Study on QoE Assessment and QoS Control in Networked Virtual Environments with Haptic Sense

力覚を利用した分散仮想環境におけるユーザ体感品質評価とサービス品質制御に関する研究

2016

Mya Sithu

Contents

Chapter 1

Introduction

In recent years, with the advancement of high speed communication networks, multimedia applications which consists of a collection of multiple media such as text, graphics, images, sound, and video became popular. Moreover, due to high performance computers, traditional 2D Computer Graphics (CG) are being replaced by high quality 3D CG in virtual environments. In networked virtual environments, users who are located around the world can interact with each other in real time [1]-[4]. Recently, in multimedia applications in the networked virtual environments, haptic media can also be used in addition to traditional visual and auditory media [5]-[9]. By using haptic interface devices [10]-[19], users can feel the haptic sense (i.e., the sense of touch) when they touch objects in the virtual environments. They can also feel the shape, weight, and softness of the objects.

There is a wide spectrum of applications that incorporate haptic technology in the areas of music performance, games, teleoperations, remote education, and so on [20]-[35]. When users employ such kinds of applications over a network, simultaneous output of media is very important not only at each terminal but also among multiple terminals depending on the types of applications. Simultaneous output of media means making the output timings (e.g., output times and speeds) of media to be the same. The media may be the same or different media at a terminal, or the same media at different terminals. For example, in joint musical performance, where users play musical instruments together over a network, sound synchronization of the users (i.e., synchronization of local sound and those received from other terminals) at each terminal is needed. As another example, in networked real-time games, media should be output simultaneously at different terminals to maintain consistency among the players' terminals. However, when the applications are used over a non-guaranteed QoS (Quality of Service) [36]-[38] network like the Internet, media may not be output simultaneously owing to network delay, delay jitter, and packet loss [39], [40]. This leads to large degradation of QoE (Quality of Experience) [41] such as deterioration of media synchronization quality, interactivity, operability of haptic interface devices, consistency, and fairness among the users. In order to clarify such kinds of influences, we should carry out QoE assessments, in which the user's perception and satisfaction are analysed by using some methodologies. Moreover, in order to maintain QoE as high as possible, it is important to carry out QoS control. In this thesis, we carry out QoE assessments to investigate the influence of network delay on QoE in the joint musical performance [21], [22], [42]-[44] and network real-time games [32]-[35] in virtual environments. We chose these areas to study deterioration of various QoE parameters caused by incapable of outputting media simultaneously at each terminal or different terminals. Then, based on results of the assessments, we clarify which QoS control should be applied to the areas to maintain QoE as high as possible.

In the followings, we describe networked virtual environments in Subsection 1.1. We present networked virtual environments with haptic sense in Subsection 1.2. Haptic sense and media are explained in Subsection 1.3. We provide a detailed explanation of QoS in Subsection 1.4, and QoS control in Subsection 1.5. Quality of Experience (QoE) is discussed in Subsection 1.6. The purpose of the thesis is written in Subsection 1.7, and the organization of the thesis is explained in Subsection 1.8.

1.1 Networked virtual environments

Networked virtual environments constructed by CG provide users an illusion of sharing a 3D virtual space with other users who are in geographically separated areas. With the evolution of haptic technology, the users get feeling of telepresence because they get haptic feedback in addition to traditional CG and sound in the virtual environments [1]-[4]. In multimedia applications of the networked virtual environments, users interact with each other or interact with objects in the virtual space together. In some applications, users in the networked virtual environment are represented by human-like agents, called avatars, and each user controls his/her own avatar in the virtual space.

In the networked virtual environments, the users normally do work together collaboratively and cooperatively in real-time. In networked real-time games of virtual environments, the users play the games competitively. Achieving realistic environment is an important factor in the environments. In traditional virtual environments, CG, video and sound support the users for high sense of realism. In recent years, with the evolution of haptic technology, users can achieve higher sense of realism in addition to traditional video and audio.

1.2 Networked virtual environments with haptic sense

In networked virtual environments with only visual and auditory senses, users capabilities are limited to interact with the other users or objects in the same virtual environment. Because of haptic technology, users are able to touch, feel and manipulate the 3D objects. Therefore, in the networked virtual environments with haptic sense, users can get more immersive experience [5]-[9].

There are various applications with haptic sense in networked virtual environments. Example applications are remote surgery simulation [27], remote education such as haptic virtual museum [7], [25], remote calligraphy [29], remote ikebana [31], and networked real time games such as haptic battle pong [32], networked fruit harvesting game [35].

1.3 Haptic media and sense

1.3.1 Characteristics of haptic media

The term “haptics” was descended from the Greek word “haptesthai” meaning “relating to the sense of touch.” It is introduced by researchers in the area of psychology at the beginning of the twentieth century to refer to touch of real objects by humans. In the late 1980s, the term was redefined to include all aspects of human-computer touch interaction [5], [6]. Nowadays, the term “haptics” is being widely used in the areas of biomechanics, psychology, engineering, and computer science for the study of human touch and force feedback.

In networked virtual environments, haptic sensations can be transmitted as media units between users’ terminals. The media units (MUs) [45] are information units for media synchronization, each of which includes the position information of haptic interface device, information about reaction force, timestamp which represents the input time, and sequence number of the media unit. There are at least two control schemes in haptic media transmission [46], [47]. One of the schemes is the position-position control scheme [48], and the other is the position-force control scheme [49]. In the position-position control scheme [48], each terminal calculates the reaction force applied to its haptic interface device or an object according to the positional information (i.e., position of haptic interface device or an object at the other terminal) provided from the other terminal and the position information of the device. In the position-force scheme [49], a terminal calculates the reaction force applied to its haptic interface device based on the information about the reaction force sent from the other terminal. The reaction force sent by the other terminal is multiplied by the gain coefficient.

1.3.2 Haptic interface devices

By using haptic interface devices in networked virtual environments, users can get the force feedback via the devices when they touch objects in the virtual environment. By comparing to video update rates, haptic devices need higher update rates for the feeling of realistic touch. Current haptic interface devices have 1 kHz or more input/output frequency [50], [51]. Because of this high frequency rate, users can feel the sense of touch naturally in the virtual environments. There are several types of haptic interface devices such as 1-DoF (Degree-of-freedom), 2-DoF, 3-DoF, and 6-DoF devices [51], [52]. Degree-of-freedom means the freedom of movement of the virtual object in 3D space. As the number of DoF increases, the device becomes more natural and intuitive, but it becomes more complicated because the number of haptic sensors and actuators increase. By using 6-DoF haptic interface devices, the virtual object can be manipulated in the directions of left/right, up/down, and forward/backward, combined with changes in orientation (i.e., pitch, yaw, and roll). Examples of haptic interface devices are Geomagic Touch [10] (see Fig. 1.1), Geomagic Touch X [11], PHANTOM Premium 1.5 [12], Omega [13], SPIDAR-GAHS [14], Falcon [15], Virtuose [16], HapticMaster [17], CyberGrasp, CyberForce [18], and so on. All the devices except CyberGrasp and CyberForce are mainly manipulated by a single interaction point, which a user operates to touch an object.

For the works in this thesis, as an haptic interface device, we employ Geomagic Touch, which is formerly known as Sensable Phantom Omni. The device can work in the workspace of 6.4 Width, 4.8 Height, and 2.8 Diagonal in inches. The force feedback can get in 3-DoF, the positional sensing can be get in 6-DoF. The maximum exertable force is 3.3 N.

1.4 Quality of Service

In 1994, the International Telecommunication Union (ITU) defined Quality of Service (QoS) as "*totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service*" [37]. In other words, we can say that QoS is a concept which represents how good the service is. QoS is characterized by several parameters which are called QoS parameters, and a QoS parameter is a scale which measures QoS quantitatively. In a traditional telephone network, QoS parameters comprises requirements on all the aspects of a connection, such as service response time, loss, signal-to-noise ratio, crosstalk, echo, interrupts, and loudness levels. In telecommunication networks, examples of QoS parameters are bit rate, frame rate, throughput, transmission speed, delay, jitter, and packet loss rate. In [38], QoS over the Internet is divided into the following six layers in descending order: user level QoS, application level QoS, end-to-end level QoS, network level QoS, node level QoS, and physical level QoS. User-level QoS is also called as Quality of Experience (QoE) [38].

Quality of service guarantees is very important for multimedia communications. However, when media streams such as voice, video, and haptic media are transmitted over a QoS non-guaranteed network like the Internet, QoS may seriously be degraded owing to network delay, delay jitter, and packet loss. Therefore, we have to carry out QoS controls to keep the quality of service as high as possible.

1.5 QoS control

There are various types of QoS control such as media synchronization control, causality control, consistency control, and error control for multimedia communications.

Figure 1.1. Geomagic Touch.

1.5.1 Media synchronization control

Media synchronization may be disturbed owing to network delays and skews, which are caused by many reasons such as the difference of time in capturing media among terminals, the difference of time in protocol processing, media interleaving, network delay jitter, packet loss, the difference of decoding time at the playout process, and clock difference. Media synchronization control is carried out to compensate for the network delay jitter. We can identify two types of media synchronization control: Object (or event-driven) and continuous synchronization control [53]. Object synchronization control means synchronization control among multimedia objects. The control adjusts the beginning output timings of media according to a scenario. Under continuous synchronization control, the output timings (e.g., output times and speeds) among media streams can be synchronized with each other. Media synchronization control is categorized into three types: Intra-stream, inter-stream [54]-[61], and group (or inter-destination) synchronization control [62]-[64].

(1) Intra-stream synchronization control

The intra-stream synchronization control is necessary for preservation of the timing relation between media units (MUs), each of which is an information unit for media synchronization, such as video frames and voice packets in a single media stream. The input intervals between MUs are disturbed owing to network delay jitter. The intra-stream synchronization control recovers the intervals when the destination outputs the MUs as shown in Fig. 1.2, where the destination outputs MUs after buffering.

There are several types of intra-stream synchronization control such as Skipping [45], Buffering [45], Adaptive Buffer Control (ABC) [27], Queue Monitoring (QM) [65], Virtual-Time Rendering (VTR) [45], and media adaptive buffering [66].

(2) Inter-stream synchronization control

Inter-stream synchronization control is required for keeping the temporal relationship among media streams. Lip-sync is a representative of the inter-stream synchronization, and it means the synchronization between spoken voice and the movement of the speaker's lips.

Media streams generally fall into a master stream and slave streams. The slave streams are synchronized with the master stream as shown in Fig. 1.3. In lip-sync, the voice is generally selected as the master stream because voice is more sensitive to intra-stream synchronization error than video. Under the inter-stream synchronization control, each slave stream is output in synchronization with the master stream. Therefore, the intra-stream synchronization quality of the slave stream may largely be degraded by the inter-stream synchronization control.

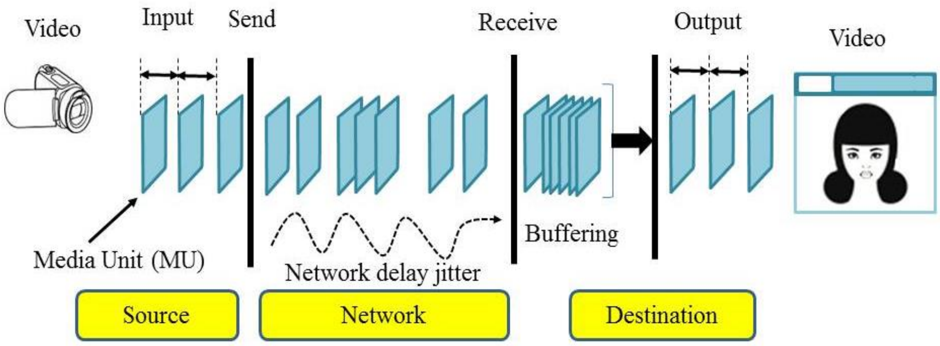


Figure 1.2. MU output under intra-stream synchronization control.

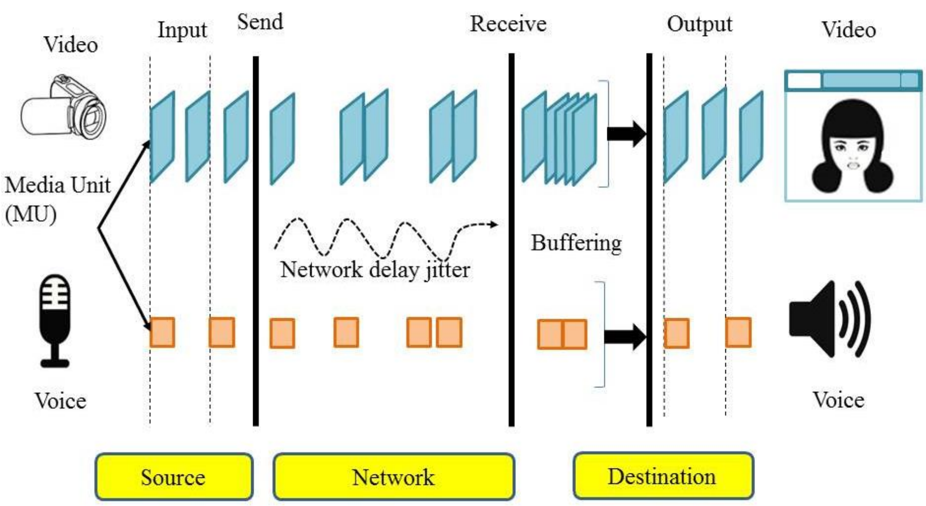


Figure 1.3. MU output under inter-stream synchronization control.

(3) Group synchronization control

Group synchronization control outputs each MU simultaneously at different destinations in multicast communication as shown in Fig. 1.4. There are three schemes for group synchronization control: The master-slave destination scheme [62], the synchronization maestro scheme [63], and the distributed control scheme [64]. The three schemes are based on the Virtual-Time Rendering (VTR) media synchronization algorithm to determine the output timing of each MU so that the timing can be the same at all the destinations.

1.5.2 Causality control

Causality means causal (i.e., temporal order) relationships among events. In networked real-time games, it is necessary to keep the causal relationships. Typical examples of causality control are the Δ-causality control [67] and adaptive Δ-causality control [68].

In the Δ-causality control, if an MU is received within a time limit (i.e., its generation time plus Δ (> 0 ms)) of the MU, the MU is saved in a buffer until the time limit, and then it is output. If the MU is received after the time limit, it is discarded since it is considered useless. In the adaptive Δ-causality control, the MU is not discarded even though it is received after the time limit. The control uses the late MU for prediction of the future position without output. The value of Δ is dynamically changed according to the maximum network delay between users' terminals.

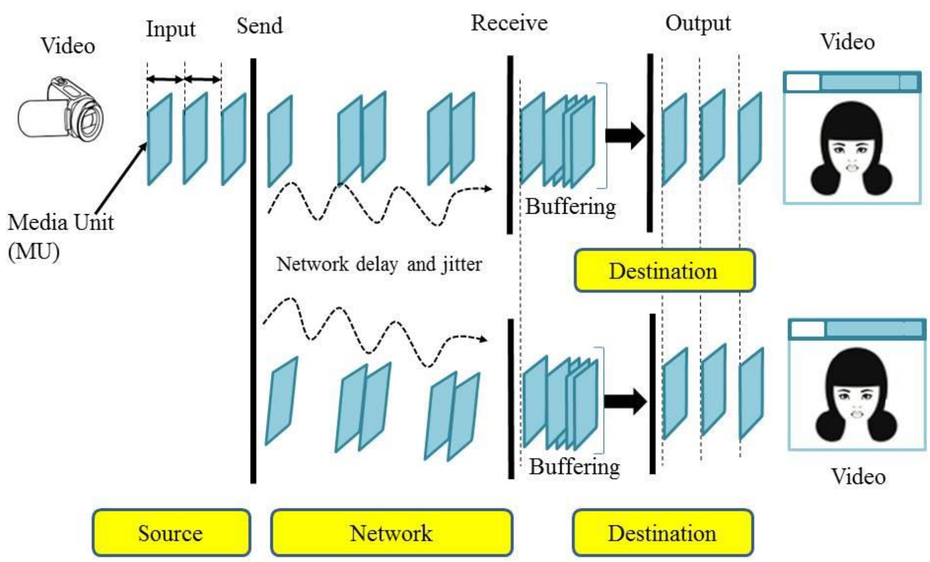


Figure 1.4. MU output under group synchronization control.

1.5.3 Local lag control

In the networked multimedia applications in virtual environments, the consistency among users' terminals is very important as the users share a 3D virtual space together. In order to keep the consistency, we can use the local lag control, which buffers local information for a constant time called the local lag according to the network delay from the local terminal to the other terminal. Therefore, when the network delay is large, it degrades the interactivity [69]. In this thesis, we employ the local lag control in a networked real-time game to keep the consistency among players. Moreover, we also use the control for joint musical performance to synchronize local sound with received sound at each terminal. In this case, we set the local lag according to the network delay from the other terminal to the local terminal.

1.6 Quality of Experience

In [39], the Quality of Experience (QoE) is defined as the overall acceptability of an application or service perceived by the end-user. Subjective QoE assessment is the most fundamental method of evaluating QoE. In the subjective QoE assessment, subjects evaluate the quality of services based on predefined evaluation criteria. However, subjective assessments are time-consuming and expensive. Therefore, objective QoE assessments are also necessary to be carried out.

1.6.1 Subjective assessment

Subjective assessment obtains user opinions about QoE (for example, the operability of haptic interface device, synchronization quality of sound, and smoothness of video). Examples of subjective assessment methods are the rating scaling method [70], SD (Semantic Differential) method [71], pair comparison method [70], constant method [72], and questionnaire method [73].

(1) Rating scale method

Rating scale method [70] is one of the most widely used methods for subjective QoE assessments. In the method, subjects carry out an assessment, and they make a judgment based on a category scale. There are two main rating scale methods known as ACR (Absolute Category Rating) and DCR (Degradation Category Rating). ACR uses five-grade quality scale and DCR uses five-grade impairment scale as shown in Table 1.1. A typical example of the rating scale method uses the five-grade scale, and subjects are asked to give a score from 1 to 5 for each stimulus. MOS (Mean Opinion Score) [74] is obtained by averaging scores of all the subjects participated in the assessment.

(2) SD method

SD method [71] is usually used to measure the connotative meaning of objects or systems. This method can be used as a QoE subjective assessment method. In the method, subjects are provided with a list of many pairs of bipolar terms such as “heavy-light,” “adequate-inadequate,” and “valuable-worthless.” For example, each subject is asked to give a score from 1 to 5 for each pair of polar terms. The appropriate bipolar terms are chosen from various points of views relating to the assessment system.

(3) Pair comparison method

In this method [70], each subject compares the quality of first stimuli to that of second stimuli, and judges which one is better. It is easy for subjects to make a judgment, but the assessment time takes longer than the other methods due to a large number of combinations.

(4) Constant method

In the constant method, the same set of stimuli is randomly presented several times in this method throughout an experiment [72]. The method can be used to measure the human perception on the smallest change of the system in each stimulus.

(5) Questionnaire survey

The questionnaire survey method collects user opinions by using a series of questions [73]. The questions in a subjective assessment should be relevant, meaningful, and easy to understand for subjects.

Table 1.1 Five grade scale

Quality scale		Impairment scale	
5	Excellent	5	Imperceptible
4	Good	4	Perceptible, but not annoying
3	Fair	3	Slightly annoying
2	Poor	2	Annoying
1	Bad	1	Very annoying

1.6.2 Objective assessment

By carrying out objective assessments at the same time as subjective assessments and analysing the relationship between subjective and objective results, users' perceptions on quality of services can be estimated from objective measures. In this thesis, we use objective assessment measures such as the number of burst balloons in a balloon bursting game, and average difference of output time of sounds between tow users. Objective assessment measures can be chosen depending on applications.

1.6.3 QoS mapping

As mentioned in Subsection 1.4, QoS over the Internet can be defined in a hierarchical structure, and QoS parameters in each level of the hierarchy are different from those in the other levels. Mapping of the QoS parameters between the levels is called QoS mapping. By performing the mapping from lower level QoS parameters to QoE parmeters, the time and cost for subjective assessment can be reduced. In this thesis, we perform QoS mappings between QoE parameters and the application level QoS parameters.

In order to investigate the relationship between subjective and objective assessment results, we use the regression analysis [75], in this thesis. Regression analysis is a popular statistical tool for the investigation of relationships between dependent variables and independent variables which are used in an experiment. The dependent variable is the output of effect, and the independent variable is the input or cause of the experiment. By the regression analysis, we understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. In QoE assessment, by making the subjective assessment results and objective assessment results as the dependent variables and independent variables, respectively, and applying regression analysis, we can clarify the relationships between user's mean opinion score and objective parameters.

1.7 Purpose of thesis

As described at the beginning of this chapter, simultaneous output of media is very important not only at each terminal but also among users' terminals. When the difference of network delay is large among the terminals, simultaneous output of media cannot be done. Therefore, QoE such as synchronization quality of media at each terminal or consistency among the terminals may seriously be degraded. In order to investigate the influences of network delay on such kinds of QoE, we focus on the joint musical performance [21], [22], [42]-[44] and network real-time games [32]-[35] in this thesis.

In the joint musical performance over a network, where multiple users play their same or different musical instruments together with the same rhythm and tempo, the synchronization quality of sound may seriously be deteriorated owing to the network delay. The local lag control [69] can be used for synchronization of sound in the joint musical performance. In [76], 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 thesis). 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. Therefore, in this thesis, we carry out subjective and objective QoE assessments to investigate the influence of network delay on the synchronization quality of sound and interactivity by handling a networked haptic drum system as an example application of the joint musical performance. We chose the networked haptic drum system because the synchronization quality of sound can be clearly understood by listening to drum beats. Moreover, by using two haptic interface devices, styli of the devices can be used instead of real drumsticks. In the system, users play a drum set in the 3D virtual space together with the same rhythm at the same tempo by using haptic interface devices as drumsticks. Based on assessment results, in order to maintain the synchronization quality of sound and interactivity in the joint musical performance, we propose the dynamic local lag control, in which the local lag is dynamically set to its optimum value according to the network delay from the other terminal to the local terminal. We also enhance the dynamic local lag control so that three or more music performers can do musical performance together with high synchronization quality of sound. Furthermore, we propose the dynamic local lag control with dynamic prediction time to get higher interactivity in networked haptic musical performance.

In the networked real-time games, the interactivity is degraded when we use the local lag control to maintain the consistency among players' terminals. This is because the control buffers local information for a constant time according to network delay from the local terminal to the other terminal. The operability of haptic interface device may also be deteriorated. Also, when the network delays are largely different between the terminals, the fairness may seriously be damaged. In [77], Brun *et al.* describe that the fairness is high when the same condition is provided to all the players. To maintain the fairness high, we can use the adaptive Δ-causality control [68], which also employs the local lag control. The adaptive Δ-causality control generally sets the local lag at each terminal to the maximum network delay among the terminals. Therefore, when the maximum network delay is large, the interactivity may seriously be degraded under the control. This leads to the severe degradation of the operability of the haptic interface device, but the fairness is maintained high. Based on this relationship of the operability and fairness, we can say that there is a trade-off relationship between the operability and fairness.

In [34] and [35], the influence of network delay on the fairness is investigated, but only the fairness is focused. Thus, the trade-off relationship between the operability and fairness is not clear. It is important to investigate the relationship of the operability and fairness in detail by QoE (operability and fairness) assessment in which we should assess the comprehensive quality (i.e., the weighted sum of the operability and fairness) to examine which types of QoE have larger contribution to the comprehensive quality. Also, we should clarify how much local lag should be set at each terminal to maintain the comprehensive quality as high as possible. Furthermore, to the best of our knowledge, there is no previous work which investigates the influence of network delay on QoE such as the operability and fairness for soft objects in virtual environments. In the case of hard objects, it is found that the objects become heavier as the network delay increases [78]; that is, their characteristics change owing to network delay. The characteristics of soft objects may change in different ways from those of hard objects; for example, they may become harder and/or heavier as the network delay increases. However, it is not clear how the characteristics change. We should carry out QoE assessments for soft objects to understand clearly how their characteristics change when the network delay increases. Therefore, in this thesis, we carry out subjective and objective QoE assessments to investigate the influence of network delay on the operability of haptic interface device and fairness among players. We handle a networked balloon bursting game, where two players burst balloons (i.e., soft objects) in a 3D virtual space competitively by using haptic interface devices. We clarify which types of QoS control should be used in networked real-time games to maintain both the operability and fairness high.

1.8 Organization of thesis

The rest of the thesis is organized as follows. In Chapter 2, we investigate influences of network delay on the synchronization quality of sound, interactivity, and comprehensive quality (i.e., the weighted sum of the synchronization quality of sound and interactivity) as QoE in the networked haptic drum system [79]. For sound synchronization at each user’s terminal, we employ the local lag control. As a result, we illustrate that QoE depends on the network delay, and there exists the optimum value of local lag according to the network delay.

In Chapter 3, we propose the dynamic local lag control which dynamically changes the local lag according to the network delay for sound synchronization in the joint musical performance by two users [80]. We also make a comparison between the dynamic local lag control and the local lag control with fixed values of local lag by QoE assessments in the networked haptic drum system. Consequently, we illustrate that the dynamic local lag control can keep the synchronization quality of sound high for the joint musical performance by two users.

In Chapter 4, we enhance the dynamic local lag control for the joint musical performance so that three or more users can play musical instruments together [81]. In order to investigate the effect of the enhanced control, we carry out subjective and objective QoE assessments in the networked haptic drum system. From assessment results, we demonstrate that the enhanced control can keep the synchronization quality of sound high for the joint musical performance by three or more users.

Although the synchronization quality of sound can be maintained by the dynamic local lag control, it still needs to improve the interactivity. In Chapter 5, in order to improve the interactivity, we use the prediction control together with the dynamic local lag control in the networked haptic drum system [82]. The scheme not only dynamically changes the local lag according to the network delay to get a high synchronization quality of sound, but also outputs the position information by predicting the future position to keep the interactivity high. Based on results of QoE assessment, we illustrate that there exists the optimum value of prediction time according to the network delay.

By taking advantage of results in Chapter 5, we propose the dynamic local lag control with dynamic control of prediction time in Chapter 6 [83]. The proposed control keeps the synchronization quality of sound and interactivity high by changing the local lag and prediction time dynamically according to the network delay. As a consequence of QoE assessment, we demonstrate that the control is effective.

In Chapter 7, we investigate the influences of network delay on the operability of haptic interface device and the fairness between players for soft objects in the balloon bursting game by carrying out QoE assessments subjectively and objectively [84], [85]. In consequence of the assessments, we show that the operability depends on the network delay from the local terminal to the other terminal, and the fairness is mainly dependent on the difference in network delay between the players' terminals. We also clarify that the contribution of the fairness is larger than that of the operability to the comprehensive quality (i.e., the weighted sum of the operability and fairness). Assessment results further show that the adaptive Δ-causality control, which adjusts the output timing of terminals to a terminal with the latest output timing, should be used as QoS control to maintain the fairness as high as possible.

Finally, we summarize and conclude the thesis in Chapter 8 and discuss future challenges and prospects based on the results in this study.

Chapter 2

Influences of network delay on QoE in joint musical performance

2.1 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 [21], [22], [42]-[44]. 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 chapter, 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 [21], [22], [76], the synchronization quality of sound may seriously be deteriorated owing to the network delay. To achieve high synchronization quality of sound in joint musical performance, we can use the local lag control [69]. As we explained in Subsection 1.5.3, the interactivity is degraded as the control buffers local information for a constant time called the local lag.

In [76], 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 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 thesis). 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.

The synchronization quality of sound at each user's terminal may be achieved both in symmetric and asymmetric delay cases by setting the local lag to the same value as the network delay from the other terminal to the local terminal. By doing this, the local sound may be synchronized with received sound from the other terminal. In order to use the local lag control for sound synchronization, we need to clarify how to change the local lag according to the network delay. In the joint musical performance, the local lag should not always be set to the same value as the network delay. This is because the interactivity is severe in the joint musical performance.

Therefore, in this chapter, we deal with a networked haptic drum system to investigate the effect of the local lag control on the synchronization quality of sound, interactivity, and comprehensive quality as QoE. 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.

The remainder of this chapter is organized as follows. Section 2.2 describes the networked haptic drum system. Section 2.3 explains our assessment environment. Assessment results are presented in Section 2.4, and Section 2.5 summarizes the chapter.

2.2 Networked haptic drum system

2.2.1 System configuration

The configuration of the networked haptic drum system is shown in Fig. 2.1, where two users 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. 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 (Geomagic Touch [10]). 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 haptic interface devices, and a sound depending on the drum component is generated.

2.2.2 Usage of system

We here present two types of usage of the networked haptic drum system. The first usage is concerned with remote teaching, and the second one is for joint performance. In both types of usage, haptic media streams are transmitted bi-directionally between the terminals.

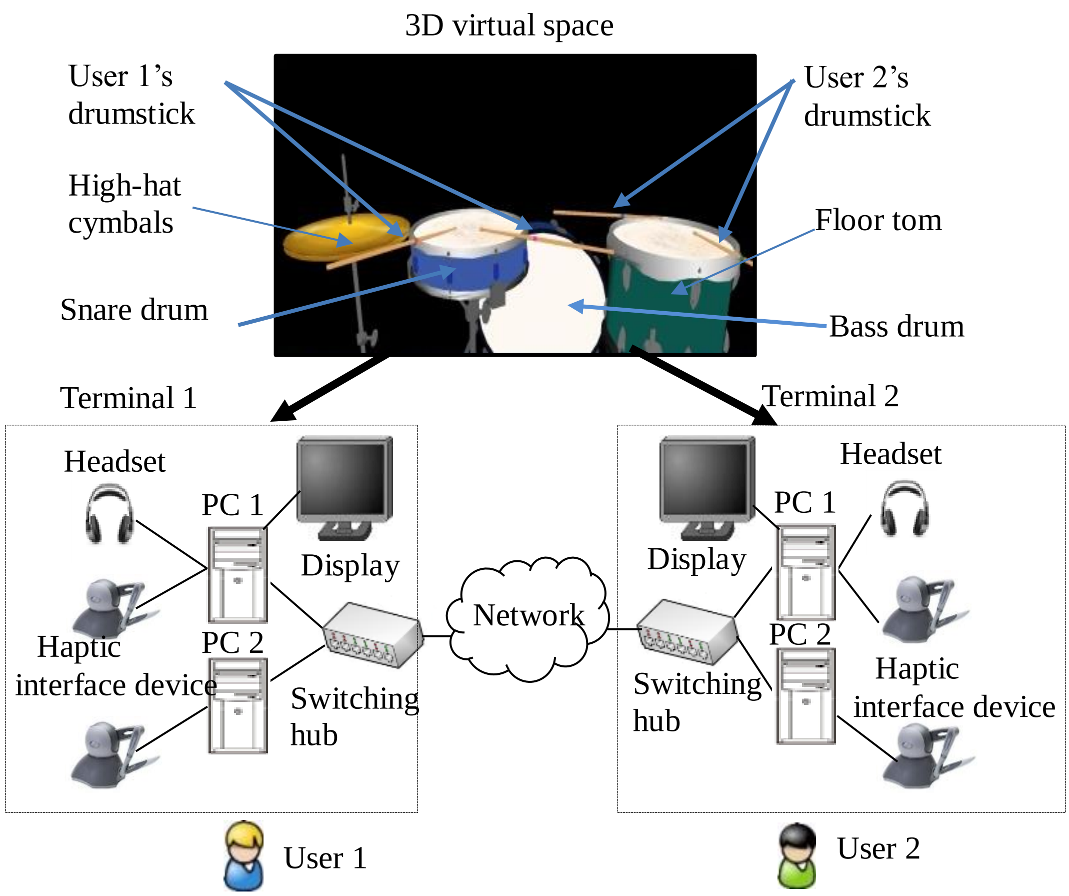


Figure 2.1. Configuration of networked haptic drum system.

(1) Remote teaching

In remote teaching, the teacher controls the student's left drumstick with his/her left one and the student's right one with his/her right one. The student is able to learn the teacher's drumstick movements while perceiving the reaction force through the haptic interface devices. There are two modes in this usage, namely, *passive mode* and *active mode*. In the passive mode, the student is passively taught by the teacher how to play the drum set. The teacher controls the student's drumsticks and hits the drum set by using his/her haptic interface devices. The student just needs to hold the drumsticks, and the haptic sensation experienced by the teacher is conveyed to the student terminal. In the active mode, the student actively plays the drum set and shows his/her drum playing ability to the teacher, who just holds his/her haptic interface devices.

(2) Joint musical performance

Joint musical performance is carried out by two users who play the drum set together by using their respective haptic interface devices. There are several types of joint musical performance in playing real-world drum sets. Sometimes users play their drum sets with the same rhythm at the same tempo. In this case, they hit their same drum set components at the same time. In another case, users beat the drum sets with their respective portions of a single tune. Occasionally, in drum battles, users play alternately with their favorite rhythms and compete to each other.

In this chapter, we handle joint musical performance in which the two users play the drum set with the same rhythm at the same tempo. In this case, it is important for the two users to hit the drum set at the same time.

2.2.3 Calculation of reaction force

There are three types of reaction force in 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. Another is produced when the user hits a drum component by a drumstick.

The reaction force $\boldsymbol{F}_1$ 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 [18] as follows: $\boldsymbol{F}_1 = -(K_s\boldsymbol{x}_1 + K_d\boldsymbol{v})$, where K_s is the spring coefficient, K_d is the damper coefficient, $\boldsymbol{x}_1$ is defined as 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. 2.2, in which the penetration depth $|\boldsymbol{x}_1|$ 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 $\boldsymbol{x}$ and $\boldsymbol{v}$ are equal to zero. The other is perceived only in remote teaching. In this case, a user (i.e., teacher or student) can get the reaction force $\boldsymbol{F}_2 = K_s\boldsymbol{x}_2$ from the other user, where $\boldsymbol{x}_2$ denotes a vector from the tip of the user's drumstick to that of the other user's drumstick (see Fig. 2.3).

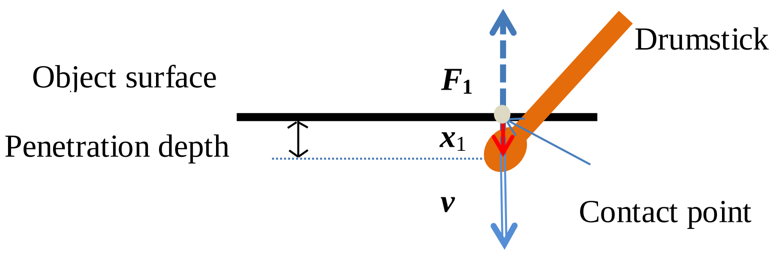


Figure 2.2. Reaction force when drum component is hitten.

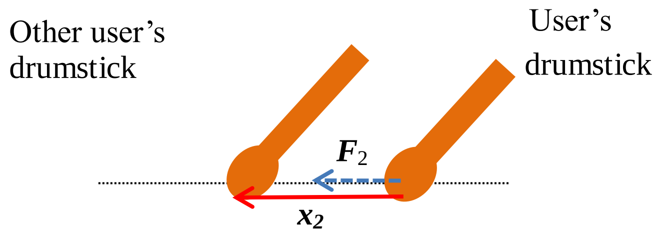


Figure 2.3. Reaction force when user remotely controls other user.

In remote teaching, each of the teacher and student feels $m_d\boldsymbol{g} + \boldsymbol{F}_1 + \boldsymbol{F}_2$. In the joint performance, each user perceives the resultant force $m_d\boldsymbol{g} + \boldsymbol{F}_1$. We carried 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/s^2, 0.2 N/mm, and 0.1 Nms/mm, respectively.

2.3 Assessment environment

2.3.1 Assessment system

Figure 2.4 shows our assessment system, where two terminals are connected to each other via a network emulator (NIST Net [86]) which is used instead of the network shown in Fig. 2.1. 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 [87]-[90], such as the Virtual-Time Rendering (VTR) algorithm [45]; 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 *delay 1* here) from terminal 2 to terminal 1 to the same value as that (*delay 2*) in the opposite direction. In the asymmetric delay case, delay 1 is not equal to delay 2. We also call the local lag at terminal 1 *local lag 1* and that at terminal 2 *local lag 2*. In this chapter, we use only PC 1 at each terminal, and each subject employs only the right drumstick for simplicity.

2.3.2 Assessment methods

We carried out subjective QoE assessment with 16 subjects (males and females) whose ages were between 20 and 28. In the symmetric delay case, we employ two rhythms (*rhythms 1 and 2*) at two tempos (slow and fast) in order to investigate the influence of drumstick movements. Rhythm 2 is more difficult than rhythm 1 since each subject need to move his/her drumstick between different drum components in rhythm 2. 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. 2.5 (a)). The high-hat cymbals are hit at all the 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 the times (see Fig. 2.5 (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. Thus, the distances of the drumstick movements are different between the two rhythms. The rhythms are used in 8 beats or 4 beats rhythms and popular in jazz and rock music. As for slow and fast tempos, each subject hits the drum set at 60 bpm (beats per minute) and 100 bpm, which are often used in fast ballads and slow rock music, respectively [91].

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

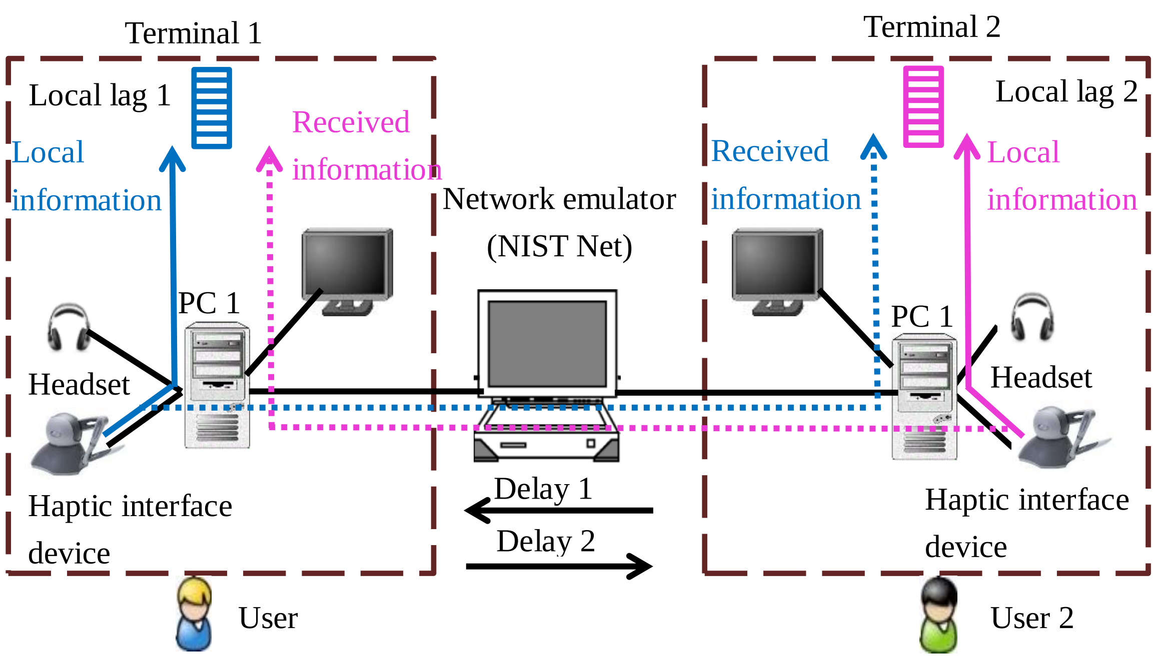
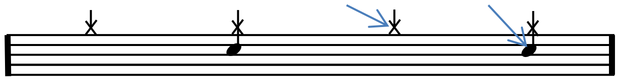
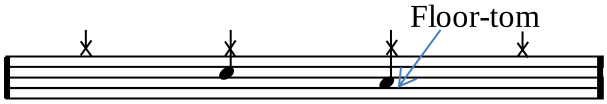


Figure 2.4. Configuration of assessment system.



(a) Rhythm 1



(b) Rhythm 2

Figure 2.5. Rhythms.

each combination of rhythm and tempo. In each combination of rhythm and tempo, the constant delay and local lags were selected in random order for each pair. 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 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*) [74] 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, rhythm 1 at the fast tempo, rhythm 2 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 twenty minutes, and the total assessment time per pair was about one and half hours including rests and practices.

In the symmetric delay case, where delay 1 is equal to delay 2, we changed the delays 1 and 2 from 0 ms to 150 ms at intervals of 50 ms, and local lags 1 and 2 were changed from 0 ms to 150 ms at intervals of 25 ms. The assessment time per pair for each combination of rhythm and tempo was about twenty minutes, and the total assessment time per pair was about one and half hours including rests and practices.

In the asymmetric delay case, we employ only rhythm 1 at the slow tempo and rhythm 2 at the fast tempo because the results of the other combinations of rhythm and tempo are almost the same as those of rhythm 1 at the slow tempo in the symmetric delay case. We carried out the assessment for several combinations of 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).

Objective assessment was also carried out at the same time as the subjective QoE assessment. We adopted the root mean square error (RMSE) [92] 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.

2.4 Assessment results

2.4.1 Subjective assessment

(1) Symmetric delay case

We plot MOS values of synchronization quality of sound for various values of the constant delay (delay 1 or 2) as a function of local lag (local lag 1 or 2) in Figs. 2.6 through 2.9. In the figures, the 95% confidence intervals are also plotted. We further show MOS values of interactivity only for the rhythm 1 at the slow tempo in Fig. 2.10 since the other combinations of rhythm and tempo had almost the same MOS values as those in Fig. 2.10. We draw MOS values of comprehensive quality in Figs. 2.11 through 2.14.

