QoE Enhancement in Audiovisual and Haptic Interactive IP Communications by Media Adaptive Intra-Stream Synchronization

Eiichi Isomura, Shuji Tasaka, and Toshiro Nunome
Department of Computer Science and Engineering, Graduate School of Engineering,
Nagoya Institute of Technology
Nagoya 466–8555, Japan

***Abstract*—For audiovisual and haptic interactive IP communications, we propose a media adaptive intra-stream synchronization control scheme, which exerts control suitable for each media. We assess *QoE (Quality of Experience*) of the following three intra-stream synchronization control schemes: 1) media adaptive buffering, 2) Skipping & buffering and 3) buffering. Schemes 1 and 2 are examples of the media adaptive intra-stream synchronization control scheme. In media adaptive buffering, we set the playout buffering time for each media according to its property. In Skipping & buffering, we apply Skipping to haptic media and playout buffering to video and audio. Scheme 3 is the conventional playout buffering, which sets the same playout buffering time to the three media. We also investigate the effect of source skipping which reduces the transmission rate of the haptic media. For subjective experiment, we designed a task whose output quality of video and haptic media dominates QoE. We assess QoE multidimensionally by the *SD (Semantic Differential*) method. As a result of the assessment, we see that the combination of scheme 1 and the source skipping can achieve higher QoE than the other schemes.**

***Index Terms*—interactive IP communications, audiovisual and haptic, QoE, intra-stream synchronization control, source skipping**

I. Introduction

Using audiovisual and haptic media together in communications is expected to improve the efficiency of remote and collaborative work. This type of usage of the three media over IP networks can produce promising new applications. However, IP networks offer best-effort services, which do not guarantee *QoS (Quality of Service)*. In other words, packet loss, network delay or delay jitter cannot be controlled; thus, the output quality of the media can seriously degrade.

In the context of networking, QoS is defined for each layer. Among them, the quality of end-user experience (i.e., *user-level QoS*) is the most important. User-level QoS is called *QoE (Quality of Experience)* in ITU-T [1]. For the service providers and users, guarantee and enhancement of QoE are ideal goals of the network services.

In the transmission of continuous media such as video and haptic media over IP networks, there are several ways to enhance QoE; e.g., intra-stream synchronization control and source skipping. Intra-stream synchronization control has a role to maintain the temporal structure in a single stream on the media receiver side. Examples of intra-stream synchronization control are *Skipping*, *playout buffering* and *VTR (virtual time rendering)* [2]. The source skipping controls the transmission rate of the media stream on the media sender side by skipping the transmission in some ways.

Iwata *et al.* investigate the effect of playout buffering control on QoE in haptic media, sound and video transmission in [3]. The haptic media is transmitted bidirectionally. In the subjective experiment, they take several values of the playout buffering time (the values of the buffering time of the three media are set to be identical). They demonstrate that as the playout buffering time increases, the subjective quality of operability of the haptic media degrades. This is because the reaction force which the users feel from the haptic interface increases as the end-to-end delay becomes longer.

Haptic media, video and audio have different properties from each other with respect to the data size, update rate and maximum allowable delay. Therefore, the most proper intra-stream synchronization control for each media can be different. Exerting intra-stream synchronization control suitable for each media can be expected to achieve QoE enhancement.

We can find a study on the application of the source skipping for haptic media in [4] where a technique of *dead-reckoning* is used in networked virtual environments. The dead-reckoning can maintain the haptic output rate at 1 kHz by prediction and convergence. The media sender compares real position coordinates coming from the haptic interface with the predicted one. If the difference between the real one and the predicted one becomes larger than a threshold value, the position information is transmitted as the haptic data. Reference [4] shows that the dead-reckoning enhances the haptic output quality during network congestion; however, it supposes a virtual environment, and QoE assessment of the audiovisual quality is not made.

