Effects of Dynamic Local Lag Control on Sound Synchronization and Interactivity in Joint Musical Performance

Mya Sithu[†], Yutaka Ishibashi (member)[†], Norishige Fukushima (member)[†]

Abstract

In this paper, we propose dynamic local lag control which dynamically changes the local lag according to the network delay for sound synchronization in joint musical performance where multiple users play musical instruments together through a network. We also make a comparison between the dynamic local lag control and the local lag control with fixed values of local lag in terms of the synchronization quality of sound, interactivity, and comprehensive quality by QoE (Quality of Experience) assessment. In the assessment, for joint musical performance, we use a networked haptic drum system in which two users play a drum set in a 3D virtual space with the same rhythm at the same tempo by using haptic interface devices. As a result, we illustrate that the dynamic local lag control is effective.

Key words: Joint musical performance, Dynamic local lag control, Quality of Experience, Networked haptic drum system

1. Introduction

A number of researchers have been directing their attention to joint musical performance in which multiple users play their respective same or different types of musical instruments together^{1)~7)}. There are several kinds of joint musical performance based on the type of music, the type of musical instrument, and the number of users. In this paper, we handle joint musical performance in which two users play the musical instruments together for simplicity. High synchronization quality of sound is generally required in joint musical performance. However, in joint musical performance over a network^{3)~7)}, the synchronization quality of sound may seriously be deteriorated owing to the network delay^{5)~7)}. To achieve high synchronization quality of sound in joint musical performance, we can use local lag control^{7)~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 5), Irie *et al.* used 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 assumed 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 6), the authors dealt with a networked haptic drum system for joint musical performance. In the system, two users at different places can play the drum set together with the same rhythm at the same tempo by using their independent haptic interface devices. Each user employs two haptic interface devices to move a pair of drumsticks in a 3D virtual space. They carried out subjective QoE (Quality of Experience) assessment to investigate the influences of network delay on the easiness of hitting and synchronization quality of sound. As a result, for example, they illustrated that the synchronization quality of sound in joint musical performance deteriorates more largely as the network delay becomes larger. Therefore, QoS (Quality of Service) control such as the local lag control is needed to achieve high synchronization quality of sound.

In 7), the authors investigated the effects of the local lag control in the joint performance of the networked haptic drum system by subjective QoE assessment. In

Received February 7, 2014; Revised May 28, 2014; Accepted July 23, 2014

[†] Graduate School of Engineering, Nagoya Institute of Technology (Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan)

the assessment, they used only one haptic interface device at each terminal as the right drumstick for simplicity. They illustrated that there exists the optimum value of local lag for the joint musical performance. They also found that the optimum value of local lag is the same as the network delay when the network delay is small, but the value is smaller than the network delay when the network delay is large. This is because the interactivity is severe in the joint musical performance. Moreover, they noticed that the optimum value of local lag is dependent on the network delay from the other terminal to the local terminal and not that in the opposite direction.

Based on the above results in 7), we here propose dynamic local lag control which dynamically changes the local lag according to the network delay in joint musical performance for both of the symmetric and asymmetric delay cases. In order to clarify the effects of the dynamic local lag control, we 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 on the synchronization quality of sound, interactivity, and comprehensive quality. We also perform objective QoE assessment at the same time as the subjective assessment. In the assessment, we use the networked haptic drum system⁶⁾ for joint musical performance. Two haptic interface devices are employed at each terminal as the left and right drumsticks.

The remainder of this paper is organized as follows. Section 2 describes the networked haptic drum system. Section 3 explains the conventional local lag control. Section 4 proposes the dynamic local lag control, and Section 5 explains 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 space⁶⁾. For three or more users, the reader is referred to 10). 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 Omni¹¹⁾). The two haptic interface devices at each terminal are used to move a

Fig. 1 Configuration of networked haptic drum system.

Fig. 2 Reaction force when drum component is hit.