In Figs. 2.6 through 2.9, we see that there exists the optimum value of local lag for each constant delay. The reason is that the synchronization quality of sound becomes higher as the difference between the local lag and the constant delay becomes smaller. From the figures, we also find that the optimum value of local lag is not always equal to the constant delay when the constant delay is large. The figures further reveal that the MOS value at the optimum value of local lag tends to decrease as the constant delay becomes larger.

In Fig. 2.10, we observe that the MOS values become smaller as the local lag becomes larger. We also notice that the differences in the MOS value among the constant delays are small. Therefore, we can say that the interactivity depends on only the local lag, and it is hardly dependent on the constant delay.

In Figs. 2.11 through 2.14, we find almost the same tendencies as those in Figs. 2.6 through 2.9. That is, 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.

(2) Asymmetric delay case

For rhythm 1 at the slow tempo, we plot MOS values of synchronization quality, those of interactivity, and those of comprehensive quality at terminal 1 in Figs. 2.15 through 2.17, respectively, for the following four combinations of constant delays (delay 1: 0 ms, 50 ms, 100 ms, and 150 ms; delay 2: 0 ms). In Fig. 2.18, we show only MOS values of comprehensive quality at terminal 2; we do not show MOS values of synchronization quality and those of interactivity since they had similar tendencies to those of comprehensive quality. In the figure, we show only results when constant delays 1 and 2 are 50 ms and 0 ms, respectively, because the other combinations of constant delays 1 and 2 had similar tendencies as that in Fig. 2.18. We also plot MOS values of comprehensive quality at terminal 1 in Fig. 2.19 for the following three combinations of constant delays (delay 1: 100 ms; delay 2: 0 ms, 50 ms, and 100 ms). For the same reason as in Fig. 2.18, we draw only MOS values of comprehensive quality at terminal 2 in Fig. 2.20 when delays 1 and 2 are 100 ms and 50 ms, respectively. For rhythm 2 at the fast tempo, we draw MOS values of comprehensive quality at terminal 1 in Fig. 2.21 for the same combinations of delays 1 and 2 as those in Figs. 2.15 through 2.17. In Fig. 2.22, we plot MOS values of comprehensive quality at terminal 2 only when delays 1 and 2 are 50 ms and 0 ms, respectively. The 95% confidence intervals are also included in the figures.

In Fig. 2.15, we see that there exists the optimum value of local lag 1 for each delay 1. The reason is that the synchronization quality of sound becomes higher as the difference between local lags 1 and 2 becomes smaller when delay 1 is small. However, from the figure, we also find that the optimum value of local lag 1 is somewhat smaller than delay 1 when delay 1 is large. The figure further reveals that the MOS value at the optimum value of local lag 1 tends to decrease as delay 1 becomes larger.

In Fig. 2.16, we observe that the MOS values decrease as local lag 1 becomes larger. We also notice that the differences in the MOS value among the values of delay 1 are very small. Therefore, the interactivity depends on only local lag 1, and it is hardly dependent on delay 1.

From Figs. 2.17 and 2.21, we find almost the same tendencies as those in Fig. 2.15. That is, there exists the optimum value of local lag 1 for each delay 1, and the MOS value at the optimum value of local lag 1 tends to decrease as delay 1 becomes larger. We also see in the figures that there are not so much differences between the combinations of rhythm and tempo.

In Fig. 2.19, we see that there exists the optimum value of local lag 1 for delay 1. We also notice that there are not so much differences among the values of delay 2.

In Figs. 2.18 and 2.22, we see that the MOS values of comprehensive quality at terminal 2 do not depend on local lag 1. The MOS values decrease as local lag 2 becomes larger. In the figures, we see that local lag 2 of 0 ms has the highest MOS values. Moreover, in Fig. 2.20, we see that local lag 2 of 50 ms has the highest MOS values. Therefore, we can say that the optimum value of local lag 2 depends on only constant delay 2.

From Figs. 2.17 through 2.22, we notice that the optimum values of local lag 2 are the same as those of local lag 1 when delay 1 is equal to delay 2.

From the above considerations, we can conclude that the optimum value of local lag 1 depends on only delay 1, and local lag 2 is dependent on only delay 2. Also, the optimum value of local lag does not depend on any combination of terminal, rhythm, and tempo.

We plot the optimum value of local lag obtained from Figs. 2.6 through 2.22 versus delay 1 or 2 in Fig. 2.23. From the figure, we see that the relationship between the optimum value of local lag and constant delay is almost linear. By regression analysis [75], we obtained the following equation:

$$\varDelta_{\text{optimum}} = 0.65\,D + 7.50, \qquad (2.1)$$

where $\varDelta_{\text{optimum}}$ is the estimated optimum value of local lag, and D is the constant delay. The contribution rate adjusted for degrees of freedom [75], which shows goodness of fit with the estimated equation, was 0.99. Figure 2.23 also reveals that the optimum value of local lag is not equal to delay 1 or 2 when the delay is larger than or equal to about 100 ms. In next chapter, we propose dynamic local lag control which dynamically changes the local lag according to Eq. (2.1).

2.4.2 Objective assessment

We show the root mean square error of sound versus the local lag only for rhythm 1 at the slow tempo in Fig. 2.24, where the 95% confidence levels are also included. This is because the other combinations of rhythm and tempo had almost the same tendencies as those in Fig. 2.24. We observe in this figure that there exists the optimum value of local lag for each constant delay. By comparing Figs. 2.6 and 2.24, 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 constant delay.

2.4.3 Relation between subjective and objective assessment results

In order to investigate the relations between the objective assessment measures and the MOS values, we carried out the regression analysis [75]. As a result, we obtained the following estimated equations for the synchronization quality of sound, interactivity, and comprehensive quality:

$$S_{mos} = -0.021E_{rms} + 4.9043, \quad (2.2)$$

$$I_{mos} = -0.020\Delta + 4.909, \quad (2.3)$$

$$C_{mos} = -0.016E_{rms} - 0.006\Delta + 4.929, \quad (2.4)$$

where S_{mos}, I_{mos}, and C_{mos} denote the estimated MOS value of synchronization quality of sound, that of interactivity, and that of comprehensive quality, respectively, and Δ is local lag 1 or 2. Also, E_{rms} denotes the root mean square error of sound. The contribution rates adjusted for degree of freedom for the equations are 0.851, 0.962, and 0.896, respectively. Figure 2.25 shows the notation employed in Figs. 2.26 through 2.28, where we add the calculated values obtained from Eqs. (2.2), (2.3), and (2.4) to the evaluated MOS values shown in Figs. 2.6 through 2.8, respectively. Actually, in the figures, we find close agreement between the evaluated values and calculated ones. Therefore, we can say that the MOS values can be estimated from the root mean square of sound and/or local lag with a high degree of accuracy.

2.5 Summary

In this chapter, we investigated the effect of the local lag control by subjective QoE assessment in joint performance of a networked haptic drum system. We also carried out objective assessment at the same time as the subjective assessment. We further examined the relationship between the MOS values and the objective performance measures. As a result, we found that there exits the optimum value of local lag, and the value is not always equal to the network delay. We also saw that the optimum value of local lag at each terminal is dependent on only the network delay from the other terminal to the terminal. We also found that MOS values can be estimated from the root mean square of sound and/or local lag with a high degree of accuracy.

In the next chapter, we propose dynamic local lag control which dynamically changes the local lag according to Eq. (2.1) in Subsection 2.4.1 and confirm the effect of the control by QoE assessment.

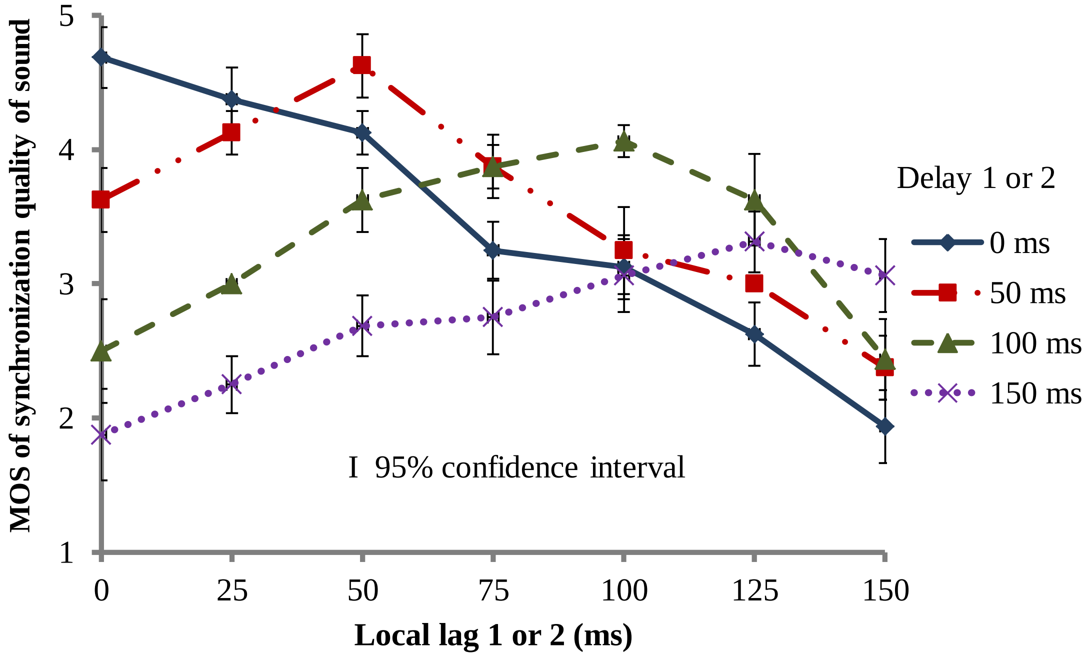


Figure 2.6. MOS of synchronization quality of sound for rhythm 1 at slow tempo in symmetric delay case.

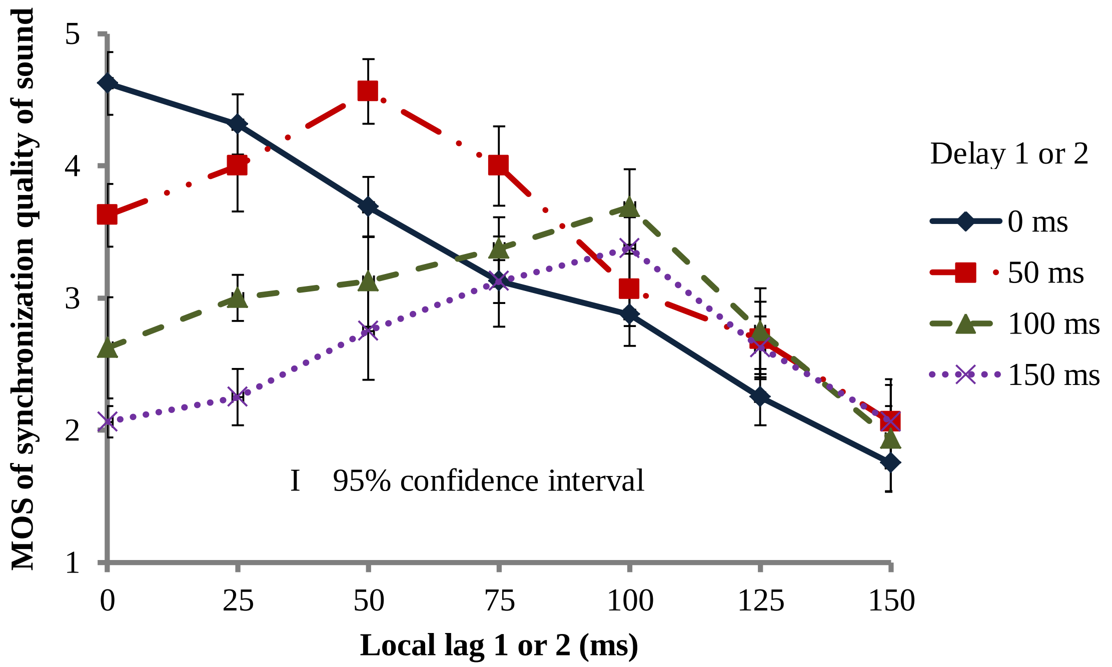


Figure 2.7. MOS of synchronization quality of sound for rhythm 1 at fast tempo in symmetric delay case.

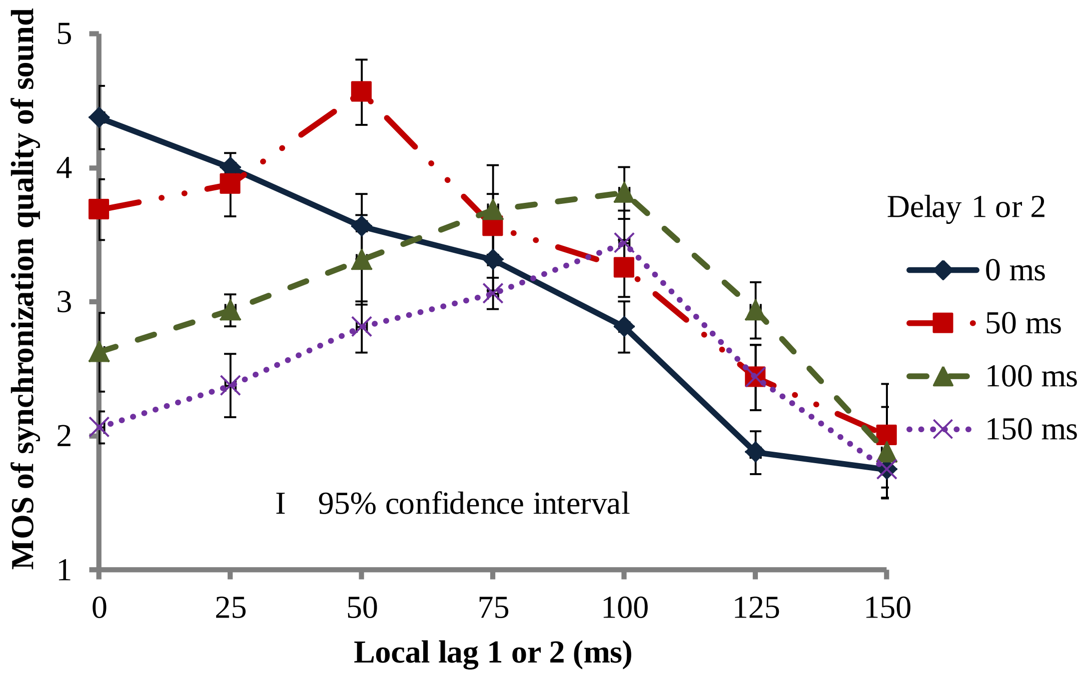


Figure 2.8. MOS of synchronization quality of sound for rhythm 2 at slow tempo in symmetric delay case.

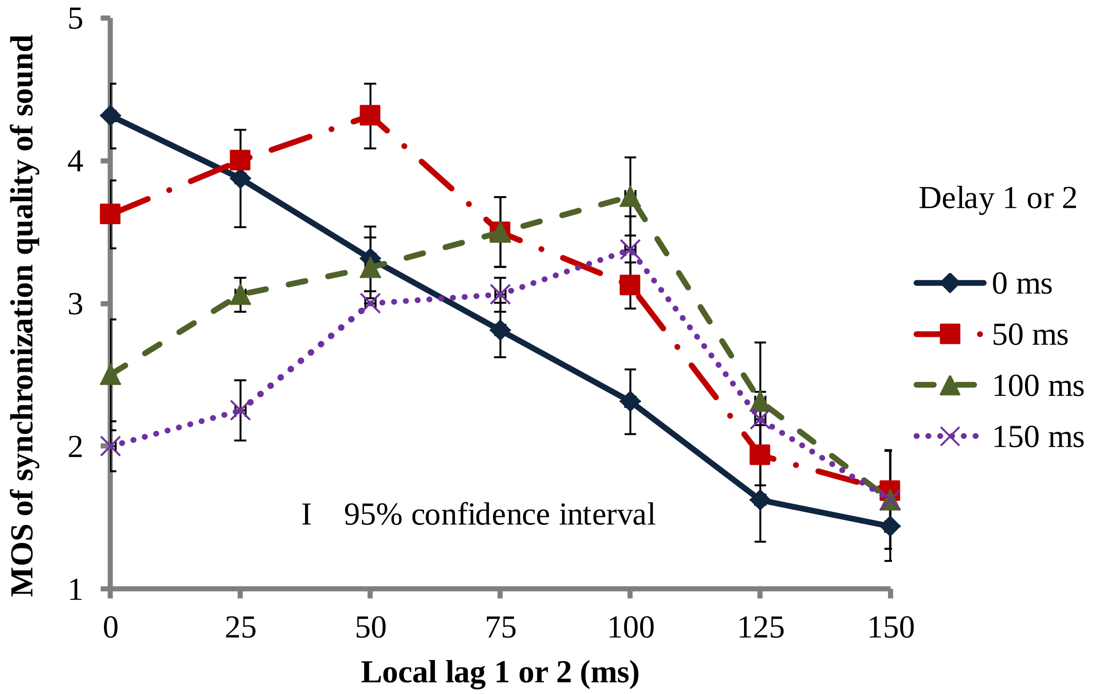


Figure 2.9. MOS of synchronization quality of sound for rhythm 2 at fast tempo in symmetric delay case.

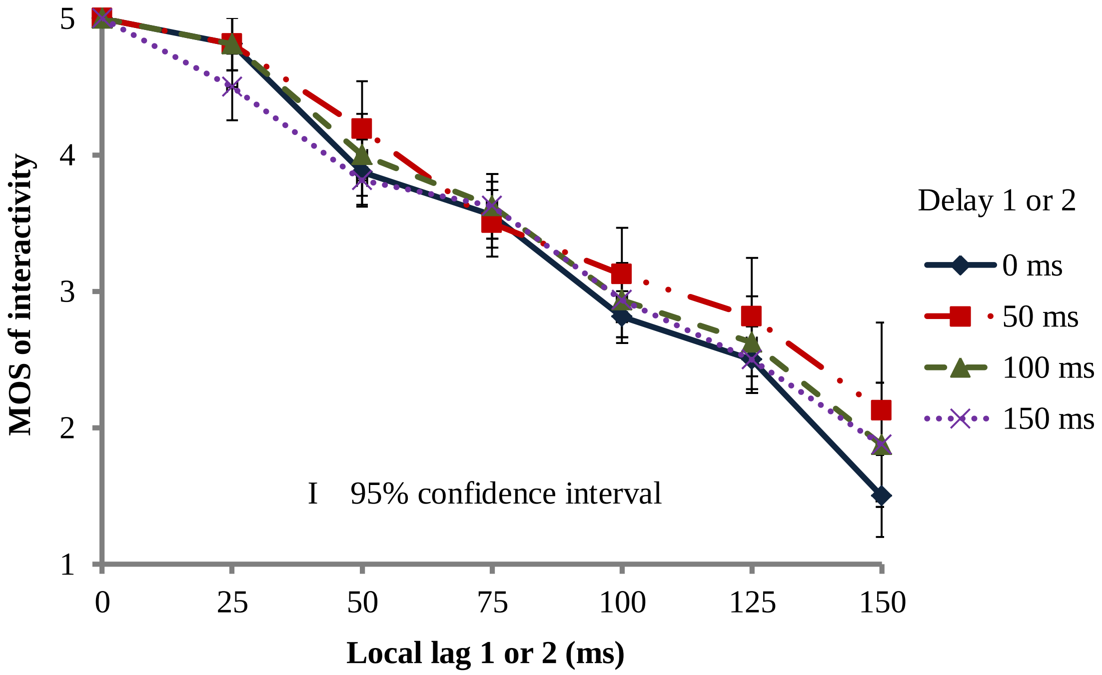


Figure 2.10. MOS of interactivity for rhythm 1 at slow tempo in symmetric delay case.

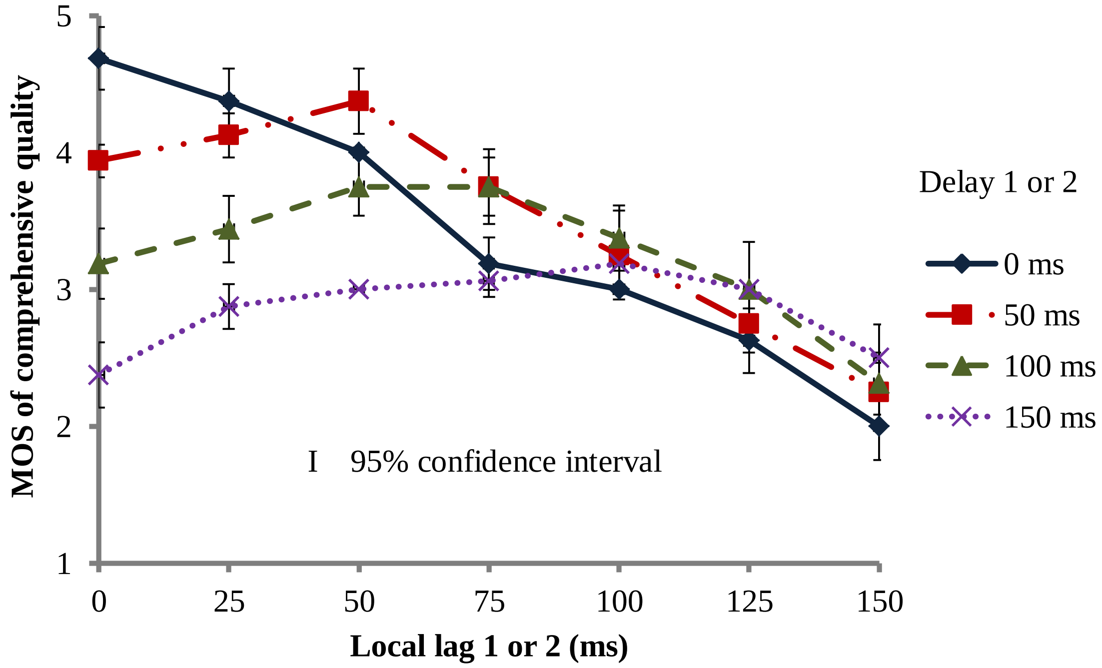


Figure 2.11. MOS of comprehensive quality of sound for rhythm 1 at slow tempo in symmetric delay case.

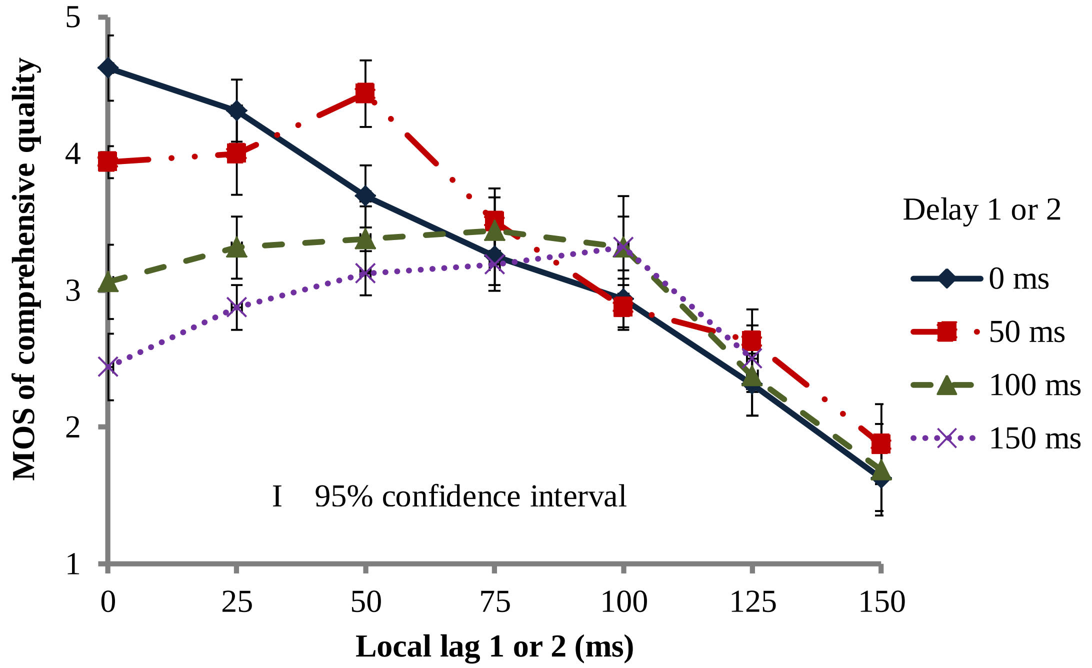


Figure 2.12. MOS of comprehensive quality of sound for rhythm 1 at fast tempo in symmetric delay case.

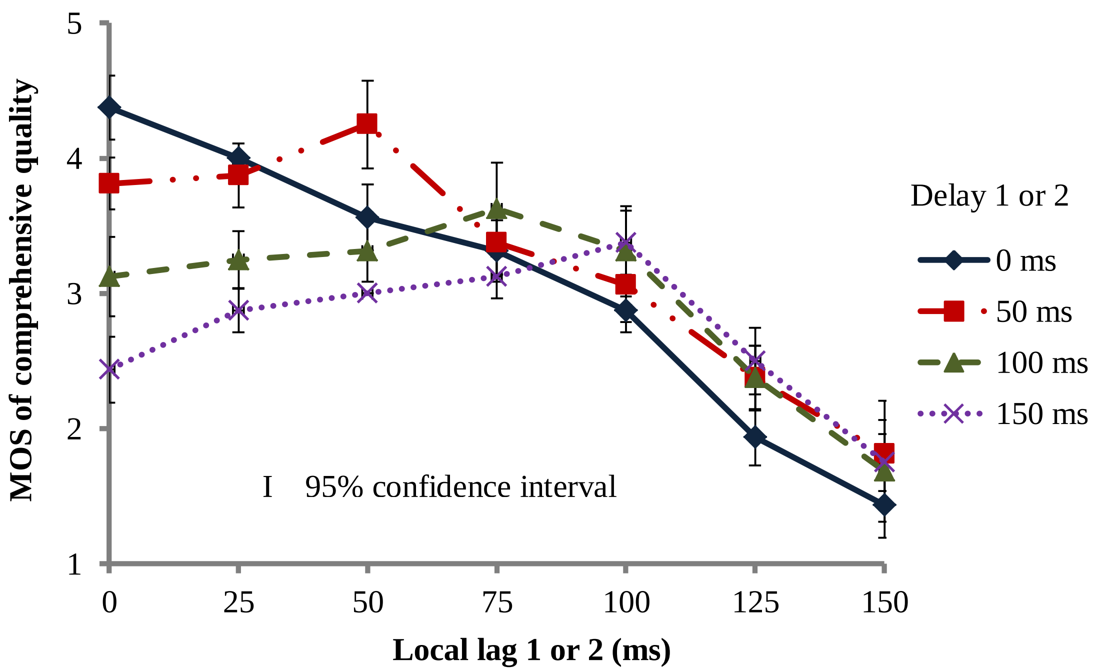


Figure 2.13. MOS of comprehensive quality of sound for rhythm 2 at slow tempo in symmetric delay case.

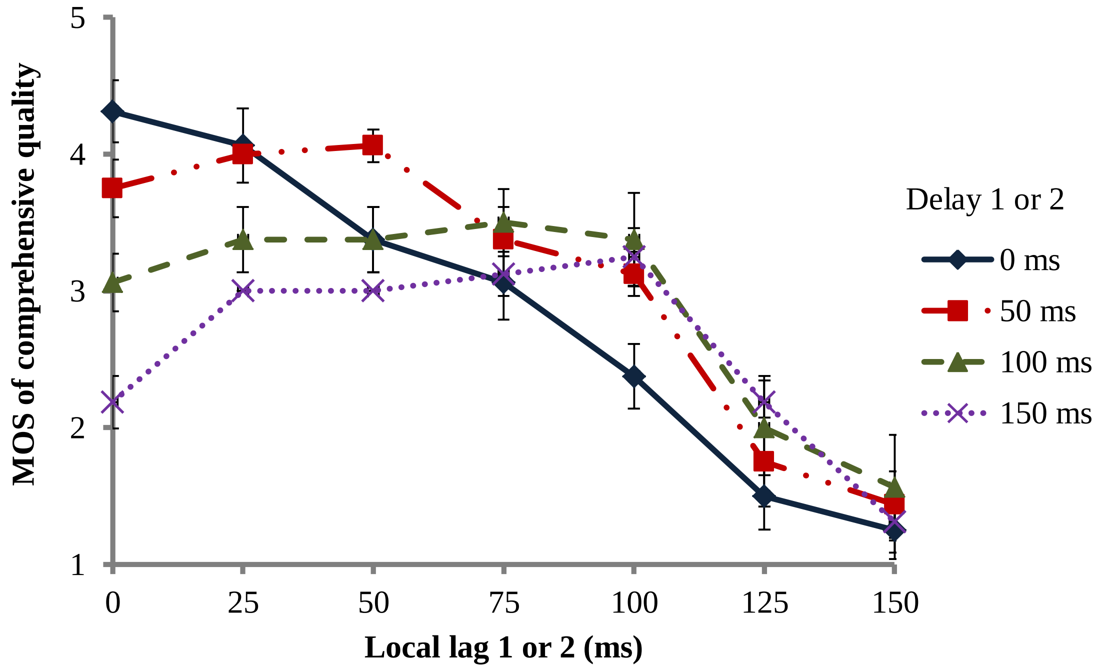


Figure 2.14. MOS of comprehensive quality of sound for rhythm 2 at fast tempo in symmetric delay case.

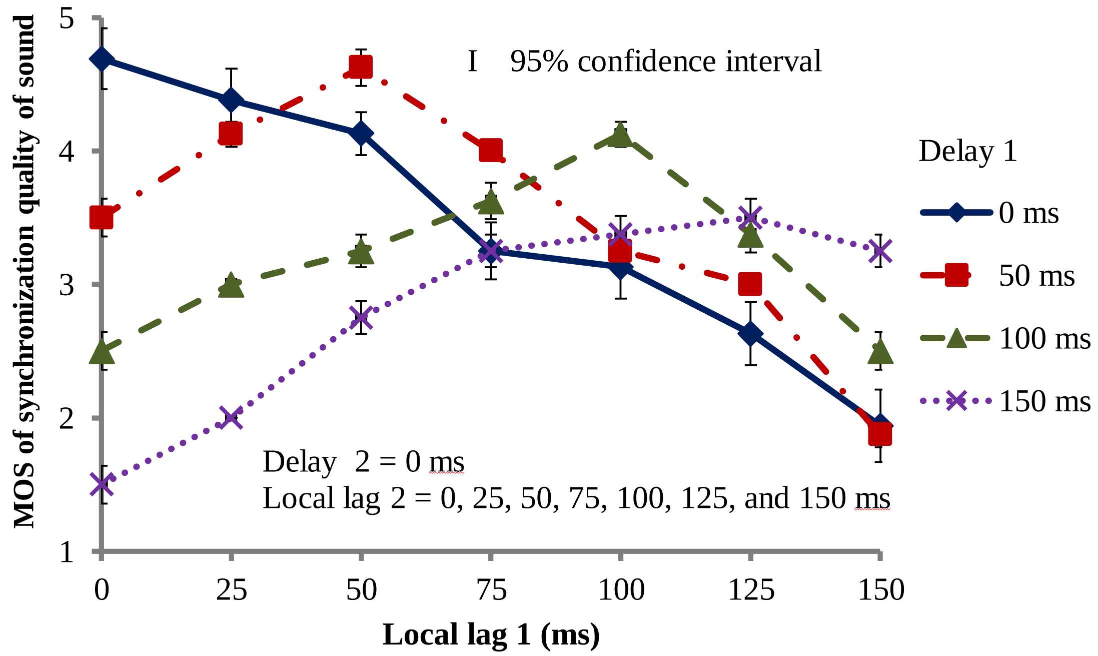


Figure 2.15. MOS of synchronization quality at terminal 1 for rhythm 1 at slow tempo in asymmetric delay case (delay 1: 0 ms, 50 ms, 100 ms, and 150 ms; delay 2: 0 ms).

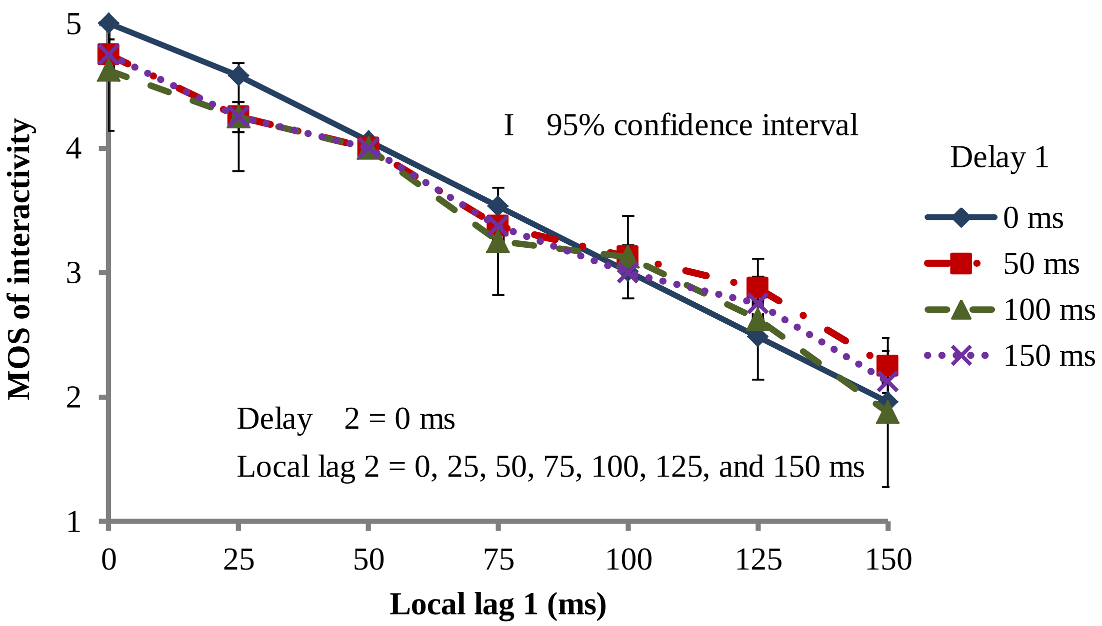


Figure 2.16. MOS of interactivity at terminal 1 for rhythm 1 at slow tempo in asymmetric delay case (delay 1: 0 ms, 50 ms, 100 ms, and 150 ms; delay 2: 0 ms).

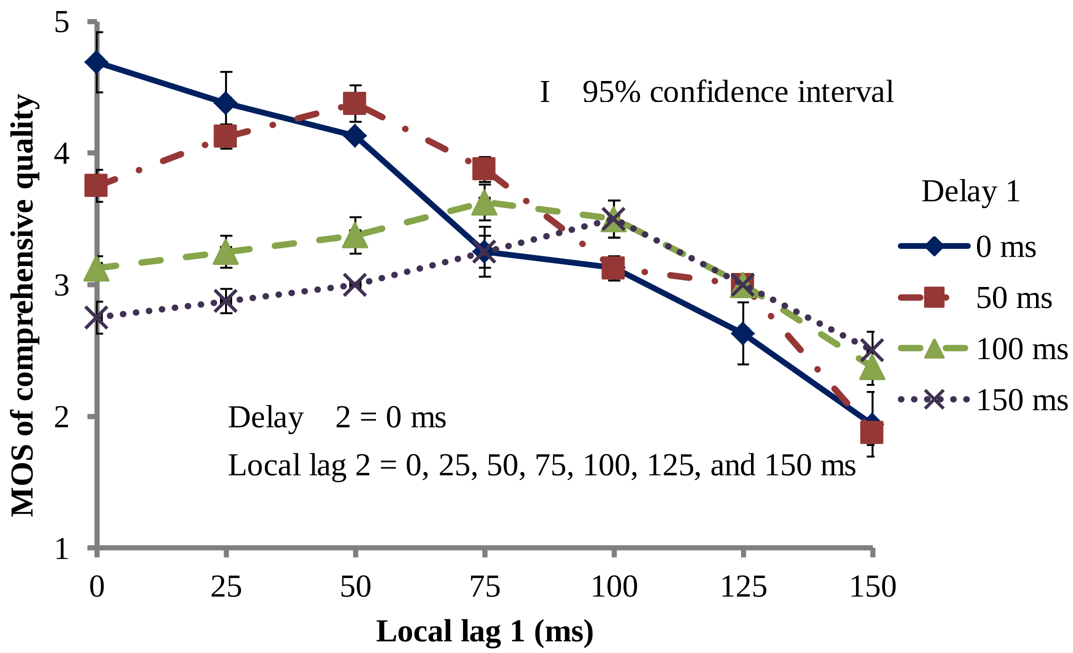


Figure 2.17. MOS of comprehensive quality at terminal 1 for rhythm 1 at slow tempo in asymmetric delay case (delay 1: 0 ms, 50 ms, 100 ms, and 150 ms; delay 2: 0 ms).

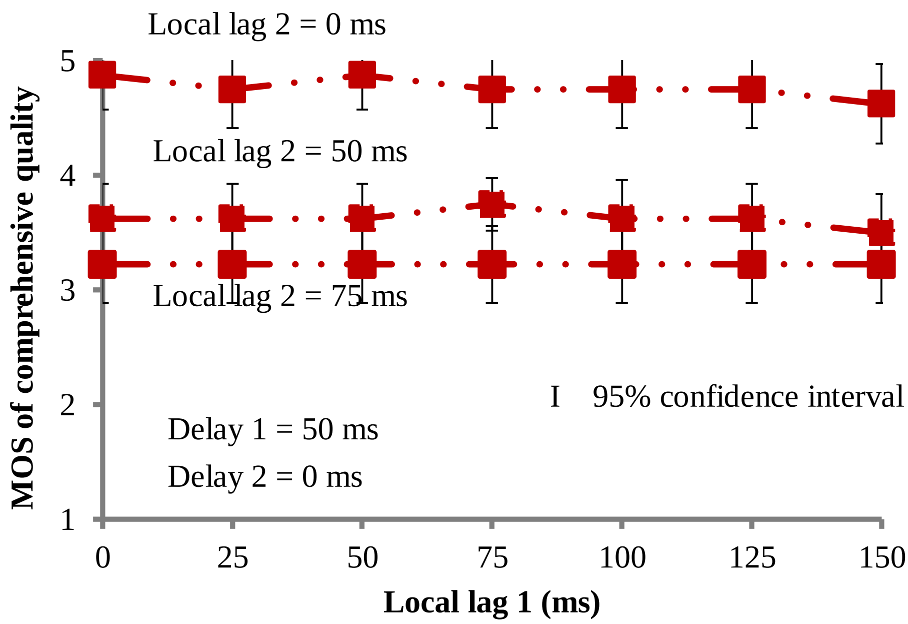


Figure. 2.18. MOS of comprehensive quality at terminal 2 for rhythm 1 at slow tempo (delays 1 and 2: 50 ms and 0 ms).

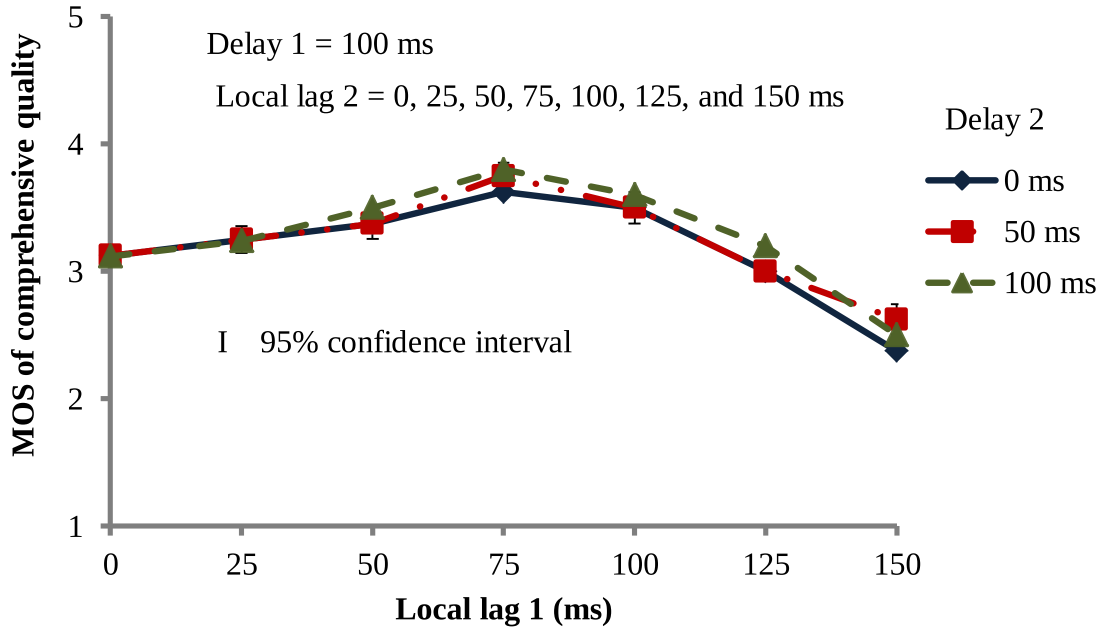


Figure. 2.19. MOS of comprehensive quality at terminal 1 for rhythm 1 at slow tempo (delay 1: 100 ms; delay 2: 0 ms, 50 ms, and 100 ms).