This paper proposes *media adaptive intra-stream synchronization control* schemes, which adopt proper intra-stream synchronization control for each media according to its property. As examples, we deal with two types of the media adaptive intra-stream synchronization control schemes: *Skipping & buffering* and *media adaptive buffering*. In Skipping & buffering, Skipping is applied to haptic media to minimize delay originated from intra-stream synchronization control, while playout buffering is applied to video and audio to maintain the output quality. In the media adaptive buffering, the playout buffering time is set for each media according to its property. QoE assessment with Skipping & buffering is made in [5].

In this paper, we study three kinds of intra-stream synchronization control schemes for comparison: Skipping & buffering, media adaptive buffering, and *buffering*. Buffering sets the same value of playout buffering time for the three media. In addition, this paper investigates the efficiency of source skipping of the haptic media in terms of QoE. We use the *SD (Semantic Differential)* method to assess QoE multidimensionally. At the same time, we also measure the *application-level QoS*.

The rest of this paper is structured as follows. Section II introduces intra-stream synchronization control including the media adaptive one. Section III indicates how to skip haptic data in our system. Section IV describes the experimental system. Section V outlines the method of QoS and QoE mea-

surement, and experimental results are presented in Section VI. Section VII concludes this paper.

II. Intra-stream synchronization control

Intra-stream synchronization control has a role to maintain the temporal structure in a single stream. We refer to the transmission unit at the application layer as a *media unit (MU)*; in this paper, we define a video frame as a video MU, a constant number of audio samples as an audio MU and positional information at the corresponding time as a haptic MU.

As a component technique for intra-stream synchronization control, we adopt Skipping and playout buffering [2]. Skipping outputs the latest MU out of a group of MUs in the case in which more than two MUs arrived at the same time; the latest one is output, and the rest are dropped. Playout buffering stores an MU in a receive-buffer until the target output time (determined by the MU birth time and buffering time). When an MU arrives after the target output time, it is either output or discarded. In this paper, if an MU is received after target output time, it is output immediately if the sequence number of the current MU is larger than that of the MU output last; otherwise the current MU is discarded.

Setting enough playout buffering time to absorb delay jitter can maintain media output quality. However, as the playout buffering time increases, output delay becomes longer. In the case of interactive services, QoE degrades because of slow response. Thus, in the playout buffering, there is a trade-off relation between the output quality and responsiveness [6].

Video, audio and haptic media have different properties (e.g., haptic media is sensitive to delay because of the increasing reaction force due to the delay). It implies that the most appropriate synchronization scheme for each media can be different from each other. This is the reason why we propose the media adaptive intra-stream synchronization control schemes:

Skipping & buffering [5] and the media adaptive buffering.

Thus, we study the following three intra-stream synchronization control schemes in the subjective experiment:

Scheme 1. Media adaptive buffering

The scheme sets proper playout buffering time depending on the media type.

Scheme 2. Skipping & buffering

The scheme applies Skipping to haptic media and playout buffering to video and audio.

Scheme 3. Buffering

The scheme adopts the same playout buffering time for the three media.

Schemes 1 and 2 can maintain good output quality of audiovisual streams and haptic manipulation; however, MU output timing can be different between the haptic media and the audio-video. Thus, inter-stream synchronization error can become larger than that of scheme 3. Note that scheme 3 can degrade audiovisual output quality when not enough playout buffering time is provided, or it can degrade the operability of the haptic media when too long playout buffering time is selected. Therefore, the audiovisual output quality and haptic operability in scheme 3 have a trade-off relation.

III. Source skipping of haptic media

In the typical haptic media transmission systems, the update rate of the haptic media is 1kHz [7]; 1000 MUs of the haptic media are transmitted every second to enhance operation accuracy of the haptic interface. However, the application which uses haptic media in real space may not require so precise operation of haptic media. Therefore, in some applications, we can expect that QoE is enhanced by source skipping of the haptic media, which reduces the amount of network traffic. In this paper, the source skipping reduces the transmission rate of the haptic media to 500 MU/s by alternating sending and skipping a haptic MU.

