The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Human-Robot Ensemble between Robot Thereminist and Human Percussionist using Coupled Oscillator Model

Takeshi Mizumoto, Takuma Otsuka, Kazuhiro Nakadai,
Toru Takahashi, Kazunori Komatani, Tetsuya Ogata and Hiroshi G. Okuno

Abstract— **This paper presents a novel synchronizing method for a human-robot ensemble using coupled oscillators. We define an ensemble as a synchronized performance produced through interactions between independent players. To attain better synchronized performance, the robot should predict the human's behavior to reduce the difference between the human's and robot's onset timings. Existing studies in such synchronization only adapts to onset intervals, thus, need a considerable time to synchronize. We use a coupled oscillator model to predict the human's behavior. Experimental results show that our method reduces the average of onset time errors; when we use a metronome, a tempo-varying metronome or a human drummer, errors are reduced by 38%, 10% or 14% on the average, respectively. These results mean that the prediction of human's behaviors is effective for the synchronized performance.**

I. Introduction

Studies on music robots that play the instrument have started from a keyboardist robot WABOT-2 in 1980's [1]. Recently, the music robots are studied from two motivations. The first is *to develop more sophisticated music robots* such as a flutist robot [2] and a saxophonist robot [3]. These studies focused on developing a music robot as a performer. The second one is *to develop an ensemble between humans and robots* such as a duet ensemble between a flutist robot and a human saxophonist [4] or a quartet ensemble between a robot drummer, a robot marimba player, a human keyboardist and a human drummer [5]. These studies focused on achieving an interaction through music. Music robots that are capable of playing in an ensemble with humans are expected to provide an interactive entertainment. In addition, we believe that ensembles have the potential of providing an entertainment beyond linguistic or cultural differences since music is almost independent of particular languages.

We aim to achieve an ensemble between humans and robots, especially, we focus on a duet ensemble between a robot thereminist and a human drummer, which is one of the simplest forms of an ensemble. We define an ensemble as "a synchronized performance produced through interactions between independent players." According to our definition, the ensemble consists of three components (1) a music-playing human, (2) a music-playing robot, and (3) a synchronization method between players. We need to develop the second and third components. Mizumoto *et al.* have developed a robot thereminist [6], which corresponds to the second component, and Otsuka *et al.* [7] developed a simple ensemble method using beat-tracking method [8] (hereafter, this ensemble robot is referred to as "*the Robot Thereminist*"), which corresponds to the third component. The Robot Thereminist is required to predict the human drummer's onset time to achieve a synchronized playing. However, existing ensemble systems such as the Robot Thereminist only adapt to a interval of the human drummer's onsets, instead of using the onset timing itself, which fact leads the limitation of the synchronization accuracy.

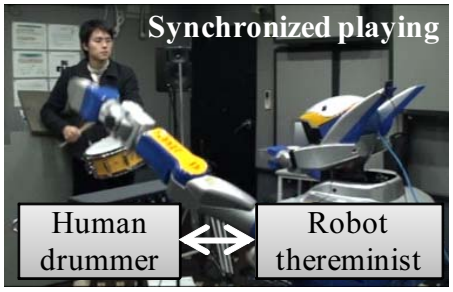


Fig. 1. Snapshot of our ensemble system

This paper presents a novel synchronization method using a coupled-oscillator model which achieves the capability of predicting the human drummer's onset timing. The behavior of coupled oscillators have been actively studied [9] from theoretical analysis to applications to explain the behaviors of various physical phenomena, e.g., frogs' calling behavior [10]. When we assume that each participant can be represented as a self-sustaining oscillator and their interactions are also represented as a coupling of oscillators, we can apply the concepts of coupled oscillators to an ensemble. Based on this idea, our synchronization method reduces the difference of two participant's onset times compared with the existing Robot Thereminist, which only adjusts the robot's playing speed according to the human's drumming speed, because the robot predicts a time of the human's drum hitting as the time of the oscillator's phase becomes zero. The robot can synchronize with the human drummer more precisely by changing the theremin's pitch on time.

Our approach has two main advantages: (1) A robot with the model can predict another participant's behavior, which is essential for synchronized motion generation, and (2) our

T. Mizumoto, T. Otsuka, T. Takahashi, K. Komatani, T. Ogata and H. G. Okuno are with Graduate School of Informatics, Kyoto University, Sakyo, Kyoto 606-8501, Japan { mizumoto, ohtsuka, tall, komatani, ogata, okuno } @kuis.kyoto-u.ac.jp

K. Nakadai is with Honda Research Institute Japan, Co., Ltd., Wako, Saitama, 351-0114, Japan, and also with Graduate School of Information Science and Engineering, Tokyo Institute of Technology. nakadai@jp.honda-ri.com

model can be applied to various ensemble situations merely by changing the parameters such as coupling strengths, which accomplishes, for example, scalability for a number of participants.

