \\ &= P(\boldsymbol{q}_n|\boldsymbol{q}_{n-1}) \prod_{i=1}^{I} P(s'_{n,i}). \end{aligned}$$

When we assume that $s'_{n,i}$ is selected at random with equal probabilities, $P(\boldsymbol{s}_1) \propto P(\boldsymbol{q}_1)$, $P(\boldsymbol{s}_n|\boldsymbol{s}_{n-1}) \propto P(\boldsymbol{q}_n|\boldsymbol{q}_{n-1})$.

Therefore, the selection of music loops is achieved with the following equation:

$$\begin{aligned} \hat{\boldsymbol{Q}} &= \underset{\boldsymbol{q}_1,\cdots,\boldsymbol{q}_N}{\operatorname{argmax}} \prod_{n=1}^{N} P(x_n|\boldsymbol{q}_n) \cdot P(\boldsymbol{q}_1) \prod_{n=2}^{N} P(\boldsymbol{q}_n|\boldsymbol{q}_{n-1}), \\ \hat{\boldsymbol{S}}' &= \underset{\boldsymbol{s}'_1,\cdots,\boldsymbol{s}'_N}{\operatorname{argmax}} \prod_{n=1}^{N} \prod_{i=1}^{I} P(x_n|s'_{n,i}). \end{aligned}$$

3.3 Estimation of the degree of excitement for music loops

When a music loop contains sound at a wide range of frequencies (from low-frequency regions to high-frequency regions) from the beginning to the end, this music loop is

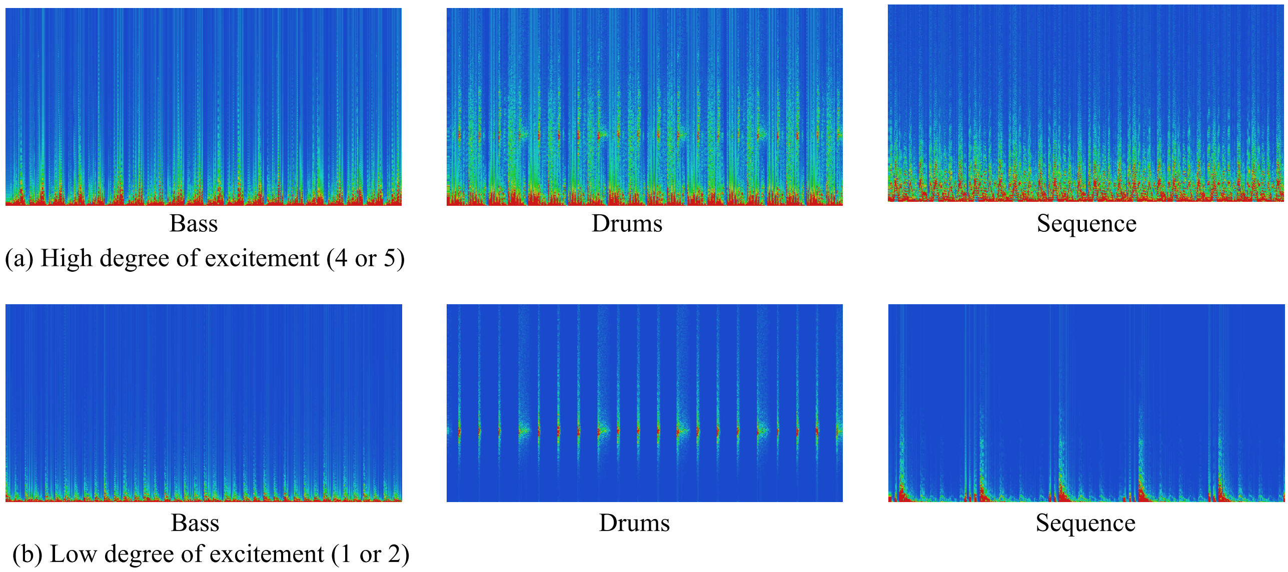


Figure 2. Examples of music loops with high and low degrees of excitement

considered to have a high degree of excitement (Figure 2). We therefore propose the following method for determining the degree of excitement. For a given music loop, a spectrogram is calculated using the short-term Fourier transform with a 4096-point Hamming window shifted by 10ms. Let $A_{t,f}(s)$ be the amplitude at the time t and the frequency f of the music loop s. The function $\sigma(A_{t,f}(s))$, ranging from 0.0 to 1.0, is calculated from $A_{t,f}(s)$ at every time and every frequency. The function $\sigma(A_{t,f}(s))$ has a value near 0.0 when $A_{t,f}(s)$ is near zero, whereas it has a value near 1.0 when $A_{t,f}(s)$ is higher than a certain value. In the current implementation, this function is calculated as follows:

$$\sigma(A_{t,f}(s)) = \begin{cases} 0.0 & (A_{t,f}(s) \leq 0.1) \\ 0.2 & (0.1 < A_{t,f}(s) \leq 0.2) \\ 0.4 & (0.2 < A_{t,f}(s) \leq 0.3) \\ 0.6 & (0.3 < A_{t,f}(s) \leq 0.4) \\ 0.8 & (0.4 < A_{t,f}(s) \leq 0.5) \\ 1.0 & (A_{t,f}(s) > 0.5) \end{cases}$$

This can be considered an approximation of the sigmoid function. After $\sigma(A_{t,f}(s))$ is calculated, its average is calculated along both the time and frequency axes:

$$R(s) = \frac{1}{TF} \sum_{t=1}^{T} \sum_{f=1}^{F} \sigma(A_{t,f}(s)).$$

It is then normalized by dividing it by the maximum value in all music loops of the same instrument part and is transformed to an integer value of 1 to d:

$$X(s) = \left\lceil d \frac{R(s)}{\max_{s' \in \mathcal{M}_i} R(s')} \right\rceil,$$

where $\lceil \cdot \rceil$ is a ceiling function.

4. IMPLEMENTATION AND EXPERIMENTS

4.1 Implementation

We implemented this system using Processing 1.5.1. Music loops were taken from *Sound PooL* [4], which consists of music loops for each of the five instrumental parts of *Sequence*, *Synth*, *Bass*, *Percussion*, and *Drums*. Of these five parts, Percussion has an unique different feature in that; it mainly include sounds that provide an accent (e.g., a crash symbal). For this reason,we place a music loop of Percussion at the beginning of every four measures, independent of the HMM-based formulation described in Section 3. The HMM-based formulation is applied to the remaining four parts of Sequence, Synth, Bass, and Drums. The parameters $P(x_n|\boldsymbol{q}_n)$, $P(\boldsymbol{q}_1)$, $P(\boldsymbol{q}_n|\boldsymbol{q}_{n-1})$ were experimentally determined.

4.2 Experimental methods

We conducted two experiments, on which evaluated the excitement estimation method through paired comparisons (Experiment 1) and the system's usability (Experiment 2).

In Experiment 1, we asked participants to listen to pairs of music loops and to answer which loop in each pair had the higher excitement. We prepared 20 pairs (i.e., four pairs for each part) at random.

In Experiment 2, we asked participants to use our system to compose a piece of background music for a silent movie. Focusing on the techno's feeling of lively motion, we used promotion movies of sport cars [1]. These movies include both slow parts and fast parts that express the evolution of excitement. The time for the composition was limited to 30 minutes, and during the 30 minutes we allowed the participants to redraw the excitemment curve and to add, remove, and change music loops as needed. After they completed each composition, we asked them to rate their

[1] 1) Daihatsu, Tokyo Motor Show 2013 KOPEN future included, https://www.youtube.com/watch?v=zca6Lv4PW8M
2) Honda, NSX Concept GT Shakedown Test, https://www.youtube.com/watch?v=Bip5j00I6Hw

Table 1. Results of Experiment 2

	Baseline	Proposed
Q1	3.13	4.02
Q2	2.54	4.46
Q3	3.94	3.59
Q4	2.39	3.15
Q5	4.98	5.31

experience with the system that they used by answering the following questions on a scale of one to seven:

Q1 How easy or difficult was it to understand the system?

Q2 How easy or difficult did you feel it was to compose this piece of music without any prior musical knowledge?

Q3 How much or how little did the composed music have the degree of excitement that you intended?

Q4 Do you feel that you could compose satisfactory music?