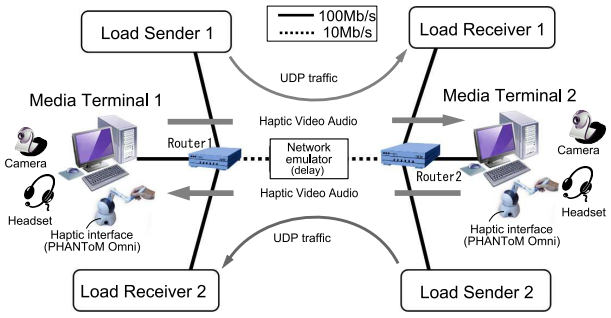

Fig. 1. System configuration.

IV. Experimental system

In this paper, we suppose that video, audio and haptic media are transmitted bidirectionally between two terminals over a best-effort IP network. As shown in Fig. 1, the experimental system consists of six PC's (Two Media Terminals, Two Load Senders and Two Load Receivers), two routers (Cisco System's Cisco 2811) and a network emulator (NIST Net[8]). We put NIST Net as a network delay generator. However, in the experiment, we set the network delay to zero for simplicity. A router and NIST Net are connected by a full-duplex Ethernet link of 10 Mb/s. All the other links are 100 Mb/s Ethernet. The link of 10 Mb/s becomes a bottleneck. A haptic interface (*PHANToM omni*), a video camera and a headset are connected to each Media Terminal.

Table I and Table II show the specifications of the three media. We use H.264 video with a resolution of 800 $\times$ 600 pixels. A video frame is divided into 15 slices, each of which forms an IP packet. When a video slice is lost, the decoder performs error concealment by using FFmpeg [9]. Audio is captured by the microphone of the headset. The haptic MU, which has coordinate data obtained from PHANToM, is transmitted bidirectionally between the two Media Terminals. This allows the user to manipulate PHANToM of the other side. We adopt the *spring-damper model*[1] [7] to calculate the reaction force which is presented to the user.

In the experiment, we take five values of the playout buffering time for video and audio of schemes 1 and 2, and the three media of scheme 3: 20, 40, 60, 100 and 150 [ms]. For the haptic media of scheme 1, the playout buffering time is kept to be 10 ms[2].

We take two values of the haptic MU rate [MU/s]: 1000 (Do *NOT* source skip) and 500 (Do source skip). Each Terminal transmits the three media streams as three separate UDP streams.

Load Sender 1 and Load Sender 2 transmit UDP load traffic to Load Receiver 1 and Load Receiver 2, respectively. Load Sender generates UDP datagram of 1472 bytes each at exponentially distributed intervals. The average bit-rate of the load traffic is 6.0 Mb/s[3].

[1]A reaction force of the haptic media can be calculated by $F = kx$, where F is the reaction force, k is a stiffness constant ($= 0.1$) and x is a displacement vector between the haptic interfaces.

[2]By a preliminary experiment, we selected 10 ms as the proper playout buffering time of the haptic media in terms of the smoothness and operability of the haptic interface.

[3]This value was selected because queuing delay on the router with the load traffic lighter than 6.0 Mb/s hardly affects the operability of the haptic media; over 6.0 Mb/s of load traffic degrades the operability of the haptic interface owing to queuing delay.

TABLE I
SPECIFICATIONS OF VIDEO AND AUDIO

	video	audio
encoding scheme	H.264 (x264) 800 × 600 pixels	Linear PCM 16kHz 8bit 1ch
average bit rate [kb/s]	2048	128
picture pattern	IPPPP	–
average MU rate [MU/s]	25	50

TABLE II
SPECIFICATION OF HAPTIC

average MU rate [MU/s]	1000	500
average bit rate [kb/s]	320	160

V. QOE ASSESSMENT METHOD

A. Task

In this paper, we aim to investigate how output quality of the video and haptic media affects QoE; for that purpose, we designed a task whose output quality of video and haptic media dominates QoE.