This paper is organized as follows: Section II describes state-of-the-art human-robot ensemble studies. Section III presents a novel ensemble system, including the music-playing robots and a coupled oscillator model for the ensemble. Section IV presents the experimental results with a metronome for one-way interaction and with a human for actual interactions. Finally, Section V concludes this paper.

II. State-of-the-art Ensemble Studies

To describe an overview of studies on ensembles, we will start from human-computer ensembles. Dannenberg's real-time accompaniment method [11] is the first work on human-computer ensembles. We have categorized these studies into two types: (1) a human leads an ensemble and a computer follows it and (2) humans and computers play an equal role. Studies on ensembles began with the first approach. Raphael proposed a probabilistic approach [12] and Simon *et al.* proposed a method of code generation based on a hidden-Markov model [13]. In the second approach, Goto *et al.* proposed a jazz system whose participants play the instrument by interacting with one another [14]. In their approach, the participants in the ensemble plays the same role, which is similar to our purpose in this paper.

The most significant difference between human-computer and -robot ensembles are embodiment, i.e., a physical body. We believe that a physical body is important for the presence as a participant, and some investigations of human supports the belief of the importance of the embodiment in ensemble, e.g., a singer's face influences an audience's judgement of the singer's emotion [15], and a pianist's playing motion has a correlation of the score that is being played [16].

We will now describe the three main related works on human-robot ensembles, which is categorized into two types: score-based and improvisational. Petersen *et al.* presented an ensemble system with a robotic flutist and a human saxophonist using a score [4]. The robot and the human played melodies alternately, instead of playing simultaneously. As our goal is to achieve a synchronized performance, we need participants to play their instruments at the same time. Otsuka *et al.* developed an ensemble system with the Robot Thereminist and a human drummer [7]. The robot changed its playing speed according to the intervals of the beat in the human's playing. They ignored the perspective of prediction, which is essential for synchronized playing. The prediction of other participants is essential because a robot needs to generate playing motions *on time*. The strategy of generating a motion after a perception is insufficient such a real-time task. Weinberg *et al.* proposed an ensemble system with two humans and two robots: robotic drum and marimba players, and human drum and keyboard players [5]. They achieved a simultaneous and improvisational performance with multiple-humans and multiple-robots. The robots played the instruments according to the human's playing, for example, they played a similar melodies or rhythms by transforming them stochastically, i.e., the robots play the similar melody to that played by the humans. Therefore, their approach is insufficient to realize the ensemble in which each player plays different melodies.

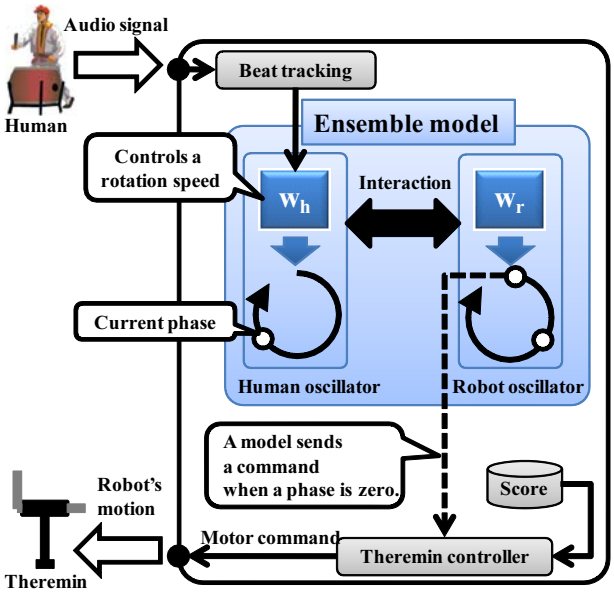


Fig. 2. Block diagram of duet ensemble system

To achieve a natural ensemble like with human players, these two types of ensembles should be combined. Players should be able to play a score, and make improvisational arrangements during the ensemble. For example, each participant would play his own musical score, and each player's speed would get slow down when other people played at slower speeds. Moreover, when an other participant played different melodies, they would change their melody through improvisation as in jazz. To achieve such an ensemble, our robot plays a given score and changes its playing speed by predicting human's onset timings.

