Effect of Dynamic Local Lag Control with Dynamic Control of Prediction Time in Joint Haptic Drum Performance

Mya Sithu, Yutaka Ishibashi, and Norishige Fukushima
Graduate School of Engineering, Nagoya Institute of Technology, Nagoya 466-8555, Japan
mya.sithu@mcl.nitech.a.cjp, ishibasi@nitech.ac.jp, fukushima@nitech.ac.jp

Abstract

In this paper, we handle joint performance of a networked haptic drum system where two users play a drum set in a 3D virtual space with the same rhythm at the same tempo. In order to improve the interactivity, we propose the dynamic local lag control with dynamic control of prediction time. The proposed control keeps the synchronization quality of sound and interactivity high. It also changes the local lag and prediction time dynamically according to the network delay. As a result, we illustrate that the control is effective.

1. Introduction

Recently, a number of researchers have been paying attention to musical performance in networked haptic environments [1]-[5]. When users at different places make the music performance together by using haptic interface devices over a QoS (Quality of Service) [6] non-guaranteed network like the Internet, the synchronization quality of sound and interactivity may seriously be deteriorated owing to the network delay [5].

In [5], the authors handle a networked haptic drum system for joint musical performance. In the system, two users at different places can play a drum set together with the same rhythm at the same tempo by using their independent haptic interface devices. The authors carry out subjective QoE (Quality of Experience) [7] assessment to investigate the influence of network delay on the synchronization quality of sound. As a result, they illustrate that the synchronization quality of sound in joint musical performance deteriorates more largely as the network delay becomes larger.

To achieve high synchronization quality of sound in joint musical performance, we can use local lag control [8], [9]. The control buffers local information for a constant time called the local lag in order to be synchronized with received information; thus, it degrades the interactivity. In [4], Irie *et al.* use the local lag control for a networked ensemble between two terminals. They set the local lag to the same value as the network delay from the local terminal to the other terminal. Therefore, the interactivity may be degraded when the network delay is large. Also, they assume that the network delay from the local terminal to the other terminal is equal to that in the opposite direction (called the *symmetric delay case* in this paper). Usually, in a network like the Internet, the network delay from the local terminal to the other terminal is different from that in the opposite direction (called the *asymmetric delay case*). In this case, high synchronization quality of sound may not be achieved.

In order to solve the above problem, the authors propose the dynamic local lag control, which dynamically changes the local lag according to the network delay from the other terminal to the local terminal, in [10]. To clarify the effect of the dynamic local lag control, they make a comparison between the dynamic local lag control and the local lag control with fixed values of local lag by carrying out subjective QoE assessment in the joint performance of the networked haptic drum system. As a result, they illustrate that the dynamic local lag control is effective. However, since the dynamic local lag control outputs not only information received from the other terminal but also information of the local terminal after buffering, the interactivity slightly deteriorates.

Prediction is one of the methods which can improve the interactivity [11]-[14]. In [12], the authors propose a group synchronization control scheme with prediction to keep the interactivity high for haptic work. They investigate the effect of the scheme in two types of work (remote haptic drawing instruction and collaborative haptic play with building blocks) by QoE assessment. As a result, they find that there is the optimum value of prediction time according to the network delay and the type of work.

In [14], the authors investigate the effect of the dynamic local lag control with prediction in the joint performance of the networked haptic drum system. As a result, they demonstrate that there exits the optimum value of prediction time according to the network delay.

Based on the above results in [14], we here propose dynamic local lag control with dynamic control of prediction time which dynamically changes the prediction time according to the network delay in the joint performance of the networked haptic drum system [5], where two users play a drum set with the same rhythm at the same tempo. We also make a comparison between the dynamic local lag control with dynamic control of prediction time and that with fixed value of prediction time on QoE.

The remainder of this paper is organized as follows. Section 2 describes the networked haptic drum system. The dynamic local lag control is outlined in Section 3. Section 4 explains the proposed control, and Section 5 describes our assessment environment. Assessment results are presented in Section 6, and Section 7 concludes the paper.