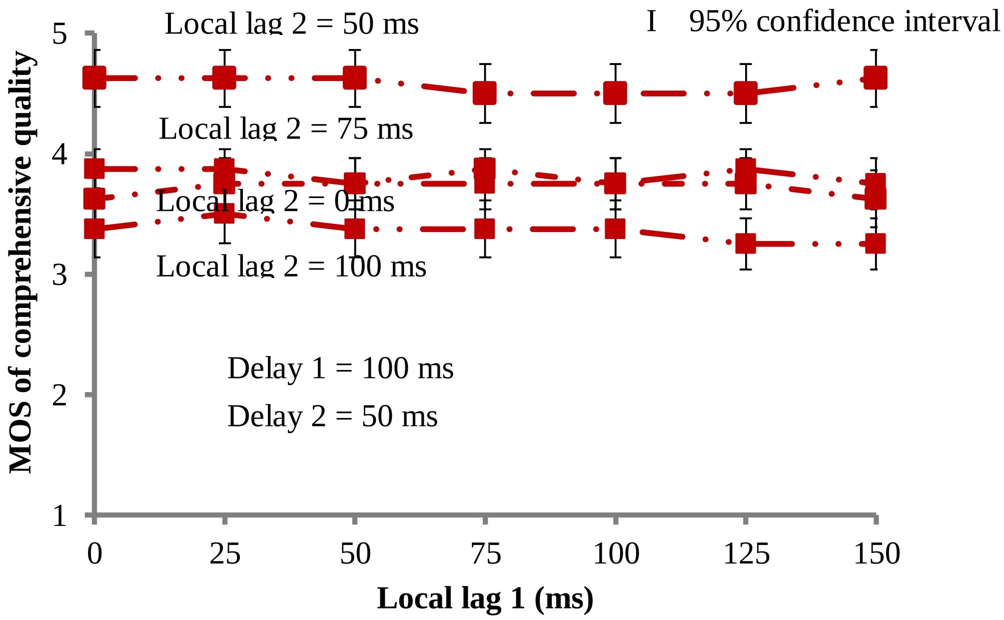


Figure. 2.20. MOS of comprehensive quality at terminal 2 for rhythm 1 at slow tempo (delays 1 and 2: 100 ms and 50 ms).

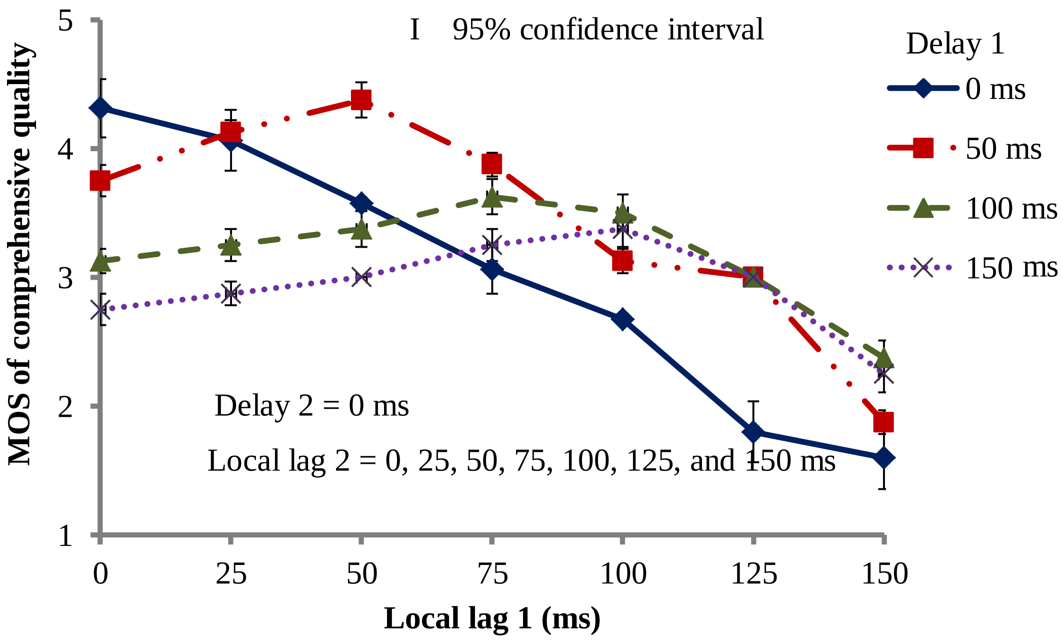


Figure. 2.21. MOS of comprehensive quality at terminal 1 for rhythm 2 at fast tempo (delay 1: 0 ms, 50 ms, 100 ms, and 150 ms; delay 2: and 0 ms).

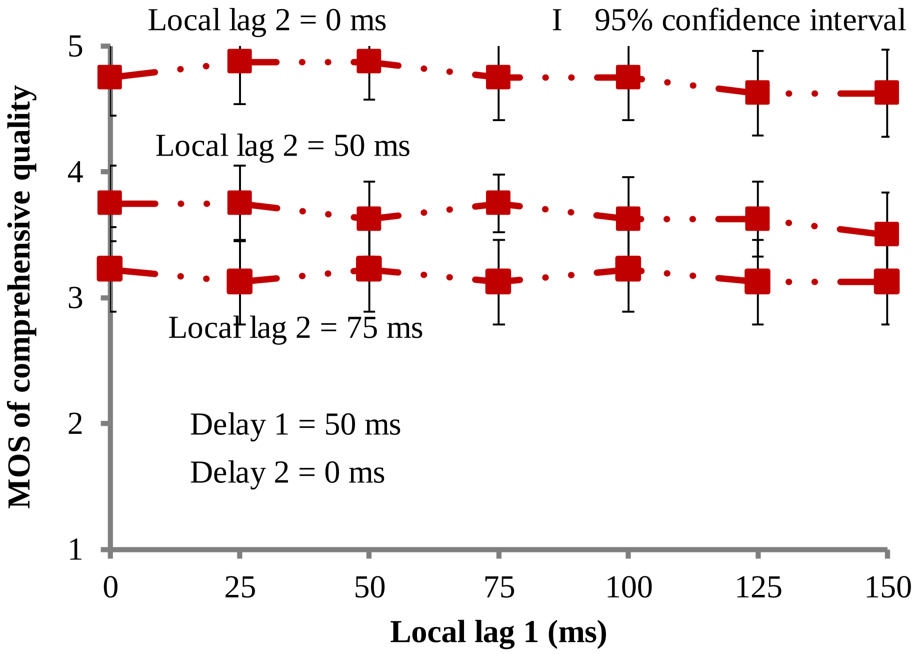


Figure. 2.22. MOS of comprehensive quality at terminal 2 for rhythm 2 at fast tempo (delays 1 and 2: 50 ms and 0 ms).

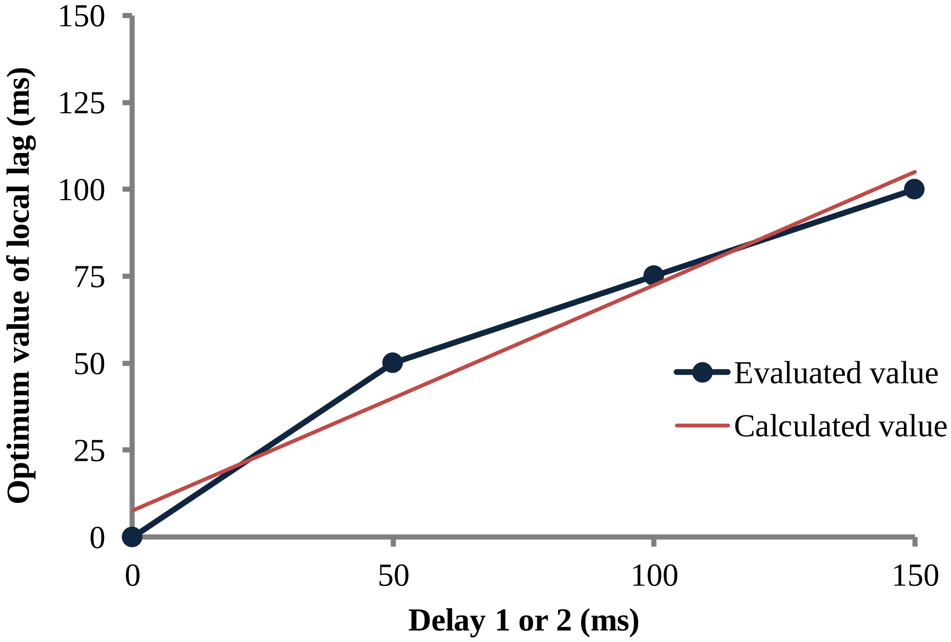


Figure 2.23. Optimum value of local lag versus constant delay.

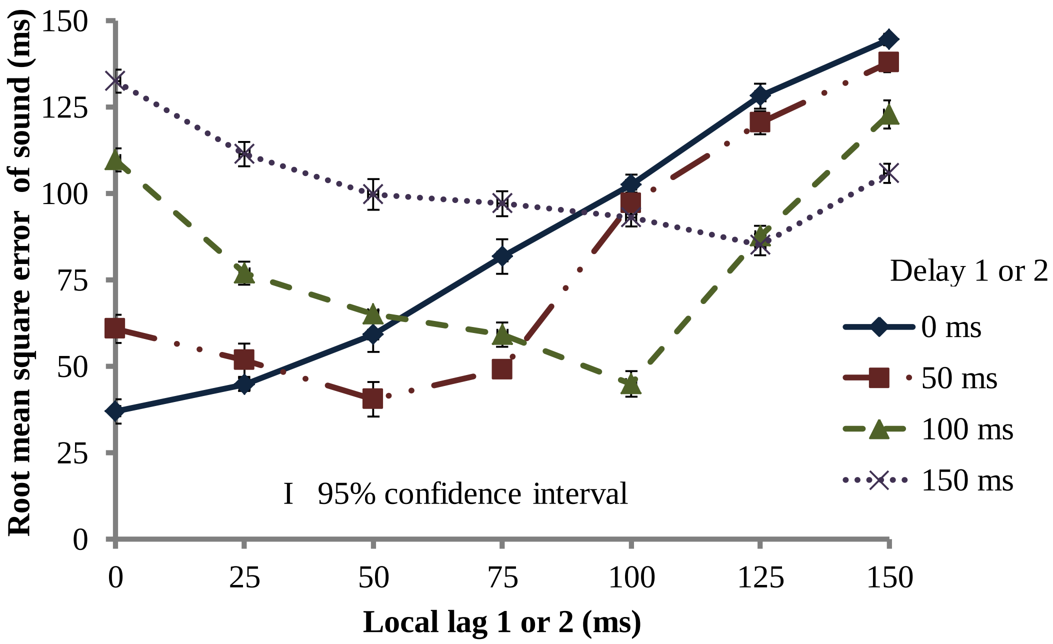


Figure 2.24. Root mean square error of sound versus local lag for rhythm 1 at slow tempo.

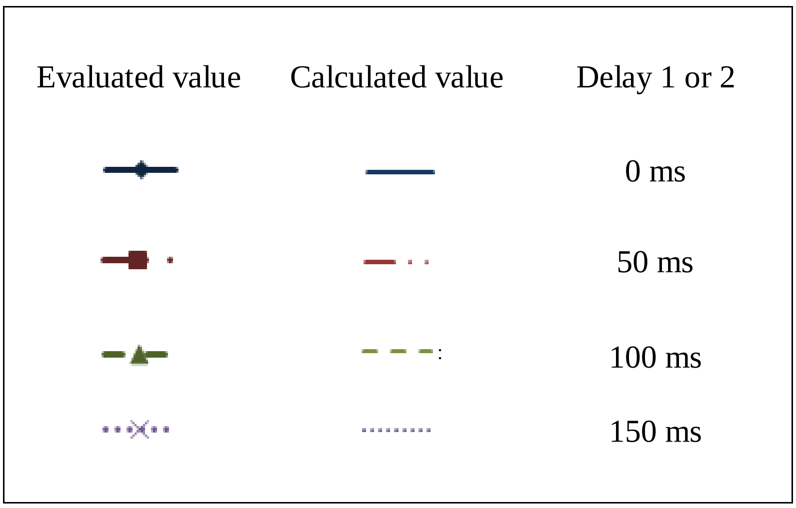


Figure 2.25. Notation of Figs. 2.26 through 2.28.

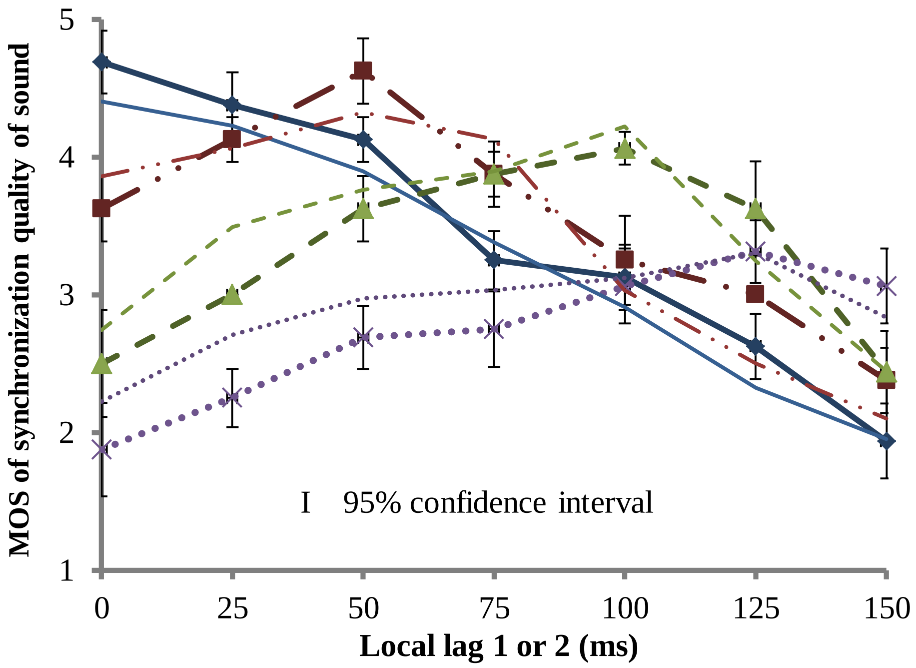


Figure. 2.26. Relation between evaluated and calculated MOS values of synchronization quality of sound for rhythm 1 at slow tempo in symmetric delay case.

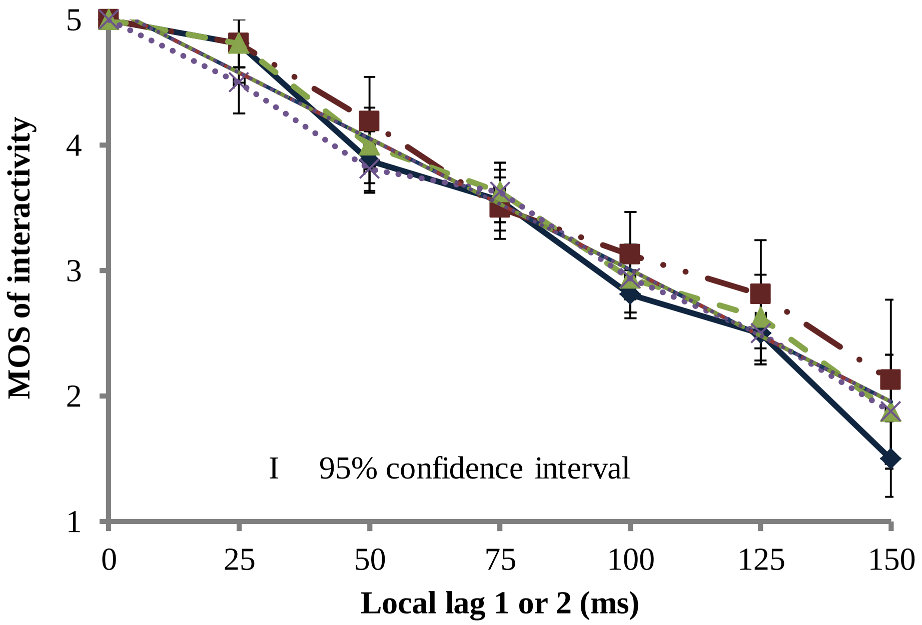


Figure. 2.27. Relation between evaluated and calculated MOS values of interactivity for rhythm 1 at slow tempo in symmetric delay case.

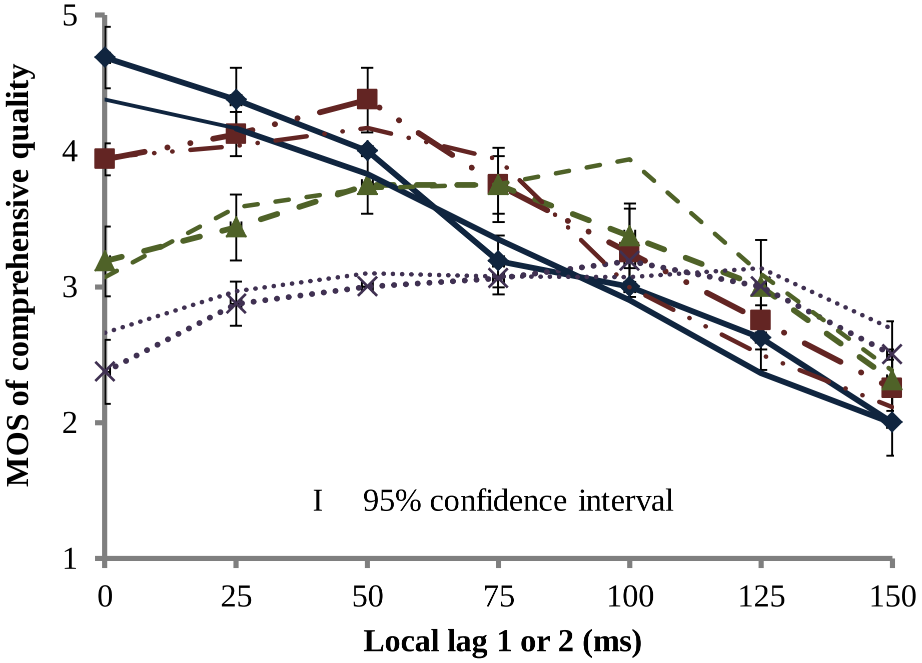


Figure. 2.28. Relation between evaluated and calculated MOS values of comprehensive quality for rhythm 1 at slow tempo in symmetric delay case.

Chapter 3

Effects of dynamic local lag control on QoE in joint musical performance

3.1 Introduction

In Chapter 2, we illustrate that there exists the optimum value of local lag for the joint musical performance. We also find 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, we notice 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, 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 system for joint musical performance. Two haptic interface devices are employed at each terminal as the left and right drumsticks.

The remainder of this chapter is organized as follows. Section 3.2 explains the conventional local lag control in the joint musical performance. Section 3.3 proposes the dynamic local lag control, and Section 3.4 presents our assessment environment. Assessment results are presented in Section 3.5, and Section 3.6 concludes the paper.

3.2 Conventional local lag control in joint musical performance

Under the conventional local lag control between two terminals, 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. 3.1. 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. 3.1, 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.

3.3 Dynamic local lag control in joint musical performance

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. 3.2, 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 $\Delta_{optimum}$ ms) obtained in Chapter 2 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:

$$\Delta_{optimum} = 0.637D + 6.578, \tag{3.1}$$

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 haptic interface device, and the sequence number of the server loop [45].

The contribution rate adjusted for degrees of freedom [75], 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. (3.1) is somewhat different from the equation of the optimum value of local lag in Chapter 2, 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 Chapter 2 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. (3.1) in the assessment where both left and right drumsticks are employed at each terminal. Therefore, to confirm that Eq. (3.1) 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 3.5, and the same equation was obtained. Therefore, Eq. (3.1) can be used in the assessment of the dynamic local lag control.

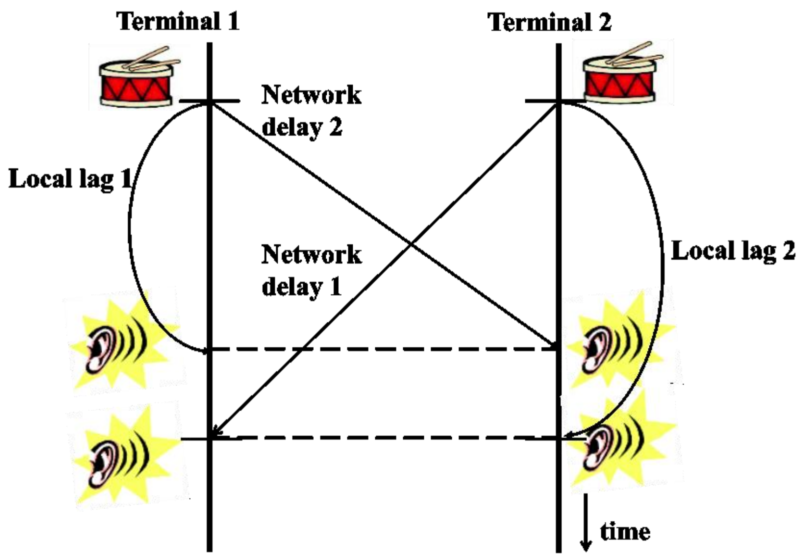


Figure 3.1. Asynchronization of sound under conventional local lag control.

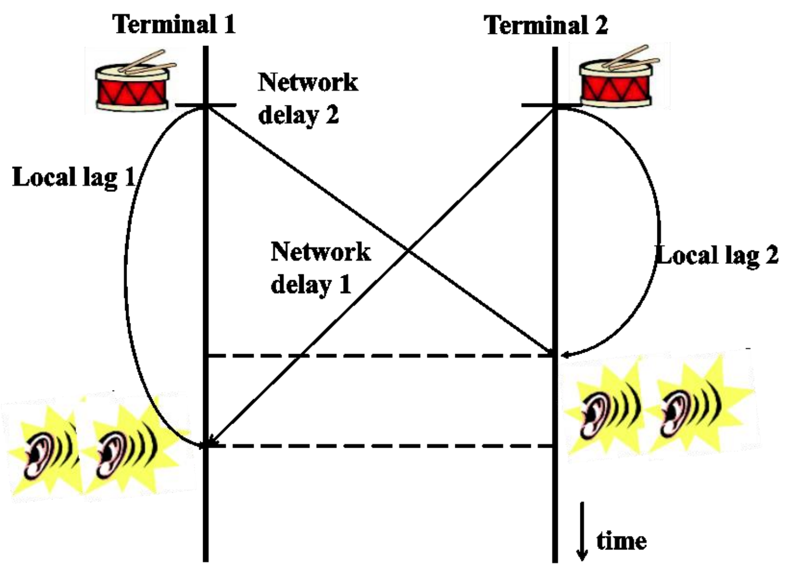


Figure 3.2 Synchronization of sound under dynamic local lag control.

3.4 Assessment environment

3.4.1 Assessment system

Figure 3.3 shows our assessment system, where two terminals are connected to each other via a network emulator (NIST Net [86]). 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 [87]-[90], such as the Virtual-Time Rendering (VTR) algorithm [45]; 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 *delay 1* here) from terminal 2 to terminal 1 to the same value as that (*delay 2*) in the opposite direction. In the asymmetric delay case, *delay 1* is not equal to *delay 2*.

3.4.2 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 delay 1 is equal to delay 2, we employed two rhythms (*rhythms 1* and *2*) at two tempos (*slow* and *fast*) as in Chapter 2 to investigate the influence of drumstick movements.

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 Chapter 2. 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,

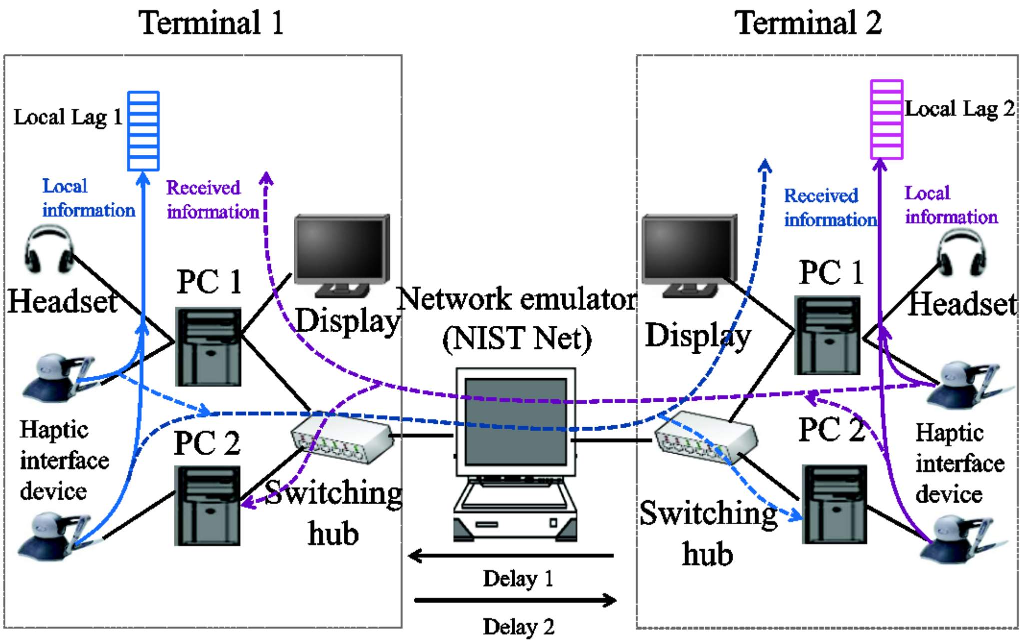


Figure 3.3 Configuration of assessment system.

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 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. (3.1). In this system, for simplicity, as in Chapter 2, we delayed only the output of sound and visual information by local lag after generating the local information.

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 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) [74] 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 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. As the assessment in Chapter 2, we adopted the root mean square error (RMSE) [92] 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.

3.5 Assessment results

3.5.1 Subjective assessment

(1) Symmetric delay case

Figure 3.4 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 (local lag 1 or 2) as a function of the constant delay (delay 1 or 2) in Figs. 3.5 through 3.7, respectively. From Figs. 3.8 through 3.10, we draw MOS values for rhythm 2 at the fast tempo. In the figures, the 95% confidence intervals are also plotted.

In Figs. 3.5 and 3.8, 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. 3.8, 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 3.6 and 3.9 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.

From Figs. 3.7 and 3.10, we find almost the same tendencies as those in Figs. 3.5 and 3.8. 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.

(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. 3.11 and 3.12, respectively, for four combinations of constant delays (delay 1: 0 ms, 50 ms, 100 ms, and 150 ms; delay 2: 0 ms). In Fig. 3.11, we draw summarized MOS values for various values of local lag 2 since the differences among the values were very small. We 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. 3.5 and 3.6, respectively, and those of synchronization quality of sound and interactivity at terminal 2 had similar tendencies to those of comprehensive quality in Fig. 3.12. We also plot MOS values of comprehensive quality at terminals 1 and 2 in Figs. 3.13 and 3.14, respectively, for three combinations of constant delays (delay 1: 100 ms; delay 2: 0 ms, 50 ms, and 100 ms). For the same reason as in Fig. 3.11, we plot summarized MOS values for various values of local lag 1 in Fig. 3.14. In Figs. 3.13 and 3.14, the notation in Fig. 3.3 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. 3.11, we see that MOS values at terminal 1 depend on only delay 1. We also note that the dynamic local lag control has the highest or second highest MOS values for each value of delay 1. In the figure, there exists the optimum value of local lag 1 for each value of 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 delay 1 when delay 1 is large. We further notice that the MOS value at the optimum value of local lag 1 tends to decrease as delay 1 becomes larger.

In Fig. 3.12, we notice that the MOS values of comprehensive quality at terminal 2 do not depend on 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 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 differences among the values of local lag 1.

Figure 3.13 reveals that the MOS values of comprehensive quality at terminal 1 do not depend on 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. 3.12, we can say that there exist the optimum values of local lags 1 and 2 according to delays 1 and 2, respectively.

In Fig. 3.14, we observe that the MOS values of the dynamic local lag control are the highest for each value of delay 2. Also, the MOS values of comprehensive quality at terminal 2 depend on only 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.

3.5.2 Objective assessment

In Fig. 3.15, we show RMSE of sound versus 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. 3.15. We observe in the figure that the dynamic local lag control has the smallest or second smallest root mean square error for each value of delay 1 or 2. By comparing Figs. 3.5 and 3.15, 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.

3.5.3 Relationship between subjective and objective assessments

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 [75]. As a result, we obtained the following three equations:

$$S_{\text{mos}} = -0.027E_{\text{rmse}} + 5.549, \tag{3.2}$$

$$I_{\text{mos}} = -0.019\Delta + 4.892, \tag{3.3}$$

$$C_{\text{mos}} = -0.019E_{\text{rmse}} - 0.005\Delta + 5.336, \tag{3.4}$$

where S_{mos}, I_{mos}, and C_{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. 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:

$$S_{\text{mos}} = -0.018|\Delta - D_c| + 4.384, \tag{3.5}$$

$$C_{\text{mos}} = -0.011|\Delta - D_c| - 0.009\Delta + 4.661, \tag{3.6}$$

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. (3.2) and (3.4), 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:

$$\hat{C}_{\mathrm{mos}} = 0.709\hat{S}_{\mathrm{mos}} + 0.276\hat{I}_{\mathrm{mos}} + 0.017 \tag{3.7}$$

where $\hat{C}_{\mathrm{mos}}$ denotes the estimated MOS value of comprehensive quality, and $\hat{S}_{\mathrm{mos}}$, and $\hat{I}_{\mathrm{mos}}$ 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. (3.5) was 0.979. From the equation, we find that since the coefficient of $\hat{S}_{\mathrm{mos}}$ is larger than that of $\hat{I}_{\mathrm{mos}}$, the contribution of the synchronization quality of sound to the comprehensive quality is larger than that of the interactivity.

3.6 Summary

In this chapter, 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 consider the dynamic local lag control for three or more terminals so that three or more users can play the joint musical performance together over a network.

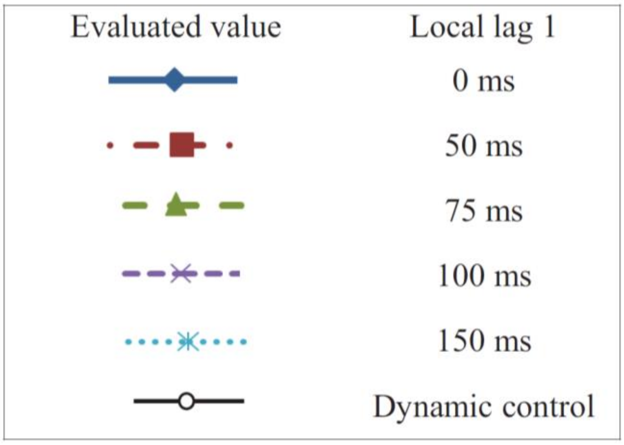


Figure 3.4. Notation in Figs. 3.5 through 3.15.

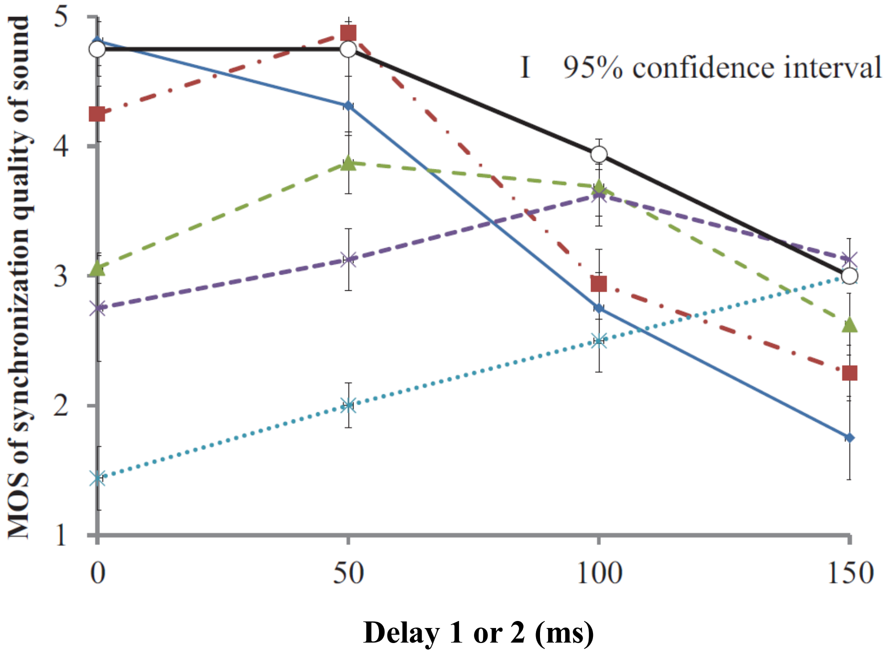


Figure 3.5. MOS of synchronization quality of sound for rhythm 1 at slow tempo in symmetric delay case.

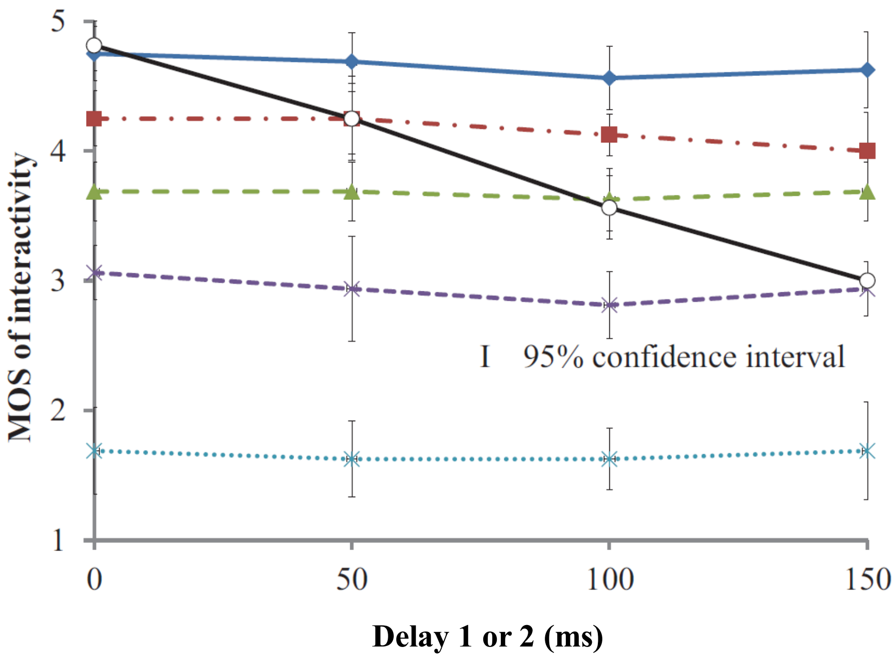


Figure 3.6. MOS of interactivity for rhythm 1 at slow tempo in symmetric delay case.

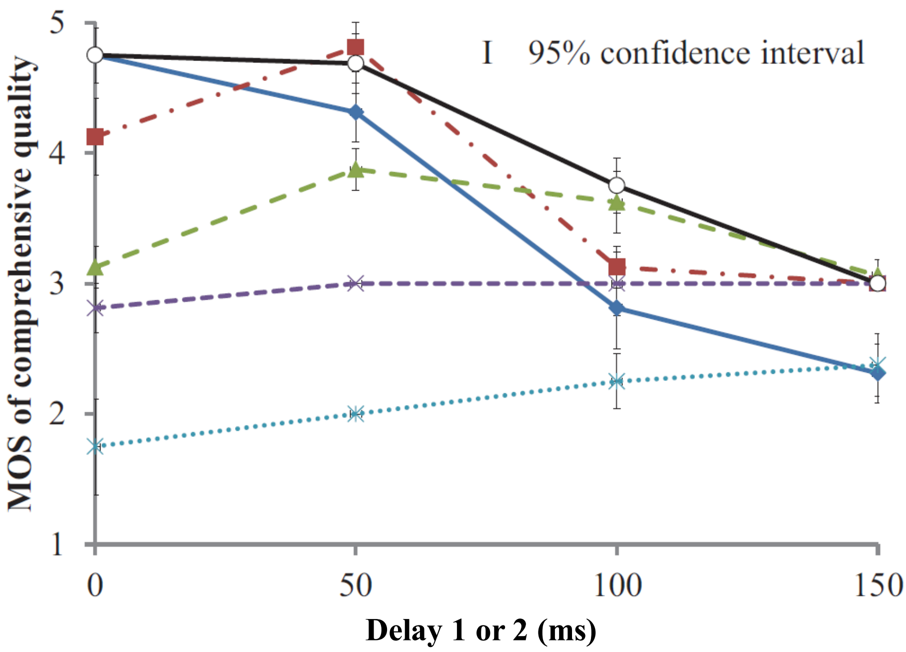


Figure 3.7. MOS of comprehensive quality for rhythm 1 at slow tempo in symmetric delay case.

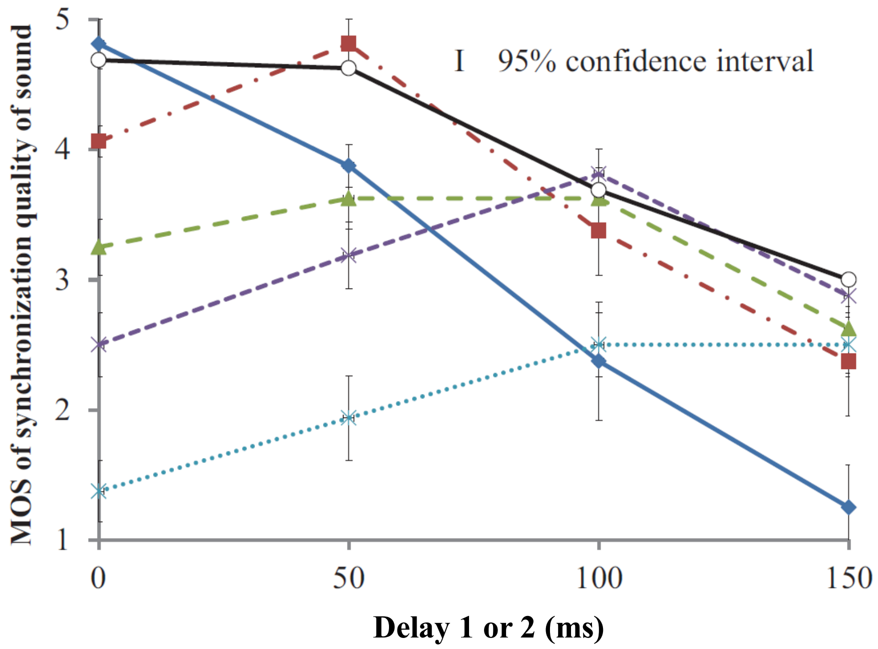


Figure 3.8. MOS of synchronization quality of sound for rhythm 2 at fast tempo in symmetric delay case.

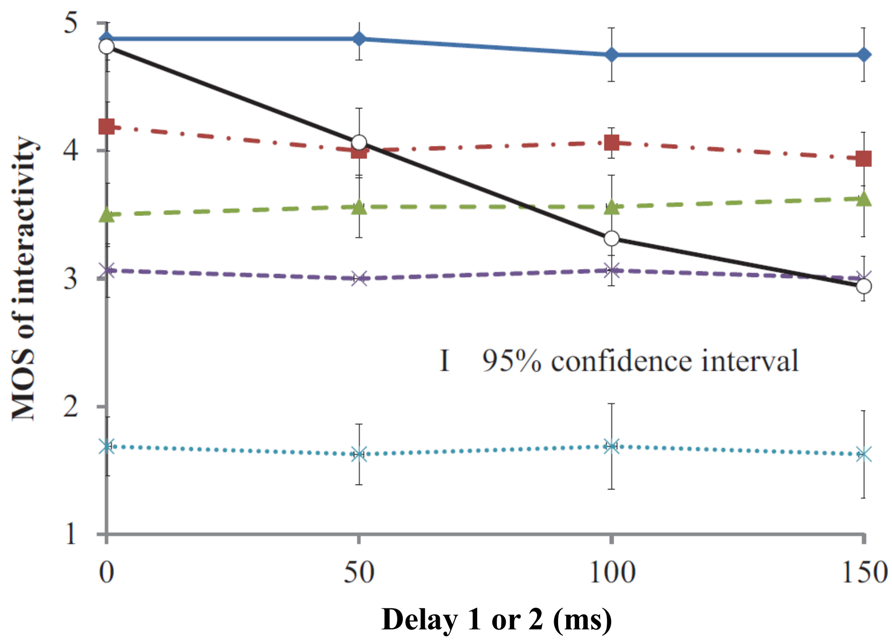


Figure 3.9. MOS of interactivity for rhythm 2 at fast tempo in symmetric delay case.

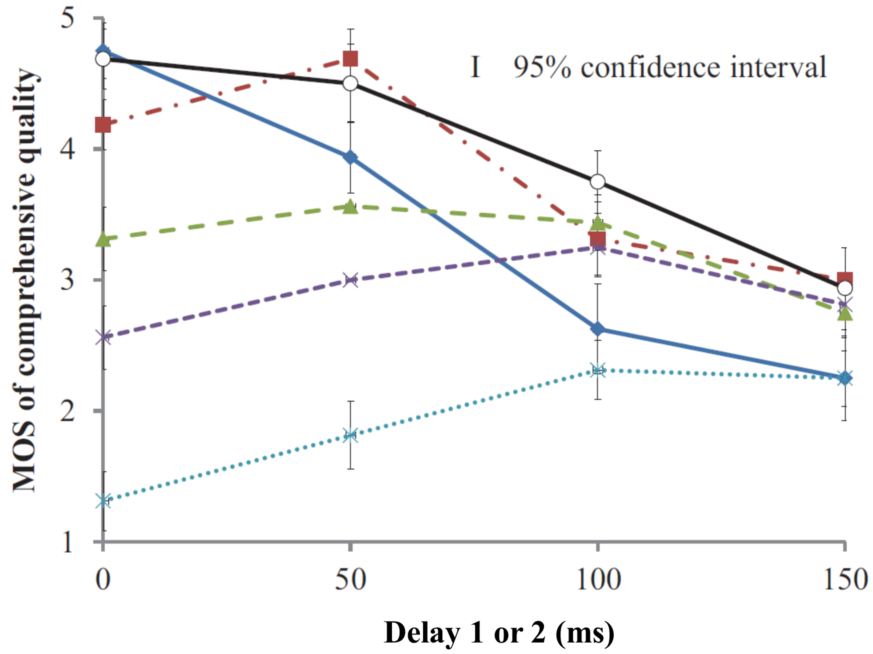


Figure 3.10. MOS of comprehensive quality for rhythm 2 at fast tempo in symmetric delay case.