III. Ensemble System Using Coupled Oscillators

This section describes a novel coupled-oscillator-model-based ensemble system between a robot thereminist and a human drummer. First, we present an overview of our system in Section III-A. Then, we explain its three main components: a real-time beat-tracking method that recognizes a human's drumming speed in Section III-B, a robot thereminist in Section III-C, and a coupled-oscillator model for synchronized performance in Section III-D.

A. Overview of our ensemble system

Figure 2 is an overview of our system, which consists of three main modules: (1) a beat-tracking module for estimating the onset of a human's playing, (2) a robot-control module for playing music, and (3) an ensemble model for predicting the human's behavior.

Our system works as follows: it records the sound of a human playing through its own microphone. Then, the beat-

tracking module estimates the beat interval in the sound of the human's playing. Our model updates the angular velocity of the human's oscillator model using the estimated interval. This model is used to simulate and predict the human's behavior. The robot waits until the phase of the robot's oscillator becomes zero by updating the human's and the robot's oscillators. When the phase becomes zero, our model commands the theremin controller to play the next musical note from a given score.

B. Real-time beat tracking

This beat-tracking algorithm has three phases: (1) estimating the tempo, (2) detecting the beat, and (3) predicting the beat time. The input is a musical signal of the human's performance.

1) Tempo estimation: Let $P(t,f)$ be the mel-scale power spectrogram of the given musical signal where t is the time index and f is the mel-filter bank bin. We use 64 banks, therefore $f = 0, 1, ..., 63$. Then, Sobel filtering is applied to $P(t,f)$ and the onset belief, $d_{inc}(t,f)$, is derived.

$$d_{inc}(t,f) = \begin{cases} d(t,f) & \text{if } d(t,f) > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

$$\begin{aligned} d(t,f) = & -P(t-1,f+1) + P(t+1,f+1) \\ & -2P(t-1,f) + 2P(t+1,f) \\ & -P(t-1,f-1) + P(t+1,f-1), \end{aligned} \quad (2)$$

where $f = 1, 2, ..., 62$. Equation (2) means the Sobel filter well-known in image processing.

The tempo is defined as the interval between two neighboring beats. This is estimated through normalized cross correlation (NCC) as Eq. (3).

$$R(t,i) = \frac{\sum_{f=1}^{62}\sum_{k=0}^{W-1} d_{inc}(t-k,f) d_{inc}(t-i-k,f)}{\sqrt{\sum_{f=1}^{62}\sum_{k=0}^{W-1} d_{inc}(t-k,f)^2 \cdot \sum_{f=1}^{62}\sum_{k=0}^{W-1} d_{inc}(t-i-k,f)^2}}, \quad (3)$$

where W is the window length for estimating the tempo and i is shift offset. The W is set to 3 [sec]. To stabilize the estimation of tempo, the local peak of $R(t,i)$ is derived as

$$R_p(t,i) = \begin{cases} R(t,i) & \text{if } R(t,i-1) < R(t,i) \\ & \quad \text{and } R(t,i+1) < R(t,i) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

For each time t, beat interval $I(t)$ is determined based on $R_p(t,i)$ in Eq. (4). The beat interval is an inverse value of the musical tempo. Basically, $I(t)$ is chosen as $I(t) = \underset{i}{\operatorname{argmax}} R_p(t,i)$. However, when a complicated drumming pattern is performed in the musical signal, the estimated tempo will fluctuate rapidly.

To prevent the beat interval from being misestimated, $I(t)$ is derived in Eq. (5). Let I_1 and I_2 be the first and second peaks in $R_p(t,i)$ when moving i.

$$I(t) = \begin{cases} 2\|I_1 - I_2\| & \text{if } (\|I_{n2} - I_1\| < \delta \\ & \quad \text{or } \|I_{n2} - I_2\| < \delta) \\ 3\|I_1 - I_2\| & \text{if } (\|I_{n3} - I_1\| < \delta \\ & \quad \text{or } \|I_{n3} - I_2\| < \delta) \\ I_1 & \text{otherwise,} \end{cases} \quad (5)$$

where $I_{n2} = 2\|I_1 - I_2\|$ and $I_{n3} = 3\|I_1 - I_2\|$. Here, δ is an error-margin parameter.

Beat interval $I(t)$ is confined to a range between 61 – 120 beats per minute (bpm). This is because this range is suitable for controlling the robot's arm.