2. Networked Haptic Drum System

The configuration of the networked haptic drum system is shown in Fig. 1, where two users (*users 1 and 2*) share a drum set which consists of high-hat cymbals, a snare drum, a bass drum, and a floor tom in a 3D virtual space [5]. The system consists of two terminals (*terminals 1 and 2*), each of which has two PCs (*PC 1 and PC 2*) connected to each other through an Ethernet switching hub (100 Mbps). Each PC has a haptic interface device (PHANToM Omni [15]). The two haptic interface devices at each terminal are used to move a pair of drumsticks in the virtual space. Also, a display and a headset are connected to PC 1 at each terminal. When each drumstick hits a drum component, the reaction force is perceived through the haptic interface device, and a sound depending on the drum component is generated.

There are two types of reaction force in the joint performance of the networked haptic drum system. One is $m_d\boldsymbol{g}$ due to the gravity, where m_d denotes the mass of a held drumstick, and $\boldsymbol{g}$ is the gravitational acceleration. The other is produced when the user hits a drum component by a drumstick. The reaction force $\boldsymbol{F}$ against the drum component is applied to him/her through his/her haptic interface device. In this case, the reaction force is calculated by using the spring-damper model as follows: $\boldsymbol{F} = -(K_s\boldsymbol{x} + K_d\boldsymbol{v})$ where K_s is the spring coefficient, K_d is the damper coefficient, $\boldsymbol{x}$ is a vector from the contact point of the object (i.e., drum component) surface to the tip of the drumstick, and $\boldsymbol{v}$ is the velocity of the drumstick (see Fig. 2). In the joint performance of the networked haptic drum system, each user perceives the resultant force $m_d\boldsymbol{g} + \boldsymbol{F}$. In order to do the joint performance comfortably in the virtual space when the network delay is negligibly small, we set the mass of a held drumstick, the gravitational acceleration, the spring coefficient, and the damper coefficient to 100 g, 0.5 m/s^2, 0.2 N/mm, and 0.1 N/mm, respectively.

3. Dynamic Local Lag Control

The dynamic local lag control dynamically changes the local lag denoted by △ (≥0 ms) according to the network delay from the other terminal to the local terminal as shown in Fig. 3. We call the network delay from terminal 2 to terminal 1 *network delay 1* and that from terminal 1 to terminal 2 *network delay 2*. Also, we call the local lag at terminal 1 *local lag 1* and that at terminal 2 *local lag 2*. In Fig. 3, where we set local lag i (i = 1 or 2) to the same value as network delay i [16], we see that sound synchronization can be achieved at each terminal. The value of local lag △ is set to the optimum value of local lag calculated from the following equation [16]:

$$\triangle = 0.637\,D + 6.578, \quad (1)$$

where D is the time interval from the moment a *media unit* (MU) [17] is generated at the other terminal until the instant the MU is output at the local terminal. An MU is the information unit for media synchronization and includes the identification number (ID) of the user, the positional information of the PHANToM cursor, and the sequence number of the servo loop [18]. Note that Eq. (1) is somewhat different from the equation of the optimum value of local lag in [16], where D denotes the constant delay generated by a network emulator. To use D instead of the constant delay, we measured D in a preliminary experiment, and we found that D is equal to the constant delay plus 4.361 ms.

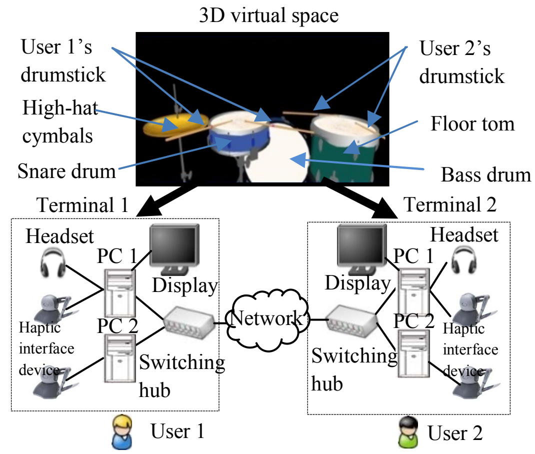

Figure 1. Configuration of networked haptic drum system.