Q5 How would you describe your level of interest in music composition?

The experiment was conducted using both the baseline system and our system. In the baseline system, the automatic music loop selection function was removed from our system. To reduce any effect of that the order of system use might have on study results, half of the participants first used our system, and the other half first used the baseline system. To reduce the effect of that participant fatigue might have on study results, we allowed the participants to have a sufficient rest between use of the two systems.

The participants were 10 university students. Of the 10 participants, seven had prior experience with playing an instrument, and two knew and/or had used a MIDI sequencer.

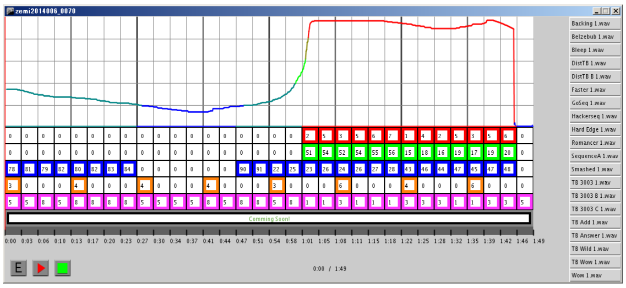
4.3 Experimental results

Experiment 1

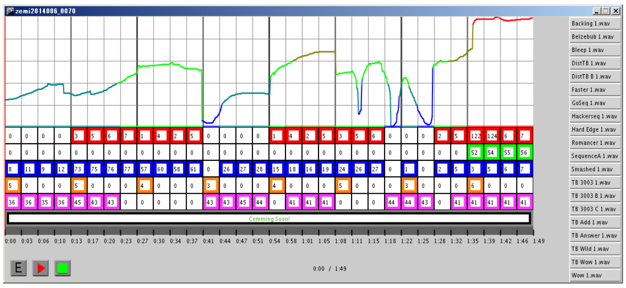
In 16.6 of the 20 pairs on average (approximately 83%), the system's estimation of the degree of excitement agreed with the participants' judgement. It was thus shown that our method for estimating the degree of excitement was appropriate.

Experiment 2

The results are listed in Table 1. For Questions 1 and 2, our system improved a mean score by 0.89 points and 1.92 points, respectively, as compared to the baseline system. This is because participants gave our system a high rating for ease of drawing a curve. On the other hand, for Question 3 our system scored 0.35 points lower than the baseline system. One participant pointed out that the timing of changes in the excitement sometimes did not match the drawn curve. Because our system places music loops measurewise, changes in the excitement occur only at changes in measures. This may explain why our system ranked slightly lower than the baseline system for Question 3. For Question 4, our system scored 0.76 points higher than the baseline system. For Question 5, our system scored 0.33 points higher than the baseline system.



(a) Participant 1



(b) Participant 10

Figure 3. Examples of the screenshots of the musical pieces composed by the participants, using our system

Figure 3 shows excerpts of the screenshots of the musical pieces that participants composed. It is apparent that participants tried to create an evolution of excitement according to the lively motion of the given movie, and that the entered excitement reflected in selection of music loops.

5. CONCLUSION

In this paper, we focused on a new musical feature, the *degree of excitement*, and proposed a loop sequencer that allows the user to directly input the degree of excitement as a curve on a computer screen. Once the degree of excitement is entered, the system determines what loops and how many loops should be placed to reflect the curve in the music. With this function, our system enabled users to quickly and easily compose techno music. Future issues that need to be addressed include reimplementing this system as a Web-based application and acquiring log data by a variety of users to improve the system's behavior.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 26240025.

6. REFERENCES

[1] S. Fukayama, K. Nakatsuma, S. Sako, T. Nishimoto, and S. Sagayama, "Automatic song composition from the lyrics exploiting prosody of the Japanese language," in *Proc. Sound and Music Computing*, 2010, pp. 299–302.

[2] D. Ando, P. Dahlstedt, M. G. Nordaxhl, and H. Iba, "Computer aided composition by means of interactive GP," in *Proc. ICMC*, 2006, pp. 254–257.

[3] G. Nierhaus, *Algorithmic Composition*. Springer, 2009.

[4] AH-Software, "Sound PooL." [Online]. Available: http://www.ah-soft.com/soundpool/