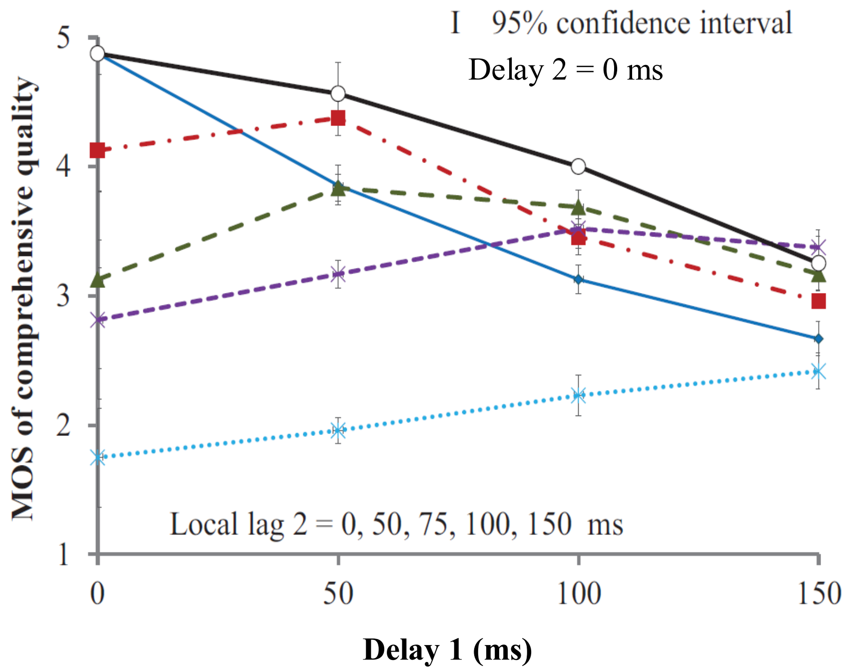


Figure 3.11. MOS of comprehensive quality at terminal 1 for rhythm 1 at slow tempo in asymmetric delay case (constant delay 1: 0 ms, 50 ms, 100 ms, and 150 ms; constant delay 2: 0 ms).

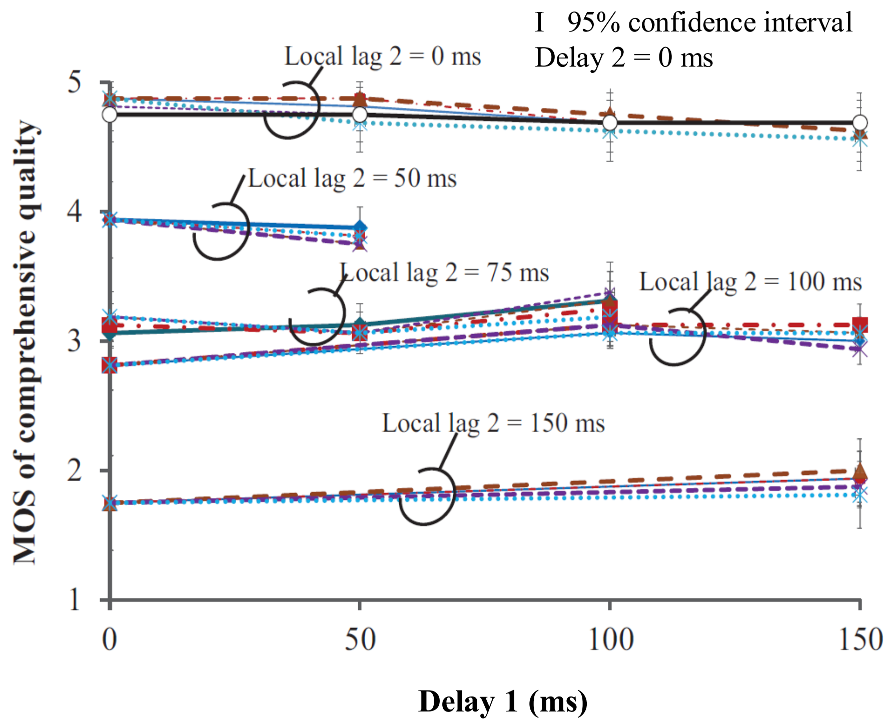


Figure 3.12. MOS of comprehensive quality at terminal 2 for rhythm 1 at slow tempo in asymmetric delay case (constant delay 1: 0 ms, 50 ms, 100 ms, and 150 ms; constant delay 2: 0 ms).

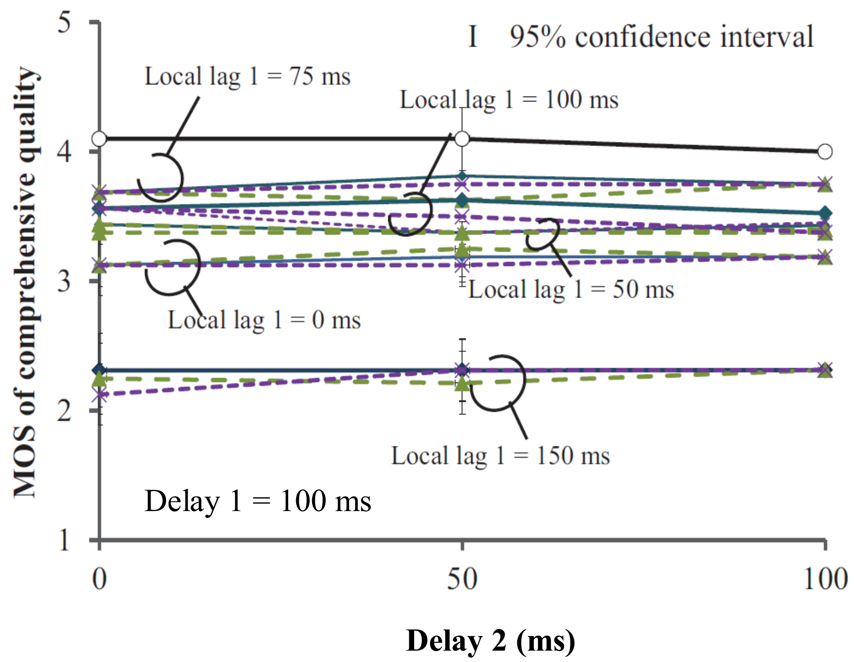


Figure 3.13. MOS of comprehensive quality at terminal 1 for rhythm 1 at slow tempo in asymmetric delay case (constant delay 1: 100 ms; constant delay 2: 0 ms, 50 ms, and 100 ms).

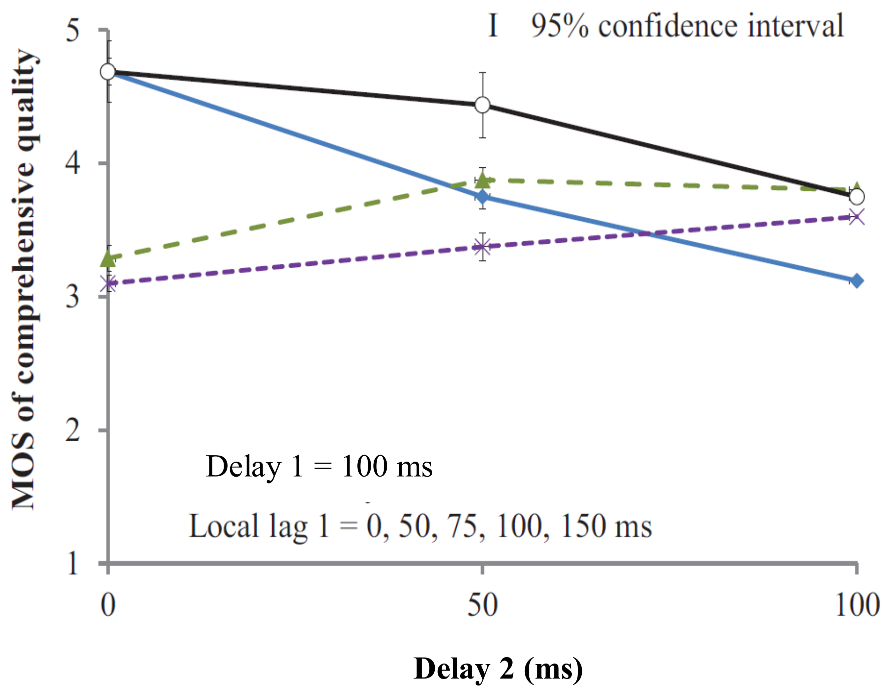


Figure 3.14. MOS of comprehensive quality at terminal 2 for rhythm 1 at slow tempo in asymmetric delay case (constant delay 1: 100 ms; constant delay 2: 0 ms, 50 ms, and 100 ms).

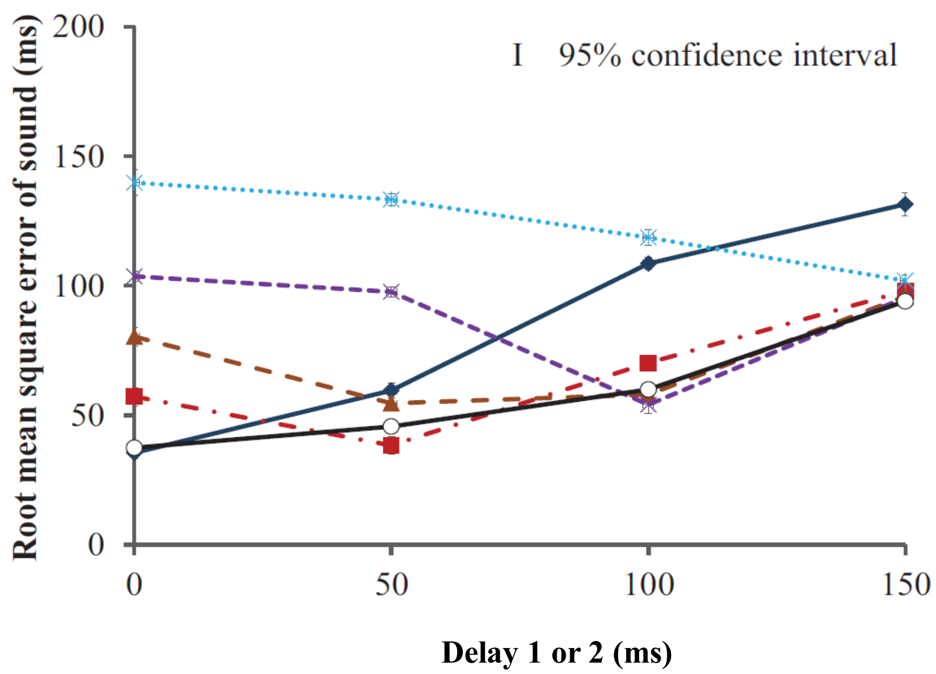


Figure 3.15. Root mean square error of sound for rhythm 1 at slow tempo in symmetric delay case.

Chapter 4

Enhancement of dynamic local lag control for more than two musical performers

4.1 Introduction

In Chapter 3, we propose the dynamic local lag control for two terminals in joint musical performance. It is important to be able to employ the control for three or more terminals so as to perform music like orchestra music.

Therefore, in this chapter, we enhance the dynamic local lag control so that three or more users can play musical instruments together through a network. In order to investigate the effect of the enhanced control, we make a comparison among the control and the following two types of control by QoE assessment subjectively and objectively. One is the adaptive Δ-causality control [68] which sets the local lag to the largest network delay among the terminals. We handle the control because it can easily adjust the output time of the local information with received information. The other is here called no control which outputs information on receiving the information. In the assessment, we use a networked haptic drum system for networked musical performance in which three users play a drum set in a 3D virtual space with the same rhythm at the same tempo.

The remainder of this paper is organized as follows. Section 4.2 describes the networked haptic drum system with three terminals. The enhanced dynamic local lag control is explained in Section 4.3. Section 4.4 describes our assessment environment. Assessment results are presented in Section 4.5, and Section 4.6 concludes the paper.

4.2 Networked haptic drum system with three terminals

The configuration of the networked haptic drum system is shown in Fig. 4.1, where we enhance the system in Chapter 2 so that three users can 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. The system consists of three terminals (*terminals 1, 2* and *3*), each of which has two PCs (*PCs 1* and *2*) connected to each other through an Ethernet switching hub

(100 Mbps). Each PC has Geomagic Touch [10] as a haptic interface device. 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 haptic interface device, and a sound depending on the drum component is generated.

4.3 Enhanced dynamic local lag control

We explain the dynamic local lag control between two terminals, and then enhance the control. The dynamic local lag control dynamically changes a buffering time (called the local lag, which is denoted by Δ ($\geq$0) ms) of local information according to the network delay from the other terminal to the local terminal. The value of Δ is calculated by the following equation as proposed in Chapter 3: $\Delta = 0.637\ D + 6.578$, where D is the *MU delay*, which is defined as the time interval from the moment an MU [45] is generated at the other terminal until the instant the MU is output at the local terminal.

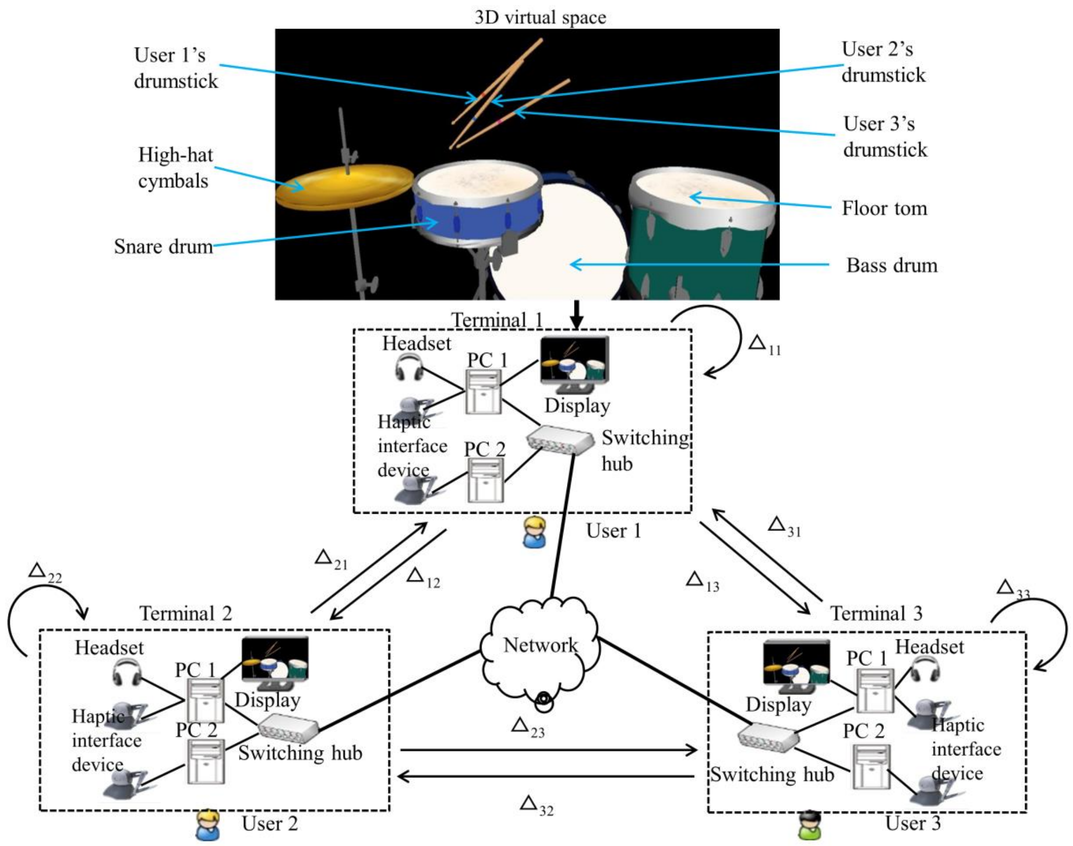


Figure 4.1. Configuration of networked haptic drum system with three terminals.

To explain the key idea of the enhanced dynamic local lag control for three terminals, we denote the MU delay from terminal i to terminal j (i, j = 1, 2, or 3) by Δ_{ij} (see Fig. 4.1). Also, we denote the local lag at terminal i by Δ_{ii}. For simplicity, let us focus on terminal 1. At terminal 1, Δ_{11} is set to the larger value of Δ_{21} and Δ_{31} according to the following equation: $\Delta_{11} = \max(\Delta_{21}, \Delta_{31})$. To synchronize the output time between received MUs, the additional buffering time is added to Δ_{j1} by $\Delta_{j1} - \Delta_{11}$ (j = 2 or 3); note that the additional buffering time of the larger one of Δ_{21} and Δ_{31} is zero. It is easy to obtain the algorithm for more than three terminals.

4.4 Assessment environment

4.4.1 Assessment system

In our assessment system, the three terminals are connected to each other via a switching hub and a network emulator (NIST Net [86]). The network emulator generates an additional constant delay D_{ij} (i, j = 1, 2 or 3) for each packet transmitted from terminal i to terminal j. The value of Δ_{ij} is equal to the value of D_{ij} plus 4.361 ms as in Chapter 3. This is because the network delay jitter can be absorbed by buffering under media synchronization control such as the Virtual-Time Rendering (VTR) [45]; we here take account of the jitter by including the buffering time in the constant delay.

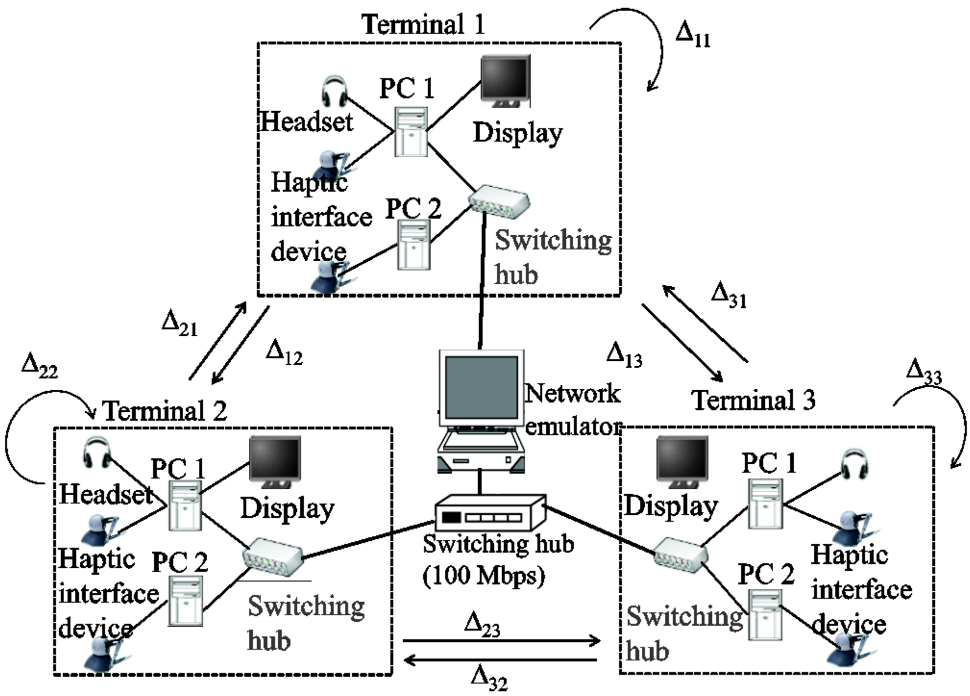


Figure 4.2. Configuration of assessment system.

4.4.2 Assessment methods

In the assessment, we compare QoE among the three types of control (the enhanced dynamic local lag control, adaptive Δ-causality control, and no control). In the adaptive Δ-causality control, the local lag at each terminal is set to the maximum MU delay among the terminals; note that the synchronization quality is perfect, but the interactivity is the poorest when the network delay is large. On the other hand, in no control, the interactivity is the highest, but the synchronization quality is most seriously damaged when the network delay is large.

In this work, 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. We carried out subjective QoE assessment with 15 subjects (males and females) whose ages were between 20 and 28.

In the assessment, we deal with three cases (*cases 1, 2,* and *3)*. In case 1, we set $D_{12} = D_{21} =$ 50 ms and $D_{23} = D_{32} =$ 0 ms. The values of D_{13} and D_{31} are simultaneously changed from 0 ms to 150 ms at intervals of 50 ms; that is, this case handles three different pairs of constant delays. In case 2, we set $D_{12} = D_{21} = D_{23} = D_{32} = 0$ ms. The values of D_{13} and D_{31} are changed from 0 ms to 150 ms at intervals of 50 ms at the same time; a pair of constant delays is different from the other pairs in this case. In case 3, we simultaneously change all the constant delays from 0 ms to 150 ms at intervals of 50 ms.

In the assessment, each trio of subjects hit the drum set with rhythm 1 at slow tempo as in the previous chapters. A different rhythm and tempo (i.e., rhythm 2 at fast tempo) were not handled in this paper because results of the other combinations of rhythm and tempo were almost the same as those of rhythm 1 at slow tempo. In this paper, we employed only one haptic interface device at each terminal as the right drumstick for simplicity (see the 3D virtual space in Fig. 4.1). Each trio of subjects practiced about two minutes under the condition that there was no constant delay. In the assessment, we selected the constant delays, cases 1, 2, and 3, the three types of control in random order for the trio. It took 30 seconds for each stimulus. After the stimulus, the subjects 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 those of the other users 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 for each stimulus. By averaging scores of all the subjects, we obtained the *mean opinion score (MOS)* [74] as a QoE parameter.

Each trio took a rest for around two minutes every about 30 minutes. The total assessment time per trio was about one and half hour including rests and practices.

Objective assessment was also carried out at the same time as the subjective assessment. We adopt the root mean square error (RMSE) [92] of sound at terminals 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 differences between the output times of three sounds (i.e., between the output times of sound at terminals 1 and 2, between those at terminals 1 and 3, and between those at terminals 2 and 3). The local lag is also employed as another objective assessment measure since it is closely related to the interactivity.

4.5 Assessment results

4.5.1 Subjective assessment

We plot MOS values of synchronization quality of sound, interactivity, and comprehensive quality at terminal 1 for case 1 in Figs. 4.3 through 4.5, respectively. In Fig. 4.6, we draw MOS values of comprehensive quality at terminal 2 for case 1. We do not show MOS values of synchronization quality of sound and interactivity at terminal 2 for case 1 since they had similar tendencies to those of comprehensive quality in Fig. 4.6. We also do not show MOS values at terminal 3 for case 1 since they had similar tendencies to those at terminal 1, which are shown in Figs. 4.3 through 4.5. For case 2, we show only MOS values of comprehensive quality at terminals 1 and 2 in Figs. 4.7 and 4.8, respectively. We do not show MOS values of synchronization quality of sound and interactivity at terminals 1 and 2 for the following reasons: MOS values of synchronization quality of sound and interactivity at terminal 1 had similar tendencies to those in Figs. 4.3 and 4.4, respectively, and MOS values of synchronization quality of sound and interactivity at terminal 2 had similar tendencies to those in Fig. 4.6. We do not also show MOS values at terminal 3 for case 2 since they had almost the same tendencies as those in Figs. 4.3 through 4.5. In Fig. 4.9, we present MOS values of comprehensive quality at all the terminals for case 3. For the same reasons as those in Figs. 4.7 and 4.8, we do not show MOS values of synchronization quality of sound and those of interactivity. In the figures, the 95% confidence intervals are also plotted.

In Fig. 4.3, we see that the MOS value of the enhanced dynamic local lag control is the highest for each value of D_{13} and D_{31}. We also notice that the MOS values of the three types of control tend to decrease as D_{13} and D_{31} become larger. The MOS value of no control is the most largely deteriorated among the three types of control since the control does not adjust the local information with the received information from the other terminals by buffering local information.

From Fig. 4.4, we find that the MOS value of no control is the highest, and it hardly depends on D_{13} and D_{31}. This is because no control does not buffer the local information. In the figure, we also notice that the MOS values of the enhanced dynamic local lag control and the adaptive Δ-causality control decrease as D_{13} and D_{31} become larger. The MOS value of the adaptive Δ-causality control is more largely deteriorated than the other two types of control. The reason is that the control sets the local lag to the largest value of the MU delays among the terminals.

Figure 4.5 reveals a similar tendency to that in Fig. 4.3. That is, the MOS value of the enhanced dynamic local lag control is the highest. The MOS values of all the three types of control decrease as D_{13} and D_{31} become larger.

In Fig. 4.6, the MOS values of the enhanced dynamic local lag control and no control are the highest or second highest. They hardly depend on D_{13} and $D_{31.}$ Furthermore, we notice that the MOS value of the adaptive Δ-causality control decreases as D_{13} and D_{31} become larger since the control sets the local lag to the maximum value of the MU delays among the terminals.

In Fig. 4.7, we see a similar tendency to that in Fig. 4.5. That is, the MOS value of the enhanced dynamic local lag control is the highest, and the MOS values of all the three types of control depend on D_{13} and D_{31}. Also, Fig. 4.8 reveals that the MOS values had almost the same tendency to that in Fig. 4.6; that is, the enhanced dynamic local lag control and no control has the highest or second highest MOS values, and the MOS value of the adaptive Δ-causality control depends on D_{13} and D_{31}. Furthermore, we note from Fig. 4.9 that the MOS values has a similar tendency to that in Fig. 4.5; that is, the MOS value of the enhanced dynamic local lag control is the highest, and those of all the three types of control decrease as D_{13} and D_{31} become larger.

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

4.5.2 Objective assessment

We show only the root mean square error of sound at terminal 1 for case 1 in Fig. 4.10, where the 95% confidence intervals are also included. We observe in the figure that the enhanced dynamic local lag control has the smallest root mean square error for each value of D_{13} and D_{31}. By comparing Figs. 4.5 and 4.10, we find that the trends of the curves are almost reverse to each other.

4.5.3 Relation between subjective and objective assessment results

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 [75]. As a result, we obtained the following three equations:

$$S_{\mathrm{mos}} = -\,0.023E_{\mathrm{rmse}} + 5.237, \tag{4.1}$$

$$I_{\mathrm{mos}} = -\,0.016\Delta + 4.689, \tag{4.2}$$

$$C_{\mathrm{mos}} = -\,0.017E_{\mathrm{rmse}} - 0.005\Delta + 5.118, \tag{4.3}$$

where S_{mos}, I_{mos}, and C_{mos} denotes the estimated MOS values of synchronization quality of sound, interactivity, and comprehensive quality, respectively. Also, E_{rmse} denotes RMSE of sound, and Δ denotes the local lag. The contribution rates adjusted for degree of freedom [75] for the equations were 0.904, 0.922, and 0.894, respectively. Since the contribution rates are high enough, 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.

4.6 Summary

In this chapter, we enhanced the dynamic local lag control so that three or more users at different terminals can play musical instruments together over a network. We investigated the effect of the control by subjective and objective QoE assessment in a networked haptic drum system for three terminals. We made a comparison among the enhanced dynamic local lag control, the adaptive Δ-causality control, and no control. As a result, we found that the enhanced control is effective.

Although the dynamic local lag control can maintain the synchronization quality of sound high, the interactivity still deteriorates slightly because the control buffers local information. It is important to consider how we can make the interactivity high.

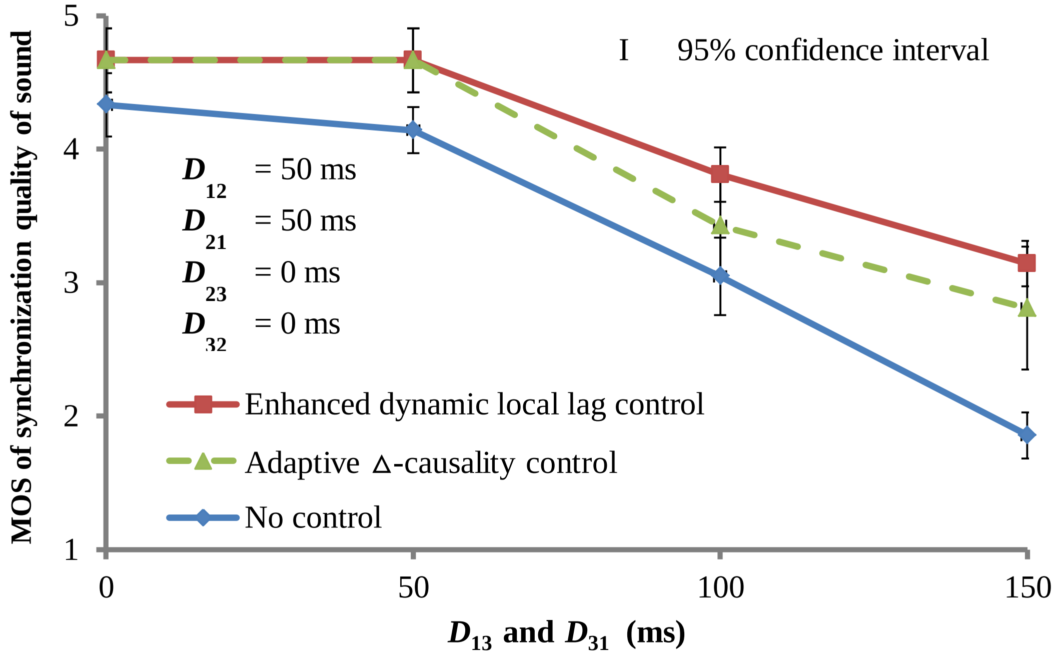


Fig. 4.3. MOS of synchronization quality of sound at terminal 1 in case 1.

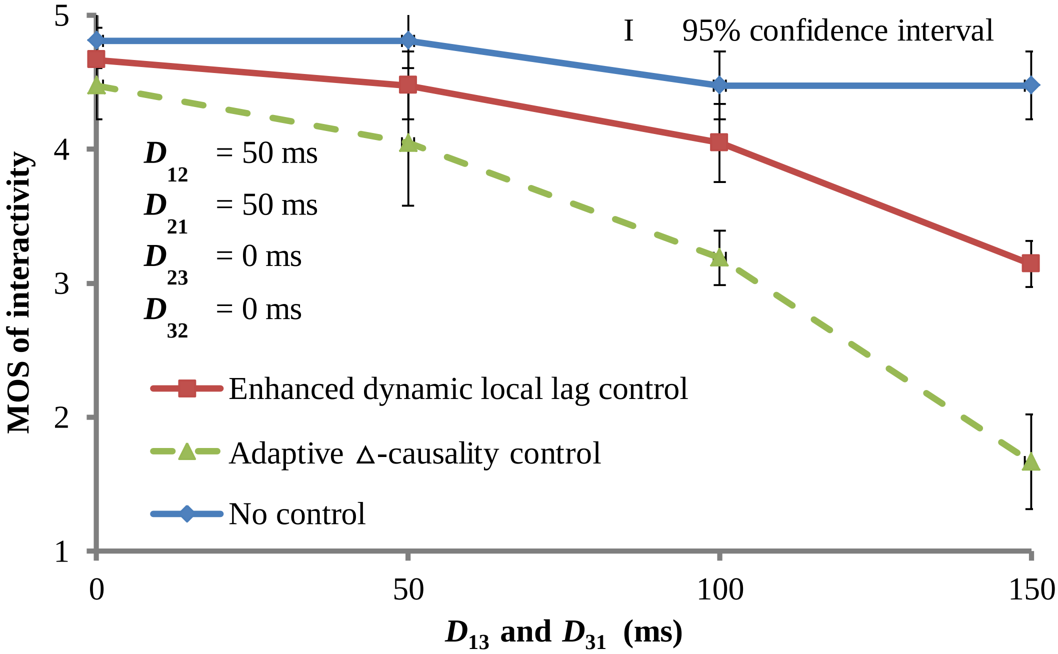


Fig. 4.4. MOS of interactivity at terminal 1 in case 1.

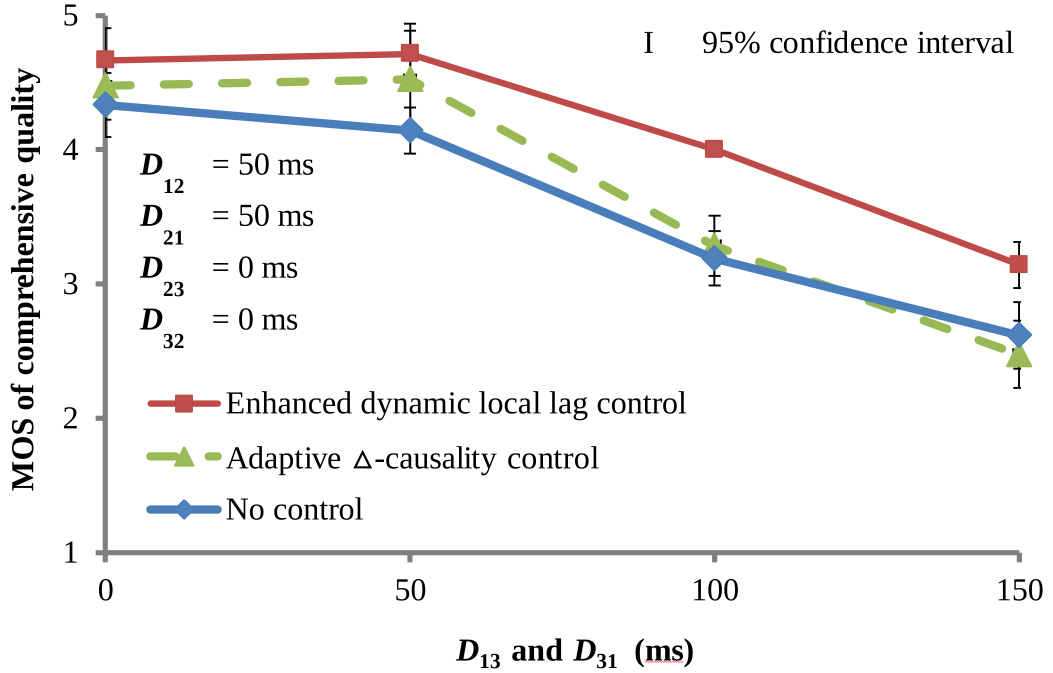


Fig. 4.5. MOS of comprehensive quality at terminal 1 in case 1.

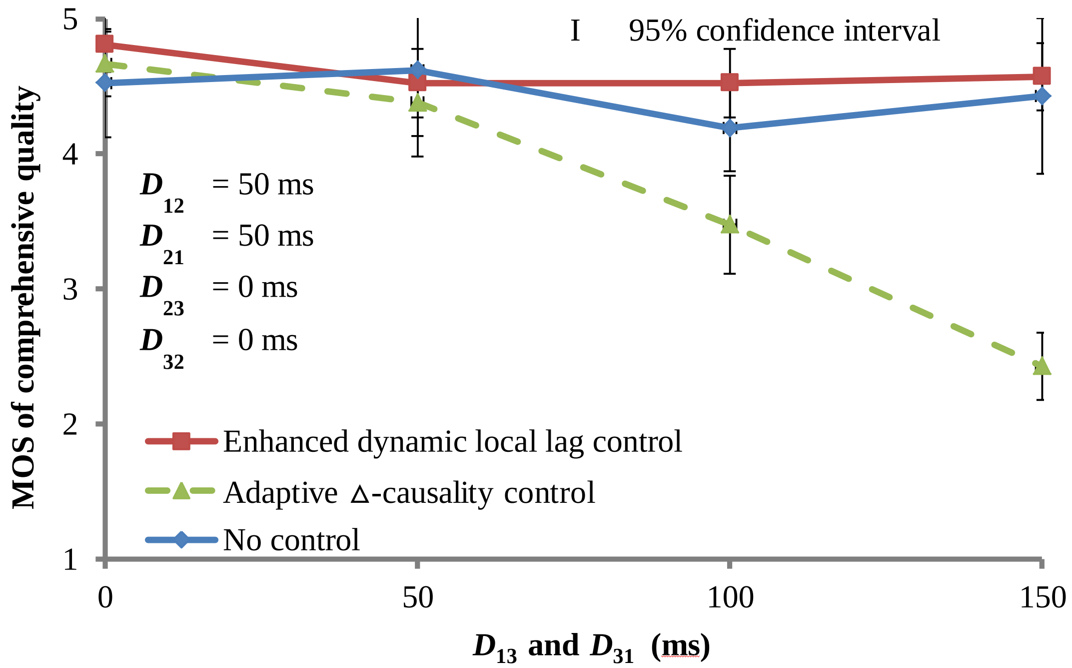


Fig. 4.6. MOS of comprehensive quality at terminal 2 in case 1.

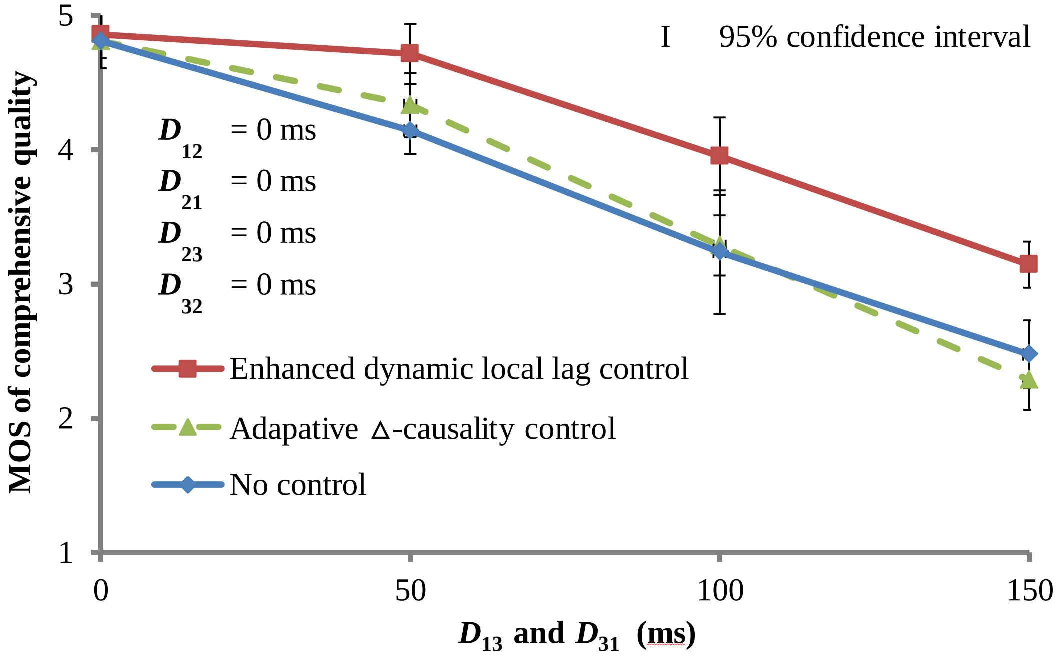


Fig. 4.7. MOS of comprehensive quality at terminal 1 in case 2.

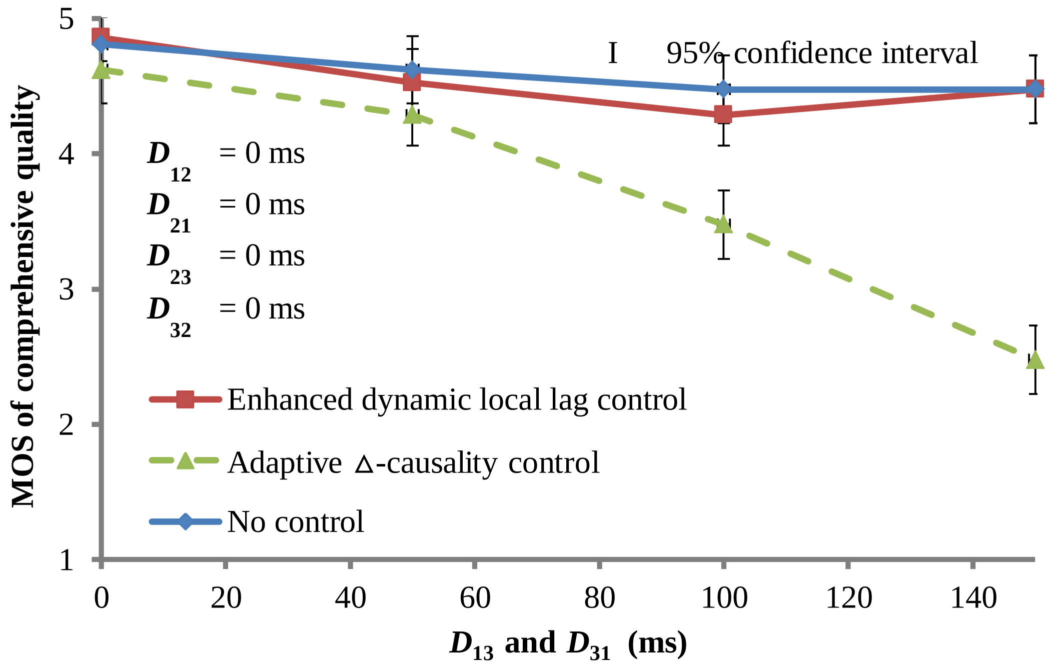


Fig. 4.8. MOS of comprehensive quality at terminal 2 in case 2.