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 $m_d \mathbf{g}$ due to the gravity, where m_d denotes the mass of a held drumstick, and $\mathbf{g}$ is the gravitational acceleration. The other is produced when the user hits a drum component by a drumstick.

The reaction force $\mathbf{F}$ 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 model¹²⁾ as follows: $\mathbf{F} = -(K_s \mathbf{x} + K_d \mathbf{v})$, where K_s is the spring coefficient, K_d is the damper coefficient, $\mathbf{x}$ is defined as a vector from the contact point of the object (i.e., drum component) surface to the tip of the drumstick, and $\mathbf{v}$ is the velocity of the drumstick (see **Fig. 2**, in which the penetration depth $|\mathbf{x}|$ is the distance between the object surface and the tip of the drumstick). If the drumstick is not in contact with the object surface, the values of $\mathbf{x}$ and $\mathbf{v}$ are equal to zero. In the joint performance of the networked haptic drum system, each user perceives the resultant force $m_d \mathbf{g} + \mathbf{F}$. We can-

ried out a preliminary experiment to determine values of the gravitational acceleration and damper coefficient as well as values of the other parameters. As a result, in order to do the joint performance comfortably 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/s², 0.2 N/mm, and 0.1 Nms/mm, respectively.

3. Coventional Local Lag Control

Under the conventional local lag control between two terminals in 5), the local lag denoted by Δ (≥ 0) ms is set to the same value as the network delay from the local terminal to the other 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*. From **Fig. 3**, we see that when network delay 1 is not equal to network delay 2, sound synchronization cannot be achieved at each terminal; that is, each user hears sound twice. In order to solve this problem, we propose the dynamic local lag control.

4. Dynamic Local Lag Control

The dynamic local lag control dynamically changes Δ according to the network delay from the other terminal to the local terminal. It should be noted that the direction of network delay is different from that in the conventional local lag control. In **Fig. 4**, where we set local lag i ($i=1$ or 2) to the same value as network delay i , we see that sound synchronization can be achieved at each terminal; that is, each user hears sound only once. In the dynamic local lag control, the value of Δ is set to the optimum value of local lag (denoted by Δ_{optimum} ms) obtained in 7) for a given network delay. We investigated the relation of the optimum value of local lag and the network delay by regression analysis, and we obtained the following equation:

$$\Delta_{\text{optimum}} = 0.637D + 6.578, \quad (1)$$

where D is the time interval from the moment a *media unit* (MU) 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 server loop¹²⁾.

Fig. 3 Asynchronization of sound under conventional local lag control.

Fig. 4 Synchronization of sound under dynamic local lag control.

The contribution rate adjusted for degrees of freedom¹³⁾, which shows goodness of fit for the equation, was 0.970. Therefore, we can get the optimum value of local lag from D with a high degree of accuracy. Note that Eq. (1) is somewhat different from the equation of the optimum value of local lag in 7), 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. The equation in 7) is obtained from the results of the assessment in which only the right drumstick is employed at each terminal. In this paper, we use Eq. (1) in the assessment where both left and right drumsticks are employed at each terminal. Therefore, to confirm that Eq. (1) can be used in the assessment, we carried out the regression analysis by using assessment results of the local lag control with fixed values of local lag in Section 6, and the same equation was obtained. Therefore, Eq. (1) can be used in the assessment of this paper.

Fig. 5 Configuration of assessment system.

5. Assessment Environment

5.1 Assessment System

Figure 5 shows our assessment system, where two terminals are connected to each other via a network emulator (NIST Net¹⁴) which is used instead of the network shown in Fig. 1. The network emulator generates an additional constant delay for each packet transmitted between the terminals. Note that the network delay jitter can be absorbed by buffering under media synchronization control^{15)~18)}, such as the Virtual-Time Rendering (VTR) algorithm¹⁹⁾; we here take account of the jitter by including the buffering time in the constant delay.