The task we have designed is the movement of an object from one position to another by manipulating the haptic interface. In the task, two subjects make a pair and are in different rooms each of which has an identical workspace. We have made three types of object: 1) a circle (the diameter is 3 cm), 2) an equilateral triangle (the length of the each side is 3 cm) and 3) a square (the length of the each side is 3 cm). The thickness of each object is about 5 mm, and it is light weight. Figure 2 illustrates the work space and the layout of the camera and PHANToM. The camera is placed above the white board (workspace), and the camera range covers the whole of the workspace.

The procedure for the task is explained below. Before the task begins, the three objects are put in the center circle. In the task, one subject plays the role of the indicator, and the other is the manipulator. First, the indicator selects an object in the center circle on his/her own side and its destination; then he/she gives the instruction to the manipulator using the headset microphone. The manipulator replies to the instruction and then manipulates the PHANToM stylus to move the object on the indicator's side to the requested destination, while watching the video and grasping the positional relation between PHANToM and the object on the indicator's side. When the object reaches the destination, the two subjects alternate the role. This work is repeated during a predetermined interval (i.e., 30 seconds). When the manipulator manipulates PHANToM, the indicator only holds his/her PHANToM stylus and surrenders him/herself to the manipulator's movement.

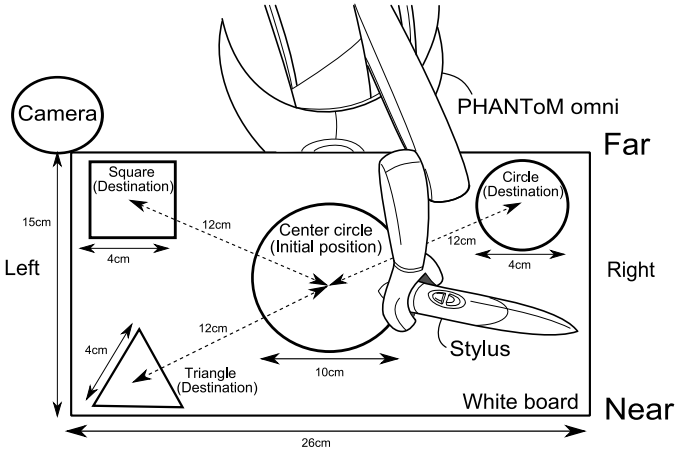

Fig. 2. Workspace and layout.

TABLE III
PAIRS OF POLAR TERMS

class	item	polar terms
Video1	Spatial quality	Video is corrupt – clear
Video2	Temporal quality	Video is jerky – smooth
Video3	Usefulness	Image is hard to grasp – easy
Haptic1	Operability	Manipulation is heavy – light
Haptic2	Smoothness	Movement is awkward – smooth
Haptic3	Stability	Manipulation is unstable – stable
Audio	Naturalness	Artificial – Natural
Inter-stream sync	Video and haptic	Out of sync – In sync
Interactive1	Response	Response is slow – rapid
Interactive2	Communicability	Hard to communicate – Easy
Interactive3	System comfort	Uncomfortable – Comfortable
Interactive4	Work difficulty	Difficult to work – Easy
Overall	Overall satisfaction	Unsatisfied – Satisfied

B. QoE measurement method

For multidimensional QoE assessment, we use the SD method [10]; it can assess an object for evaluation, which is referred to as a stimulus, from many points of view with many *pairs of polar terms*. Table III shows the polar terms for the experiment; these terms can be classified into six classes: video, haptic media, audio, inter-stream synchronization, interaction and overall satisfaction.

For each pair of polar terms, the subject gives a score to the stimulus by the *rating scale method* [11] with five grades. The best grade (score 5) represents the positive adjective (the right-hand side one in each pair in Table III), while the worst grade (score 1) means the negative adjective. The middle grade (score 3) is neutral.