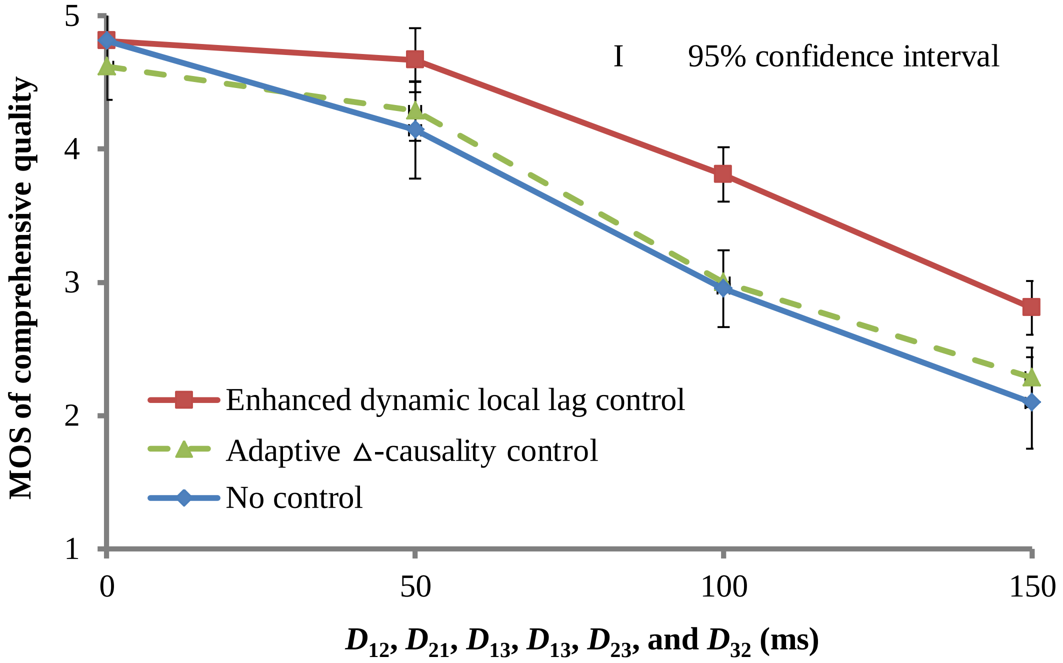


Fig. 4.9. MOS of comprehensive quality at all terminals in case 3.

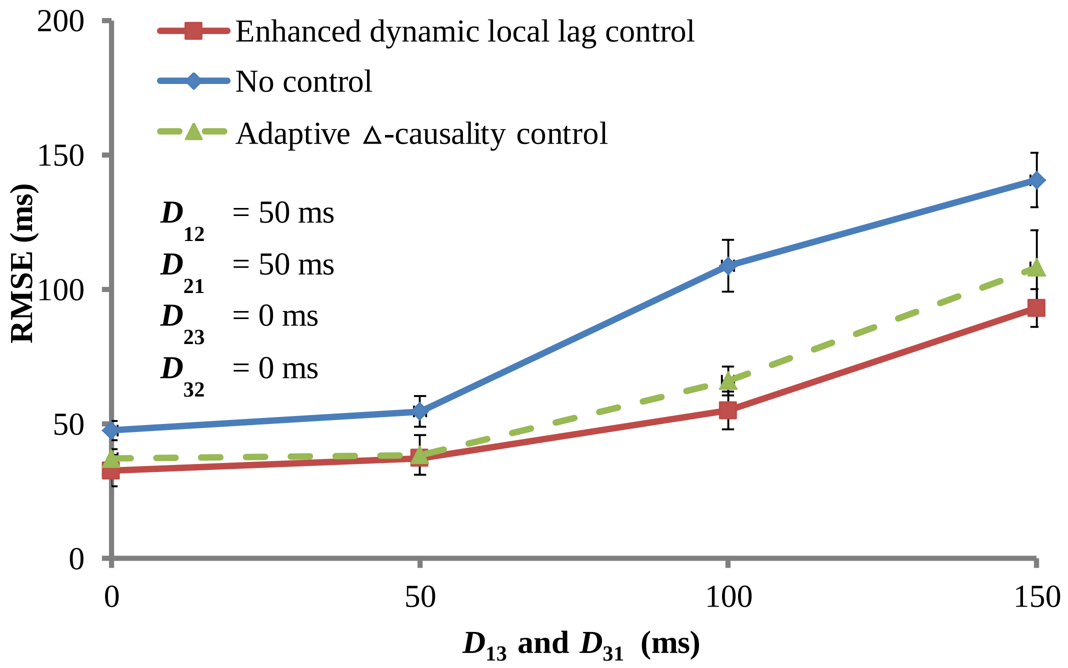


Fig. 4.10. Root mean square error of sound at terminal 1 in case 1.

Chapter 5

QoE assessment of dynamic local lag control with prediction in joint haptic drum performance

5.1 Introduction

In order to keep the synchronization quality of sound high, we propose the dynamic local lag control in Chapters 3 and 4. To clarify the effect 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 in the joint performance of the networked haptic drum system. As a result, we 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 [93]-[95]. In [94], 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 order to use the prediction control with the dynamic local lag control in the networked haptic drum system, we need to clarify how to change the prediction time dynamically according to the network delay. However, the change method of the prediction time has not been studied yet.

In this chapter, we investigate the effect of the dynamic local lag control with prediction on QoE in the joint performance of the networked haptic drum system.

The remainder of this chapter is organized as follows. Subsection 5.2 explains the prediction control, and Subsection 5.3 describes our assessment environment.

Assessment results are presented in Subsection 5.4, and Subsection 5.5 concludes the chapter.

5.2 Prediction control

The prediction control outputs the position information by predicting the future position later than the output time of an MU by the prediction time $T_{\text{predict}}\,(\geq 0\text{ ms})$ to keep the interactivity high. For simplicity, the first-order prediction [96] is used in this chapter. 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 $\boldsymbol{P}(t + T_{\text{predict}})$ 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 $\boldsymbol{P}(t + T_{\text{predict}})$ is calculated as follows:

$$\boldsymbol{P}(t + T_{\text{predict}}) = \boldsymbol{P}(t) + \{\, \boldsymbol{P}(t) - \boldsymbol{P}(t - \delta)\,\}\; T_{\text{predict}}/\delta, \tag{5.1}$$

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 $t - \delta$.

5.3 Assessment environment

5.3.1 Assessment system

As shown in Fig. 5.1, in our assessment system, two terminals are connected to each other via a network emulator (NIST Net [86]). The emulator generates an additional constant delay for each packet transmitted between the terminals. We take account of the jitter by including the buffering time in the constant delay as in assessments of previous chapters. We handle the symmetric delay case in the assessment. We set the constant delay (*delay 1* here) from terminal 2 to terminal 1 to the same value as that (*delay 2*) in the opposite direction.

5.3.2 Assessment methods

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 the previous chapters to investigate the influence of drumstick movements.

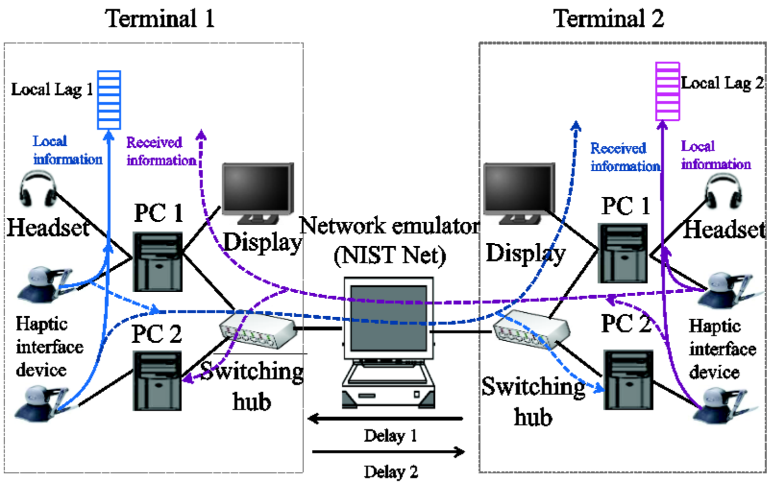


Figure 5.1. Configuration of assessment system.

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. In the assessment, we changed the constant delay from 0 ms to 150 ms at intervals of 50 ms, and 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 each combination of rhythm and tempo, the constant delay and T_{predict} were selected in random order for the pair.

It took 30 seconds for each stimulus. After each stimulus, the pair of subjects was asked to base their judgments about the 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 [74]). 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 to obtain the *mean opinion score (MOS)* [74] as a 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 then rhythm 2 at the fast tempo. The total assessment time per pair was about one and half hours including rests and practices.

5.4 Assessment results

Figure 5.2 explains the notation employed in the following figures. We show QoE assessment results for rhythm 1 at the slow tempo in Figs. 5.3 through 5.7, 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, for various values of delay 1 or 2 as a function of T_{predict}. For rhythm 2 at the fast tempo, we show the MOS values in Figs. 5.8 to 5.12. In the figures, the 95% confidence intervals are also plotted.

In the figures, we see that the QoE assessment results of rhythm 1 at the slow tempo have similar tendencies to those of rhythm 2 at the fast tempo.

In Figs. 5.3 and 5.8, we find that the MOS values of delay 1 or 2 excluding 150 ms start to become smaller when T_{predict} exceeds some values. We also notice that the MOS values increase as delay 1 or 2 becomes larger. The reason is as follows. When T_{predict} becomes longer than the buffering time, the position information of own drumstick is ouput by using prediction. Thus, as T_{predict} increases, the prediction error becomes larger.

In Figs. 5.4 and 5.9, we observe that the MOS values become smaller as T_{predict} becomes larger. We also notice that the differences in the MOS values of delay 1 or 2 are very small. 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.

In Figs. 5.5 and 5.10, we find that the MOS values become smaller as delay 1 or 2 becomes larger. We also see that the MOS values hardly depend on T_{predict}, and they depend on only delay 1 or 2.

Figures 5.6 and 5.11 reveal that the MOS values tend to become smaller as delay 1 or 2 becomes larger when T_{predict} is small. When delay 1 or 2 is from 50 ms to 150 ms, the MOS values increase until T_{predict} increases up to some value. The reason is as follows. The prediction control advances the output time of local information by the amount of T_{predict}; thus, the interactivity becomes better as T_{predict} becomes larger.

From Figs. 5.7 and 5.12, we observe that there are optimum values of T_{predict} for each value of delay 1 or 2. We notice that the optimum values become larger as delay 1 or 2 increases.

We plot the optimum values of prediction time obtained from Figs. 5.7 and 5.12 versus delay 1 or 2 in Fig. 5.13. By regression analysis [92], we obtained the following equation:

$$T_{\text{predict}} = \max\left(0.007D^2 - 0.327D + 1.818,\ 0\right) \tag{5.2}$$

where T_{predict} ($\geq$0 ms) is the optimum value of prediction time; note that if T_{predict} is smaller than 0 ms, $T_{predict}$ is set to 0 ms. To use D as in Eq. (3.1) instead of the constant delay, we measured D, and we found that D is equal to the constant delay plus 4.361 ms. The contribution rate adjusted for degrees of freedom [92], which shows goodness of fit with the estimated equation, was 0.952. Actually, we find close agreement between the optimal values and calculated ones in Fig. 5.13.

5.5 Summary

In this chapter, we investigated the effect of the dynamic local lag control with prediction by subjective QoE assessment in the joint performance of a networked haptic drum system. As a result, we found that there exits the optimum value of prediction time according to the network delay.

As the next chapter, we will assess the effect of dynamic local lag control with dynamic control of prediction time which dynamically changes the prediction time according to Eq. (5.2) in Subsection 5.4.

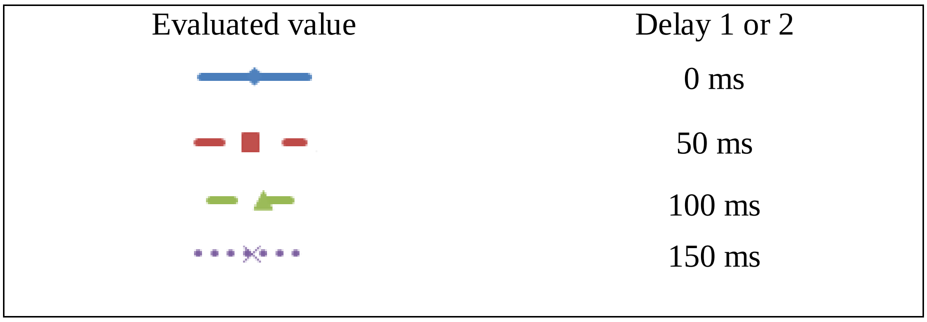


Figure 5.2. Notation in Figs. 5.3 through 5.12.

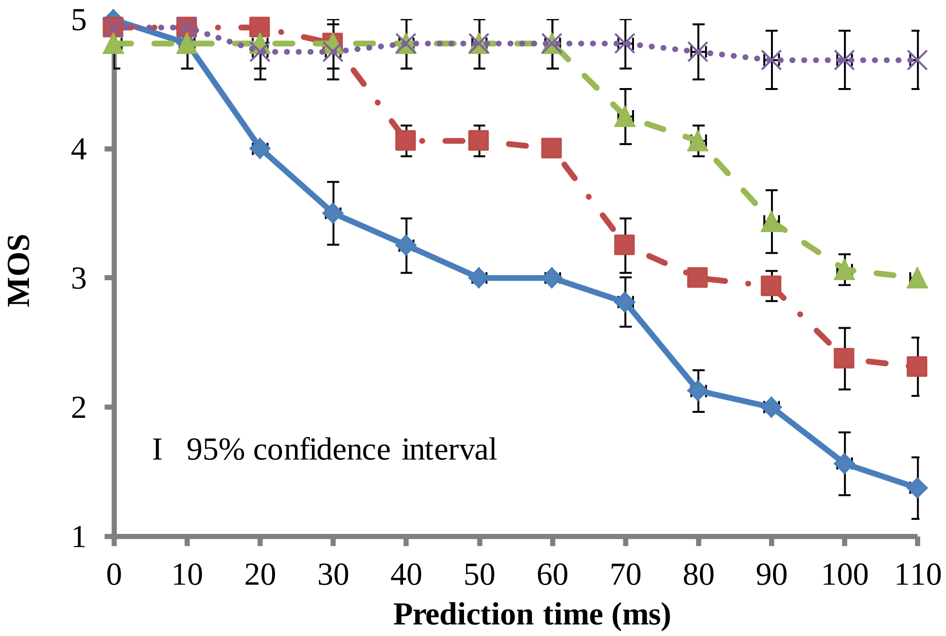


Figure 5.3. MOS of output quality of own drumstick for rhythm 1 at slow tempo.

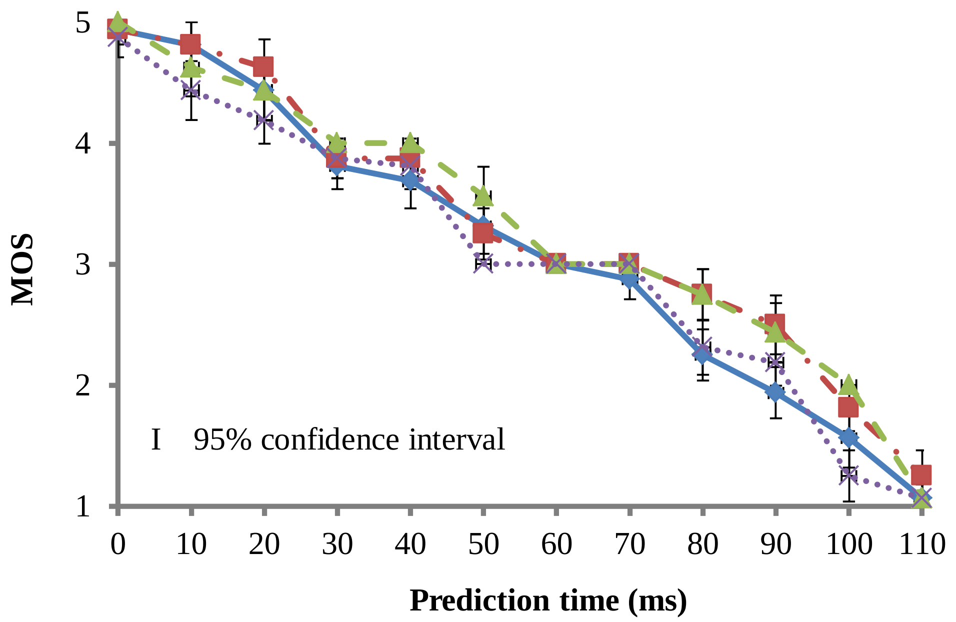


Figure 5.4. MOS of output quality of other drumstick for rhythm 1 at slow tempo.

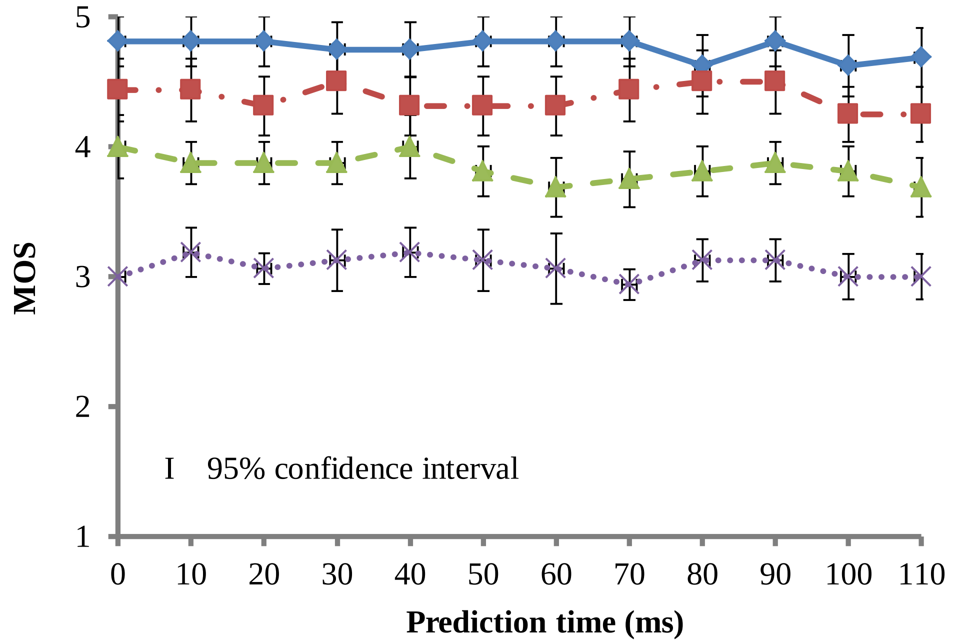


Figure 5.5. MOS of synchronization quality of sound for rhythm 1 at slow tempo.

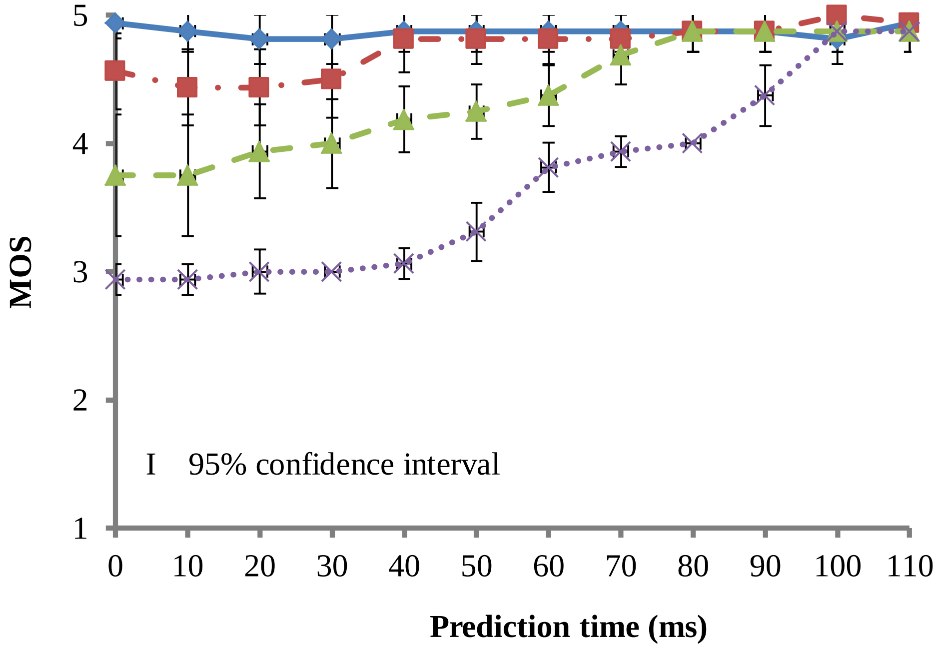


Figure 5.6. MOS of interactivity for rhythm 1 at slow tempo.

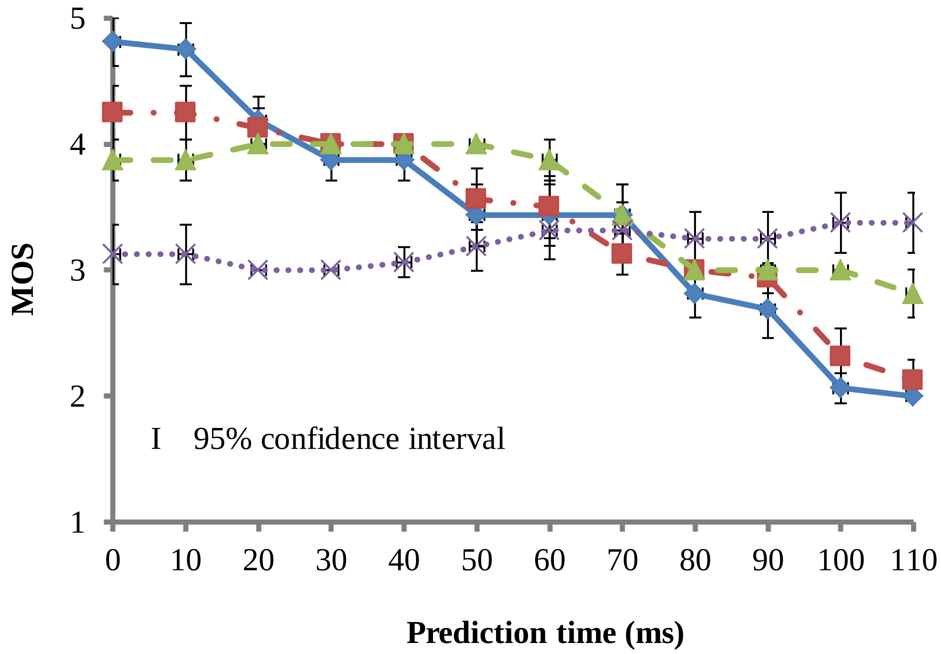


Figure 5.7. MOS of comprehensive quality for rhythm 1 at slow tempo.

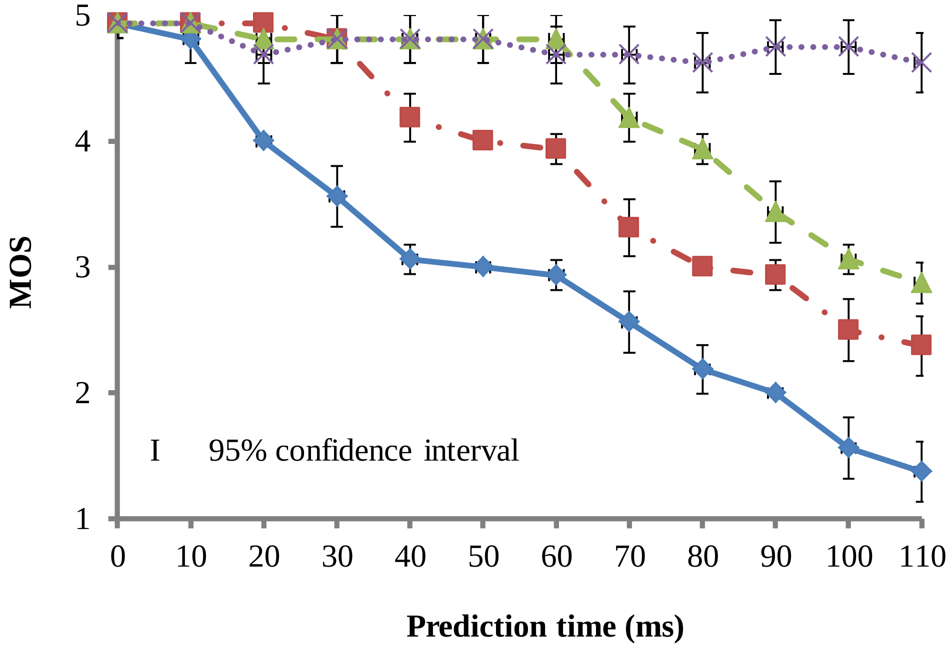


Figure 5.8. MOS of output quality of own drumstick for rhythm 2 at fast tempo.

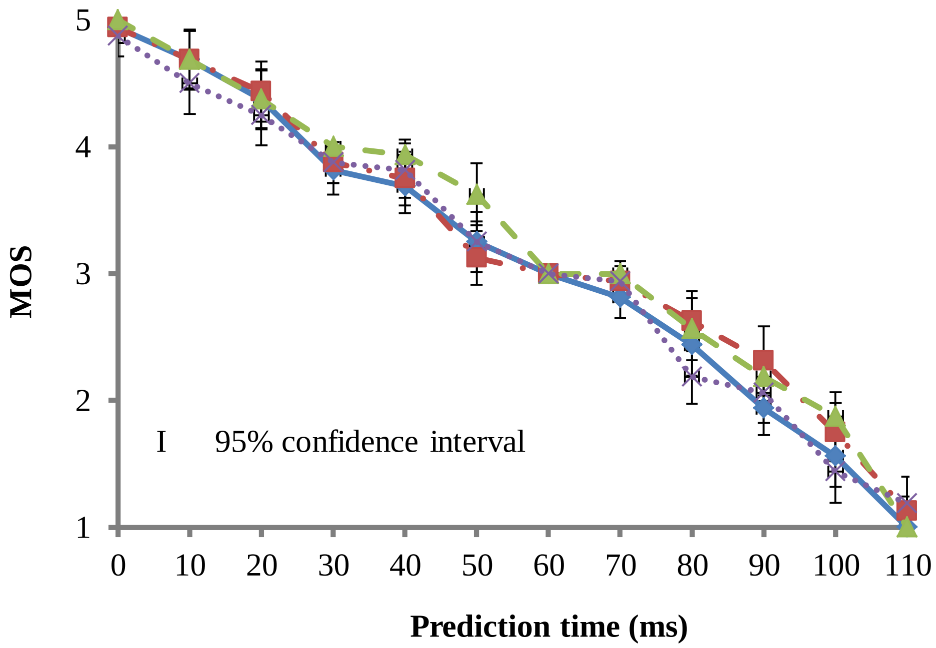


Figure 5.9. MOS of output quality of other drumstick for rhythm 1 at slow tempo.

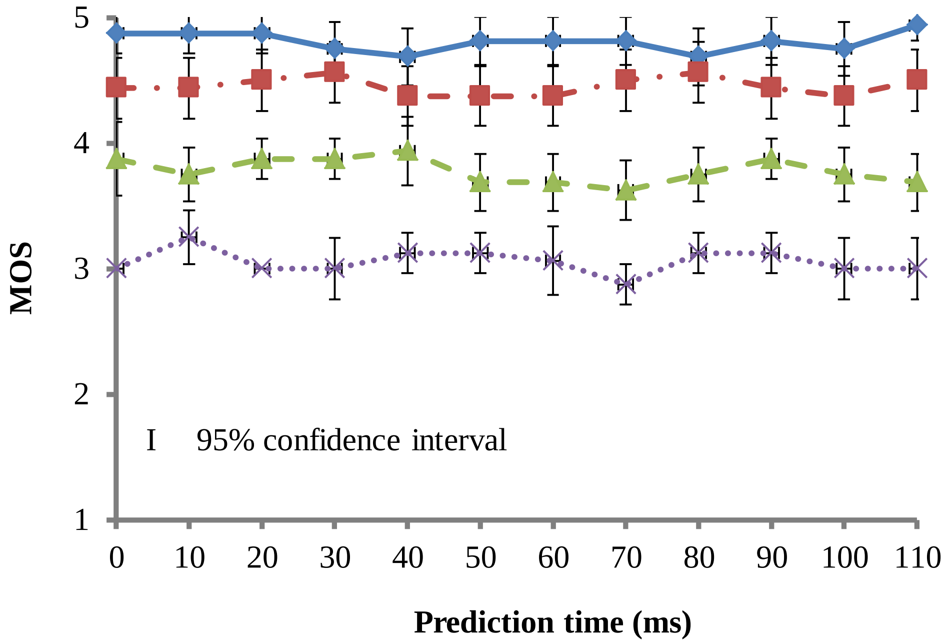


Figure 5.10. MOS of synchronization quality of sound for rhythm 1 at slow tempo.

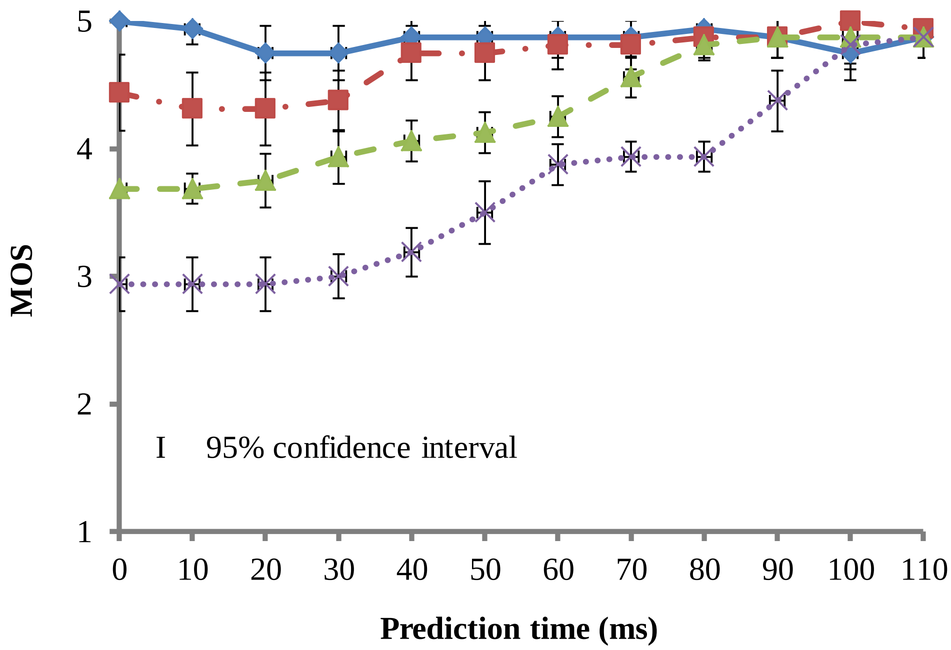


Figure 5.11. MOS of interactivity for rhythm 1 at slow tempo.

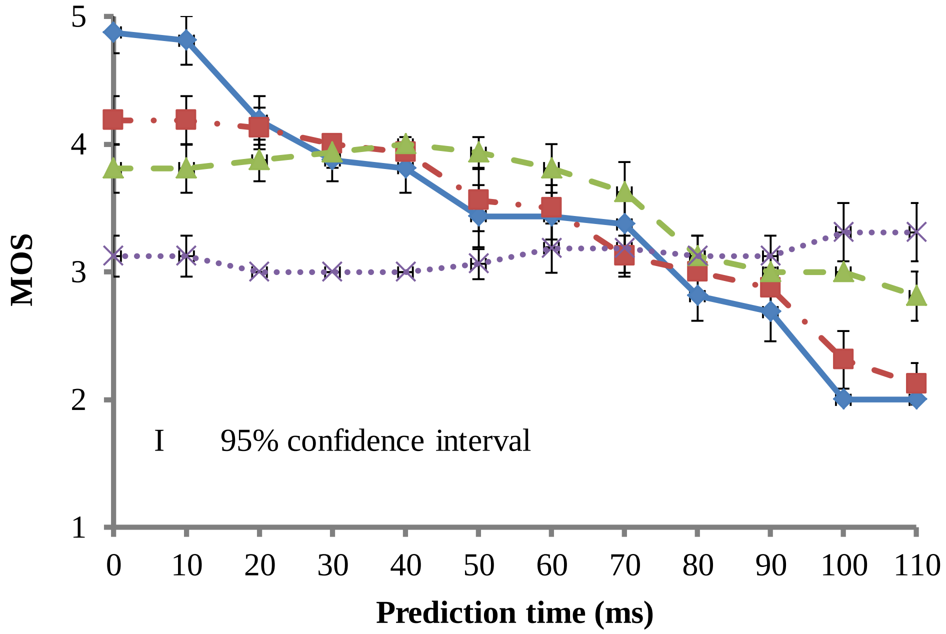


Figure 5.12. MOS of comprehensive quality for rhythm 1 at slow tempo.

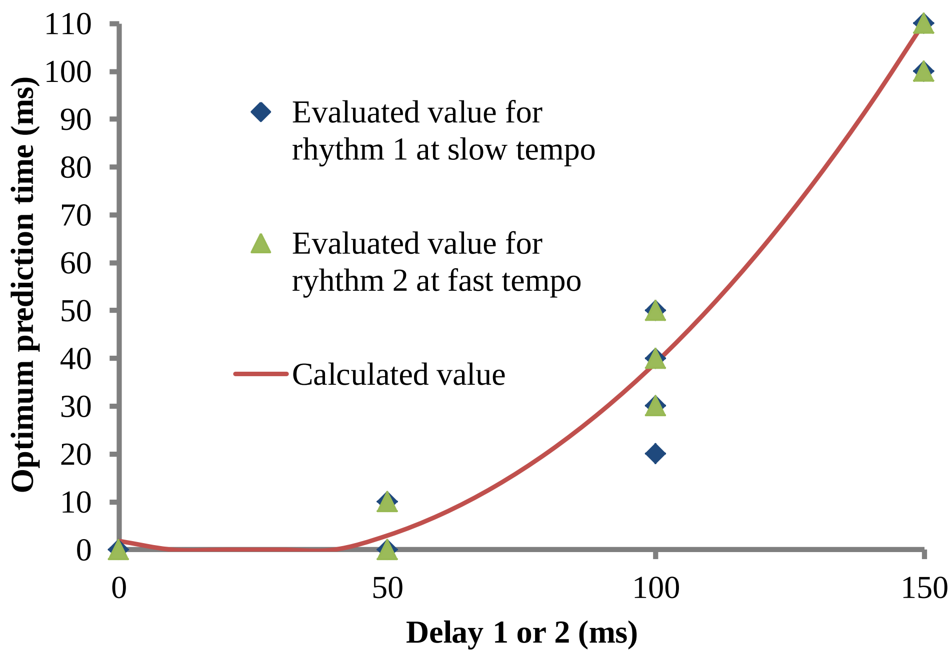


Figure 5.13. Optimum prediction time versus constant delay

Chapter 6

Effects of dynamic local lag control with dynamic control of prediction time in joint haptic drum performance

6.1 Introduction

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

Based on the results in Chapter 5, 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. 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 6.2 explains the proposed dynamic local lag control with dynamic control of prediction time. Section 6.3 describes our assessment environment. Assessment results are presented in Section 6.4, and Section 6.5 concludes the chapter.

6.2 Dynamic local lag control with dynamic control of prediction time

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} ($\geq$0 ms) to keep the interactivity high. For simplicity, the first-order prediction [96] is used in this work. 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 $\boldsymbol{P}(t + T_{\mathrm{predict}})$ 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 $\boldsymbol{P}(t + T_{\mathrm{predict}})$ is calculated as follows:

$$\boldsymbol{P}(t + T_{\mathrm{predict}}) = \boldsymbol{P}(t) + \{ \boldsymbol{P}(t) - \mathrm{P}(t - \delta) \} \, T_{\mathrm{predict}}/\delta, \tag{6.1}$$

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 Chapter 5. We investigated the relation between T_{predict} and the network delay by regression analysis [75], and we obtained the following equation:

$$T_{\mathrm{predict}} = \max\,(0.007D^2 - 0.327D + 1.818,\ 0), \tag{6.2}$$

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 [92], 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.

6.3 Assessment environment

6.3.1 Assessment system

In our assessment system, two terminals are connected to each other via a network emulator (NIST Net [86]). The emulator generates an additional constant delay for each packet transmitted between the terminals. As in the assessment environments of previous chapters, 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.

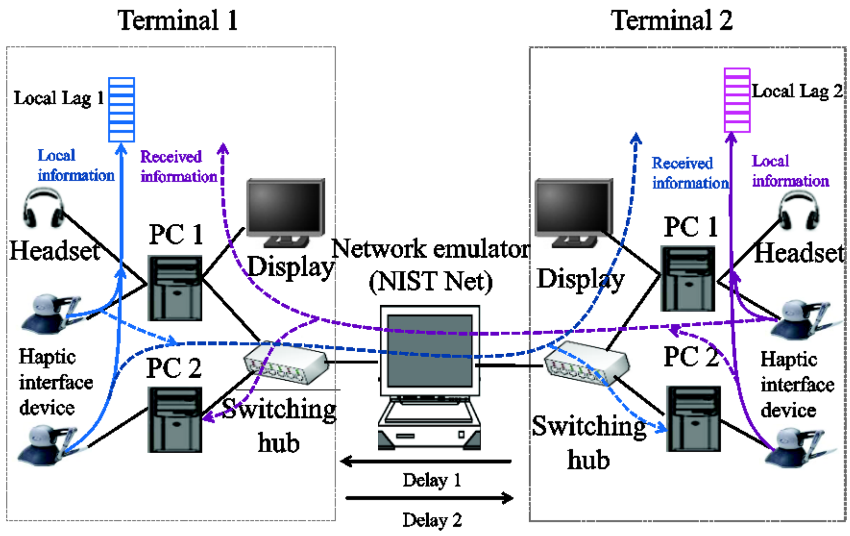


Figure 6.1. Configuration of assessment system.

6.3.2 Assessment methods

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. In this chapter, 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 chapter 2.

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. (6.2).

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 [74]).

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 MOS [74] 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.

6.4 Assessment results

Figure 6.2 explains the notation employed in the following figures. We show QoE assessment results for rhythm 1 at the slow tempo in Figs. 6.3 through 6.7, 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 delay 1 or 2. For rhythm 2 at the fast tempo, we show only MOS values of comprehensive quality in Fig. 6.8. We do not show other MOS values of rhythm 2 at fast tempo since they had similar tendencies to those in Figs. 6.3 through 6.7, respectively. In the figure, the 95% confidence intervals are also plotted.

In Fig. 6.3, we find that the MOS value of dynamic control of prediction time is hardly dependent on delay 1 or 2 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 delay 1 or 2 is small. When T_{predict} is between 20 ms and 110 ms, the MOS values increase until delay 1 or 2 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. 6.4, we observe that the MOS value of dynamic control of prediction time decreases as delay 1 or 2 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 delay 1 or 2. 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 6.5 reveals that the MOS values become smaller as delay 1 or 2 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. 6.6, we find that the MOS value of dynamic control of prediction time is high for any value of delay 1 or 2. Also, in the figure, the MOS values decrease as delay 1 or 2 becomes larger when T_{predict} is fixed; the MOS values decrease more largely as T_{predict} becomes smaller.

In Figs. 6.7 and 6.8, 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.

6.5 Summary

In this chapter, 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. It is also important to enhance the proposed control so that three or more users can perform joint musical performance.

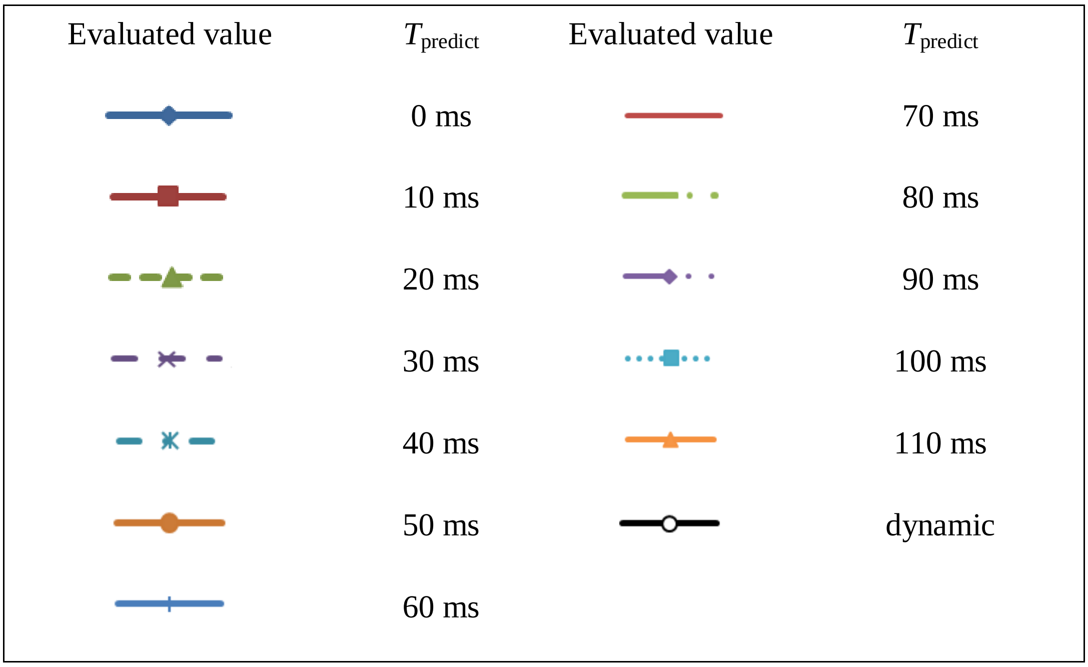


Figure 6.2. Notation in Figs. 6.3 through 6.7.