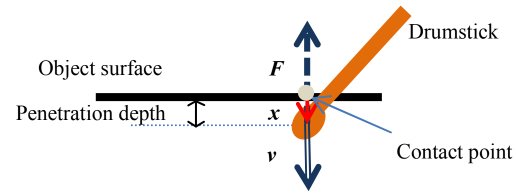

Figure 2. Reaction force when drum component is hitten.

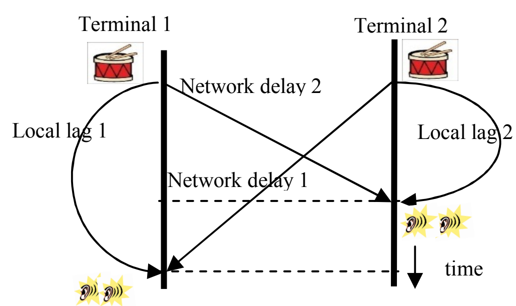

Figure 3. Synchronization of sound under dynamic local lag control.

4. Proposed Control

The dynamic local lag control with dynamic control of prediction time outputs the position information by predicting the future position later than the output time of an MU by the prediction time T_{predict} (≥0 ms) to keep the interactivity high. For simplicity, the first-order prediction [19] is used in this paper. The control also advances the output time of each MU at the local terminal by $T_{predict}$ ms. However, if there does not exist an MU which should be output after $T_{predict}$ ms, the MU is output by prediction.

In the first-order prediction, the predicted position $\boldsymbol{P}(t + T_{predict})$ is the position at the future time which is $T_{predict}$ ms later than the output time t of the last-received MU. The predicted position $\boldsymbol{P}(t + T_{predict})$ is calculated as follows:

$$\boldsymbol{P}(t + T_{predict}) = \boldsymbol{P}(t) + \{ \boldsymbol{P}(t) - \mathrm{P}(t - \delta)\}\, T_{predict}/\delta, \quad (2)$$

where δ is the difference in timestamp between succeeding two MUs, and $\boldsymbol{P}(t)$ is the position included in an MU whose output time is t. $\boldsymbol{P}(t - \delta)$ is the position included in an MU whose output time is $t - \delta$.

In the dynamic local lag control with dynamic control of prediction time, $T_{predict}$ is changed dynamically according to the network delay. This is because we found that there exists the optimum value of $T_{predict}$ according to the network delay in [14]. We investigated the relation between $T_{predict}$ and the network delay by regression analysis [20], and we obtained the following equation [14]:

$$T_{predict} = \max(0.007D^2 - 0.327D + 1.818, 0), \quad (3)$$

where $T_{predict}$ (≥ 0 ms) is the optimum value of prediction time; note that when $T_{predict}$ is smaller than 0 ms, $T_{predict}$ is set to 0 ms. The contribution rate adjusted for degrees of freedom [20], which shows goodness of fit with the estimated equation, was 0.952. Therefore, we can get $T_{predict}$ from D with a high degree of accuracy.

5. Assessment Environment

5.1. Assessment System

In our assessment system, two terminals are connected to each other via a network emulator (NIST Net [21]) which is used instead of the network shown in Fig. 1. The emulator generates an additional constant delay for each packet transmitted between the terminals. Note that the network delay jitter can be absolved by buffering under media synchronization control such as the Virtual-Time Rendering (VTR) algorithm [17]; we take account of the jitter by including the buffering time in the constant delay. In order to focus mainly on the dynamic control of prediction time, we here handle only the symmetric delay case; that is, we set the constant delay from terminal 2 to terminal 1 to the same value as that in the opposite direction.