2) Beat detection: Each beat time is estimated using onset belief $d_{inc}(t,f)$ and beat interval $I(t)$. Two kinds of beat reliabilities are defined: the reliability of the neighboring beat and that of the continuous beat. Neighboring-beat reliability $S_n(t,i)$ defined in Eq. (6) is the belief that the adjacent beat lies in the $I(t)$ interval.

$$S_n(t,i) = \begin{cases} \sum_{f=1}^{62} (d_{inc}(t-i,f) + d_{inc}(t-i-I(t),f)) \\ \qquad \text{if } (i \le I(t)), \\ 0 \\ \qquad \text{if } (i > I(t)) \end{cases} \quad (6)$$

Continuous-beat reliability $S_c(t,i)$ defined in Eq. (7) is the belief that the sequence of musical beats lies in the estimated beat intervals.

$$S_c(t,i) = \sum_{m=0}^{N_{beats}} S_n(T_p(t,m),i), \quad (7)$$

$$T_p(t,m) = \begin{cases} t - I(t) & \text{if } m = 0, \\ T_p(t,m-1) - I(T_p(t,m)) & \text{if } m \ge 1, \end{cases}$$

where $T_p(t,m)$ is the m-th previous beat time at time t, and N_{beats} is the number of beats used to calculate continuous-beat reliability.

These two reliabilities are then integrated into the beat reliability $S(t)$ as

$$S(t) = \sum_i S_n(t-i,i) \cdot S_c(t-i,i). \quad (8)$$

The latest beat time, $T(n+1)$, is one of the peaks in $S(t)$ that is the closest to $T(n) + I(t)$, where $T(n)$ is the n-th beat time.

3) Prediction of beat time: Predicted beat time T' is obtained by extrapolation using latest beat time $T(n)$ and current beat interval $I(t)$.

$$T' = \begin{cases} T_{tmp} & \text{if } T_{tmp} \ge \frac{3}{2} I(t) + t, \\ T_{tmp} + I(t) & \text{otherwise,} \end{cases} \quad (9)$$

$$\begin{aligned} T_{tmp} = & \; T(n) + I(t) + (t - T(n)) \\ & - \{(t - T(m)) \mod I(t)\} \end{aligned} \quad (10)$$

C. Method of controlling the Thereminist Robot

We used the robot thereminist developed by Mizumoto *et al.* [6] as a participant in an ensemble. It had a portable robot-control system because its model-based control was

designed to make the system independent of a particular hardware. This is an important feature for an ensemble system because we can easily extend it to a multiple-robot ensemble system by only importing the system to new robots. Actually, Mizumoto *et al.* [6] implements the control system on two different robots.

One approach to controlling robot motion is parametric-model-based feedforward control. A robot estimates a set of model parameters that describes the relationship between its arm and the theremin's pitch, i.e., pitch characteristics. Then, the robot starts moving by using an inverse model. A system with this approach works in two phases: a calibration and a performance phase. The robot in the calibration phase moves its arm and records the theremin's sound. Then, the parameters of the model are estimated using a Levenberg-Marquardt method, which is a method of nonlinear optimization [17].

The model of the theremin's pitch characteristics, M_p, is:

$$\hat{p} = M_p(x_p; \boldsymbol{\theta}) = \frac{\theta_2}{(\theta_0 - x_p)^{\theta_1}} + \theta_3, \tag{11}$$

where x_p denotes the robot's arm position, $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3)$ denotes the model parameters, and $\hat{p}$ denotes the estimated pitch with M_p. The θ_3 means the theremin's pitch when the robot's arm is far enough away from the theremin's antenna. When the robot's arm moves closer to the antenna, the theremin's pitch increases. The first term in Eq. 11 denotes how the pitch increases.

We can obtain an inverse model analytically as:

$$\hat{x}_p = M_p^{-1}(p, \boldsymbol{\theta}) = \theta_0 - \left(\frac{\theta_2}{p - \theta_3}\right)^{1/\theta_1}, \tag{12}$$

After the model parameters have been estimated, the robot plays a melody according to a score. We define a score as a sequence of two values: (1) the name of the note, e.g., C3♯ and D4♭ and (2) the duration of the note. Let p_i and d_i be the name of the i-th note and duration, and N be the number of notes to be played. The robot converts a note to a target pitch in hertz using an equal temperament in the performance phase. Then, it converts the target pitch to a target arm position using an inverse model. The robot estimates its parameters and plays a given score by following this procedure.

D. Coupled-oscillator model for synchronized ensemble

This section describes coupled oscillators and its application to the task model for a synchronized ensemble, which we call the ensemble model, hereafter. First, we explain the oscillator model in Section III-D.1 and its application to the ensemble system in Section III-D.2.

The key advantage is that the robot which has the model can know the time when a human hits the drum through the model. Due to this advantage, we can reduce the time delay of the robot's motion because it can move its arm on time without waiting for a human's drum onset.