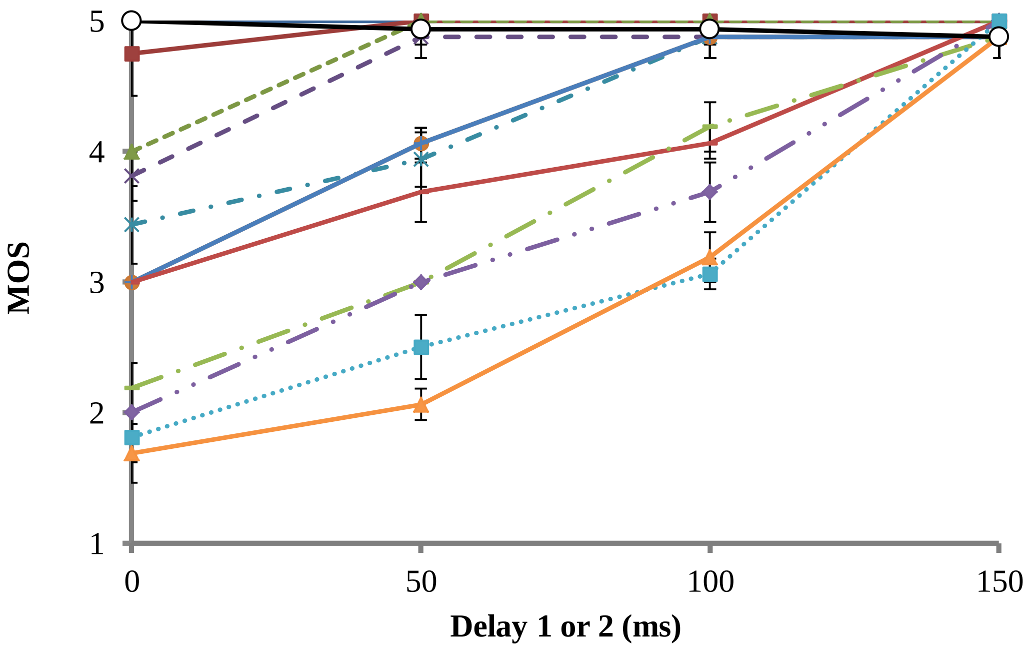


Figure 6.3. MOS of output quality of own drumstick for rhythm 1 at slow tempo.

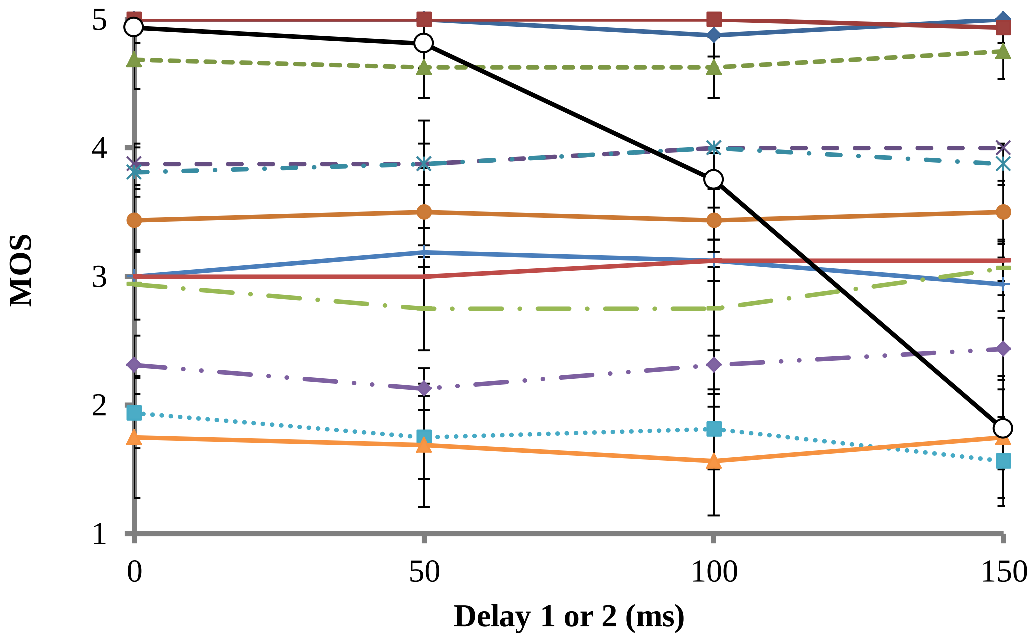


Figure 6.4. MOS of output quality of other drumstick for rhythm 1 at slow tempo.

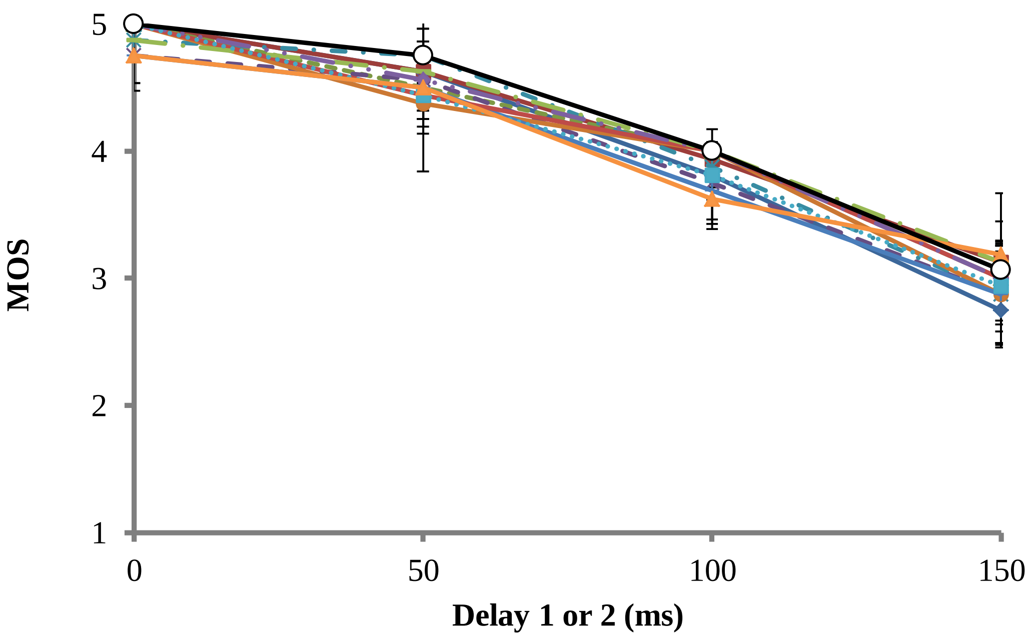


Figure 6.5. MOS of synchronization quality of sound for rhythm 1 at slow tempo.

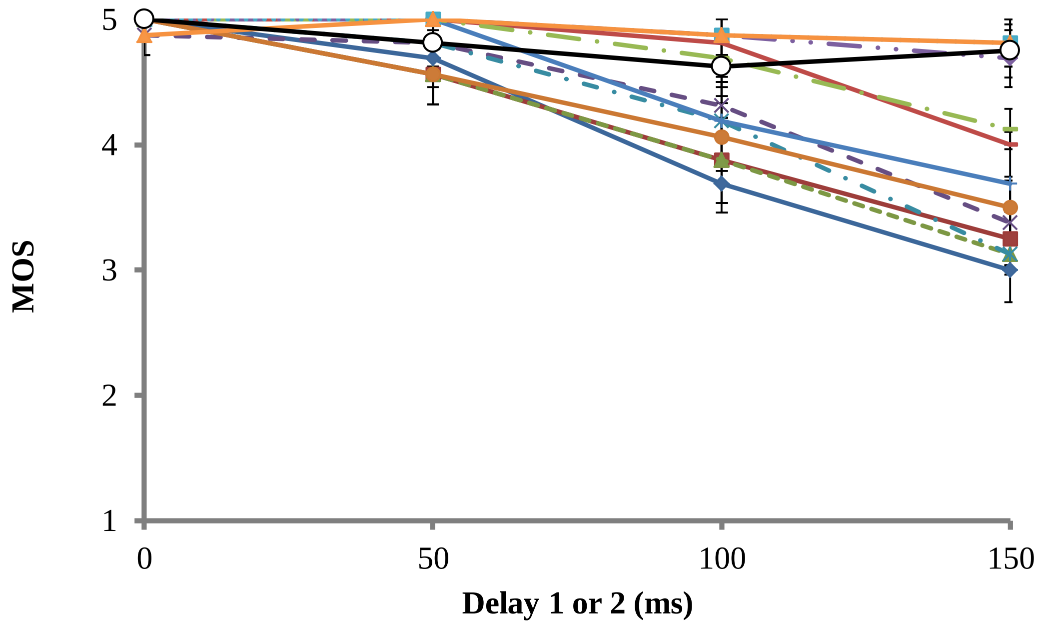


Figure 6.6. MOS of interactivity for rhythm 1 at slow tempo.

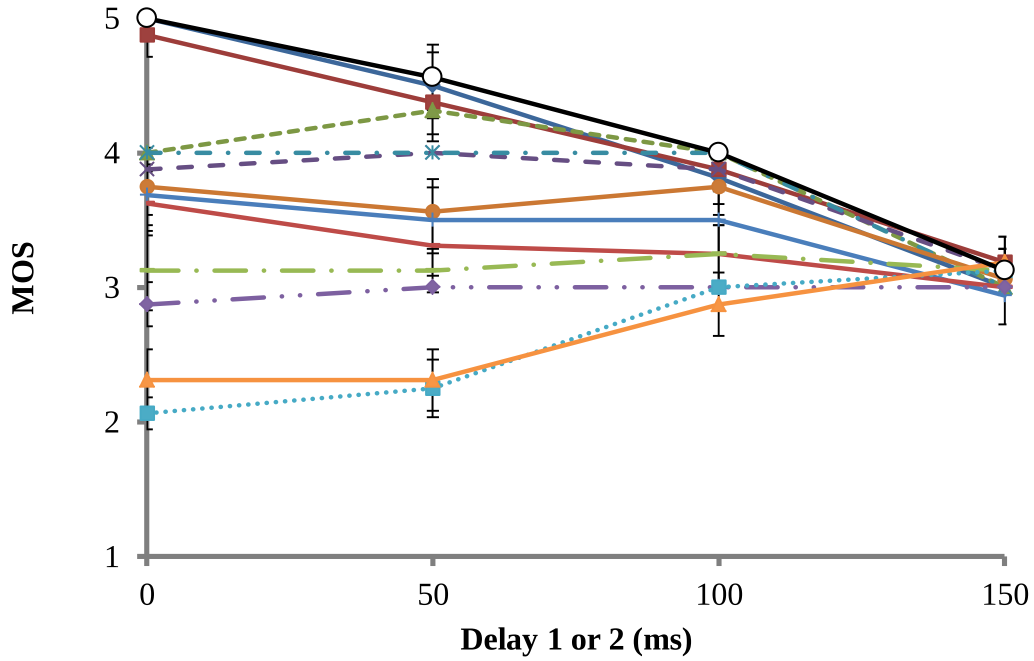


Figure 6.7. MOS of comprehensive quality for rhythm 1 at slow tempo.

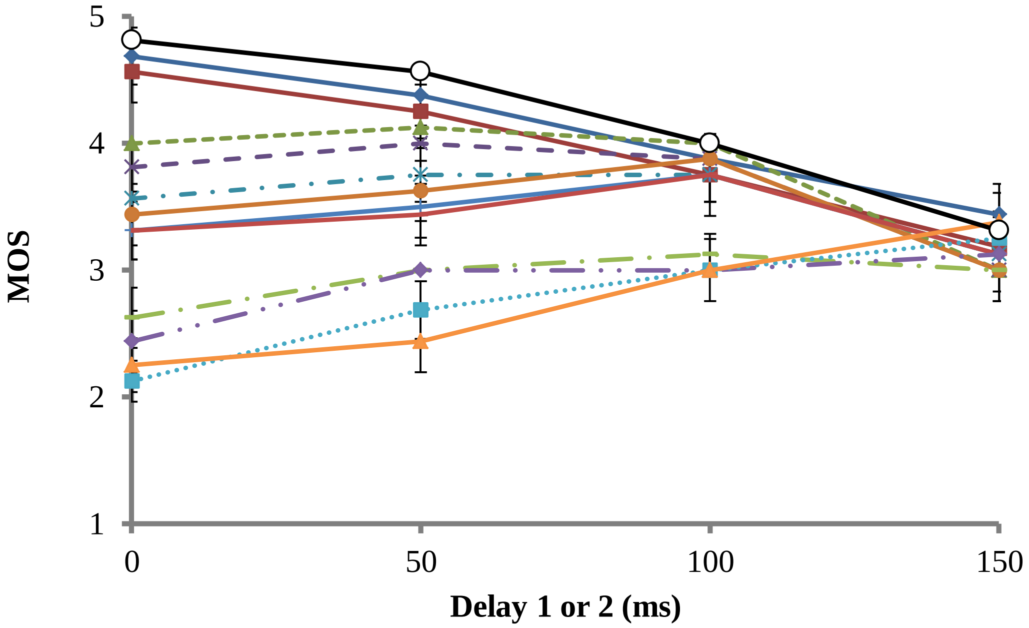


Figure 6.8. MOS of comprehensive quality for rhythm 2 at fast tempo.

Chapter 7

Influences of network delay on QoE for soft objects in networked haptic real-time game

7.1 Introduction

A number of researchers have been paying their attention to networked real-time games with haptic sense [32]-[35]. Players can perceive a higher sense of immersion by using haptic interface devices in networked real-time games. When such a game is played over a network where QoS is not guaranteed like the Internet, the consistency (e.g., the positions of an object at different terminals are the same) of objects in a 3D virtual space at the players' terminals may be disturbed owing to the network delay, delay jitter, and packet loss.

To keep the consistency high at the terminals, we can employ the local lag control [69], which buffers the local information for a constant time called the local lag according to the network delay from the local terminal to the other terminal. Thus, the interactivity is degraded when the local lag is large. The operability of haptic interface device may also be deteriorated. If the network delay from the local terminal to the other terminal is about the same to that from the other terminal to the local terminal, the fairness is not deteriorated although the operability is deteriorated. However, when the difference in network delay between the terminals is large; the fairness among the players is largely damaged. For example, in a networked real-time game when two players play, when the network delay from a terminal to the other terminal is large, and that in the opposite direction is small, the operability is seriously degraded only at the terminal with the larger network delay under the local lag control. This leads to unfairness between the players. In [77], Brun *et al.* describe that the fairness is high when the same condition is provided to all the players.

To maintain the fairness high, we can use the adaptive Δ-causality control [68], which also employs the local lag control. The adaptive Δ-causality control generally sets the local lag at each terminal to the maximum network delay among the terminals. Therefore, when the maximum network delay is large, the interactivity may seriously be degraded under the control. This leads to the severe degradation of the operability of the haptic interface device, but the fairness is maintained high. Based on this relationship of the operability and fairness, we can say that there is a trade-off relationship between the operability and fairness.

In previous works, for example [34] and [35], the influence of network delay on the fairness is investigated using hard objects in a 3D virtual space. In [34] a networked real-time game in which each of two players try to contain a target (a sphere) competitively with his/her object (a rigid cube) by manipulating his/her haptic interface device is dealt with in a 3D virtual space. Assessment results show that the players feel unfairness when the difference in network delay between them is larger than about 30 ms. Also, in [35], the influence of the time it takes for a smell to reach a player on the fairness is investigated in a fruit harvesting game where two players play in a 3D virtual space. In the game, all the fruit objects in the 3D virtual space are hard ones. As a result, it is illustrated that the fairness is hardly damaged when the time is smaller than about 500 ms.

In [34] and [35], since only the fairness is investigated, the relationship between the operability of haptic interface device and the fairness between players is not clear. It is important to investigate the relationship of the operability and fairness in detail by QoE (operability and fairness) assessment in which we should assess the comprehensive quality (i.e., the weighted sum of the operability and fairness) to examine which QoE has larger contribution to the comprehensive quality. Also, we should clarify how much local lag should be set at each terminal to maintain the comprehensive quality as high as possible. For example, if the fairness has larger contribution to the comprehensive quality, the comprehensive quality may be maintained high by adjusting the output timing of the terminals to the terminal which has the slowest timing as in the adaptive Δ-causality control. Furthermore, to the best of our knowledge, there is no previous work which investigates the influence of network delay on QoE such as the operability and fairness for soft objects in virtual environments. In the case of hard objects, it is found that the objects become heavier as the network delay increases [78]; that is, their characteristics change owing to network delay. The characteristics of soft objects may change in the same way as those of hard objects; for example, they become harder and/or heavier as the network delay increases. However, it is not clear how the characteristics change. We should carry out QoE assessments for soft objects to understand clearly how their characteristics change when the network delay increases.

In this chapter, we investigate the influences of network delay on the operability of the haptic interface device and the fairness between players by carrying out QoE assessments subjectively in a balloon bursting game, where two players burst balloons (i.e., soft objects) in a 3D virtual space. We also perform objective QoE assessment at the same time as the subjective assessment. We further investigate the relationship between subjective and objective assessment results.

7.2 Balloon bursting game

7.2.1 System configuration

The system configuration of the balloon bursting game is shown in Fig.1, where each of two players (*players 1* and *2*) bursts balloons with his/her respective stylus in a 3D virtual space, and the two players compete with each other for the number of burst balloons. The system consists of two terminals (*terminals 1* and 2), each of which has a PC, a haptic interface device (Geometric Touch [10]), a display, and a headset. A player at each terminal employs the haptic interface device to move the virtual stylus (a CG image of the stylus of the haptic interface device) in the virtual space. When the player touches the balloon with the tip of the stylus, the reaction force is perceived through the haptic interface device; he/she can feel the softness of the balloon. The balloon is distorted when the player pushes the balloon with the stylus. If he/she pushes it strongly, the balloon is largely distorted, and it is burst and disappeared. Then, he/she hears a sound of bursting it via the headset.

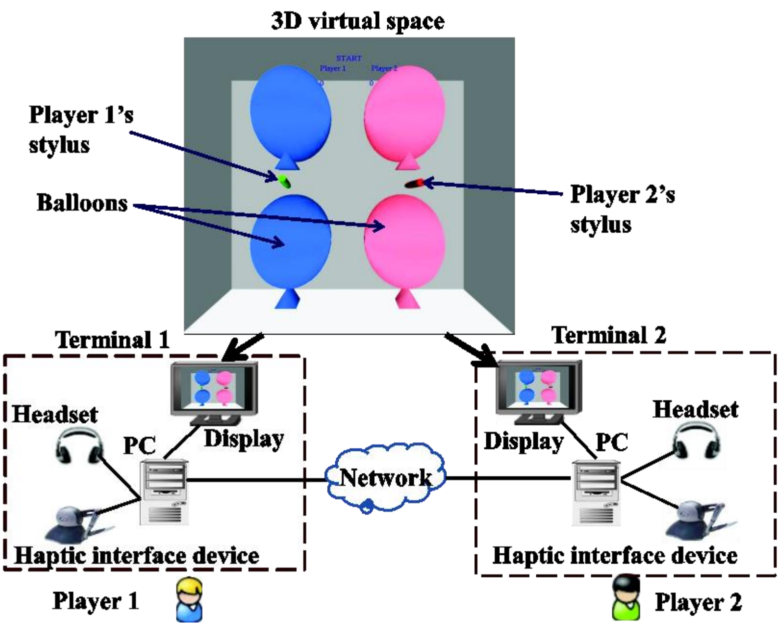


Figure 7.1. System configuration of balloon bursting game.

7.2.2 Playing method

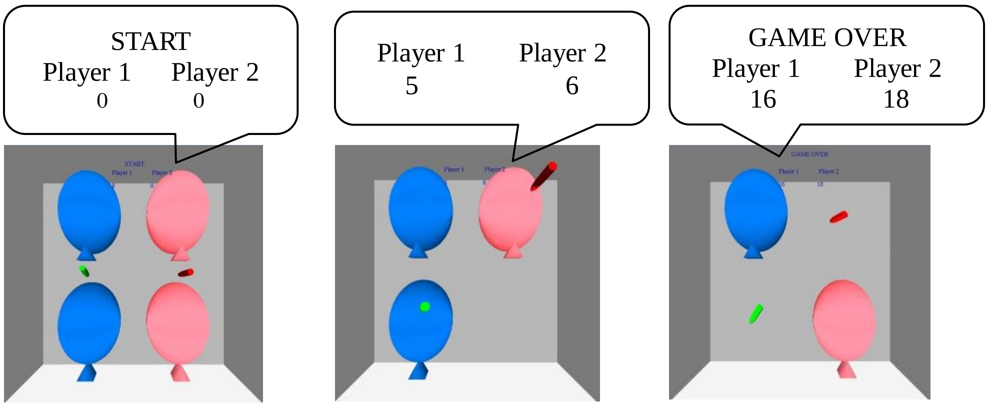
In this chapter, as shown in Fig. 7.1, there are four balloons in the 3D virtual space. Player 1 bursts two blue balloons alternately on the left side of the virtual space, and player 2 bursts two pink balloons on the right side. This purpose is to avoid the situation of trying to burst the same balloon simultaneously at the two terminals for simplicity. If both players try to burst the same balloon at the same time, we can determine which player bursts the balloon earlier than the other user by using AtoZ [97] and Count Down Protocol [98] as in [99]. Before the start of the game, the players stand ready by placing their styli at their respective original positions (see Fig. 7.2 (a)). The players start to burst the balloons when "START" message is displayed on the screen and a buzzer sound is output (see Fig. 7.2 (a)). During the game, the numbers of burst balloons of the two players are displayed on the screen (see Fig. 7.2 (b)). The players stop the game when "GAME OVER" message appears on the screen 30 seconds after the beginning of the game (see Fig. 7.2 (c)). The buzzer sound also alerts the players to stop the game at that time. A player who has burst more balloons than the other player wins the game; player 2 wins in Fig. 7.2 (c). When a balloon is burst and disappeared, a new balloon automatically appears at the location of the burst balloon. Both players try to burst their respective balloons from the front side of the balloon as fast as they can.

7.2.3 Calculation of reaction force

The reaction force applied to the haptic interface device is generated by the haptic rendering engine [18], which uses the object shape and material properties such as stiffness and friction for calculation of the reaction force. In order to get the reaction force against a player via the haptic interface device, we use the following function of the OpenHaptics HLAPI (Haptic Library Application Programing Interface) [18]:

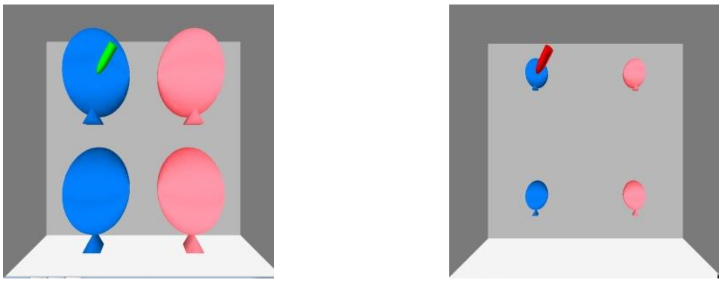
hlGetDoublev (HL_REACTION_FORCE, r_force),

where hlGetDoublev() is a function used for getting the reaction force (HL_REACTION_FORCE) from the haptic rendering engine, and r_force is a user-defined parameter which receives a returned value of HL_REACTION_FORCE. The value of the force applied to a balloon when a player pushes the balloon with a stylus is received from the parameter HL_CURRENT _FORCE. In a preliminary experiment, we confirmed that the value of HL_CURRENT_ FORCE is almost equal to that of HL_REACTION_FORCE; that is, a player feels almost the same reaction force to the force pushed by him/her. The player feels larger reaction force as the penetration



(a) Start of game (b) During game (c) End of game

Figure 7.2. Displayed images in virtual space.



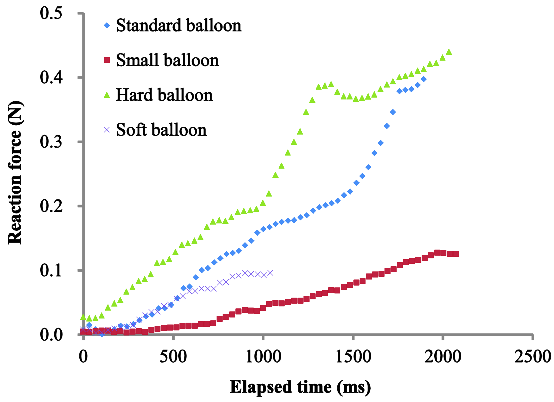
(a) Standard, hard, and soft balloon cases (b) Small balloon case

Figure 7.3. Displayed images of balloons in standard, small, hard, and soft balloon cases.

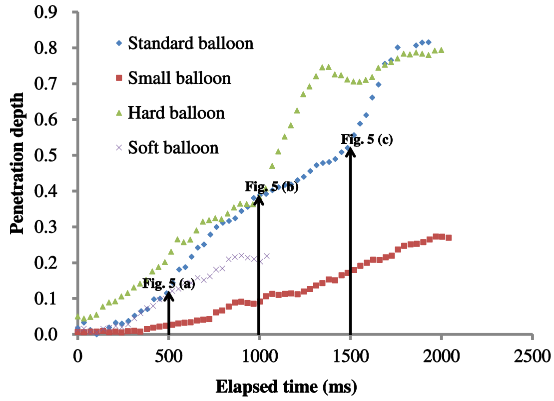
depth of the stylus becomes larger. The penetration depth of the stylus is the distance from the surface of the balloon to the tip of the stylus. When the balloon is distorted by the stylus, its volume is varied. We can obtain the volume from the haptic rendering engine.

In order to clarify the relations among the reaction force, the penetration depth, and the volume of a balloon, we here explain four cases (called the *standard balloon case*, *small balloon case*, *hard balloon case*, and *soft balloon case*). In the standard balloon case, the radii of three dimensional axes (x, y, and z) of the balloon are 1.1, 1.5 and 1.1, respectively (see Fig. 7.3 (a)), where we assume that the length of the stylus is 1.0. In the small balloon case, we use a balloon which has the radii of 0.35, 0.48, and

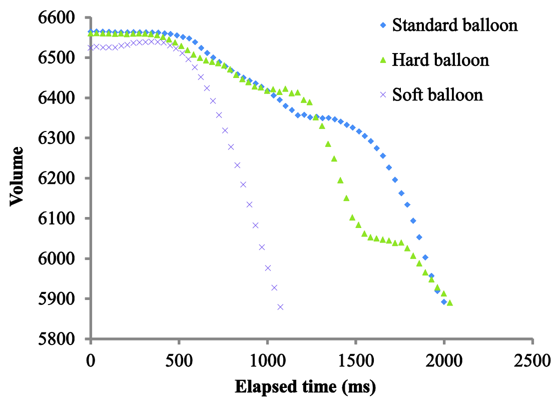
0.35, respectively (see Fig. 7.3 (b)). Balloons in the hard and soft balloon cases are two times harder and softer, respectively, than those in the standard balloon case; the sizes of the balloons are the same as that in the standard balloon case. In Figs. 7.4 (a) through (d), we plot the reaction force, penetration depth, and volume versus the elapsed time for the four cases. The elapsed time is defined as the time interval from the moment the surface of a balloon is touched with the stylus until the instant the balloon is burst. In Fig. 7.5, we show displayed images of a balloon for the standard balloon case when the balloon is distorted; Figures 7.5 (a), (b), and (c) corresponds to the standard balloon at the elapsed times of 500 ms, 1000 ms, and 1500 ms, respectively, in Fig. 7.4 (b).



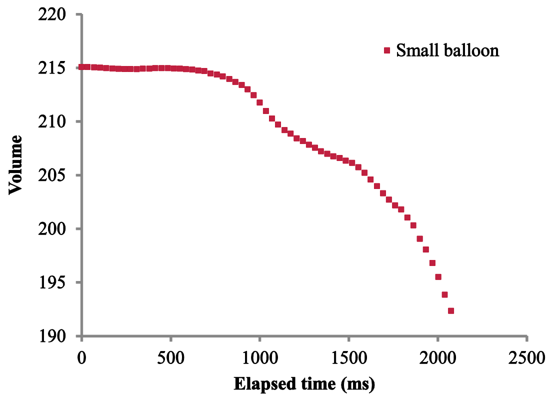
(a) Reaction force



(b) Penetration depth to balloon

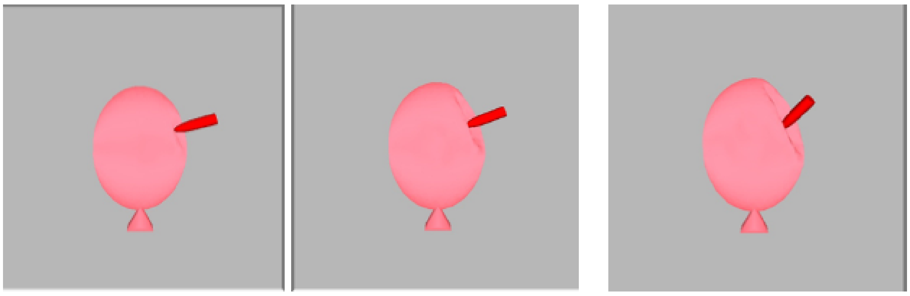


(c) Volume of balloon in standard, hard, and soft balloon cases



(d) Volume of balloon in small balloon case

Figure 7.4. Relations among reaction force, penetration depth, and volume of balloon.



(a) Elapsed time: 500 ms (b) Elapsed time: 1000 ms (c) Elapsed time: 1500 ms

Figure 7.5. Displayed images of distorted balloon.

7.2.4 Judgment of bursting

There may be several methods of judgment of bursting a balloon in a virtual space. For example, a balloon can be burst according to the volume of the balloon and/or the force pushed with the stylus. In the former method, the balloon is burst when the volume of the balloon is smaller than a threshold value. In a preliminary experiment where we used the former method, we set the threshold value to 90% of the initial volume of the balloon.

In the latter method, when the force applied to the balloon from the cursor of the haptic interface device is larger than another threshold value, the balloon is burst. We set the threshold value to 0.4 N in our preliminary experiment.

As the other method, if either of the conditions of the above methods is met, the balloon can be burst. In our assessment in this chapter, we use a method in which a balloon is burst when the volume of a balloon reaches a threshold value for simplicity.

7.3 Assessment environment

7.3.1 Assessment system

In our assessment system, the two terminals are connected to each other via a network emulator (NIST Net [86]). The network emulator generates an additional constant delay for each packet transmitted between the terminals. We can here take account of network delay jitter as in the previous chapters. We call the constant delay from terminal 1 to 2 *delay 1*, and that from terminal 2 to 1 *delay 2* (see Fig. 7.6). We call the local lag at terminal 1 *local lag 1* and that at terminal 2 *local lag* 2. There are two cases in our assessment (called the *symmetric delay case* and the *asymmetric delay case*). In the symmetric delay case, the constant delay from terminal 1 to terminal 2 is always equal to that in the opposite direction. In the asymmetric delay case, the constant delay from terminal 1 to terminal 2 can be different from that in the opposite direction.

7.3.2 Assessment methods

(1) Symmetric delay case

In this assessment, we investigated the influence of network delay on only the operability of haptic interface device. We did not investigate the fairness between players because the fairness is perfect when the values of network delay are the same between the terminals. In order to focus on the relation between the network delay and the operability of haptic interface device, we only used terminal 1 in this case. Local lag 1 was set to the same values as delay 1. We carried out QoE assessment with 16 subjects (males and females) whose ages were between 20 and 30. Each subject practiced about two minutes under the condition that the constant delays were set to 0 ms before the assessment for each case. In the assessment, each subject burst the two blue balloons alternately which were located on the left side of the virtual space. When a balloon was burst, a new balloon always automatically appeared for the burst balloon. In order to investigate the influence of network delay on the different types of balloons, we carried out four assessments using four types of balloons which are described in subsection 7.2.3. The assessments in this case are called the *standard balloon assessment*, *small balloon assessment*, *hard balloon assessment*, and *soft balloon assessment*, respectively, according to the types of balloons that we use.

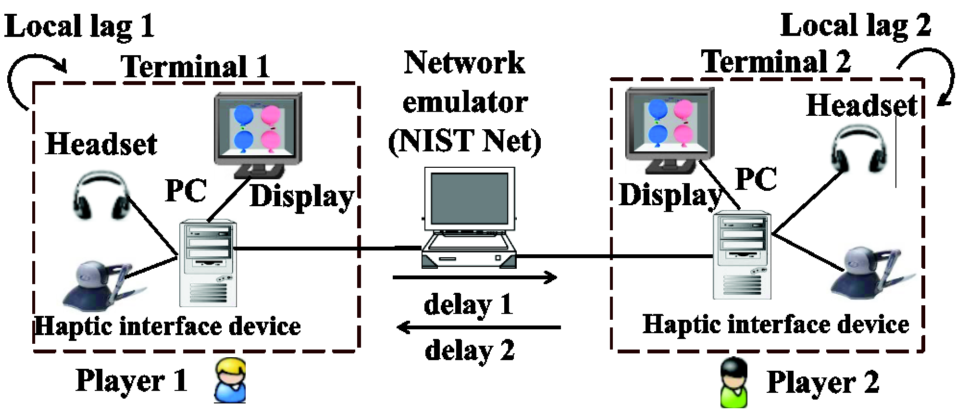


Figure 7.6. Configuration of assessment system.

In each assessment, we changed the delays in random order from 0 ms to 500 ms at intervals of 50 ms for each subject. The order of assessments was also selected in random order for each subject. The subject burst the balloons with the stylus continuously for 30 seconds. After each stimulus, he/she was asked to base his/her judgment about the easiness of bursting based on the five-grade impairment scales (5: Imperceptible, 4: Perceptible, but not annoying, 3: Slightly annoying, 2: Annoying, 1: Very annoying) [74]. By averaging the scores of all the subjects, we obtained Mean Opinion Score (MOS) [74]. As objective assessment measure, we adopted the number of burst balloons. The total assessment time for each case was about 20 minutes per subject.

(2) Asymmetric delay case

In this case, we carried out QoE assessment to clarify the influences of network delay on the operability, fairness, and comprehensive quality. In the assessment, two subjects played the balloon bursting game together. Before the assessment, each pair of subjects played the balloon bursting game for three times to get used to the game on the condition that delays 1 and 2 are set to 0 ms; that is, the same condition is provided to the pair. By practicing, each subject knows how to burst a balloon by using a haptic interface device. In this case, we carried out the assessment by setting delay 1 to 0 ms or 200 ms, and setting delay 2 to 50 ms, 100 ms, 200 ms, 300 ms, or 500 ms. We carried out QoE assessment with 16 subjects (males and females) whose ages were between 20 and 30.

Assessments in which delay 1 was set to 0 ms and 200 ms are referred to as *assessments 1* and *2*, respectively. In order to know the suitable local lag values at the two terminals to maintain both operability and fairness, we carried out the assessments by setting local lags 1 and 2 to different values. Local lag 1 was changed from 0 ms to 500 ms at intervals of 50 ms, and local lag 2 was set to the same value as delay 2. In each assessment, the order of combinations of delays and the local lags were changed in random order for the pair. It took 30 seconds for each stimulus. After each stimulus, each subject was asked to base his/her judgment about the operability, fairness, and comprehensive quality in terms of wording used to define the five-grade impairment scale (5: Imperceptible, 4: Perceptible, but not annoying, 3: Slightly annoying, 2: Annoying, 1: Very annoying) [74]. In each stimulus, if the pair obtained almost the same results about victory or defeat as those in the practice, the pair regarded the fairness as high and valued the score at 5. The comprehensive quality is the weighted sum of the operability and interactivity; thus, the comprehensive quality is the most important. Each subject gave a score from 1 through 5 to each stimulus. By averaging the scores of all the subjects, we obtained MOS. We also adopted the number of burst balloons as an objective assessment measure. The total assessment time for each assessment was about two hours per pair.

7.4 Assessment results

7.4.1 Subjective assessment

(1) Symmetric delay case

We show the MOS values of operability versus delay 1 for four assessments in Fig. 7.7. In Fig. 7.7, we see that the MOS values of all the assessments decrease as delay 1 becomes larger. This is because the local information at terminal 1 is buffered for a time of delay 1; thus, since the interactivity is degraded, each subject feels that the balloon becomes harder and more slippery, and it is difficult to burst the balloon. From the figure, we also find that the MOS value of the small balloon assessment is smaller than those of the other cases when delay 1 is larger than or equal to about 100 ms. The reason is that it is more difficult to operate a virtual stylus so as not to slip from the small balloon when the delay 1 is large. Moreover, we observe in the figure that when delay 1 is smaller than or equal to about 150 ms, the MOS value is higher than 3.5; this means that the deterioration in QoE is allowable for network delays smaller than or equal to about 150 ms according to [100].

(2) Asymmetric delay case

We show the MOS values of operability at terminals 1 and 2 for assessment 1 as a function of local lag 1 in Figs. 7.8 and 7.9, respectively. We also plot the MOS values of fairness at terminals 1 and 2 for assessment 1 in Figs. 7.10 and 7.11, respectively. In Figs. 7.12 and 7.13, we draw the MOS values of operability at terminals 1 and 2 for assessment 2, respectively. The MOS values of fairness at terminals 1 and 2 for assessment 2 are presented in Figs. 7.14 and 7.15, respectively. The 95% confidence intervals are also plotted in the figures.

In Figs. 7.8 and 7.12, we see that the MOS values of operability for all the values of delay 2 at terminal 1 decrease as local lag 1 increases as in the symmetric delay case. This is because the local information at terminal 1 is buffered for a time of local lag 1; thus, since the interactivity is degraded, each subject feels that the balloon becomes harder and more slippery, and it is difficult to burst the balloon. From Figs. 7.9 and 7.13, we find that the MOS values at terminal 2 hardly depend on local lag 1, and they depend on mainly delay 2 or local lag 2. This is because the local information at terminal 2 is buffered for a time of delay 2.

In Figs. 7.10, 7.11, 7.14 and 7.15, we notice that the MOS values of fairness for all the values of delay 2 increase as local lag 1 increases. This is because each subject feels fairness strongly when the absolute difference of local lags 1 and 2 becomes smaller.

By comparing Figs. 7.8 and 7.10 or Figs. 7.12 and 7.14, we confirm that there is a trade-off relationship between the MOS value of operability and that of fairness at terminal 1; that is, as local lag 1 increases, the MOS value of operability becomes smaller, but that of fairness becomes larger. Also, from Figs. 7.9, 7.11, 7.13, and 7.15, we note that the MOS value of operability at terminal 2 hardly changes, but the MOS value of fairness becomes larger.

To clarify the optimum values of local lag at both terminals, we have calculated the average MOS values of two terminals for assessments 1 and 2, which are shown in Figs. 7.16 through 7.19. In Fig. 7.16 and 7.18, we see that the average MOS value of operability for two terminals decreases as local lag 1 becomes larger for each value of delay 2. In the figure, we also find that delay 2 of 50 ms has the highest MOS value, and delay 2 of 500 ms has the lowest for each value of local lag 1. In Figs. 7.17 and 7.19, we notice that the average MOS values increase as local lag 1 becomes larger for all the values of delay 2. From the average MOS values of the two terminals, we further confirm that the trade-off relationship exists between the MOS value of operability and that of fairness. From Figs. 7.17 and 7.19, we also observe that the MOS value is larger than 3.5 when the absolute value of local lag difference (local lag 1 minus local lag 2) is smaller than or equal to around 75 ms. Thus, the allowable range is around 75 ms according to [100]. This information of allowable range of difference in local lag difference may be helpful for networked applications in which only fairness is focused to be maintained.

We show the MOS values of comprehensive quality at terminals 1 and 2 for assessment 1 in Figs. 7.20 and 7.21, respectively. We also plot the MOS values of comprehensive quality at terminals 1 and 2 for assessment 2 in Figs. 7.22 and 7.23, respectively. The average MOS value of comprehensive quality of two terminals for assessments 1 and 2 are also plotted in Figs. 7.24 and 7.25, respectively. Figures 7.20 and 7.22 reveal that the MOS values of comprehensive quality for all the values of delay 2 at terminal 1 hardly depend on local lag 1. In Figs. 7.21 and 7.23, we find that the MOS value hardly depends on local lag 1 when delay 2 is 50 ms or 100 ms. For the other values of delay 2, the MOS values increase as local lag 1 becomes larger. From Figs. 7.24 and 7.25, we note that the average MOS value for delay 2 of 50 ms hardly depends on local lag 1. For the other values of delay 2, the average MOS values tend to increase slightly as local lag 1 becomes larger. When delay 2 is 100 ms, the optimum value of local lag 1 is from about 75 ms to around 100 ms. When delay 2 is 200 ms, 300 ms or 500 ms, the optimum value of local lag 1 is within the range from about 125 ms to around 200 ms, from about 225 ms to around 300 ms, or from about 425 ms to around 500 ms, respectively.

From Figs. 7.20, 7.21, 7.22 and 7.23, we notice that the contribution of the fairness is larger than that of the operability to the comprehensive quality. From Figs. 7.24 and 7.25, we can also clarify how the local lags should be set to maintain the comprehensive quality as high as possible at both terminals. Local lags 1 and 2 should be set to the same values of delays 1 and 2, respectively, if the difference in network delays between the terminals is smaller than or equal to about 50 ms. When the difference is larger than about 50 ms, for simplicity, the local lags can be set to the larger value of network delays between the terminals as in the adaptive Δ-causality control.

