ABSTRACT

Musical sessions are often co-created through the interaction among plural players. We developed a music interaction system between human and machine rhythm performance, which is composed of rhythm maps with an association mechanism. To achieve musical interaction, a many-to-many association should be considered, because a preferable rhythm performance with partner’s rhythms is not unique. The proposed association mechanism tuned by actual human performances allows the system to generate various rhythms that fit a partner’s rhythm. We adopted a 3D acceleration sensor as an interface for human performance. We observed interactions with a human and virtual player to evaluate the effectiveness of the proposed system.

1. INTRODUCTION

A music session is often co-created through the interaction among multiple players. Recently, many systems for musical sessions between humans and machines have been proposed in order to achieve a musical performance, accompaniment, or improvisation session. Raphael [1],[2] proposed a real-time accompaniment system for a human performance. Goto et al. [3] achieved a Jazz improvisation session among a human and multiple virtual players, and Hamanaka et al. [4] proposed a Jam session system with virtual players which can imitate the personalities of various human players. We proposed a piano session system which can exchange initiative among human and computers [5]. However, these systems require the human to possess skills to play specific instruments.

Playing rhythms is the simplest and most essential performance as musical expression. In rhythm interaction, players often generate their rhythm based on association with other players. Nishijima et al. [6] proposed a Neuro-Drummer that generates rhythms as a reaction to the human player performance. This system uses neural networks to make a one-to-one mapping among the rhythms. However in real sessions, interaction requires not only a one-to-one rhythm association but also a many-to-many association.

We propose a rhythm interaction system based on a rhythm association which can realize a many-to-many relationship among players. The proposed system is composed of the associated rhythm maps and user interface. The rhythm association is realized by the connection between rhythm maps. Each rhythm map and the corresponding connection are trained according to actual human performances [7]. The many-to-many relationship represented by this association system resulted in a more suitable combination of plural players.

On the other hand, in actual rhythm performance, there are limitations of action and space due to the physical structure of the instruments. Hence, we have proposed a 3D acceleration sensor as an unrestricted interface device for music control [8]. We also employ a commercially available 3D acceleration sensor for rhythm manipulation in the proposed system.

2. PROPOSED RHYTHM INTERACTION SYSTEM

2.1. System Overview

The overview of our interaction system is shown in Fig. 1. Humans can generate rhythms (Human performance) by shaking the 3D acceleration sensor. The system recognizes the rhythm played by human, and then generates rhythms (Machine Performance) based on the rhythm association. The rhythm association system is realized utilizing two rhythm maps and the map associator trained based on performances by experienced players in advance.

2.2. Rhythm Similarity

In this study, the rhythm at each bar is represented as rhythm vector \( V = (v_1, v_2, ..., v_{16}) \), \( 0 \leq v_i \leq 127 \), which is composed of sixteen sampled velocities (intensity) quantized in 128 steps. Therefore, the temporal resolution is a sixteenth note.

The learning of the rhythm map is based on human’s acoustic similarity of the rhythm vectors considering the
As the four-dimensional matrix of connection weights between units on two rhythm maps, $R_1$ and $R_2$. In the following explanations, positions on the rhythm map $R_1$ and map $R_2$ are denoted as $X_1$ and $X_2$ respectively. The rhythm vectors stored in rhythm map $R_t$ are denoted as $R_t(X_i)$, and the connection weight between two units is expressed as $w(X_1, X_2)$, $(0 \leq w \leq 1)$.

The map link is trained to reinforce the connection of the two rhythms played at the same time. Moreover, connections of similar rhythms are also reinforced according to the distance in the maps. When the rhythm vector $V_1$ and $V_2$ are played at the same time, map links are updated by the following equations, supposing $\hat{X}_1$ as a BMU of $V_1$ and $\hat{X}_2$ as a BMU of $V_2$.

$$w_{t+1}(X_1, X_2) = 1 - \{(1 - w_t(X_1, X_2))(1 - K(X_1, X_2))\}$$

(3)

$$K(X_1, X_2) = \exp\left(\frac{|X_1 - \hat{X}_1|^2}{2\gamma_1^2(X_1)}\right) \exp\left(\frac{|X_2 - \hat{X}_2|^2}{2\gamma_2^2(X_2)}\right)$$

(4)

$$\gamma_i(X_i) = \frac{m}{8} \sum_{X_j=8\text{neighbors of }X_i} S(R_t(X_i), R_t(X_j))$$

(5)

where, $i = 1, 2$, and $\gamma$ on each BMU is the parameter to express the area of update for neighbor units. $\gamma$ is determined by the average similarities among the 8-neighbors. $m$ is constant. The Fig-4 illustrates the rhythm association system and its training procedure.

### 2.5. Human Performance

We adopted the Wii™ Remote [10] (produced by NINTENDO) as a 3D acceleration sensor shown in the Fig. 5. The acceleration sensor of the Wii Remote measures the acceleration, and is used to detect ONSET (i.e. time to sound) and VELOCITY (i.e. the strength of the beat). In order to use the Wii Remote in Max/MSP/Jitter, we utilized the “aka.wiremote” object developed by Akamatsu [11].