1) General description of coupled-oscillator model: A coupled-oscillator model consists of two components: an oscillator and its interactions. The oscillator is a self-sustaining system, which keeps working repeatedly by itself. For example, a pendulum clock and a drummer who maintains the same speed can be considered to be oscillators. We can define an oscillator's phase $\phi(t)$ with

$$\phi(t) = (\phi_0 + 2\pi t / T_{osc}) \bmod 2\pi, \tag{13}$$

where t denotes the time, T_{osc} denotes the period of the oscillator, and ϕ_0 denotes the initial phase. $\phi(t) = 2\pi n$ denotes the same state in an oscillator. The oscillator's dynamics is described by the differential equation of its phase:

$$\frac{d\phi_1}{dt} = \omega_1, \tag{14}$$

where ω_1 denotes an angular frequency of the oscillator. When two oscillators interact, we call them coupled. A coupling is represented by adding a 2π-periodic function to Eq. 14. We show a coupled two oscillators below:

$$\frac{d\phi_1}{dt} = \omega_1 + K_1 Q(\phi_2 - \phi_1) \quad \text{and} \tag{15}$$

$$\frac{d\phi_2}{dt} = \omega_2 + K_2 Q(\phi_1 - \phi_2), \tag{16}$$

where ϕ_1 and ϕ_2 are the phases of the coupled oscillators, the Q is a coupling term which is a 2π-periodic function of the phase difference, K_1 and K_2 are positive coupling strengths, and ω_1 and ω_2 are natural frequencies.

We present the Kuramoto model, which is a basic oscillator model [18].

$$\frac{d\phi_1}{dt} = \omega_1 + K_1 \sin(\phi_2 - \phi_1) \quad \text{and} \tag{17}$$

$$\frac{d\phi_2}{dt} = \omega_2 + K_2 \sin(\phi_1 - \phi_2). \tag{18}$$

The key feature of this model is that a sinusoidal function is used as a coupling term. We can hence analyze the behavior of these two oscillators. First, we define the phase difference, $\phi = \phi_1 - \phi_2$. Then, the dynamics of ϕ is described as:

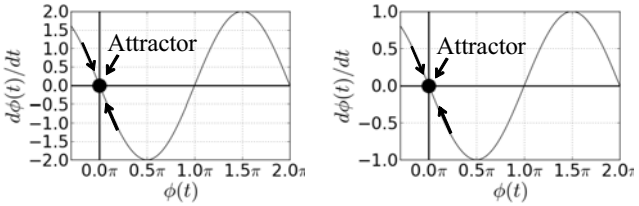
$$\frac{d\phi}{dt} = \omega_1 - \omega_2 + K_1 \sin(-\phi) - K_2 \sin(\phi) \tag{19}$$

$$= (\omega_1 - \omega_2) - (K_1 + K_2) \sin(\phi) \tag{20}$$

Assuming the natural frequencies of two oscillators are the same, we can determine the behavior of them by plotting a graph of Eq. 20.

Figure 3 plots the behaviors of the two oscillators with two parameters, K_1 and K_2. The vertical axis denotes the differential coefficient of the phase difference and the horizontal axis denotes the phase difference.

Figure 3 (a) plots the situation when two oscillators are coupled equally ($K_1 = K_2 = 1$). In this situation, two oscillators are synchronized when the phase difference is zero. Figure 3 (b) shows that even if only the second oscillator is influenced ($K_1 = 0, K_2 = 1$), the attractor is at the same place.



(a) $K_1 = K_2 = 1$ (b) $K_1 = 0, K_2 = 1$

Fig. 3. Attractors in Kuramoto model

2) Application to ensemble system: We use two assumptions to apply the oscillator model to an ensemble system.

1) A participant has an internal oscillator. He plays music so that when the phase is zero, there is an onset in the played sound. For example, a drummer hits a drum when his internal oscillator's phase is zero.
2) A participant knows the other's phase when its onset begins. For example, the partner of the drummer knows the drummer's phase when he hits the drum.

We focus on a duet ensemble between a human drummer and the Robot Thereminist. We then define the rule for a onset timing, i.e., the time when a phase becomes zero, as follows. For a drum sound, it is the onset when the drum is hit. The onset timings for the theremin playing is defined as the time when the pitch is changed. Calculating the theremin's onset timings from the pitch trajectory is not trivial because the trajectory has continuous value unlike a piano. We hence rounded the trajectory down to the nearest 100 [cent] to emphasize the onset timings.

We use the Kuramoto model in Eqs. 17 and 18 as the oscillator model. In addition, we add an update rule to reduce the robot's natural frequency to that of a human. This is because a drum usually dominates the rhythm of an ensemble.