5.2. Assessment Method

In the assessment, we made a comparison between the dynamic local lag control with dynamic control of prediction time and that with prediction, in which $T_{predict}$ is fixed. We carried out subjective QoE assessment with 16 subjects (males and females) whose ages were between 20 and 28. We employed the same two rhythms (*rhythms 1 and 2*) and two tempos (slow and fast tempos) as those in [5] to investigate the influence of drumstick movements. In rhythm 1, each subject hits the high-hat cymbals by his/her left drumstick and the snare drum by his/her right one repeatedly (see Fig. 4 (a)). The high-hat cymbals are hitten at all four beats, and the snare drum is done at the second and fourth beats. In rhythm 2, the subject plays the snare drum and floor-tom by his/her right drumstick at the second and third beats, respectively, while hitting the high-hat cymbals by his/her left drumstick at all times (see Fig. 4 (b)). In rhythm 1, because the subject hits the same drum components repeatedly, he/she does not need to move their drumsticks to the other drum components. On the other hand, in rhythm 2, he/she needs to move the right drumstick between the snare drum and the floor-tom. Rhythm 2 is more difficult than rhythm 1 since each subject need to move his/her right drumstick between the two different drum components. The rhythms are used in 8 beats or 4 beats rhythms and popular in jazz and rock music. As for the slow and fast tempos, each subject hits the drum set at 60 beats per minute (bpm) and 100 bpm, which are often used in fast ballads and slow rock music, respectively [22]. Here we employ only rhythm 1 at the slow tempo and rhythm 2 at the fast tempo because we found that the results of the other combinations of rhythm and tempo are almost the same as those of rhythm 1 at the slow tempo in [5].

Each pair of subjects practiced about two minutes under the condition that there was no constant delay and prediction time was set to 0 ms before the assessment of each combination of rhythm and tempo. The quality at this time was the standard in the assessment. The constant delay and the two types of control (i.e., the dynamic local lag control with dynamic control of prediction time and that with fixed value of prediction time) were presented in random order for the pair; the prediction time was also done in random order in the control where T_{predict} is fixed. We changed the constant delay from 0 ms to 150 ms at intervals of 50 ms in both types of control. In the control where T_{predict} is fixed, T_{predict} was changed from 0 ms to 110 ms at intervals of 10 ms. Note that when T_{predict} is 0 ms, the prediction is not used; that is, only the dynamic local lag control is carried out. In the dynamic local lag control with dynamic control of prediction time, T_{predict} was dynamically changed according to Eq. (2).

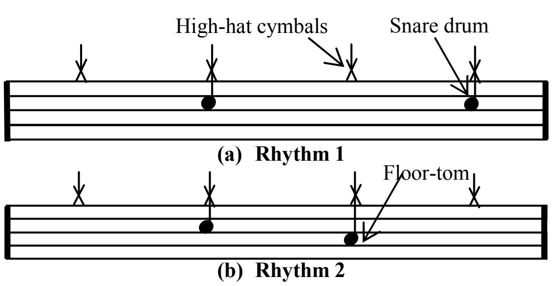

Figure 4. Two Rhythms.

It took 30 seconds for each stimulus. After each stimulus, the pair of subjects was asked to base their judgments about the output quality of own drumstick, output quality of other drumstick, synchronization quality of sound, interactivity, and comprehensive quality based on the five-grade impairment scales (5: Imperceptible, 4: Perceptible, but not annoying, 3: Slightly annoying, 2: Annoying, 1: Very Annoying [23]). The comprehensive quality is a synthesis of the output quality of own drumstick, output quality of other drumstick, synchronization quality of sound, and interactivity. Each subject gave a score from 1 through 5 to each stimulus according to the degree of deterioration on the condition that there exist the constant delay and prediction time. We obtained the *mean opinion score* (MOS) [23] as a QoE parameter by averaging the scores of all the subjects.

Each pair of subjects took a rest for about two minutes before we changed the combination of rhythm and tempo. The pair played the drum set in the following order: Rhythm 1 at the slow tempo, and then rhythm 2 at the fast tempo. The total assessment time per pair was about one and half hours including rests and practices.

6. Assessment Results

Figure 5 explains the notation employed in the following figures. We show QoE assessment results for rhythm 1 at the slow tempo in Figs. 6 through 10, where we plot MOS values of output quality of own drumstick, those of output quality of other drumstick, those of synchronization quality of sound, those of interactivity, and those of comprehensive quality, respectively, as a function of the constant delay. For rhythm 2 at the fast tempo, we show only MOS values of comprehensive quality in Fig. 11. We do not show other MOS values of rhythm 2 at fast tempo since they had similar tendencies to those in Figs. 6 through 10, respectively. In the figure, the 95% confidence intervals are also plotted.