The QoE measure adopted in this paper is the *psychological scale*, which is an *interval scale* in the *psychometric methods* [11]. Note that the QoE measure mainly used in ITU-T/R recommendations and many of technical papers is the *MOS (Mean Opinion Score)*, which is an *ordinal scale*. Since the interval scale can represent the human subjectivity more accurately than the ordinal scale, we use the psychological scale instead of MOS.

The interval scale can be calculated by the method of *successive categories* [12], which is composed of the rating-scale method and the law of categorical judgment. We apply the law of categorical judgment to the measurement result by the rating-scale method in order to obtain the interval scale [12]. We have to confirm the goodness of fit for the obtained scale. For a test of goodness of fit, we conduct Mosteller's test [13]. Once the goodness of fit has been confirmed, we use the interval scale as the psychological scale.

The subjects in the experiment were male and female students in their teens or twenties; the number of the subjects is 57 (28 males and 29 females). Each pair of the subjects assessed 30 *stimuli* because of three kinds of intra-stream synchronization controls, five values of the playout buffering time, and two values of the haptic MU rate. These stimuli were presented in random order. It took about 60 minutes for a subject to assess all the stimuli.

VI. EXPERIMENTAL RESULTS

A. Application-level QoS parameters

In this paper, we pick up the *video slice arrival ratio, MU output rate,* and *mean square error of inter-stream synchronization* as application-level QoS parameters.

The video slice arrival ratio is the ratio of the number of output video slices to the total number of transmitted video slices. The MU output rate is the average number of MUs output per second. The mean square error of inter-stream synchronization between the video and the haptic media is defined as the average square of the difference between the output time difference of the video and the corresponding haptic media and their time stamp difference.

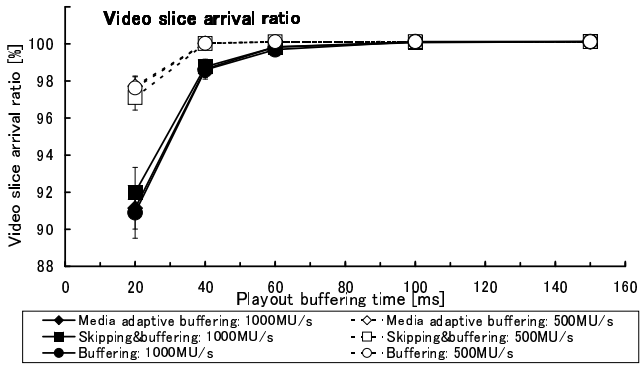

Fig. 3. Video slice arrival ratio.

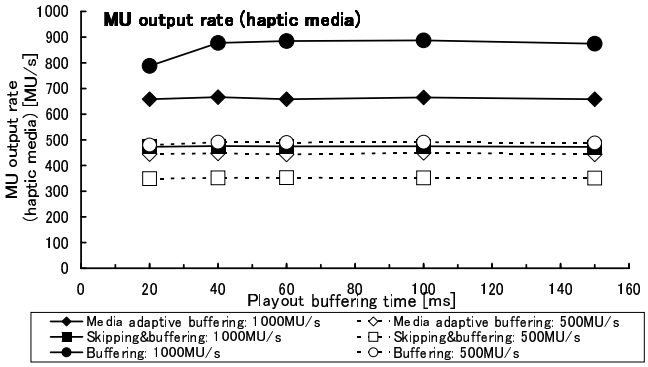

Fig. 4. MU output rate (haptic media).