Our ensemble model is summarized as follows:

The phase dynamics of two oscillators are:

$$\frac{d\phi_h}{dt} = \omega_h + K_h \sin(\phi_r - \phi_h) and \quad (21)$$

$$\frac{d\phi_r}{dt} = \omega_r + K_r \sin(\phi_h - \phi_r). \quad (22)$$

The update rule of ω_r is:

$$\omega_r \leftarrow \omega_r + \mu(\omega_r - \omega_h), \quad (23)$$

where ϕ_r and ϕ_h denote the robot's and human's phases, K_r and K_h denote the robot's and human's coupling strength, and μ denotes a learning coefficient.

IV. Experiments

We evaluate how accurately a robot plays the theremin with a partner using three experiments. The ensemble partner is different for each experiments: (1) a metronome as a completely accurate drummer, (2) a tempo-varying metronome that simulates a drummer with fluctuation without interactions (3) a human drummer which has fluctuations and interactions. Note that we evaluate only the onset error of the total ensemble, because estimating the onsets of a drum is easy for the beat tracking, which is already evaluated by K. Murata *et al.* [8].

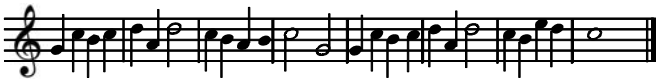


Fig. 4. Score of Aura Lee

A. Configurations

For all three experiments, we use a humanoid robot, HRP-2, as the platform for the robot thereminist, and Etherwave Theremin of Moog Music as the instrument. The distance between the robot and theremin is 50 cm. We use an American folk song "Aura Lee" as the music to play. Figure 4 has the score for the song. We empirically set four model parameters of four the oscillator model as: K_r = 0.4, $\omega_h = \omega_r = 2\pi/700$ and, $\mu = 0.01$. ϕ_r, ϕ_h and ω_r are updated at interval of 50 msec. We set $K_h = 0$ in the first and second experiments because the metronome never be influenced by the human drumming. In contrast, we empirically set $K_h = 0.4$ in the third experiment because the human drummer is influenced by the Robot Thereminist. The value of $K_h = 0.4$ is the same as K_r, which means that the human and the robot is influenced by each other with the same strength.

We compare the efficiency of our oscillator-model-based ensemble with the Robot Thereminist, which is a baseline method. The robot with the baseline method adapts its playing speed according to the estimated onset intervals of the beat-tracking.

We use four different metronome tempi: 66, 80, 100 and 112 bpm. These tempi covers the possible speed of the beat tracking method. Three trials are conducted for each speed. We then evaluate with a mean onset error which is the time difference between the theremin and the drum onset. The mean onset error is defined as follows:

$$error = \frac{1}{N}\sum_{j=1}^{N} \min_{i=1,\ldots,M} \left|onset_t(i) - onset_d(j)\right|, \quad (24)$$

where N denotes the number of drum onsets, M denotes the number of theremin onsets, and $onset_t(i)$ and $onset_d(j)$ denote the theremin's i-th and the drum's j-th onset, respectively.

B. Experiment 1: Ensemble with a metronome

As the first experiment, we evaluate how a robot plays a music according to a metronome, which is the "perfect" drummer, instead of using the real human drummer.

Figure 5 shows the result. The horizontal axis denotes the tempo of the metronome. The vertical axis denotes the mean onset error. The red bars denote the errors with our method, the black bars denote those of the baseline method and the white bars denote the worst error that could happen, i.e., the half of a beat interval. The heights of the white bars decrease when a tempo gets faster because the beat interval shortens.

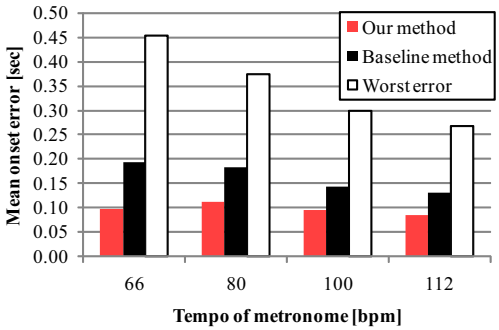


Fig. 5. Results 1: Onset error with various tempi

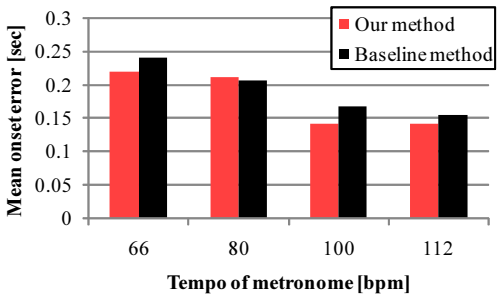


Fig. 6. Results 2: Onset error with various fluctuating tempi

According to the result, our method reduced onset error than the baseline method by 39%, which means that our method realized to play a melody which matches with the human's onset more accurately.