In Fig. 6, we find that the MOS value of dynamic control of prediction time is hardly dependent on the constant delay and is the highest or the second highest. In the control where $T_{predict}$ is fixed, the MOS values tend to become smaller as $T_{predict}$ becomes larger when the constant delay is small. When $T_{predict}$ is between 20 ms and 110 ms, the MOS values increase until the constant delay up to some value. The reason is as follows. When $T_{predict}$ becomes longer than the buffering time, the position information of own drumstick is output by using prediction. Thus, as $T_{predict}$ increases, the prediction error becomes larger.

In Fig. 7, we observe that the MOS value of dynamic control of prediction time decreases as the constant delay becomes larger. In the control where $T_{predict}$ is fixed, the MOS values become smaller as $T_{predict}$ becomes larger. We also notice that the MOS values hardly depend on the constant delay. Because the position information of the other drumstick is output by using prediction, the prediction error becomes larger as $T_{predict}$ increases; thus, the output quality deteriorates.

Figure 8 reveals that the MOS values become smaller as the constant delay becomes larger in both types of control. We also notice that the differences in the MOS value among the values of $T_{predict}$ are very small.

In Fig. 9, we find that the MOS value of dynamic control of prediction time is high for any constant delay. Also, in the figure, the MOS values decrease as the constant delay becomes larger when $T_{predict}$ is fixed; the MOS values decrease more largely as $T_{predict}$ becomes smaller.

In Figs. 10 and 11, we observe that the MOS values of dynamic control of prediction time are the highest or the second highest. Therefore, we can say that the dynamic local lag control with dynamic control of prediction time is effective. Also, in the figures, we see that there are not so much differences between the combinations of rhythm and tempo.

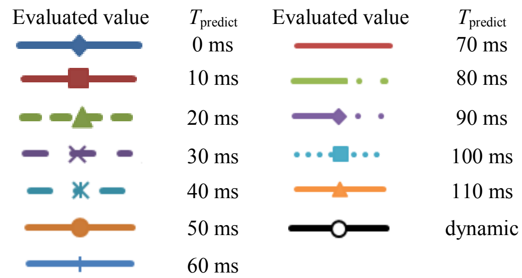

Figure 5. Notation in Figs. 5 through 14.

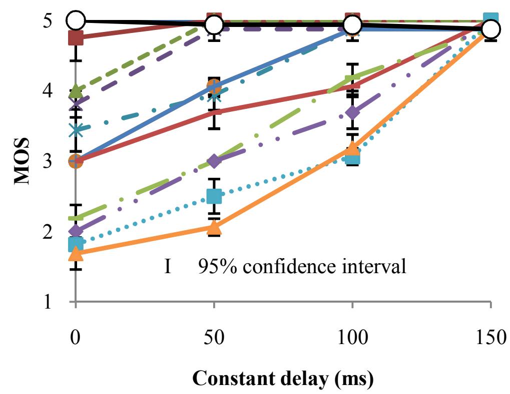

Figure 6. MOS of output quality of own drumstick for rhythm 1 at slow tempo.

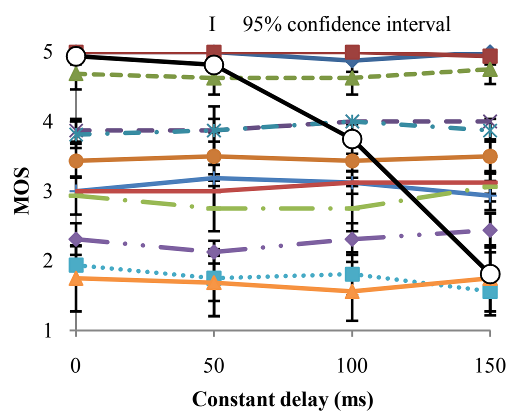

Figure 7. MOS of output quality of other drumstick for rhythm 1 at slow tempo.