7.4.2 Objective Assessment Results

(1) Symmetric delay case

We also show the average number of burst balloons versus delay 1 in Fig. 7.26. Figure 7.26 reveals that the average numbers of burst balloons in all the assessments become smaller as delay 1 increases. Form the figure, we also observe that the average number of burst balloons in the soft balloon assessment is the highest. The average number of burst balloons in the hard balloon assessment is the smallest when delay 1 is smaller than around 250 ms. When delay 1 is larger than about 300 ms, the number of burst balloons in the small balloon assessment is the smallest.

(2) Asymmetric delay case

We show the average numbers of burst balloons at terminals 1 and 2 for assessment 1 in Figs. 7.27 and 7.28, respectively. We also plot the average numbers of burst balloons at terminals 1 and 2 for assessment 2 in Figs. 7.29 and 7.30, respectively. The average difference in the number of burst balloons for two terminals in assessments 1 and 2 are shown in Fig. 7.31 and 7.32, respectively. The 95% confidence intervals are also plotted in the figures.

In Figs. 7.27 and 7.29, we see that the average number of burst balloons at terminal 1 becomes smaller as local lag 1 increases for each value of delay 2. In Figs. 7.28 and 7.30, we find that the average number of burst balloons at terminal 2 hardly depends on local lag 1. By comparing Figs. 7.27 and 7.28 with Figs. 7.8 and 7.9, or comparing Figs. 7.29 and 7.30 with Figs. 7.12 and 7.13, respectively, we can say that the MOS value of operability decreases as the average number of burst balloons becomes smaller for each value of delay 2. From Figs. 7.31 and 7.32, we notice that the average difference in the number of burst balloons becomes smaller as local lag 1 increases for each value of delay 2. This is because the local lag difference between the two terminals becomes smaller as local lag 1 increases. By comparing Figs. 7.31 and 7.32 to Fig. 7.10 or 7.11, and Figs. 7.14 or 7.15, we see that the tendencies of the curves are inverses of each other; that is, the highest MOS value can be obtained when the average difference in the number of burst balloons is the smallest for each value of delay 2.

7.4.3 Relations between Objective and Subjective Results

As described in Section 7.4.2, by comparing the objective results to subjective results, we found that they are related to each other. We should clarify how they are related to each other in detail, and examine whether the subjective results can be estimated from the objective results or not. To investigate the relationship between the average number of burst balloons (or the local lag) and the MOS value of operability, we carried out the regression analysis [75]. As a result, we obtained estimated equations shown in Table 7.1 for both terminals of the two assessments. In Table 7.1, O_{MOS} denotes the estimated MOS value of operability, N_{burst} is the average number of burst balloons, Δ is the local lag, and R^2 is the contribution rate adjusted for degrees of freedom [75], which shows goodness of fit with the estimated equation. Since the contribution rates are very high in Table 7.1, we can say that the MOS value of operability can be estimated with a high degree of accuracy from the average number of burst balloons or the local lag.

We also carried out regression analysis to investigate the relationship between the MOS value of fairness and the absolute value of average difference in the number of burst balloons, and that between the MOS value of fairness and the absolute value of local lag difference. As a result, we obtained equations shown in Table 7.2 for both terminals in the two cases. In Table 7.2, F_{MOS} is the estimated MOS value of fairness,

$D_{N_{\mathrm{burst}}}$ is the average difference in the number of burst balloons, and D_{Δ} is the local lag difference. From Table 7.2, we find that the contribution rates are high. Thus, we can say that the MOS value of fairness can be estimated with a high degree of accuracy from the absolute value of average difference in the number of burst balloons or the absolute value of the local lag difference. Furthermore, according to high contribution rates shown in Table 7.3, where C_{MOS} is the estimated MOS value of comprehensive quality, we can say that we can estimate the MOS value of comprehensive quality with a high degree of accuracy from the average number of burst balloons and the absolute value of average difference in the number of burst balloons, or from the lcoal lag and the absolute value of local lag difference.

In addition, we examined the relationship between the average MOS values of two terminals and the objective results. The estimated equations for average MOS value of operability are shown in Table 7.4, where $A_{O_{\mathrm{MOS}}}$ is the estimated average MOS value of operability, N_{burst1} is the average number of burst balloons at terminal 1, and N_{burst2} is the average number of burst balloons at terminal 2. From Table 7.4, we notice that the averge MOS value of operability can be estimated from the average number of burst balloons at two terminals to a large extent since the contribution rates in the table are high. We also found that the estimated equations for average MOS value of fairness are almost the same as those in Table 7.2. This means that the average MOS value of fairness can be estimated with a high degree of accuracy from the absolute value of average difference in number of burst balloons or the absolute value of the local lag difference. The estimated equations for average MOS value of comprehensive quality are shown in Table 7.5, where $A_{C_{\mathrm{MOS}}}$ is the average MOS value of comprehensive quality, Δ_1 is the local lag at terminal 1, and Δ_2 is the local lag at terminal 2. According to the high contribution rates shown in Table 7.5, we can say that the average MOS value of comprehensive quality can be estimated from the average number of burst balloons at two terminals and the absolute value of average difference in the number of burst balloons, or from the lcoal lags at two terminals.

Table 7.1. Estimated equations for MOS of operability.

Equation	R^2
$O_{\mathrm{MOS}} = 0.228N_{\mathrm{burst}} - 0.448$	0.968
$O_{\mathrm{MOS}} = -0.007\Delta + 4.634$	0.968

Table 7.2. Estimated equations for MOS of fairness.

Equation	R^2
$F_{\mathrm{MOS}} = -0.211\left\|D_{N_{\mathrm{burst}}}\right\| + 4.356$	0.891
$F_{\mathrm{MOS}} = -0.007\|D_{\Delta}\| + 4.315$	0.936

Table 7.3. Estimated equations for MOS of comprehensive quality.

Equation	R^2
$C_{\mathrm{MOS}} = 0.112N_{\mathrm{burst}} - 0.107\left\|D_{N_{\mathrm{burst}}}\right\| + 1.891$	0.931
$C_{\mathrm{MOS}} = -0.003\Delta - 0.003\|D_{\Delta}\| + 4.196$	0.928

Table 7.4. Estimated equations for average MOS of operability.

Equation	R^2
$A_{O_{\mathrm{MOS}}} = 0.106N_{\mathrm{burst1}} + 0.125N_{\mathrm{burst2}} - 0.515$	0.977
$A_{O_{\mathrm{MOS}}} = -0.004\Delta_1 - 0.003\Delta_2 + 4.590$	0.978

Table 7.5. Estimated equations for average MOS of comprehensive quality.

Equation	R^2
$A_{C_{\mathrm{MOS}}} = 0.022N_{\mathrm{burst1}} + 0.106N_{\mathrm{burst2}} - 0.071\left\|D_{N_{\mathrm{burst}}}\right\| + 1.679$	0.966
$A_{C_{\mathrm{MOS}}} = 0.001\Delta_1 - 0.004\Delta_2 + 4.180$	0.956

7.5 Summary

In this chapter, we investigated the influences of network delay on QoE such as the operability and fairness for a balloon bursting game in a networked virtual environment with haptic sense by carrying out subjective and objective QoE assessments. As a result, we found that the operability of haptic interface device strongly depends on the local lag. On the other hand, we can say that the operability depends on the network delay from the local terminal to the other terminal. We also observed that the soft objects become harder and more slippery as the network delay increases. Also, we examined that the fairness between players depends on the difference in network delay between the terminals or the local lag difference between the terminals. We further confirmed that there exists a trade-off relationship between the operability and fairness. Moreover, we clarified how the local lags should be set at each terminal to be able to keep both operability and fairness as follows. The local lag at each terminal can be set to the same value as the network delay from the local terminal to the other terminal when the absolute difference in network delay between the terminals is smaller than or equal to about 50 ms. When the difference is larger than about 50 ms, we can set the local lags to the larger value of network delay between the terminals.

Additionally, we examined the relationships between subjective and objective assessment results. We found that the MOS value of operability can be estimated from the average number of burst balloons or local lag with a high degree of accuracy. We also illustrated that the MOS value of fairness can roughly be estimated from the average difference in number of burst balloons or the absolute value of local lag difference. Furthermore, we noted that the MOS value of comprehensive quality can be estimated from the average number of burst balloons and the absolute value of average difference in number of burst balloons, or from the local lag and the absolute value of local lag difference to a large extent.

As our future work, we will carry out QoE assessments with other judgments of bursting balloons in the balloon bursting game; for example, a balloon can be burst when the force applied to the balloon reaches a threshold. We will also confirm the trade-off relationship of the operability and fairness in other networked real-time games. We need to confirm whether the MOS value of comprehensive quality can be kept high or not at both terminals by setting the local lags according to the difference in network delay between the terminals as described in Subsection 7.4.2. Furthermore, it is important to carry out the confirmation in networked real-time games over the Internet.

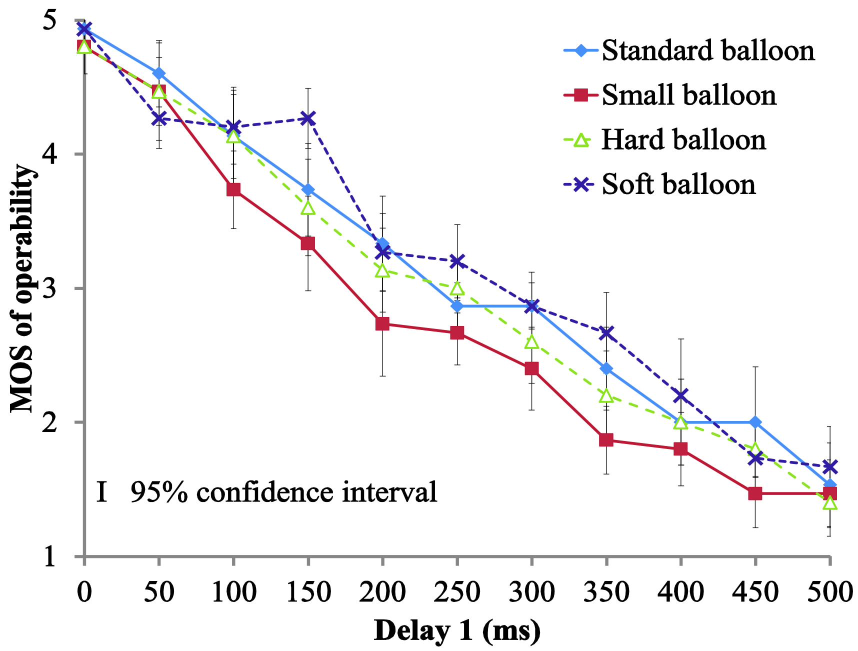


Figure 7.7. MOS of operability in symmetric delay case.

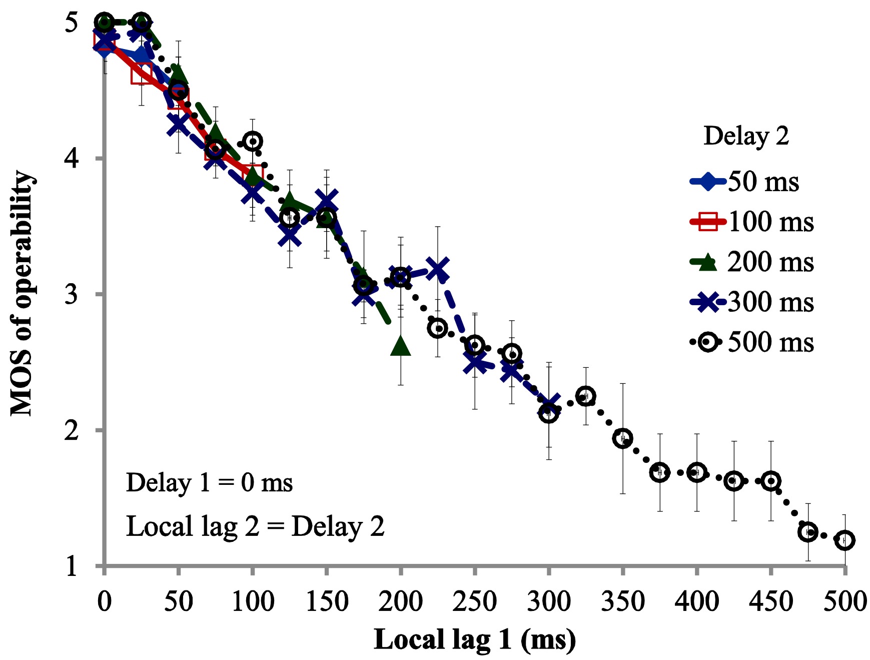


Figure 7.8. MOS of operability at terminal 1 for assessment 1 in asymmetric delay case.

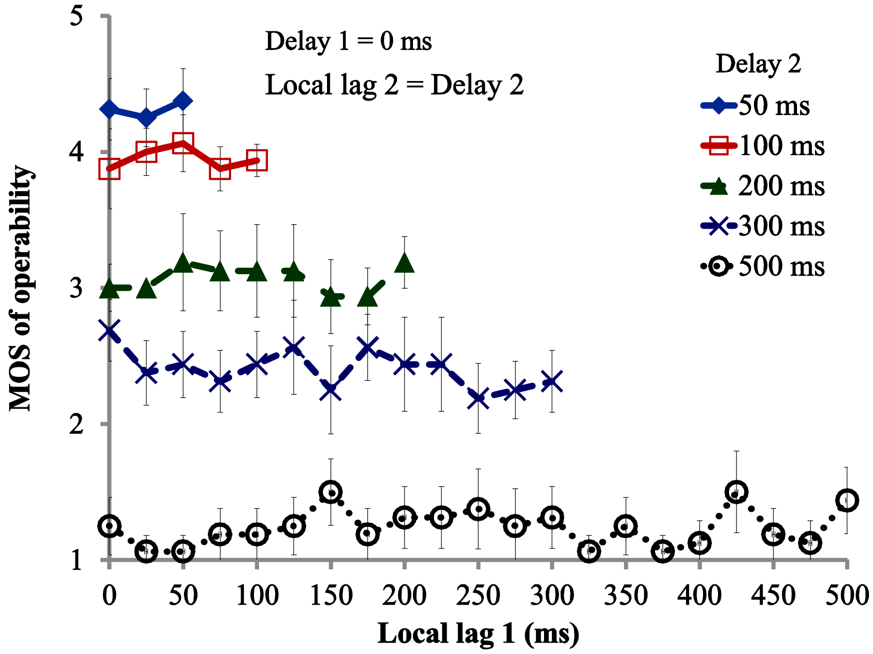


Figure 7.9. MOS of operability at terminal 2 for assessment 1 in asymmetric delay case.

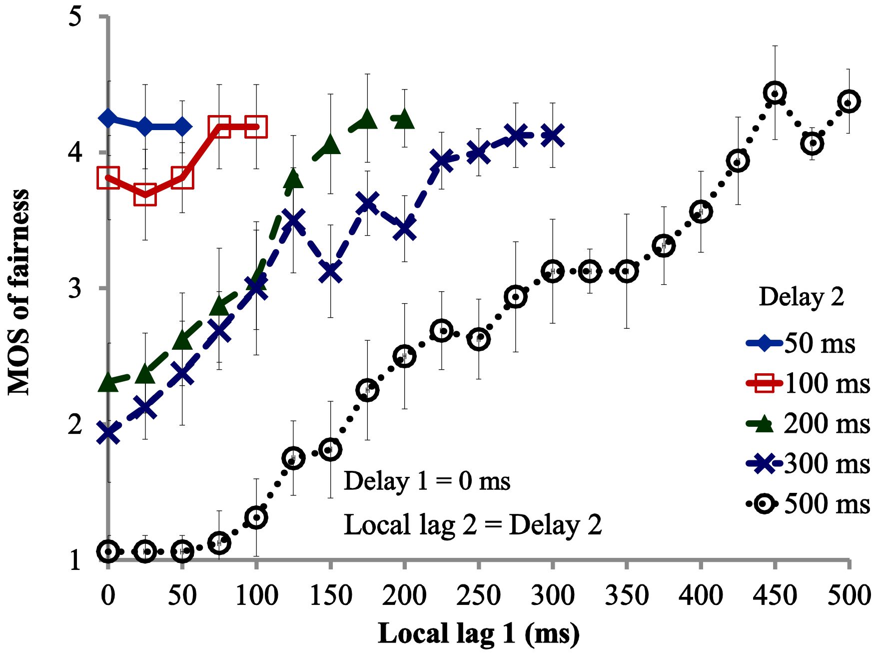


Figure 7.10. MOS of fairness at terminal 1 for assessment 1 in asymmetric delay case.

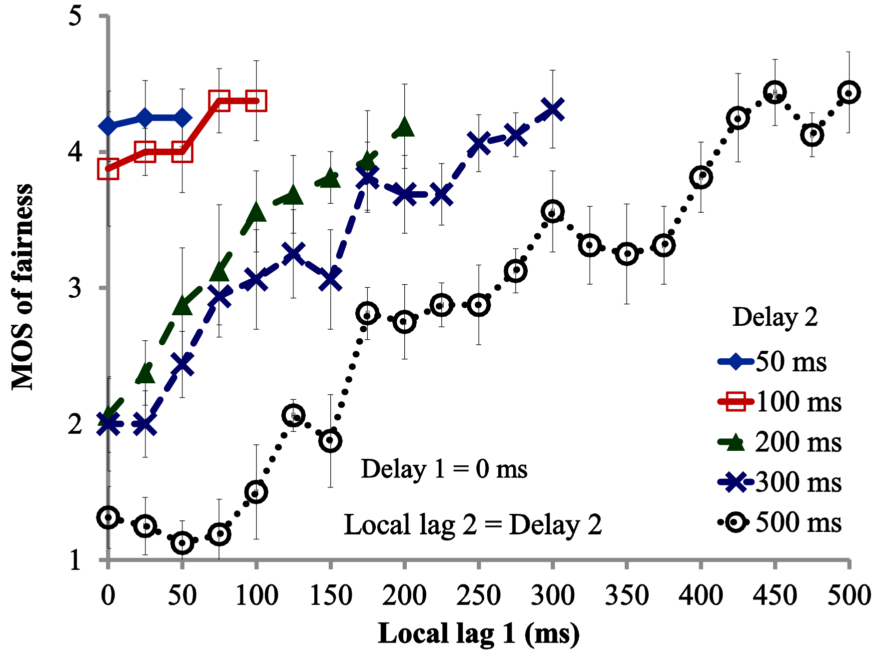


Figure 7.11. MOS of fairness at terminal 2 for assessment 1 in asymmetric delay case.

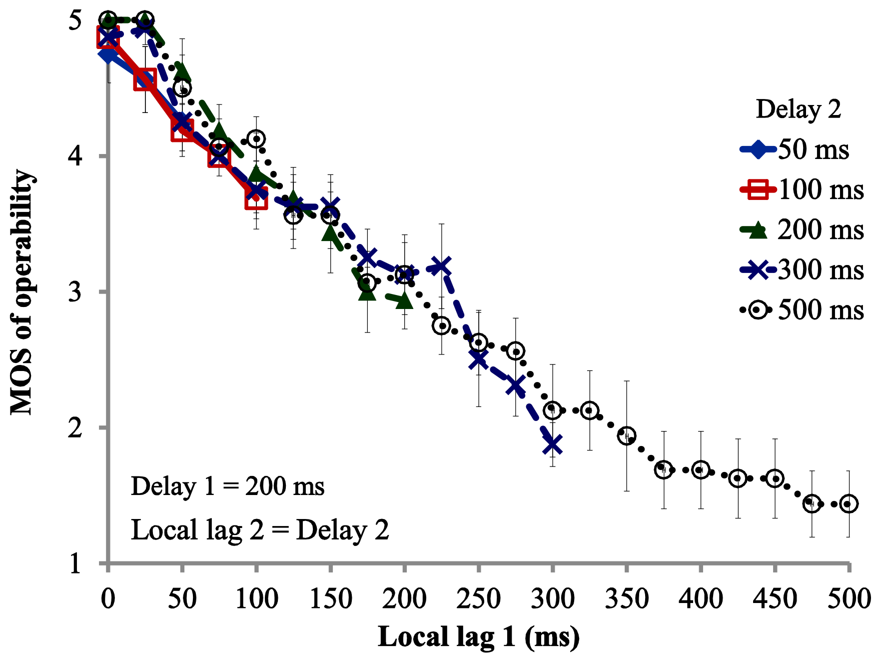


Figure 7.12. MOS of operability at terminal 1 for assessment 2 in asymmetric delay case.

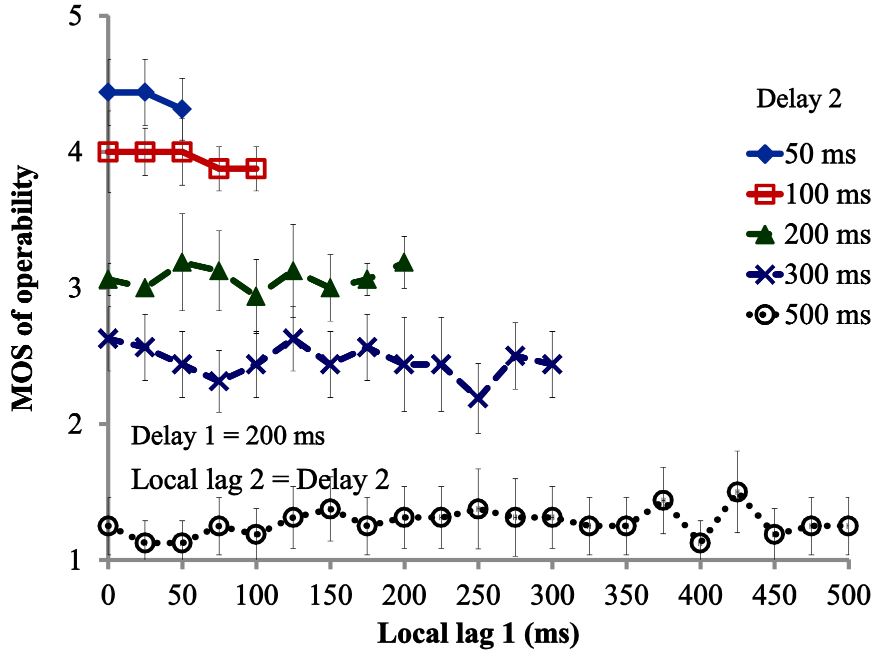


Figure 7.13. MOS of operability at terminal 2 for assessment 2 in asymmetric delay case.

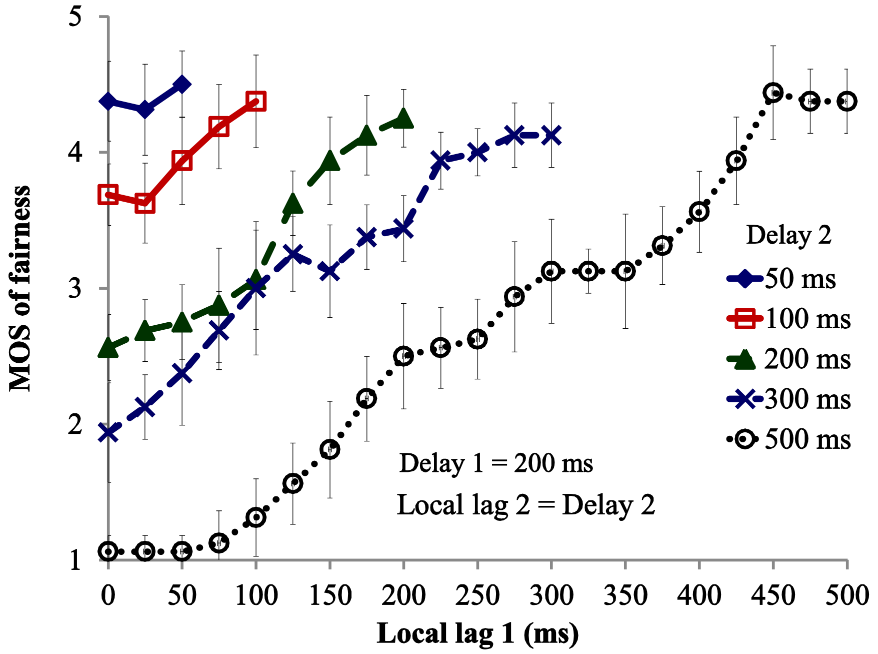


Figure 7.14. MOS of fairness at terminal 1 for assessment 2 in asymmetric delay case.

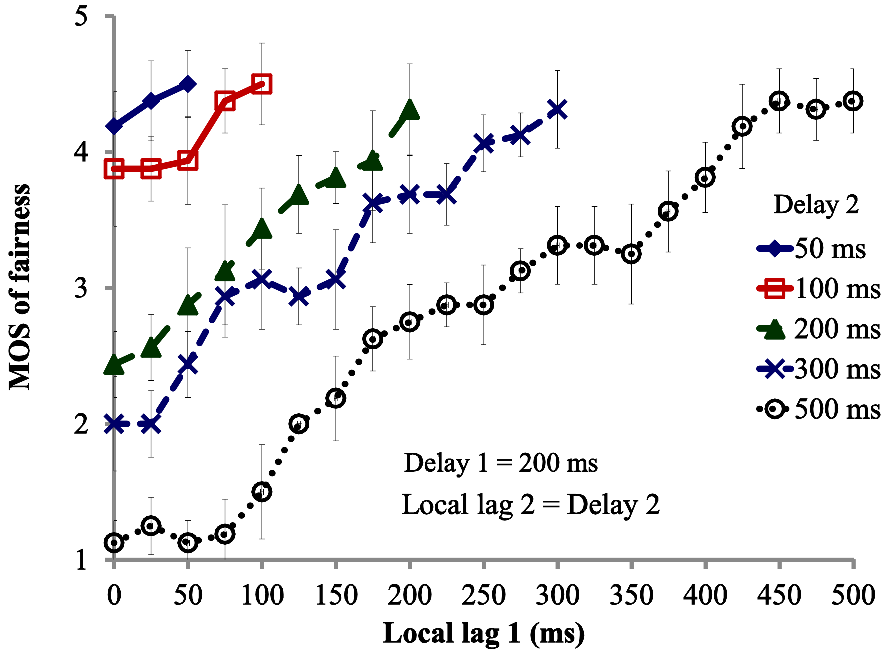


Figure 7.15. MOS of fairness at terminal 2 for assessment 2 in asymmetric delay case.

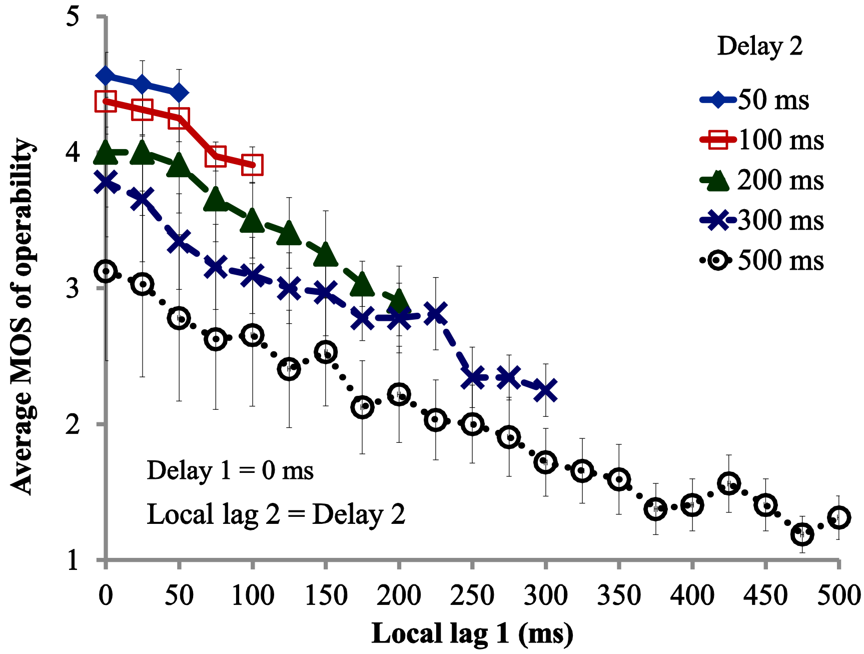


Figure 7.16. Average MOS of operability for two terminals for assessment 1 in asymmetric delay case.

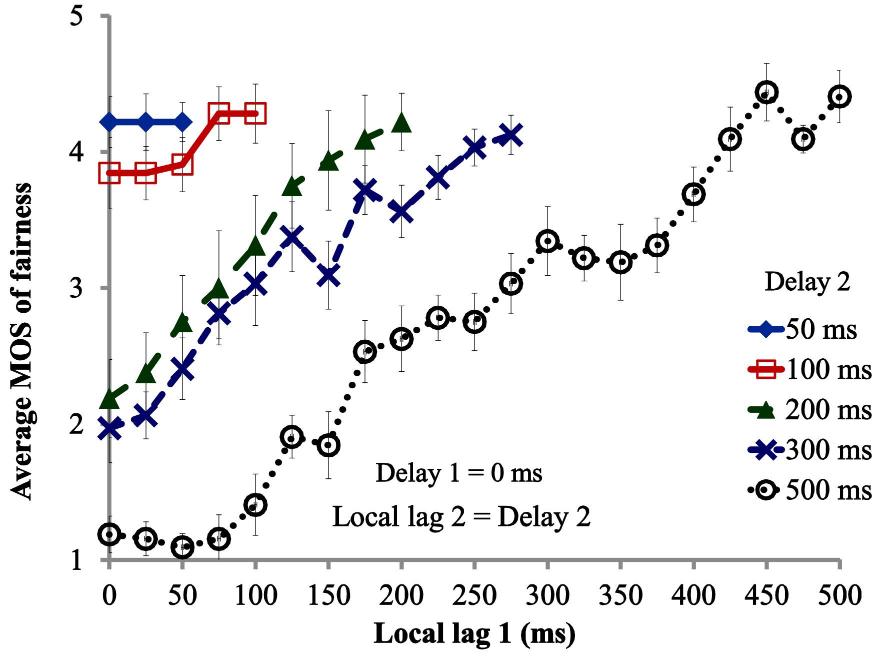


Figure 7.17. Average MOS of fairness for two terminals for assessment 1 in asymmetric delay case.

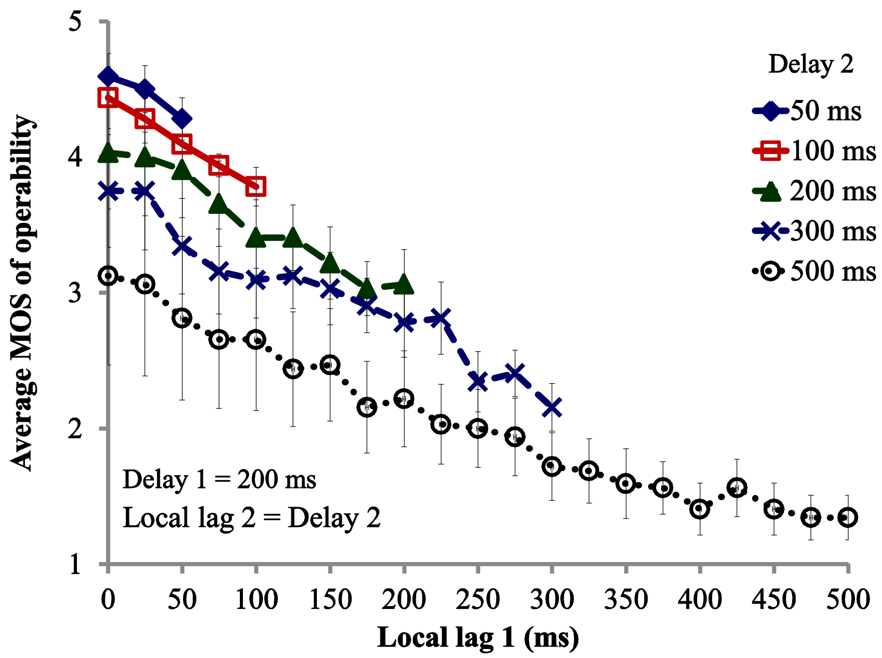


Figure 7.18. Average MOS of operability for two terminals for assessment 2 in asymmetric delay case.

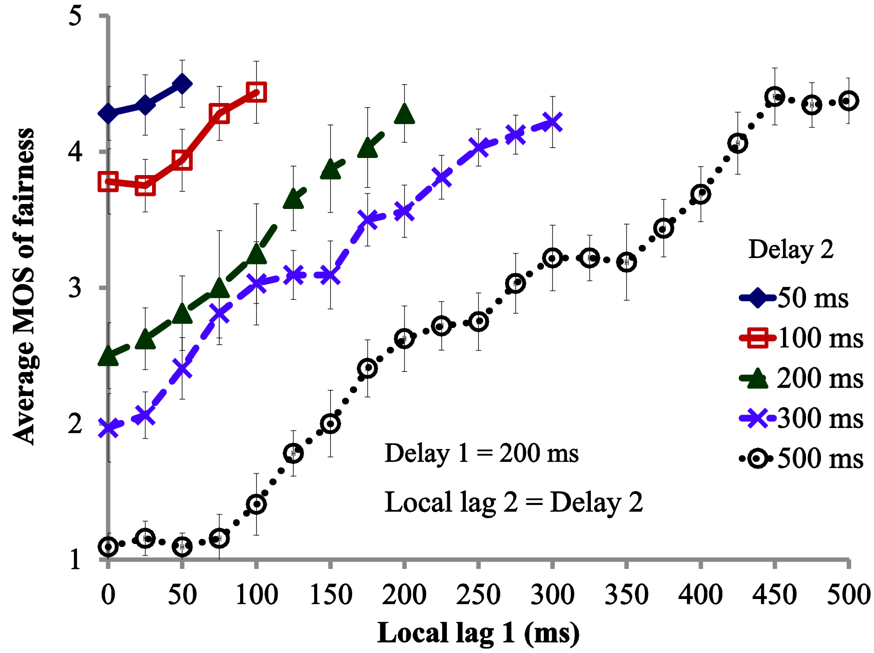


Figure 7.19. Average MOS of fairness for two terminals for assessment 2 in asymmetric delay case.

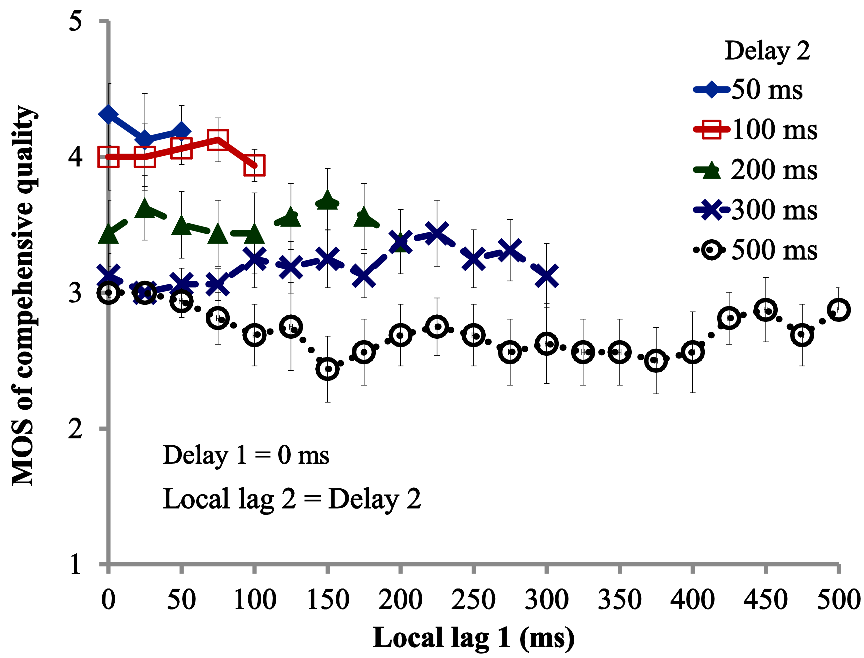


Figure 7.20. MOS of comprehensive quality at terminal 1 for assessment 1 in asymmetric delay case.

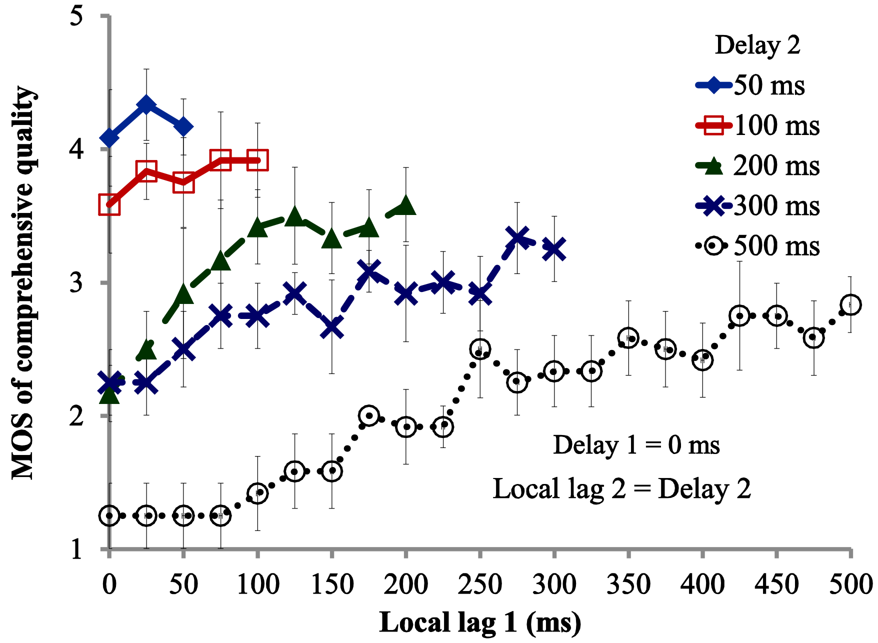


Figure 7.21. MOS of comprehensive quality at terminal 2 for assessment 1 in asymmetric delay case.

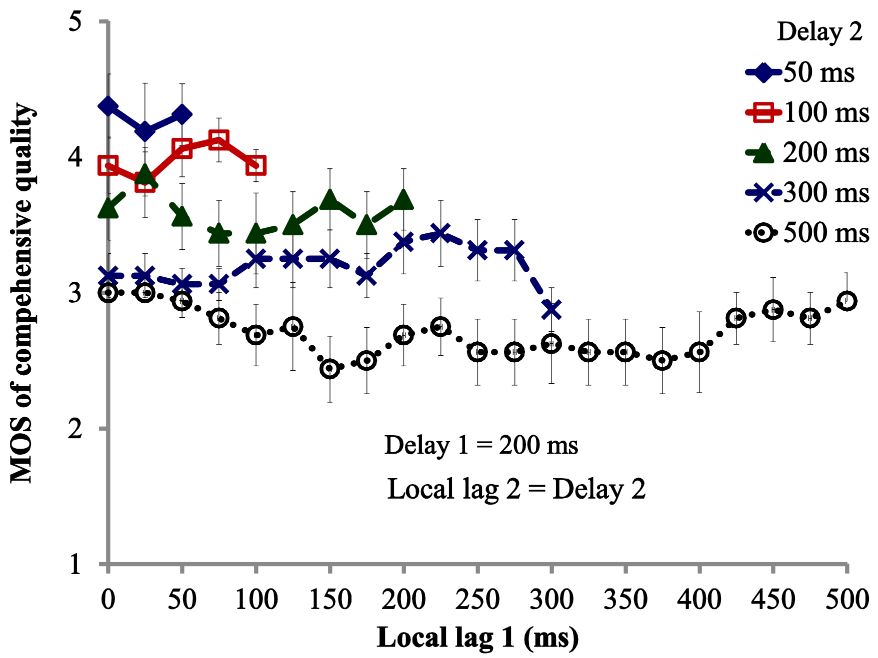


Figure 7.22. MOS of comprehensive quality at terminal 1 for assessment 2 in asymmetric delay case.

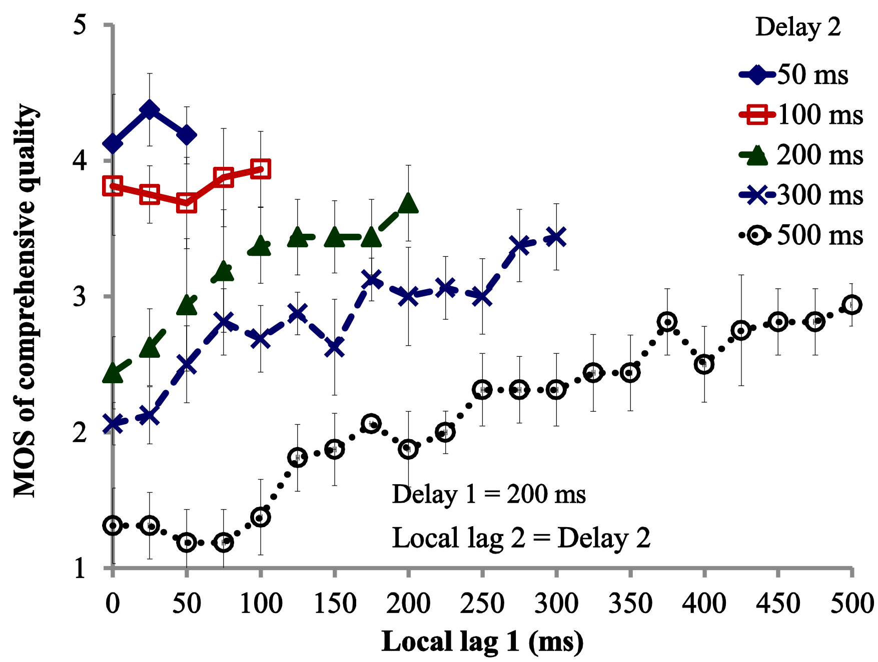


Figure 7.23. MOS of comprehensive quality at terminal 2 for assessment 2 in asymmetric delay case.