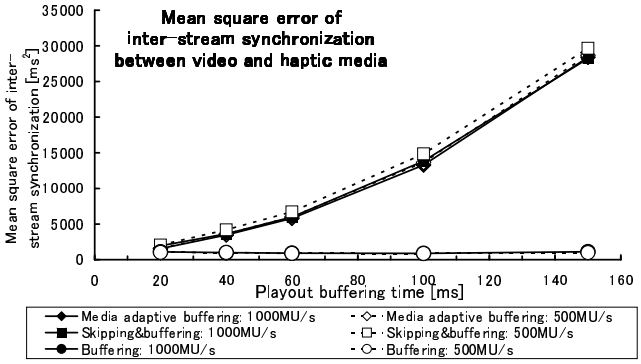

Fig. 5. Mean square error of inter-stream synchronization between video and haptic media.

Figures 3 through 5 plot the video slice arrival ratio, the MU output rate of the haptic media, and the mean square error of inter-stream synchronization between the video and the haptic media, respectively, as a function of the playout buffering time of audio and video. Note that the playout buffering time of the haptic media of scheme 1 (Media adaptive buffering) is set to 10 ms and that the haptic media of scheme 2 (Skipping & buffering) adopts Skipping. The figures display the two kinds of the transmission rate of the haptic media: 1000 MU/s (the source skipping is NOT applied) and 500 MU/s (the source skipping is applied). The figures also depict 95 percent confidence intervals of the measurement values.

Fig. 3 plots the video slice arrival ratio. When the playout buffering time is 100 ms or 150 ms, the video arrival ratio is 100% regardless of the source skipping. When the playout buffering time is 20 ms though 60 ms, using the source skipping of the haptic media can improve the video slice arrival ratio. This is because that the delay jitter of the video is decreased by the source skipping of the haptic media.

From Fig. 4, we find that scheme 3 (Buffering) takes the highest MU output rate of the haptic media among the three schemes, while scheme 2 takes the lowest value of the MU output rate. In scheme 3, it is not likely to lose haptic MUs because the playout buffering time of the haptic media is set to 20 ms and larger values. In scheme 2, Skipping is used for the haptic media; therefore, it is likely to lose the MUs because Skipping outputs the latest MU only and the rests are dropped. In scheme 1 (media adaptive buffering), since we set the playout buffering time to 10 ms for the haptic media, the haptic MU output rate becomes higher than scheme 2, which uses Skipping for the haptic media.

Fig. 5 shows the mean square error of inter-stream synchronization between video and haptic media. Since the same value of the playout buffering time is set for the three media in scheme 3, the synchronization error is small. However, in scheme 1 and scheme 2, the synchronization error increases as the video playout buffering time increases.

B. QoE

In this paper, we pick up several QoE measures: the video spatial quality, operability of the haptic media, inter-stream synchronization quality between video and haptic media, and overall satisfaction.

For the experimental results, we calculated the interval scale from each pair of the polar terms. In addition, we carried out Mosteller's test for a test of the goodness of fit of the interval scale. As a result of the test with a significance level of 0.05, we saw that the hypothesis can not be rejected. Thus, we use the calculated interval scales as the psychological scales.

We can select an arbitrary origin because the psychological scale is an interval scale. Then, we set the origin so that it can make the lower boundary of Category 2 become 1.00.

Figures 6 through 9 plot the psychological scales of the video spatial quality, operability of haptic media, inter-stream synchronization quality between video and haptic media, and overall satisfaction, respectively, as a function of the playout buffering time. The lower boundaries of the categories are also plotted as dotted lines parallel to the abscissa.

In Fig. 6, when the playout buffering time is 20 ms, the psychological scale values are low because of lost video slices (see Fig. 3). This is because the playout buffering time is not long enough to absorb delay jitter. If the playout buffering time is long enough, the video spatial quality is kept high. Therefore, when the playout buffering time is 100 ms or 150 ms, the psychological scale values of the video spatial quality are high. When the playout buffering time is 20 ms through 60 ms, the psychological scale values can be enhanced by applying the source skipping to the haptic media.

Fig. 7 shows the psychological scale of the operability of the haptic media. Since the output delay of haptic media can be reduced by using short playout buffering time or Skipping, scheme 1 and scheme 2 can achieve high QoE as a whole. On the other hand, in scheme 3, as the playout buffering time of the haptic media increases, the operability degrades since the reaction force becomes larger.