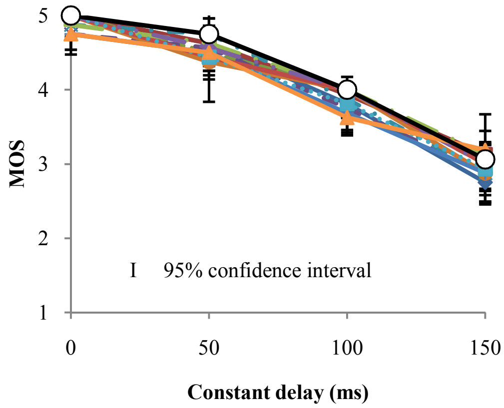

Figure 8. MOS of synchronization quality of sound for rhythm 1 at slow tempo.

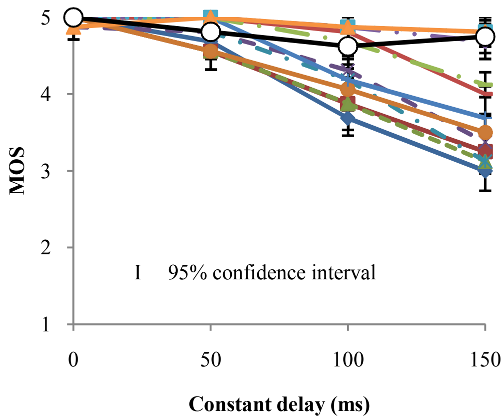

Figure 9. MOS of interactivity for rhythm 1 at slow tempo.

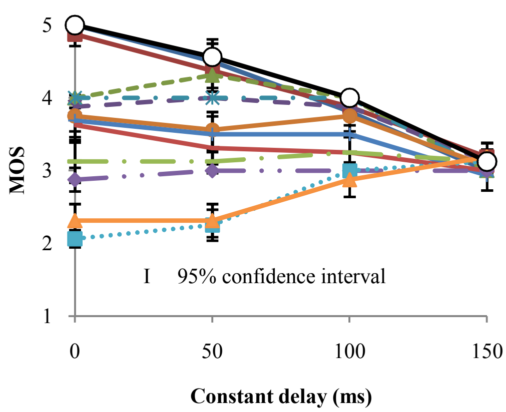

Figure 10. MOS of comprehensive quality for rhythm 1 at slow tempo.

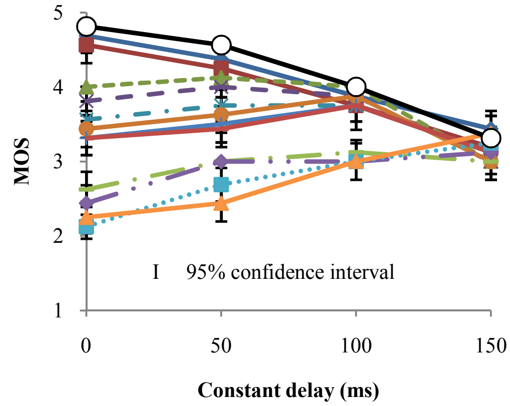

Figure 11. MOS of comprehensive quality for rhythm 2 at fast tempo.

7. Conclusions

In this paper, we proposed the dynamic local lag control with dynamic control of prediction time in order to keep the interactivity and synchronization quality of sound high. We also investigated the effect of the proposed control by subjective QoE assessment in the joint performance of a networked haptic drum system. As a result, we found that the dynamic local lag control with dynamic control of prediction time is effective.

As the next step of our research, we will enhance the proposed control so that three or more users can perform joint musical performance.

References

[1] K. Huang, E. Do, and T. Starner, “PianoTouch: A wearable haptic piano instruction system for passive learning of piano skills,” Proc. the 12th IEEE International Symposium on Wearable Computers, pp. 41-44, Sept. 2008.