The acceleration signal is sent from the Wii Remote to the laptop computer by Bluetooth. When a human gestures with the Remote, i.e. like beating a drum, the acceleration signal is represented as shown in the Fig. 5. Then, the onset and its velocity in the beating gesture are detected by the following method.
From the acceleration waveform $f(t)$, the time $t_s$ and $t_f$ are determined by the threshold $Ac'$ (the Fig. 5). Where, $t_s$ and $t_f$ are defined as the time the swing starts and the time it finishes, respectively. The onset moment of the action is set to $t_f$, and its velocity $v(t_f)$ is set by following equation,

$$v(t_f) = M \int_{t_s}^{t_f} \{ Ac' - f(t) \} dt$$  \hspace{1cm} (6)

where, $t_f$ is corresponded to the sixteenth note $i$ in the bar, and $v(t_f)$ is regarded as $v_i$. The velocity $v_i$ is 0 when the sixteenth note $i$ does not correspond to onset time $t_f$.

The sound for onset is generated by a MIDI sequencer, and the user can select timbre of sound. (the Fig. 6)

### 2.6. Machine Performance

When a human performs a rhythm $V$, its BMU $\hat{X}_1$ on the rhythm map 1 is determined. Then, the relevance ratio (fitness potential) between $\hat{X}_1$ and units on the map 2 is determined by the map link as shown in the Fig. 7. By adopting the relevance ratio on the map 2, we introduce an agent that moves on the map 2 as a virtual player. Then, the agent could understand the dynamics of the system and recognize his/her role in the performance, such as “the user plays the solo part, and the computer follows”, or “the computer plays actively as the main player and the user follows with simple rhythms.” In the Fig. 8, from index 15 to 33, human played low velocity rhythm and computer played louder performance.

We evaluated the proposed rhythm interaction system. We observed the behavior of the agent and human performance while examinees tried a performance, and we received some comments about the system. The number of examinees was five, with three of them having experience with playing instruments and the others having no experience.

To collect the training data for system learning, 500 pairs of rhythm vectors are sampled from actual performances by a skilled player. The size of the rhythm maps is $20 \times 20$. Therefore, the size of the table of map links is $20 \times 20 \times 20 \times 20$. The rhythm map is trained $200 \times 500$ times. The map link is trained 500 times. The initial parameter $\sigma(0)$ for the update region is set as 10, and the parameter $m$ is set as 2.0. $Ac'$, $\alpha$, $\beta$, and $M$ were determined as 0.2 [G], $3.0 \times 10^{-3}$, 0.5, and 250.

The trajectory of performances during the interaction is shown in the Fig. 8. $I_1(X)$ and $I_2(X)$ represent average velocity in the rhythm vector on rhythm map $R_1(X)$ and $R_2(X)$. If the average velocity of rhythm are large, luminance gets darker. Positions of indexes on $I_1(X)$ and $I_2(X)$ represent the BMU position on $R_1$ and position of agent on $R_2$, and number corresponds the sequence of bar, when the position on map has changed. The reports from examinees and observation are as follows.

The human players who had experience with music soon became accustomed to performing with the system. They could understand the dynamics of the system and recognized his/her role in the performance, such as “the user plays the solo part, and the computer follows”, or “the computer plays actively as the main player and the user follows with simple rhythms.”
Based on some feedback from examinees, the following future works are being considered: the introduction of other gestures for sound control to express sustain or effect, the extension for rhythm and melody or music and image association, and the implementation for non-MIDI based ensemble.

Acknowledgment

This research is supported by CREST project “Foundation of technology supporting the creation of digital media contents” of JST and the 21st Century Center of Excellence Program, “The innovative research on symbiosis technologies for human and robots in the elderly dominated society,” Waseda University.

References


Figure 8. Trajectory of performances during the interaction

On the other hand, it was initially difficult for the inexperienced users to play various rhythms at first. But soon, they learned to play after being inspired by machine performance. Even though the required time depends on individual users, they could eventually perform interaction sessions like a skilled player.

At 28-36, When the agent reached the top of the fitness potential, the position of the agent became unstable, which made the machine performance more exciting and inspiring to the human performance. And at time index9-13 and index36-37, even thought human played similar rhythms, virtual player performed different rhythms, this is achieved by many-to-many association.

The system can gradually generate rhythms that fit the human performance. However, it cannot respond to sudden changes of human rhythm because the agent does not jump over the neighborhoods. In order to catch up with the human performance when sudden change is occurred, an external force should be employed in the agent behavior model to allow a jump on the map. It means a sort of changing motif or restart of the virtual player.

4. CONCLUSION

We proposed a music interaction system between human and machine rhythm performance, which is composed of rhythm maps with an association mechanism. We also adopted a 3D acceleration sensor as an interface for human performance. For system evaluation, we conducted some experiments with examinees including both musically experienced and inexperienced people. We received comments about our system such as; “It is easy to play music even for inexperienced people,” “Users can feel creative interaction with computer after some practice”.

From the results, we gathered that users can feel a creative rhythm interaction with our system. Moreover, a non-unique reaction to a human performance was achieved with a many-to-many association and with the dynamic behavior of the agent. Users could also perform with unrestricted actions by using the 3D acceleration sensor.