We handle the symmetric and asymmetric delay cases in the assessment. In the symmetric delay case, we set the constant delay (called *constant delay 1* here) from terminal 2 to terminal 1 to the same value as that (*constant delay 2*) in the opposite direction. In the asymmetric delay case, constant delay 1 is not equal to constant delay 2.

5.2 Assessment Methods

In the symmetric and asymmetric delay cases, we made a comparison between the dynamic local lag control and the local lag control with fixed values of local lag. We carried out subjective QoE assessment with 16 subjects (males and females) whose ages were between 20 and 28.

In the symmetric delay case, where constant delay 1 is equal to constant delay 2, we employed two rhythms (*rhythms 1 and 2*) at two tempos (*slow and fast*) as

Fig. 6 Two rhythms.

in 6) 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. 6 (a)). The high-hat cymbals are hit at all the four beats, and the snare drum is hit 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 the times (see Fig. 6 (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 needs to move one of his/her drumsticks 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, respectively²⁰⁾.

In the subjective QoE assessment, we employed only rhythm 1 at the slow tempo and rhythm 2 at the fast tempo in the symmetric delay case. This is because we found that assessment results of the other combinations of rhythm and tempo are almost the same as those of rhythm 1 at the slow tempo in 6). Each pair of subjects practiced about two minutes under the condition that there was no constant delay, and local lags 1 and 2 were set to 0 ms before the assessment of each combination of rhythm and tempo. In the assessment, constant delays 1 and 2, local lags 1 and 2, and the two types of control (i.e., the dynamic local lag control and the local lag control with fixed values of local lag) were selected in random order for the pair. We changed constant delays 1 and 2 from 0 ms to 150 ms at intervals of 50 ms in both types of control. In the local lag control with fixed values of local lag, we set local lag 1 to the same value as local lag 2, and the values were set to 0 ms, 50 ms, 75 ms, 100 ms, and 150 ms. In the dynamic local lag control, the value of local lag was dynamically changed according to Eq. (1). In this paper, for simplicity, we delayed only the output of sound and visual information by local lag after generating the local information. The reaction force was perceived without delay when a drumstick hit the drum set. There are other possibilities, for example, outputting the visual information and haptic media simultaneously and delaying only the output of sound. It is also important to examine the effect of the dynamic local lag control for other possibilities in the joint musical performance. However, this is for further study.

It took 30 seconds for each stimulus. After each stimulus, the pair were asked to base their judgments about the synchronization quality of sound, interactivity, and comprehensive quality based on the five-grade quality scales (5: Excellent, 4: Good, 3: Fair, 2: Poor, 1: Bad). The synchronization quality means how much simultaneously the sound of one user and the that of the other user are outputted. The interactivity is the time difference from the moment a user hits a drum component until the instant the user hears a sound of the component. The comprehensive quality is the weighted sum of the synchronization quality of sound and interactivity; thus, the comprehensive quality is the most important. Each subject gave a score from 1 through 5 to each stimulus. By averaging scores of all the subjects, we

obtained the *mean opinion score (MOS)*²¹⁾ as a subjective QoE parameter.

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 rhythm 2 at the fast tempo. The assessment time per pair for each combination of rhythm and tempo was about half an hour, and the total assessment time per pair was about an hour including rests and practices.

In the asymmetric delay case, we employed only rhythm 1 at the slow tempo because there were not so much differences between the combinations of rhythm and tempo in the symmetric delay case. We carried out the assessment for several combinations of constant delays 1 and 2 (50 ms and 0 ms, 100 ms and 0 ms, 150 ms and 0 ms, and 100 ms and 50 ms). The total assessment time per pair was about an hour including a practice.

Objective assessment was also carried out at the same time as the subjective QoE assessment. We adopted the root mean square error (RMSE)²²⁾ of sound at a terminal as an objective assessment measure. The root mean square error is defined as the square root of the mean square error, which denotes the average of squared difference between the output times of two sounds (sound generated at the local terminal and that received from the other terminal). The root mean square error of sound at terminal 1 is equal to that at terminal 2 in the symmetric delay case, but the root mean square errors at terminals 1 and 2 are different from each other in the asymmetric delay case. The local lag is also employed as an objective assessment measure since it is closely related to the interactivity.