[2] I. Arimoto, K. Hikichi, K. Sezaki, and Y. Yasuda, “Influence of network delay on ensemble application,” Proc. IEEE International Workshop on Haptic Audio

Visual Environments and Their Applications (HAVE), pp. 19-22, Oct. 2005.
[3] Q. Zeng, Y. Ishibashi, N. Fukushima, S. Sugawara, and K. E. Psaniss, "Influence of network delay on QoE in networked ensemble with stero video, sound, and haptic media," Record of 2012 Tokai-Section Joint Conference on Electrical and Related Engineering, P1-8, Sept. 2012.
[4] Y. Irie, S. Aoyagi, T. Takada, K. Hirata, K. Kaji, S. Katagiri, and M. Ohsaki, "Development of assistant system for ensemble in t-Room," (in Japanese), IPSJ SIG Technical Report, vol. 2009-DPS-141, no. 23, Nov. 2009.
[5] M. Sithu, P. Huang, and Y. Ishibashi, "Influence of network delay in networked haptic drum performance," IEICE Technical Report, MVE2012-93, Jan. 2013.
[6] ITU-T Rec. E. 800: Definitions of terms related to quality of service, Sept. 2008.
[7] ITU-T Rec. G.100/P.10 Amendment 1: New appendix I- Definition of quality of experience (QoE), Jan. 2007.
[8] M. Mauve, J. Vogel, and W. Effelsberg, "Local lag and timewrap: Providing consistency for replicated continuous applications," IEEE Trans. on Multimedia, vol. 6, no. 1, Feb. 2004.
[9] D. Stuckel and C. Gutwin, "The effects of local lag on tightly-coupled interaction in distributed groupware," Proc. ACM Computer Supported Cooperative Work, pp. 447-456, Nov. 2008.
[10] M. Sithu, Y. Ishibashi, and N. Fukushima, "Dynamic local lag control for sound synchronization in joint musical performance," Proc. ACM the 12th Annual Workshop on Network and Systems Support for games (NetGames), Dec. 2013.
[11] L. A. Zadeh and J. R. Ragazzini, "An extension of Wiener's theory of prediction," Journal of Applied Physics, vol. 21, pp. 645-655, 1950.
[12] P. Huang, Y. Ishibashi, N. Fukushima, and S. Sugawara, "QoE assessment of group synchronization control scheme with prediction in work using haptic media," International Journal of Communications, Network and System Sciences (IJCNS), vol. 5, no. 6, pp. 321-331, June 2012.
[13] Y. Hara, Y. Ishibashi, N. Fukushima, and S. Sugawara, "Adaptive delta-causality control scheme with dynamic control of prediction time in networked haptic game," Proc. ACM The 11th Annual Workshop on Network and Systems Support for Games (NetGames), Nov. 2012.
[14] M. Sithu, Y. Ishibashi, and N. Fukushima, "QoE assessment of dynamic local lag control with Prediction in Joint Haptic Drum Performance," IEICE Technical Report, CQ2013-67, Jan. 2014.
[15] J. Salisbury and M. Srinivasan, "Phantom-based haptic interaction with virtual objects," IEEE Computer Graphics and Applications, vol. 17, no. 5, pp. 6-10, Sept./Oct. 1997.
[16] M. Sithu and Y. Ishibashi, "QoE assessment of joint haptic drum performance: Effect of local lag control," Proc. the 2nd IEEE Global Conference on Consumer Electronics (GCCE), pp. 461-465, Oct. 2013.
[17] Y. Ishibashi, S. Tasaka, and T. Hasegawa, "The Virtual-Time Rendering algorithm for haptic media synchronization in networked virtual environments," Proc. the 16th International Workshop on Communications Quality and Reliability (CQR), pp. 213-217, May 2002.
[18] SensAble Technologies, Inc., "Openhaptics toolkit programmer's guide," version 3.0, 2009.
[19] S. K. Singhal and D. R. Cheritton, "Using a position history-based protocol for distributed object visualization," Technical Report, Standard University, Department of Computer Science, 1994.
[20] S. Chatterjee, A. S. Hadi, and B. Price, "Regression analysis by example," John Wiley & Sons, Hoboken, 1999.
[21] M. Carson and D. Santay, "NIST Net – A Linux-based network emulation tool," ACM SIGCOMM, vol. 33, no. 3, pp. 111-126, July 2003.
[22] http:// www.redheadmusic.com/ pdf/ listening_room/ Bruce_Felter_CD. pdf.
[23] ITU-R BT. 500-12, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, Sept. 2009.