In Fig. 8, we find the following result. Although the video and haptic media can be out of synchronization in scheme 1 and scheme 2, the psychological scale values of the synchronization are high as a whole. This is because almost all the subjects hardly perceived the asynchrony in the experimental task (namely, the object movement).

Fig. 9 reveals that the overall satisfaction is the highest with the combination of scheme 1 and the source skipping for haptic media. This is because the audiovisual output quality and haptic manipulation quality are maintained better than the other schemes. When the playout buffering time is 20 ms or 40 ms, the overall satisfaction with the source skipping is clearly higher than that without the source skipping. This is because

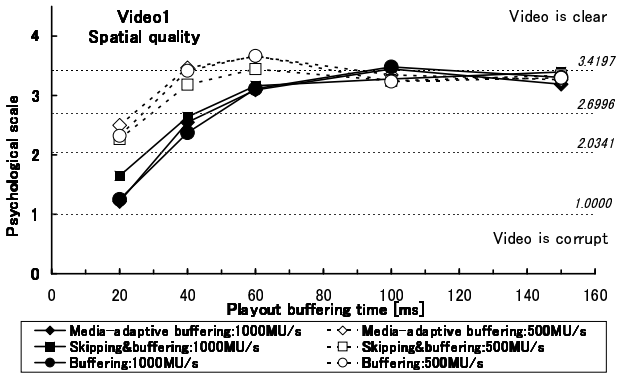

Fig. 6. Psychological scale versus playout buffering time (video spatial quality).

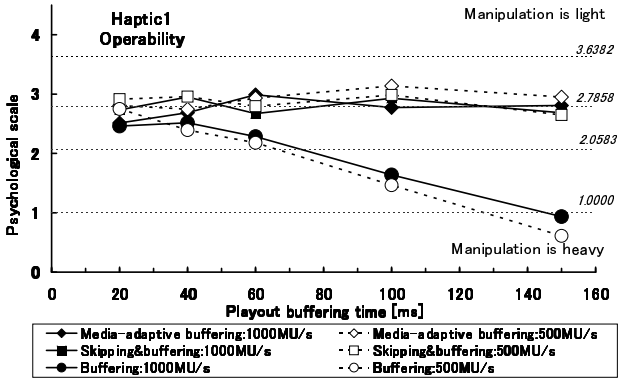

Fig. 7. Psychological scale versus playout buffering time (operability of haptic media).

the video output quality is improved by the source skipping (see Fig. 6). In scheme 3, we can find that there exists the optimum playout buffering time, which maximizes the psychological scale value. The reason is as follows. Audiovisual output quality degrades owing to the delay jitter when the playout buffering time is short. When the playout buffering time is long, the operability of the haptic media degrades because of increase in the output delay. Therefore, in scheme 3, there is a trade-off relation between audiovisual output quality and the operability of the haptic media. On the other hand, in scheme 1 and scheme 2, it is difficult to find the optimum playout buffering time because the haptic manipulation quality is almost constant as seen in Fig. 7.

VII. Conclusions

We proposed the media adaptive intra-stream synchronization control and assessed QoE and application-level QoS. As a result, we observed that the combination of scheme 1 and the source skipping for the haptic media can achieve higher QoE than the other schemes because of high video slice arrival ratios and enhanced output quality of the haptic media.

As future work, we should investigate the system without video error concealment in order to examine the effect of video quality on QoE. We also plan to study how to select proper playout buffering time for haptic media for non-zero values of network delays by NIST Net and other values of the load traffic. The proper MU rate of the haptic media according to task types and application of dead-reckoning are also issues to be examined.

Acknowledgment

The authors thank Prof. Y. Ishibashi of Nagoya Institute of Technology for his valuable advice on the experimental system. This work was supported by the Grant-In-Aid for Scientific Research of Japan Society for the Promotion of Science under Grant 21360183.