6. Assessment Results

6.1 Subjective Assessment

(1) Symmetric Delay Case

Figure 7 shows the notation employed in the following figures. For rhythm 1 at the slow tempo, we plot MOS values of synchronization quality, those of interactivity, and those of comprehensive quality for various values of local lag 1 (or 2) as a function of constant delay 1 (or 2) in Figs. 8 through 10, respectively. From Figs. 11 through 13, we draw MOS values for rhythm 2 at the fast tempo. In the figures, the 95% confidence intervals are also plotted.

In **Figs. 8 and 11**, we see that the MOS values of

the dynamic local lag control are the highest or second highest for each constant delay. We also confirm that there exists the optimum value of local lag for each constant delay when the local lag is fixed. The reason is that the synchronization quality of sound becomes higher as the difference between the local lag and constant delay becomes smaller when the constant delay is small. From the figures, we find that the optimum value of local lag is the same as the constant delay when the constant delay is smaller than or equal to about 100 ms, but that of local lag is smaller than the constant delay when the constant delay is larger than about 100 ms. For example, in Fig. 8, the optimum value of local lag is 100 ms when the constant delay is 150 ms. We further note in the figures that the MOS value at the optimum value of local lag tends to decrease as the constant delay becomes larger.

Figures 9 and 12 reveal that the MOS values of interactivity hardly depend on the constant delay when the local lag is fixed. The MOS values become smaller as the local lag becomes larger. The MOS value of the dynamic local lag control is larger than that in the case where we set the local lag to the same value of the constant delay when the constant delay is larger than or equal to about 100 ms. The MOS value of the dynamic local lag control decreases almost linearly as the constant delay becomes larger. Because 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. In order to keep the interactivity high, we improve the dynamic local lag control using prediction control in 23) as in 24). In 23), we illustrate that there is the optimal prediction time according to the network delay. Therefore, we propose the dynamic local lag control with dynamic control of prediction time which dynamically changes the prediction time according to the network delay in 25) as in 26).

From **Figs. 10 and 13**, we find almost the same tendencies as those in Figs. 8 and 11. That is, the MOS values of the dynamic local lag control are the highest or second highest for each constant delay. Therefore, the dynamic local lag control is effective. Also, when the local lag is fixed, there exists the optimum value of local lag for each constant delay, and the MOS value at the optimum value of local lag tends to decrease as the constant delay becomes larger. We also see in the figures that there are not so much differences between the combinations of rhythm and tempo.

Fig. 7 Notation in Figs. 8 through 15 and Fig. 18.

Fig. 8 MOS of synchronization quality of sound for rhythm 1 at slow tempo in symmetric delay case.

Fig. 9 MOS of interactivity for rhythm 1 at slow tempo in symmetric delay case.

(2) Asymmetric Delay Case

In the asymmetric delay case, we show only MOS values of comprehensive quality at terminals 1 and 2 for rhythm 1 at the slow tempo in **Figs. 14 and 15**, respectively, for four combinations of constant delays (constant delay 1: 0 ms, 50 ms, 100 ms, and 150 ms; constant delay 2: 0 ms)²⁷⁾. In Fig. 14, we draw summarized MOS values for various values of local lag 2 since the differences among the values were very small. We

Fig. 10 MOS of comprehensive quality for rhythm 1 at slow tempo in symmetric delay case.

Fig. 13 MOS of comprehensive quality for rhythm 2 at fast tempo in symmetric delay case.

Fig. 11 MOS of synchronization quality of sound for rhythm 2 at fast tempo in symmetric delay case.