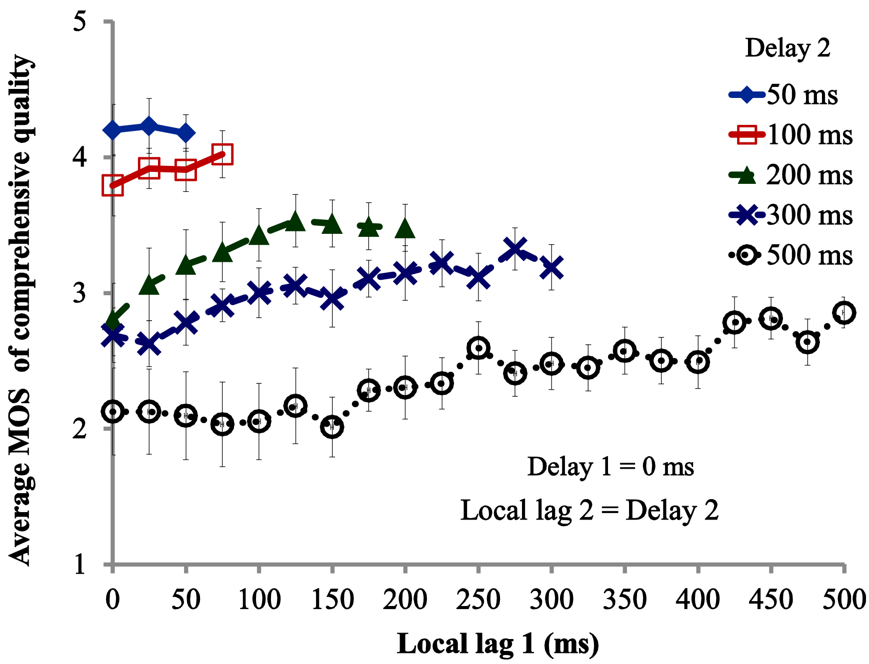


Figure 7.24. Average MOS of comprehensive quality for two terminals for assessment 1 in asymmetric delay case.

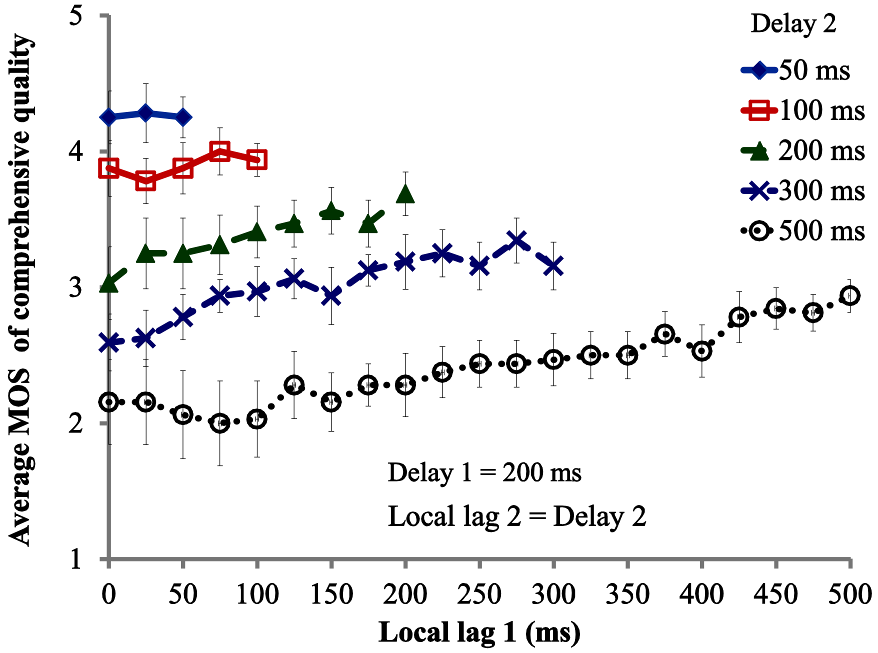


Figure 7.25. Average MOS of comprehensive quality for two terminals for assessment 2 in asymmetric delay case.

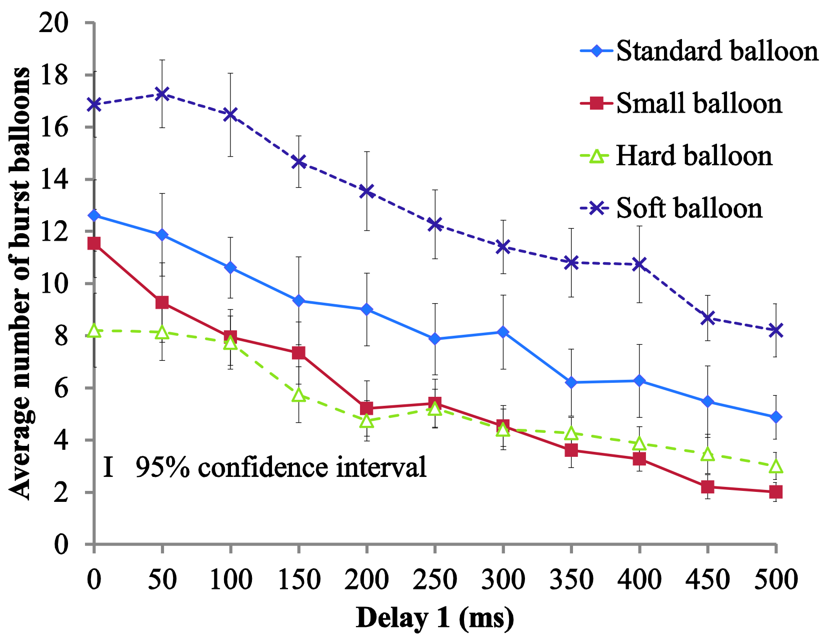


Figure 7.26. Number of burst balloons in symmetric delay case.

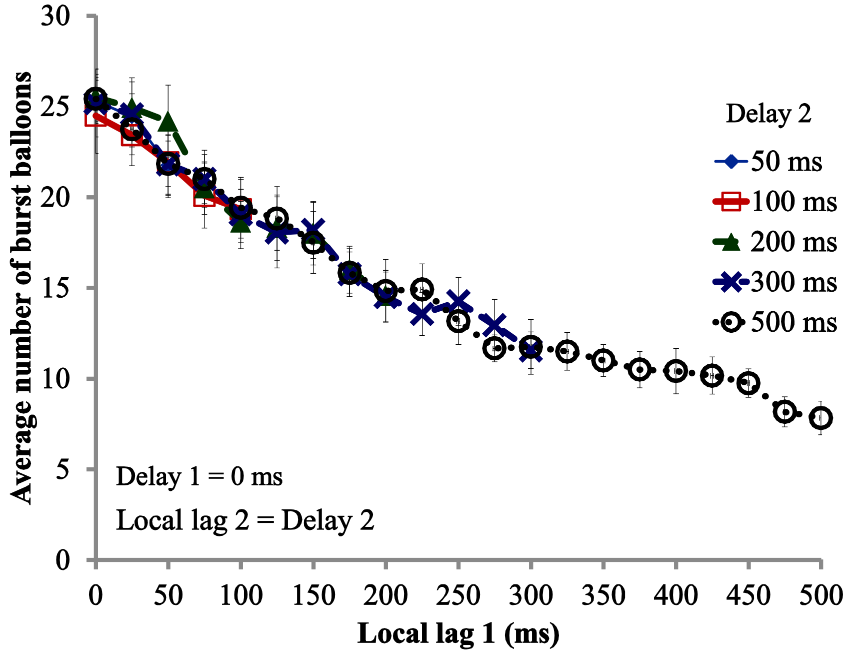


Figure 7.27. Average number of burst balloons at terminal 1 for assessment 1 in asymmetric delay case.

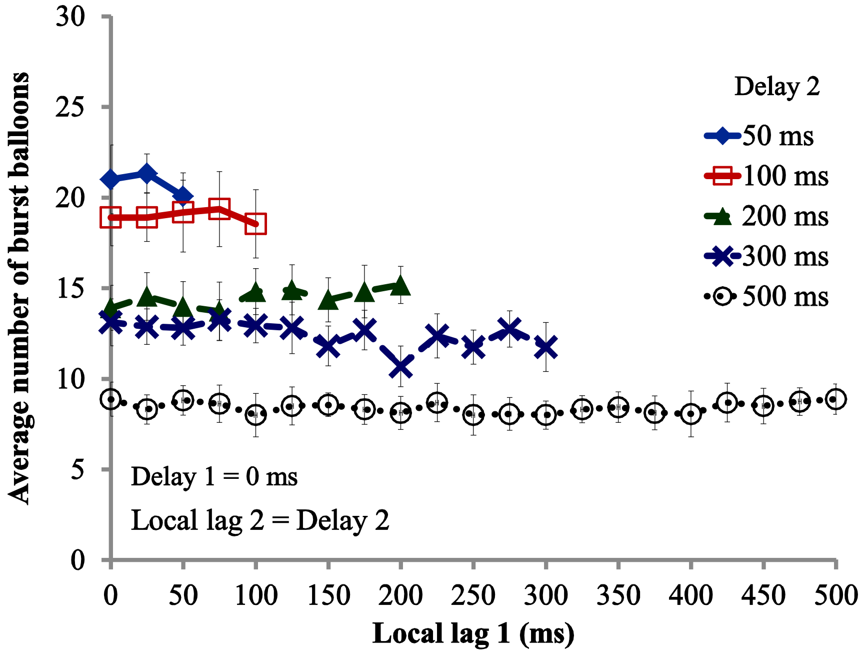


Figure 7.28. Average number of burst balloons at terminal 2 for assessment 1 in asymmetric delay case.

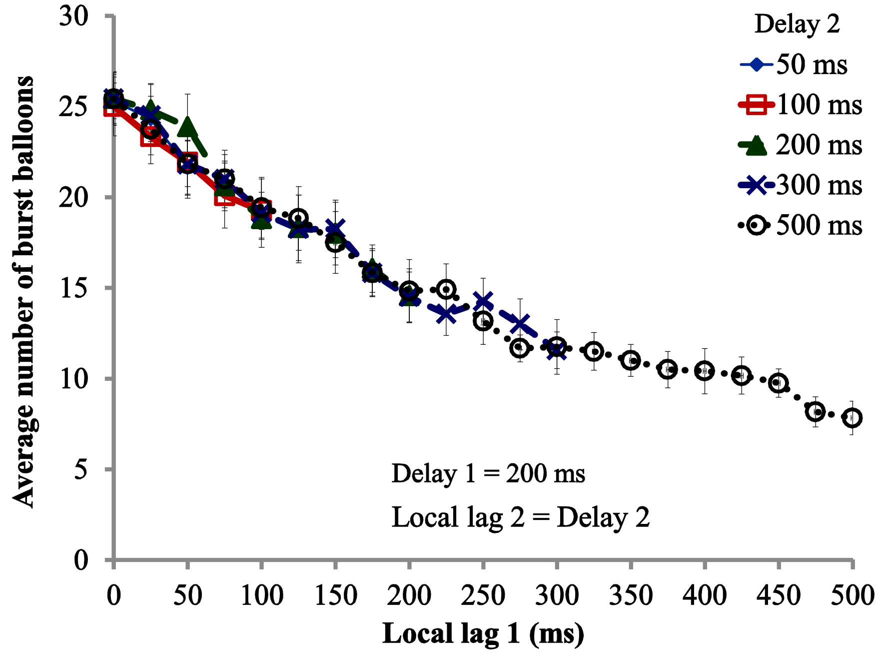


Figure 7.29. Average number of burst balloons at terminal 1 for assessment 2 in asymmetric delay case.

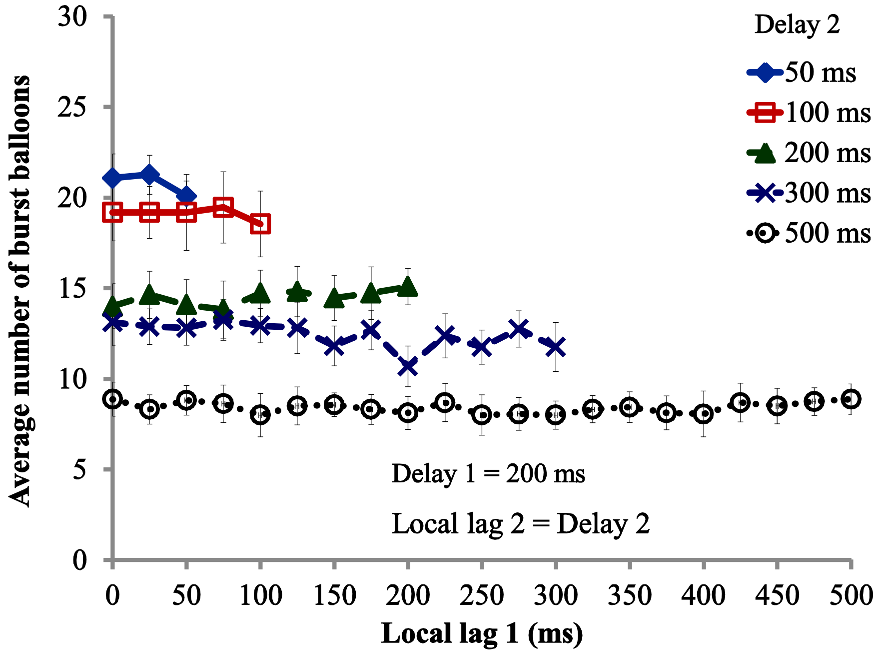


Figure 7.30. Average number of burst balloons at terminal 2 for assessment 2 in asymmetric delay case.

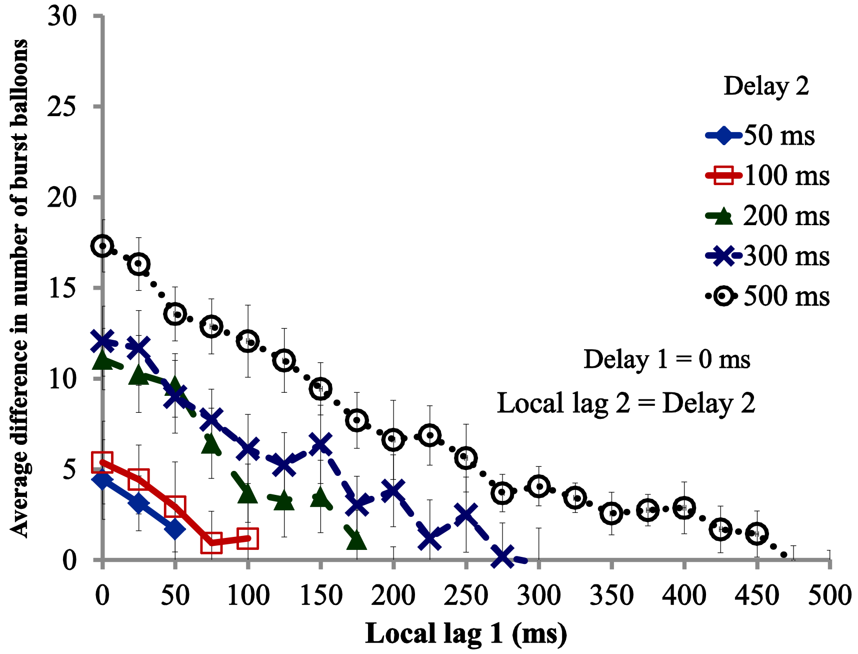


Fig. 7.31. Average difference in number of burst balloons between terminals for assessment 1 in asymmetric delay case.

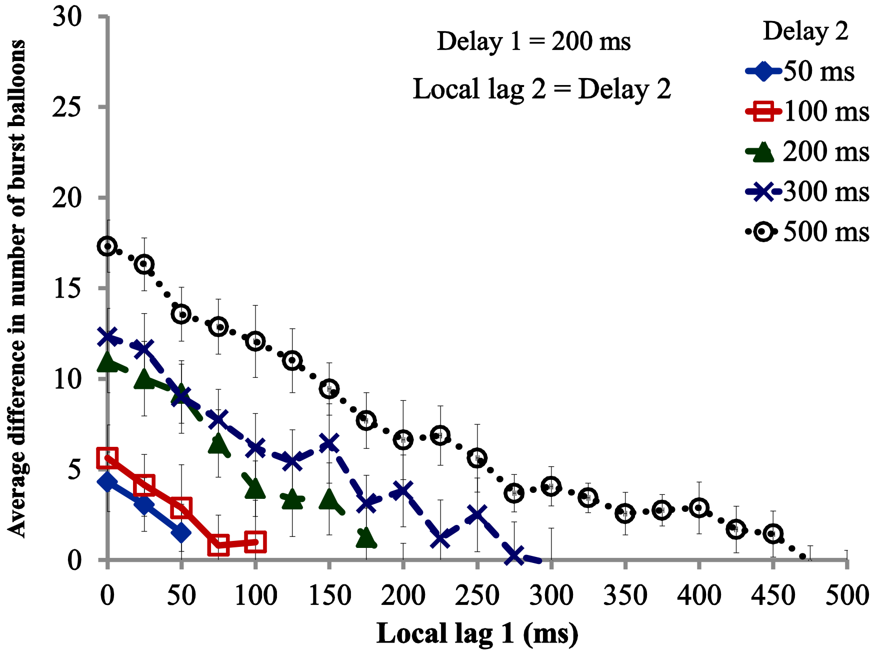


Fig. 7.32. Average difference in number of burst balloons between terminals for assessment 2 in asymmetric delay case.

Chapter 8

Conclusions

In this thesis, we investigated the influences of network delay on QoE by subjective and objective QoE assessments in areas of joint musical performance and networked real-time games. We chose the two areas to investigate how QoE is deteriorated when simultaneous output of media is incapable at each terminal or different terminals because of large difference in network delay among the terminals. We handled a networked haptic drum system and a networked balloon bursting game as example applications. In the joint musical performance, we investigated the influences of network delay on the synchronization quality of sound, interactivity, and the comprehensive quality (i.e., weighted sum of the synchronization quality of sound and interactivity) by QoE assessments. Then, we proposed the dynamic local lag control to keep the synchronization quality of sound high based on the assessment results. We also proposed the dynamic local lag control with dynamic prediction of time for better interactivity. In the networked real-time games, we investigated the influences of network delay on the operability of haptic interface devices, and fairness among players by QoE assessments. We clarified that there is a trade-off relationship between the operability and fairness. We found that the adaptive Δ-causality control can be used in networked real-time games to maintain the comprehensive quality (i.e., weighted sum of the operability and fairness) as high as possible.

In Chapter 2, we investigated the influences of network delay on the synchronization quality of sound, interactivity, and comprehensive quality under the local lag control by subjective QoE assessment in the networked haptic drum system. We also carried out objective assessment at the same time as the subjective assessment. We further examined the relationship between the MOS values and the objective performance measures. Consequently, we found the following results:

- There exits the optimum value of local lag, and the value is not always equal to the network delay.
- The optimum value of local lag at each terminal is dependent on only the network delay from the other terminal to the local terminal.
- The MOS values can be estimated from the root mean square of sound and/or local lag with a high degree of accuracy.

In Chapter 3, we proposed the dynamic local lag control for sound synchronization in the 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 the networked haptic drum system. We further examined the relationship between the MOS values and the objective performance measures. As a consequence, we found the following results:

- The dynamic local lag control can keep the synchronization quality of sound high.
- The dynamic local lag control can keep the interactivity better than the traditional local lag control.
- The MOS values can be estimated from the root mean square of sound and/or the local lag with a high degree of accuracy.

In Chapter 4, we enhanced the dynamic local lag control so that three or more users at different terminals can play musical instruments together over a network. We investigated the effect of the enhanced dynamic local lag control by subjective and objective QoE assessment in the networked haptic drum system for three terminals. We made a comparison among the enhanced control, the adaptive Δ-causality control, and no control. As a result, we found that the enhanced control can keep the synchronization quality of sound high for three or more terminals.

Although the dynamic local lag control keeps the interactivity higher than the traditional local lag control, the interactivity is still deteriorated as local information is output after buffering under the control. Therefore, in Chapter 5, in order to keep the interactivity higher, we enhanced the control by combining with prediction. We investigated the effect of the dynamic local lag control with prediction by subjective QoE assessment in the joint musical performance of the networked haptic drum system. As a result, we found that there exits the optimum value of prediction time according to the network delay.

In Chapter 6, we proposed the dynamic local lag control with dynamic control of prediction time in order to keep both the synchronization quality of sound and interactivity 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.

In Chapter 7, we investigated the influences of network delay on QoE such as the operability and fairness for the networked balloon bursting game by carrying out subjective and objective QoE assessments. We clarified how the local lags should be set at each terminal to be able to keep the comprehensive quality high. Also, we examined the relationships between subjective and objective assessment results. We found the following outcomes from the assessment results:

- The soft objects become harder and more slippery as the network delay increases.

- The operability of haptic interface device strongly depends on the local lag or the network delay from the local terminal to the other terminal.
- The fairness between players depends on the difference in network delay between the terminals or the local lag difference between the terminals.
- There exists a trade-off relationship between the operability and fairness.
- The local lag at each terminal can be set to the larger value of network delay between the terminals.
- The MOS values can be estimated from objective assessment measures such as the average number of burst balloons, local lag, the average difference in number of burst balloons or the absolute value of local lag difference with a high degree of accuracy.

As our future work, we will consider the influences of other QoS parameters such as packet loss and delay jitter. Although we focused on only the two areas in this thesis, it is also important to investigate the influences of the network delay, delay jitter, and packet loss on QoE in other areas. This kind of study is important not only in virtual environments, but also in real environments. For example, we should apply the dynamic local lag control to joint musical performance in real environments, where multiple users play the same or different types of musical instruments together, and investigate the effects of the control. It is also important to confirm the effects of the adaptive Δ-causality control on the fairness among multiple players in real-time networked games over the internet.

The results found in this thesis can be helpful to researchers, developers, and technicians in academia and industries. Researchers who work to solve the problems of QoE deteriorations in various areas can take into account the effects of the proposed control in this thesis when they consider solutions how to maintain QoE high. Also, researchers who study about human perceptibility and sensation in various kinds of applications can also compare the results in this thesis to their results as human perceptions of the imperfections occurred from QoS non-guaranteed networks may be different according to different types of applications. The results are also effective to services in entertainment industries such as music entertainment industry and game industry. For example, in music entertainment industry, the effects of the dynamic local lag control can be taken into account for high quality music performance over the Internet. Also, developers in game industry can consider the trade-off relationship between the operability and the fairness for their new games. The results of the allowable range of network delay for the operability and fairness found in this thesis may also be helpful to operability oriented applications and fairness oriented applications.

Acknowledgements

First and foremost, I would like to express my deep gratitude to my supervisor Prof. Yutaka Ishibashi. I greatly appreciate his valuable guidance and support throughout these years. His suggestions, encouragement, mentoring and comments helped me to steer this research in the right direction.

I also wish to extend special thanks to Prof. Norishige Fukushima and Dr. Pingguao Huang for their general guidance, comments and suggestions.

I am grateful to Professor Hiroshi Matsuo and Professor Toshiro Nunome for their invaluable comments and suggestions on my thesis.

My sincere gratitude goes to the Japanese and Myanmar Governments for giving me such an opportunity to get the Monbukagakusho scholarship. With their support, I became to be able to study at Nagoya Institute of Technology.

I would like to acknowledge the faculties and staffs of International Student Center at Nagoya Institute of Technology for their hospitability and kindness.

And also, sincere thanks to all my colleagues for their encouragement and help throughout my doctoral work.

The last but not least, I am grateful to my family members, especially my mother, who specifically offered strong moral and physical support, care and kindness, during the year of my Ph.D. study.

References

[1] M. Eid, M. Orozco, and A. El Saddik, “A guided tour in haptic audio visual environment and applications,” International Journal of Advanced Media and Communication, vol. 1, no. 3, pp. 265-297, June 2007.

[2] S. Singhal and M. Zyda, “Networked virtual environments: Design and implementation,” Addison Wisley Publishing, July 1999.

[3] C. Joslin, I. Pandzic, and N. Thalmann, “Trends in networked collaborative virtual environments”, Computer Communication Journal, vol. 26, no. 5, pp. 430-437, Mar. 2003.

[4] T. Ogi, K. Tamagawa, T. Yamada, and M. Hirose, “Collaborative scientific visualization in networked immersive virtual environment,” Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC), no. 6, pp. 86-91, Oct. 1999.

[5] A. Marshall and Y. Wai, “Haptic virtual environment performance over IP networks: A case study,” Proc. IEEE International Symposium on Distributed Simulation and Real-Time Applications, pp. 181-189, Oct. 2003.

[6] A. El Saddik, “The potential of haptics technologies,” IEEE Instrumentation & Measurement Magazine, vol. 10, issue 1, pp. 10-17, Feb. 2007.

[7] M. L. McLaughlin, G. Sukhatme, J. Hespanha, A. Ortega, and G. Mekioni, “The haptic Museum,” Proc. Conference on Electronic Imaging and the Visual Arts (EVA), 2000.

[8] Y. Kurokawa, Y. Ishibashi and T. Asano, " Group synchronization conrtrol in a remote haptic drawing system," Proc. the IEEE International Conference on Multimedia and Expo (ICME), pp. 572-575, July 2007.

[9] 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.

[10] http://geomagic.com/en/products/phantom-omni/overview.

[11] http://geomagic.com/en/products/phantom-desktop/overview.

[12] http://geomagic.com/en/products/phantom-premium/overview.

[13] http://www.forcedimension.com.

[14] M. Sato, “Development of string-based force display : SPIDAR,” Proc. the 8th International Conference on Virtual System and Multi Media (VSMM), pp. 1034-1039, Sep. 2002.
[15] Novint Technologies, Inc., “Haptic device abstraction layer programmer's guide,” version 1.1.9 Beta, Sep. 2007.
[16] F. Gosselin and A. Riwan, “Design of virtuose 3D: A new haptic interface for teleoperation and virtual reality,” Proc. ICAR, pp. 205-212, Aug. 2001.
[17] R. Q. Vanderlinde, P. Lammertse, E. Frederiksen, and B. Ruiter, “The HapticMaster, a new high-performance haptic interface,” Proc. EuroHaptics, pp. 1-5, 2002.
[18] SensAble Technologies, Inc., “Openhaptics toolkit programmer’s guide,” version 3.0, 2009.
[19] Immersion Corporation: CyberGrasp, “http:// www. immersion.com /3d/products/cyber grasp.php.”
[20] 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.
[21] 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.
[22] 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.
[23] T. Asano, Y. Ishibashi and S. Kameyama, “Interactive haptic transmission for remote control systems,” Proc. the IEEE International Conference on Multimedia and Expo (ICME), pp. 2113-2116, July 2006.
[24] T. Miyasato, H. Noma, and F. Kishino, “Subjective evaluation of perception of delay time between visual information and tactile information,” IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, vol. E79-A, no. 5, pp. 655-657, May 1996.
[25] T. Asano and Y. Ishibashi, “Guidance services for a haptic museum in distributed virtual environments,” Proc. ICAT Workshop on VR Applications and Entertainment Technology, pp. 39-42, Nov./Dec. 2004.
[26] T. Harada, S. Ohno, and M. Sato, “The inuence of multi-modal sensory information display on dribbling of a basket-ball in a virtual workspace,” Proc.

the 4th International Conference on Virtual System and Multi-media, vol. 21, no. 4, pp. 393-404, Apr. 1997.

[27] O. Wongwirat and S. Ohara, "Haptic media synchronization for remote surgery through simulation," IEEE MultiMedia, vol. 13, no. 3, pp. 62-69, July 2006.

[28] S. Lee, S. Monn, and J. Kim, "A network-adaptive transport scheme for hapticbased collaborative virtual environments," Proc. ACM the 5th Annual Workshop on Network and Systems Support for Games (NetGames), Oct. 2006.

[29] K. Hikichi, Y. Yasuda, A. Fukuda, and K. Sezaki, "The effect of network delay on remote calligraphic teaching with haptic interfaces," Proc. ACM the 5th Annual Workshop on Network and Systems Support for Games (NetGames), Oct. 2006.

[30] T. Asano and Y. Ishibashi, "A remote haptic painting lesson system," Proc. SPIE Opitcs East, Multimedia Systems and Applications X, vol. 6777, no. 28, Sep. 2007.

[31] P. Huang, Y. Ishibashi, N. Fukushima, and S. Sugawara, "QoE assessment of olfactory media in remote ikebana with haptic media," Proc. IEEE 2012 International Communications Quality and Reliability (CQR) Workshop, May 2012.

[32] D. Morris, N. Joshi, and K. Salisbury, "Haptic Battle Pong: High-degree-of-freedom haptics in a multiplayer gaming environment," Proc. Experimental Gameplay Workshop, Game Developers Conference, Mar. 2004.

[33] S. Andrews, J. Mora, J. Lang, and W. S. Lee, "HaptiCast: A physically-based 3D game with haptic feedback," Proc. Future Play Conference, Oct. 2006.

[34] Y. Ishibashi and H. Kaneoka, "Fairness among game players in networked haptic environments: Influence of network latency," Proc. IEEE International Conference on Multimedia and Expo (ICME), July 2005.

[35] Y. Ishibashi, S. Hoshino, Q. Zeng, N. Fukushima, and S. Sugawara, "QoE assessment of fairness in networked game with olfaction: Influence of time it takes for smell to reach player," Springer's Multimedia Systems Journal (MMSJ), Special Issue on Network and Systems Support for Games, vol. 20, issue 5, pp. 621-631, May 2014.

[36] ITU-T Rec. I. 350, "General aspects of quality of service and network performance in digital networks, including ISDNs," Mar. 1993.

[37] ITU E. 800, "Terms and definitions related to quality of service and network performance including dependability," Aug. 1994.

[38] S. Tasaka, and Y. Ishibashi, "QoS in Distributed Multimedia Applications and QoS Mapping," (in Japanese), CQ2001-40, pp. 7-14, July 2001.
[39] P. Huang and Y. Ishibashi, "QoS control and QoE assessment in multi-sensory communications with haptics," IEICE Trans. Communications, vol. E96-B, no. 2, Feb. 2013.
[40] M. Sithu and Y. Ishibashi, "Media synchronization control in multimedia communication," IGI Global, Emerging Research on Networked Multimedia Communication Systems, Chapter 2, pp. 25-61, 2015.
[41] ITU-T Rec. P. 10/G. 100 Amendment 1, "New appendix I – Definition of quality of experience (QoE)," International Telecommunication Union, Jan. 2007.
[42] 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, vol. 25, no. 7, pp. 1049-1061, July 2013.
[43] P. Keller, "Joint action in music performance," Journal of Emerging Communication, 21, 6, pp. 454-461, July 2006.
[44] 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), IPSJ SIG Technical Report, 2009-DPS-141, 23, Nov. 2009.
[45] 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 Communication Quality & Reliability (CQR), pp. 213–217, May 2002.
[46] Y. Yoo, M. Sung, N. Kim, and K. Jun, "An experiemntal study on the performance of haptic data transmission in networked haptic collaboration," Proc. International Conference on Advanced Communication Technology, pp. 657-662, Feb. 2007.
[47] J. Osada, Y. Ishibashi, N. Fukushima, and S. Sugawara, "QoE comaprison of haptic control schemes in remote instruction system with haptic media, video, and voice," Proc. International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), July 2012.
[48] B. Khademian and K. Hashtrudi-Zaad, "Performace issues in collaborative haptic training," Proc. IEEE International Conference on Robotics and Automation (ICRA), pp. 3275-3263, April 2007.
[49] B. Willaert, B. Corteville, D. Reynaerts, H. Brussel, and E. Poorten, "Bounded environment passivity of the calssical position-force teleoperation controller," Proc. IEEE/RSJ IROS, Oct. 2009.

[50] M. Lin and K. Salisbury, "Haptic rendering: Beyond visual computing," IEEE Computer Graphics Applications, vol. 24, no. 2, pp. 22-23, Mar. 2004.

[51] K. Salisbury, F. Conti, and F. Barbgli, "Haptic rendering: Introductory concepts," IEEE Computer Graphics Applications, vol. 24, no. 2, pp. 24-32, Mar. 2004.

[52] R. Weller and G. Zachmann, "User performance in complex bimanual haptic manipulation with 3DoFs vs. 6DoFs," Proc. IEEE Haptics, Mar. 2012.

[53] A. Campbell, G. Coulson, F. Garcia, and D. Hutchison, "A continuous media transport and orchestration service," Proc. ACM SIGCOMM, pp. 99-110, Aug. 1992.

[54] E. Biersack, W. Geyer, and C. Bernhardt, "Intra- and inter-stream synchronization for stored multimedia streams," Proc. of IEEE Multimedia Systems, pp. 372-381, 1996.

[55] L. Ehley, B. Furht, and M. Ilyas, "Evaluation of multimedia synchronization techniques," Proc. IEEE Multimedia Systems, pp. 514-519, May 1994.

[56] Q. Zeng, Y. Ishibashi, N. Fukushima, S. Sugawara, and K. E. Psannis, "Influences of inter-stream synchronization errors among haptic media, sound, and video on quality of experience in networked ensenmble," Proc. the 2nd IEEE Global Conference on Consumer Electronics (GCCE), pp. 466-470, Oct. 2013.

[57] Y. Ishibashi and S. Tasaka, "A synchronization mechanism for continuous media in multimedia communications," Proc. IEEE International Conference on Computer Communications (INFOCOM), pp. 1010-1019, Apr. 1995.

[58] Y. Ishibashi, S. Tasaka, and H. Ogawa, "Media synchronization quality of reactive control schemes. IEICE Transactions on Communications," vol. E86-B, no. 10, pp. 3103-3113, Oct. 2003.

[59] H. Sannomiya, J. Osada, Y. Ishibashi, N. Fukushima, and S. Sugawara, "Inter-stream synchronization control with group synchronization algorithm," Proc. IEEE Global Conference on Consumer Electronics (GCCE), pp. 520-524, Oct. 2013.

[60] Y. Ishibashi, T. Kanbara, and S. Tasaka, "Inter-stream synchronization between haptic media and voice in collaborative virtual environments," Proc. of ACM Multimedia, pp. 604-611, Oct. 2004.

[61] C. C. Yang, "Design of the application-level protocol for synchronized multimedia sessions," Proc. of IEEE International Conference on Communications (ICC), pp. 2518–2522, 2002.

[62] Y. Ishibashi, A. Tsuji, and S. Tasaka, “A group synchronization mechanism for stored media in multicast communications,” Proc. IEEE International Conference on Computer Communications (INFOCOM), pp. 692-700, Apr. 1997.
[63] Y. Ishibashi, and S. Tasaka, “A group synchronization mechanism for live media in multicast communications,” Conference Record of Global Communications Conference (GLOBECOM), pp. 746-752, Nov. 1997.
[64] Y. Ishibashi, and S. Tasaka, “A distributed control scheme for group synchronization in multicast communications,” Proc. International Symposium on Communications (ISCOM), pp. 317-323, 1999.
[65] K. Hikichi, H. Morino, I. Arimoto, K. Sezaki, and Y. Yasuda, “The evaluation of delay jitter for haptic collaboration over the Internet,” Conference Record of IEEE Global Communications Conference (GLOBECOM) , pp. 1492-1496, Nov. 2002.
[66] E. Isomura, S. Tasaka, and T. Nonume, “QoE enhancement in audiovisual and haptic interactive IP communications by media adaptive intra-stream synchronization,” Proc. IEEE TENCON, pp. 1085-1089, Nov. 2011.
[67] Y. Ishibashi and S. Tasaka, “Causality and media synchronization control for networked multimedia games: Centralized versus distributed,” Proc. the 2nd ACM Workshop on Network and Systems Support for Games (NetGames), pp. 34-43, May 2003.
[68] Y. Ishibashi, Y. Hashimoto, T. Ikedo, and S. Sugawara, “Adaptive Δ-causality control with adaptive dead-reckoning in networked games,” Proc. the 6th ACM Workshop on Network and Systems Support for Games (NetGames), pp. 75- 80, Sept. 2007.
[69] M. Mauve, J. Vogel, and W. Effelsberg, “Local lag and timewrap: Providing consistency for replicated continuous applications,” IEEE Transactions on Multimedia, vol. 6, no. 1, pp. 47-57, Feb. 2004.
[70] ITU-T Rec. P. 911, “Subjective audiovisual quality assessment methods for multimedia applications,” Dec. 1998.
[71] C. E. Osgood, “The nature and measurement of meaning,” Psychological Bullentin, vol. 49, no. 3, pp. 197-237, May 1952.
[72] H. Ohnishi and K. Mochizuki, “Effect of delay of feedback force on perception of elastic force: A psychophysical approach,” IEICE Trans. Commun., vol. E90-B, no. 1, pp. 12-20, Jan. 2007.
[73] P. Brooks and B. Hestnes, “User measures of quality of experience: Why being objective and quantitive is important,” IEEE Netw., vol. 24, no.2, pp. 8-13, March/April 2010.

[74] ITU-R BT. 500-12, “Methodology for the subjective assessment of the quality of television pictures,” International Telecommunication Union, Sept. 2009.

[75] S. Chatterjee, A. S. Hadi, and B. Price, “Regression analysis by example,” John Wiley & Sons, Hoboken, 1999.

[76] Y. Irie, S. Aoyagi, T. Takada, K. Hirata, K. Kaji, S. Katagiri, and M. Ohsaki, “Development of assistant system for ensemble in t-Room,” IPSJ SIG Technical Report, (in Japanese), vol. 2009-DPS-141, no. 23, Nov. 2009.

[77] J. Brun, F. Safaei, and P. Boustead, “Managing latency and fairness in networked games,” Communications of ACM, vol. 49, no. 11, pp. 46-51, Nov. 2006.

[78] M. Fujimoto and Y. Ishibashi, “The effect of stereoscopic viewing of a virtual space on a networked game using haptic media,” Proc. ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE), pp. 317-320, June 2004.

[79] 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.

[80] M. Sithu, Y. Ishibashi, and N. Fukushima, “Effects of dynamic local lag control on sound synchronization and interactivity in joint musical performance,” ITE Trans. Media Technology and Applications, Special Section on Multimedia Transmission System and Services, vol. 2, no. 4, pp. 299-309, Oct. 2014.

[81] M. Sithu, Y. Ishibashi, and N. Fukushima, “Enhancement of dynamic local lag control for networked musical performance,” Proc. The 3rd IEEE Global Conference on Consumer Electronics (GCCE), pp. 674-678, Oct. 2014.

[82] 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.

[83] M. Sithu, Y. Ishibashi, and N. Fukushima, “Effect of dynamic local lag control with dynamic control of prediction time in joint haptic drum performance,” Proc. The 12th International Conference on Computer Applications (ICCA), pp. 343-350, Feb. 2014.

[84] M. Sithu, Y. Ishibashi, P. Huang, and N. Fukushima, “Influences of network delay on quality of experience for soft objects in networked real-time game with haptic sense,” International Journal of Communications, Network and System Sciences (IJCNS), 2015.

[85] M. Sithu, Y. Ishibashi, P. Huang, and N. Fukushima, “Trade-off relationship between operability and fairness in networked balloon bursting game using haptic

interface devices," in Proc. The Ninth International Conference on Genetic and Evolutionary Computing (ICGEC), pp. 127-138, Aug. 2015.

[86] M. Carson and D. Santay, "NIST Net – A Linux-based network emulation tool," ACM SIGCOMM, vol. 33, no. 3, pp. 111-126, July 2003.

[87] Z. Huang, K. Nahrstedt, and R. Steinmetz, "Evolution of temporal multimedia synchronization principles: A historical viewpoint," ACM Trans. on Multimedia Computing, Communications, and Applications, vol. 9, no. 34, Oct. 2013.

[88] F. Boronat, J. Lloret, and M. Garcia, "Multimedia group and inter-stream synchronization techniques: A comparative study," Journal of Information System, vol. 34, no.1, pp. 108-131, Mar. 2009.

[89] G. Blalowski, and R. Steinmetz, "A media synchronization survey: Reference model, specification, and case studies," Journal of IEEE Selected Areas in Communications, vol. 14, no. 1, pp. 5-35, Jan. 1996.

[90] 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.

[91] http://www.redheadmusic.com/pdf/listening_room/ Bruce_Felter_CD.pdf.

[92] J. F. Kenny and E. S. Keeping, "Mathematics of statistics," NJ: Van Nostrand, pp. 59-60, 1962.

[93] 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.

[94] 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.

[95] 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.

[96] 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.

[97] Y. Kawano, D. Hanawa, and T. Yonekura, "An approach to the P2P type real-time network virtual ball-game with the mutual prediction protocol and AtoZ," (in Japanese), Trans. of the Virtual Reality Society of Japan, vol. 9, no. 2, pp. 141- 150, June 2004.

[98] Y. Kawano and T. Yonekura, “Realization of avoidance of the critical case in P2P type virtual ball game by Count Down Protocol.” IEICE Trans. on Information and Systems (Japanese Edition), vol. J89-D, no. 10, pp. 2219- 2228, Oct. 2006.

[99] Y. Kusunose, Y. Ishibashi, N. Fukushima, and S. Sugawara, “QoE comparison of competition avoidance methods for management of shared object in networked real-time game with haptic media,” Proc. 21st International Conference on Artificial Reality and Telexistence (ICAT), Nov. 2011.

[100] ITU-R BT. 1359-1, “Relative timing of sound and vision for broadcasting,” 1998.