We also evaluate onset errors in a time difference between onset times of the metronome and the human drummer as an upper limit. The mean onset error for the human is around 1 msec. Since the mean onset error of our method is around 10 msec, there is room for improvement.

C. Experiment 2: Ensemble with a time-varying metronome

We add tempo fluctuations to the metronome in this experiment to simulate a human's drumming without interactions. The standard deviation of the tempo fluctuation is 10% of the mean value.

Figure 6 shows a bar char of the results. The vertical and horizontal axes have the same meaning as in Figure 5. The results reveal that our method worsen in performance than the first experiment although our method performs still better than the baseline method.

This performance degradation is caused by the oscillator's fast adaptation to the fluctuated tempo. The model predicts the onset timing more stable when we set smaller the learning coefficient, μ, however, the small μ slows the adaptation speed down. We need more investigation of μ to analyze this trade-off.

We also evaluated the onset error using a human drummer. The mean onset error was 120 msec. Although the error is better than our result, the error increased than that of Experiment 1. The result suggests that the task of rhythm synchronization with a fluctuated drummer is too difficult to our method than the no-fluctuated drummer because our model assumes that a drummer hits at a almost same interval. Therefore, the difference between our method and baseline was small decreased because the difficulty of the task.

D. Experiment 3: Ensemble with human

We evaluate how our robot thereminist synchronizes to the human drummer using the average of time differences between the theremin's and the drum's onset times. This experiment is conducted as follows. In prior to each trial, the human listens to a sound of metronome as an initial tempo. Then, he starts drumming according to the sound. When the robot starts playing, the metronome is stopped in order to ensure that the interaction is based on the sounds being played by the human and the robot.

TABLE I

RESULT 3: THE MAXIMUM AND MINIMUM ERROR FOR EACH CONDITION

	Maximum error		Minimum error	
Condition	Proposed	Baseline	Proposed	Baseline
66 bpm	0.376	0.431	-0.426	-0.451
80 bpm	0.350	0.400	-0.361	-0.386
100 bpm	0.309	0.312	-0.294	-0.308
112 bpm	0.238	0.254	-0.280	-0.284

Fig. 7 shows the result. The vertical axis denotes the mean onset error and the horizontal axis denotes the initial tempo. The results reveal that our method reduces the onset error by 14% on the average. When the initial tempo is slow, 66 and 80 [bpm], our method reduced the onset error by 20%. In contrast, when the initial tempo is fast, our method reduced the onset error by 8%.

Table I shows the maximum and minimum errors of the our method and the baseline method for each tempo. The table shows that our method reduces also the maximum and minimum onset error. The improvements of these errors decrease with the initial tempo increase.

We discuss the reason of the tendency of improvement degradation which depends on the initial tempo, which is commonly seen in the mean, maximum, and minimum onset errors. When the tempo is fast, the human's tempo fluctuation decreases because the beat interval is shorten. Also, the human easily keep the tempo constant when the tempo is fast because the drumming motion is more rhythmic. Thus, the difference of performance between the baseline and our prediction methods decreased. On the other hand, when the tempo is slow, the human influenced by the motion and the sound of the Robot Thereminist. Therefore, our method that simulates the human's interactions works well.

V. CONCLUSION AND FUTURE WORKS

We presented a novel model for a human-robot ensemble using a coupled-oscillator model that made a robot capable of predicting a human's behavior. We used the Kuramoto model, which is a basic coupled-oscillator model and added an update method of a natural frequency to give the robot the capability of adaptation to the human's playing speed. We implemented the ensemble system between a human drummer and a robot theremin player. The robot

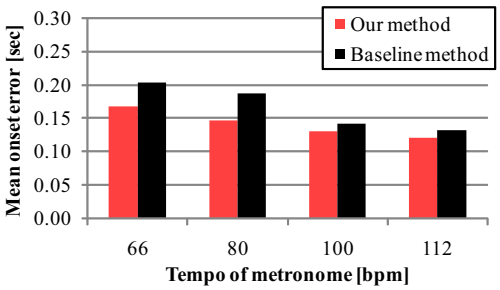


Fig. 7. Results 3: Onset error with various initial tempi

predicted the human's drumming time and the robot played the theremin according to the prediction. The experimental results revealed that our system could reduce onset error more than a method that only adapted to the theremin's playing speed.