Fig. 12 MOS of interactivity for rhythm 2 at fast tempo in symmetric delay case.

do not show MOS values of synchronization quality of sound and those of interactivity at terminals 1 and 2. The reason is that the MOS values of synchronization quality of sound and interactivity at terminal 1 had similar tendencies to those in Figs. 8 and 9, respectively,

and those of synchronization quality of sound and interactivity at terminal 2 had similar tendencies to those of comprehensive quality in Fig. 15. We also plot MOS values of comprehensive quality at terminals 1 and 2 in **Figs. 16 and 17**, respectively, for three combinations of constant delays (constant delay 1: 100 ms; constant delay 2: 0 ms, 50 ms, and 100 ms). For the same reason as in Fig. 14, we plot summarized MOS values for various values of local lag 1 in Fig. 17. In Figs. 16 and 17, the notation in Fig. 7 is used by replacing local lag 1 with local lag 2. In the figures, we use the MOS values obtained in the symmetric delay case when both constant delays are 0 ms and 100 ms. The 95% confidence intervals are also included in the figures.

In Fig. 14, we see that MOS values at terminal 1 depend on only constant delay 1. We also note that the dynamic local lag control has the highest or second highest MOS values for each value of constant delay 1. In the figure, there exists the optimum value of local lag 1 for each value of constant delay 1 when local lag 1 is fixed. However, from the figure, we find that the optimum value of local lag 1 is somewhat smaller than constant delay 1 when constant delay 1 is large. We further notice that the MOS value at the optimum value of local lag 1 tends to decrease as constant delay 1 becomes larger.

In Fig. 15, we notice that the MOS values of comprehensive quality at terminal 2 do not depend on constant delay 1. We also see that the MOS values of the dynamic local lag control are the highest or second highest for each value of constant delay 1. Furthermore, we observe that the MOS values decrease as local lag 2 becomes larger; that is, local lag 2 of 0 ms has the highest MOS values. We also find that there are not so much

Fig. 14 MOS of comprehensive quality at terminal 1 for rhythm 1 at slow tempo in asymmetric delay case (constant delay 1: 0 ms, 50 ms, 100 ms, and 150 ms; constant delay 2: 0 ms).

Fig. 15 MOS of comprehensive quality at terminal 2 for rhythm 1 at slow tempo in asymmetric delay case (constant delay 1: 0 ms, 50 ms, 100 ms, and 150 ms; constant delay 2: 0 ms).

differences among the values of local lag 1.

Figure 16 reveals that the MOS values of comprehensive quality at terminal 1 do not depend on constant delay 2. In the figure, the MOS values of the dynamic local lag control are the highest. Also, local lag 1 of 75 ms has the second highest MOS values; therefore, by taking account of the results in Fig. 15, we can say that there exist the optimum values of local lags 1 and 2 according to constant delays 1 and 2, respectively.

In Fig. 17, we observe that the MOS values of the dynamic local lag control are the highest for each constant delay. Also, the MOS values of comprehensive quality at terminal 2 depend on only constant delay 2, and they do not depend on local lag 1.

From the above considerations, we can conclude that the dynamic local lag control is effective.

6.2 Objective Assessment

In **Fig. 18**, we show RMSE of sound versus constant

Fig. 16 MOS of comprehensive quality at terminal 1 for rhythm 1 at slow tempo in asymmetric delay case (constant delay 1: 100 ms; constant delay 2: 0 ms, 50 ms, and 100 ms).

Fig. 17 MOS of comprehensive quality at terminal 2 for rhythm 1 at slow tempo in asymmetric delay case (constant delay 1: 100 ms; constant delay 2: 0 ms, 50 ms, and 100 ms).

Fig. 18 Root mean square error of sound for rhythm 1 at slow tempo in symmetric delay case.

delay 1 (or 2) only for rhythm 1 at the slow tempo in the symmetric delay case, where the 95% confidence intervals are also included. We do not show RMSE of sound for rhythm 2 at the fast tempo in the symmetric

delay case and that for rhythm 1 at the slow tempo in the asymmetric delay case since they had almost the same tendencies as that in Fig. 18. We observe in the figure that the dynamic local lag control has the smallest or second smallest root mean square error for each constant delay. By comparing Figs. 8 and 18, we find that the trends of the curves are reverse to each other; that is, the highest MOS value can be obtained when the root mean square error is the smallest for each local lag value.