As future work, we need to evaluate and discuss our ensemble model more strictly by comparing with the observations of a human-human ensemble since we only evaluated the onset errors of robot-human ensemble in this paper. We have also three research projects planned for the future. First, we should extend our ensemble model, for example, it may be more suitable to use a relaxation oscillator, whose oscillations emit spikes like a drum sound. Second, we need extend our ensemble system into multiple-robots and multiple-humans to evaluate our model's scalability. Third, we need to develop a visual-cue recognition system, e.g., one that can identify gestures. This is important because an ensemble involves multi-modal interactions. When we add visual information to a system, we need to consider how to integrate audio and visual cues.

ACKNOWLEDGEMENTS

We wish to thank Ikkyu Aihara for his helpful discussion on oscillator models. This study was partially supported by a Grant-in-Aid for Scientific Research (S) (No. 19100003), a Grant-in-Aid for Scientific Research on Innovative Areas (No. 22118502), and the Global COE Program.

REFERENCES

[1] S. Sugano and I. Kato. WABOT-2: Autonomous robot with dexterous finger-arm – finger-arm coordination control in keyboard performance –. In *Proc. of ICRA*, pages 90–97, 1987.

[2] J. Solis, K. Taniguchi, T. Ninomiya, T. Yamamoto, and A. Takahashi. Development of Waseda flutist robot WF-4RIV: Implementation of auditory feedback system. In *Proc. of ICRA*, pages 3654–3659, 2008.

[3] J. Solis, K. Petersen, T. Ninomiya, M. Takeuchi, and A. Takanishi. Development of anthropomorphic musical performance robots: From understanding the nature of music performance to its application to entertainment robotics. In *Proc. of IROS*, pages 2309–2314, 2009.

[4] K. Petersen, J. Solis, and A. Takanishi. Development of a aural real-time rhythmical and harmonic tracking to enable the musical interaction with the waseda flutist robot. In *Proc. of IROS*, pages 2303–2308, 2009.

[5] G. Weinberg, B. Blosser, T. Mallikarjuna, and A. Raman. The creation of a multi-human, multi-robot interactive jam session. In *Proc. of NIME*, pages 70–73, 2009.

[6] T. Mizumoto, H. Tsujino, T. Takahashi, T. Ogata, and H. G. Okuno. Thereminist robot: Development of a robot theremin player with feedforward and feedback arm control based on a theremin's pitch model. In *Proc. of IROS*, pages 2297–2302, 2009.

[7] T. Otsuka, T. Mizumoto, K. Nakadai, T. Takahashi, K. Komatani, T. Ogata, and H. G. Okuno. Music-ensemble robot that is capable of playing the theremin while listening to the accompanied music. In *Proc. of IEA/AIE*, pages 102–112, 2010.

[8] K. Murata, K. Nakadai, K. Yoshii, R. Takeda, T. Torii, and H. G. Okuno. A robot uses its own microphone to synchronize its steps to musical beats while scatting and singing. In *Proc. of IROS*, pages 2459–2464, 2008.

[9] S. H. Strogatz. *SYNC: The Emerging Science of Spontaneous Order*. Hyperion, 2003.

[10] I. Aihara. Modeling synchronized calling behavior of japanese tree frogs. *Phys. Rev. E*, 8:011918–011925, 2009.

[11] R. B. Dannenberg. An on-line algorithm for real-time accompaniment. In *Proc. of ICMC*, pages 193–198, 1984.

[12] C. Raphael. A probabilistic expert system for automatic musical accompaniment. *Journal of Computational and Graphical Statistics*, 10(3):487–512, 2001.

[13] I. Simon, D. Morris, and S. Basu. MySong: Automatic accompaniment generation for vocal melodies. In *Proc. of ACM CHI*, pages 725–734, 2008.

[14] M. Goto, I. Hidaka, H. Matsumoto, Y. Kuroda, and Y. Muraoka. A jazz session system for interplay among all players – VirJa session (virtual jazz session system). In *Proc. of ICMC*, pages 346–349, 1996.

[15] W. F. Thompson, F. A. Russo, and L. Quinto. Audio-visual integration of emotional cues in song. *Cognition and Emotion*, 22(8):1457–1470, 2008.

[16] J. Mac Ritchie, B. Buck, and N. J. Bailey. Visualizing musical structure through performance gesture. In *Proc. of ISMIR*, pages 237–242, 2009.

[17] D.W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11(2):431–441, 1963.

[18] Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Dover Publications, 2003.