6.3 Mapping from Objective Assessment Measures to MOS Values

In order to investigate the relations between the objective assessment measures (i.e., the root mean square error of sound and the local lag) and the MOS values, we carried out the regression analysis. As a result, we obtained the following three equations:

$$\hat{S}_{\text{mos}} = -0.027E_{\text{rmse}} + 5.549, \quad (2)$$

$$\hat{I}_{\text{mos}} = -0.019\Delta + 4.892, \quad (3)$$

$$\hat{C}_{\text{mos}} = -0.019E_{\text{rmse}} - 0.005\Delta + 5.336, \quad (4)$$

where $\hat{S}_{\text{mos}}$, $\hat{I}_{\text{mos}}$, and $\hat{C}_{\text{mos}}$ denotes the estimated MOS value of synchronization quality of sound, that of interactivity, and that of comprehensive quality, respectively. Also, E_{rmse} denotes RMSE of sound, and Δ denotes local lag 1 or 2. The contribution rates adjusted for degree of freedom for the equations were 0.904, 0.964, and 0.922, respectively. **Figure 19** shows the notation employed in **Figs. 20 through 22**, where we add the calculated values obtained from Eqs. (2), (3), and (4) to the evaluated MOS values shown in Figs. 8 through 10, respectively. We do not plot the calculated MOS values for the other figures since they had similar tendencies to those in Figs. 20 through 22. In the figures, we actually confirm close agreement between the evaluated values and calculated ones. Therefore, we can say that the MOS values can be estimated from RMSE of sound and/or the local lag with a high degree of accuracy.

We also conducted the regression analysis to investigate the relation between the absolute difference of the local lag and the constant delay (from the other terminal to the local terminal), and the MOS values (S_{mos} and C_{mos}). As a result, we obtained the following equations:

$$\hat{S}_{\text{mos}} = -0.018|\Delta - D_c| + 4.384, \quad (5)$$

$$\hat{C}_{\text{mos}} = -0.011|\Delta - D_c| - 0.009\Delta + 4.661, \quad (6)$$

Fig. 19 Notation in Figs. 20 through 22.

Fig. 20 Relation between evaluated and calculated MOS values of synchronization quality of sound for rhythm 1 at slow tempo in symmetric delay case .

where D_c is the constant delay from the other terminal to the local terminal. The contribution rates adjusted for degree of freedom for the equations were 0.771 and 0.866, respectively. Therefore, by comparing with the contribution rates of Eqs. (2) and (4), estimating S_{mos} and C_{mos} from E_{rmse} and/or Δ can get higher accuracy than estimating S_{mos} and C_{mos} from Δ and D_c .

Furthermore, in order to investigate the relations among the evaluated MOS values, we carried out the regression analysis, and obtained the following equation:

$$\hat{C}_{\text{mos}} = 0.709S_{\text{mos}} + 0.276I_{\text{mos}} + 0.017 \quad (7)$$

where $\hat{C}_{\text{mos}}$ denotes the estimated MOS value of comprehensive quality, and S_{mos} , and I_{mos} denotes the evaluated MOS value of synchronization quality of sound and that of interactivity, respectively. The contribution rate adjusted for degree of freedom for Eq. (5) was 0.979. From the equation, we find that since the coefficient of S_{mos} is larger than that of I_{mos} , the contribution of the synchronization quality of sound to the comprehensive quality is larger than that of the interactivity.

Fig. 21 Relation between evaluated and calculated MOS values of interactivity for rhythm 1 at slow tempo in symmetric delay case.

Fig. 22 Relation between evaluated and calculated MOS values of comprehensive quality for rhythm 1 at the slow tempo in symmetric delay case.

7. Conclusions

In this paper, we proposed dynamic local lag control for sound synchronization in joint musical performance. We also made a comparison between the dynamic local lag control and the local lag control with fixed values of local lag by subjective and objective QoE assessments in joint performance of a networked haptic drum system. We further examined the relationship between the MOS values and the objective performance measures. As a result, we found that the dynamic local lag control is effective. Moreover, we noted that MOS values can be estimated from the root mean square of sound and/or the local lag with a high degree of accuracy.

As the next step of our research, we will investigate the influence of packet loss by QoE assessment. We also need to examine the effect of combination of the dynamic local lag control and media synchronization

control which absorbs the network delay jitter. Furthermore, it is important to examine the effect of the dynamic local lag control over the Internet.

References

- 1) J. Loehr, D. Kourtis, C. Vesper, N. Sebanz, and G. Knoblich: "Monitoring individual and joint action outcomes in duet music performance," *Journal of Cognitive Neuroscience*, **25**, 7, pp. 1049-1061 (July 2013)
- 2) P. Keller: "Joint action in music performance," *Journal of Emerging Communication*, **21**, 6, pp. 454-461 (July 2006)
- 3) I. Arimoto, K. Hikichi, K. Sezaki, and Y. Yasuda: "Influence of network delay on ensemble application," in *Proc. IEEE International Workshop on Haptic Audio Visual Environments and Their Applications (HAVE)*, pp. 19-22 (Oct. 2005)
- 4) Q. Zeng, Y. Ishibashi, N. Fukushima, S. Sugawara, and K. E. Psannis: "Influence of network delay on QoE in networked ensemble with stereo, video, sound, and haptic media," *Record of 2012 Tokai-Section Joint Conference on Electrical and Related Engineering*, P.1-8 (Sept. 2012)
- 5) Y. Irie, S. Aoyagi, T. Takada, K. Hirata, K. Kaji, S. Katagiri, and M. Ohasaki: "Development of assistant system for ensemble in t-Room," (in Japanese), *IPSI SIG Technical Report*, **2009-DPS-141**, 23 (Nov. 2009)
- 6) M. Sithu, P. Huang, and Y. Ishibashi: "Influence of network delay in networked haptic drum performance," *IEICE Technical Report*, MVE2012-93 (Jan. 2013)
- 7) M. Sithu and Y. Ishibashi: "QoE assessment of joint haptic drum performance: Effect of local lag control," in *Proc. the 2nd IEEE Global Conference on Consumer Electronics (GCCE)*, pp. 461-465 (Oct. 2013)
- 8) M. Mauve, J. Vogel, and W. Effelsberg: "Local lag and timewrap: Providing consistency for replicated continuous applications," *IEEE Trans. on Multimedia*, **6**, 1 (Feb. 2004)
- 9) D. Stuckel and C. Gutwin: "The effects of local lag on tightly-coupled interaction in distributed groupware," in *Proc. ACM Computer Supported Cooperative Work*, pp. 447-456 (Nov. 2008)
- 10) M. Sithu, Y. Ishibashi, and N. Fukushima: "Enhancement of dynamic local lag control for networked musical performance," in *Proc. the 3rd IEEE Global Conference on Consumer Electronics (GCCE)* (Oct. 2014)
- 11) J. Salisbury and M. Srinivasan: "Phantom-based haptic interaction with virtual objects," *IEEE Computer Graphics and Applications*, **17**, 5, pp. 6-10 (Sept./Oct. 1997)
- 12) SensAble Technologies, Inc.: "Openhaptics toolkit programmers' guide," **3.0** (2009)
- 13) S. Chatterjee, A. S. Hadi, and B. Price: "Regression analysis by example," John Wiley & Sons, Hoboken (1999)
- 14) M. Carson and D. Santay: "NIST Net - A Linux-based network emulation tool," *ACM SIGCOMM*, **33**, 3, pp. 111-126 (July 2003)
- 15) Z. Huang, K. Nahrstedt, and R. Steinmetz: "Evolution of temporal multimedia synchronization principles: A historical viewpoint," *ACM Trans. on Multimedia Computing, Communications, and Applications*, **9**, 1, no. 34 (Oct. 2002)
- 16) F. Boronat, J. Lloret, and M. Garcia: "Multimedia group and inter-stream synchronization techniques: A comparative study," *Journal of Information Systems*, **34**, 1, pp. 108-131 (Mar. 2009)
- 17) G. Blalowski, and R. Steinmetz: "A media synchronization survey: Reference model, specification, and case studies," *Journal of IEEE Selected Areas in Communications*, **14**, 1, pp. 5-35 (Jan. 1995)
- 18) Y. Ishibashi and S. Tasaka: "A comparative survey of synchronization algorithms for continuous media in network environments," in *Proc. IEEE Local Computer Network (LCN)*, pp. 337-348 (Nov. 2000)
- 19) Y. Ishibashi, S. Tasaka, and T. Hasegawa: "The Virtual-Time Rendering algorithm for haptic media synchronization in networked virtual environments," in *Proc. the 16th International Workshop on Communications Quality and Reliability (CQR)*, pp. 213-217 (May 2002)
- 20) http://www.redheadmusic.com/pdf/listening/_room/Bruce_Felter_CD.pdf.
- 21) ITU-R BT. 500-12: "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union (Sept. 2009)
- 22) J. F. Kenny and E. S. Keeping: "Mathematics of statistics," NJ:

- Van Nostrand, pp. 59-60 (1962)
- 23) 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)
 - 24) 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), **5**, 6, pp. 321-331 (June 2012)
 - 25) M. Sithu, Y. Ishibashi, and N. Fukushima: "Effect of dynamic local lag control with dynamic control of prediction time in joint haptic drum performance," to appear in Proc. the 12th International Conference on Computer Applications (ICCA) (Feb. 2014)
 - 26) Y. Hara, Y. Ishibashi, N. Fukushima, and S. Sugawara: "Adaptive delta-causality control scheme with dynamic control of prediction time in networked haptic game," in Proc. ACM the 11th Annual Workshop on Network and Systems Support for Games (NetGames) (Nov. 2012)
 - 27) M. Sithu, Y. Ishibashi, and N. Fukushima: "Dynamic local lag control for sound synchronization in joint musical performance," ACM the 12th Annual Workshop on Network and Systems Support for Games (NetGames) (Dec. 2013)

Mya Sithu received B.C.Tech., and M.C.Tech. degrees from University of Computer Studies, Yangon, Myanmar, in 2004 and 2008, respectively. Now she is a Ph.D candidate in Nagoya Institute of Technology, Japan. Her research interests are QoS control and QoE assessment in haptic media communications. She is a IEEE student member.

Yutaka Ishibashi received the B.E., M.E., and Dr.E. degrees from Nagoya Institute of Technology, Nagoya, Japan, in 1981, 1983, and 1990, respectively. In 1983, he joined the Musashino Electrical Communication Laboratory of NTT. From 1993 to 2001, he served as an Associate Professor of Department of Electrical and Computer Engineering, Faculty of Engineering, Nagoya Institute of Technology. Currently, he is a Professor of Department of Scientific and Engineering Simulation, Graduate School of Engineering, Nagoya Institute of Technology. His research interests include networked multimedia, QoS control, and media synchronization. He is a fellow of IEICE and a member of IEEE, ACM, IPSJ, and VRSJ.

Norishige Fukushima received the B.E., M.E., and Ph.D. degree from Nagoya University, Japan, in 2004, 2006, and 2009, respectively. Since 2009, he has been an assistant professor at Graduate School of Engineering, Nagoya Institute of Technology, Japan. His research interests are multi view image capturing, calibration, processing, and coding.