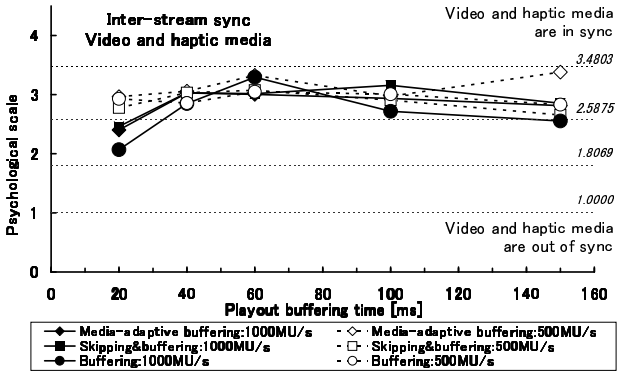

Fig. 8. Psychological scale versus playout buffering time (inter-stream sync quality between video and haptic media).

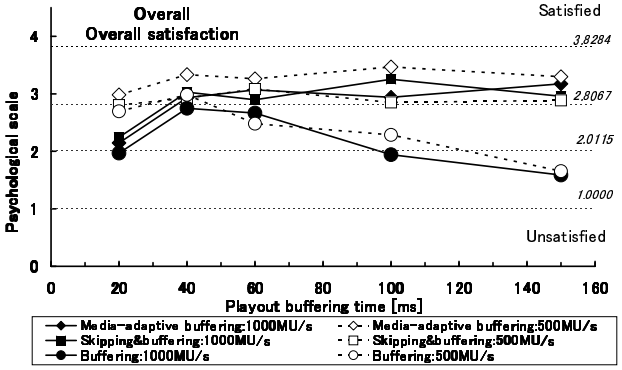

Fig. 9. Psychological scale versus playout buffering time (overall satisfaction).

References

[1] ITU–T Rec. P.10/G.100 Amendment 2, "Amendment 2: New definitions for inclusion in Recommendation ITU-T P.10/G.100," July 2008.

[2] S. Sun, T. Fujimoto, Y. Ishibashi, and S. Sugawara, "A comparison of output quality among haptic media synchronization algorithms," in *Proc. ICAT'08*, pp. 43-50, Dec. 2008.

[3] K. Iwata, Y. Ishibashi, N. Fukushima, and S. Sugawara, "QoE assessment in haptic media, sound, and video transmission: Effect of playout buffering control," in *Proc. ACE'10*, Nov. 2010.

[4] T. Kanbara, Y. Ishibashi, and S. Tasaka, "Haptic media synchronization control with dead-reckoning in networked virtual environments," in *Proc. SCI'04*, vol. III, pp. 158-163, July 2004.

[5] Y. Ito, S. Tasaka, T. Nunome and Y. Ishibashi, "Effect of playout buffering time on QoE in audiovisual and haptic interactive IP communications," (in Japanese) *IEICE Technical Report*, CQ2010-64, Nov. 2010.

[6] S. Tasaka and N. Misaki, "Maximizing QoE of interactive services with audio-video transmission over bandwidth guaranteed IP networks," in *Conf. Rec. IEEE GLOBECOM 2009*, CQPRM-09-6, Dec. 2009.

[7] SensAble Technologies, Inc., "OpenHaptics Toolkit programmer's guide," Version 2.0, 2005.

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

[9] FFmpeg, http://ffmpeg.org/.

[10] C. E. Osgood, "The nature and measurement of meaning," *Psychological Bulletin*, vol. 49, no. 3, pp. 197–237, May 1952.

[11] J. P. Guilford, *Psychometric methods*, McGraw–Hill, N. Y., 1954.

[12] S. Tasaka and Y. Ito, "Psychometric analysis of the mutually compensatory property of multimedia QoS," in *Conf. Rec. IEEE ICC2003*, pp.1880–1886, May 2003.

[13] F. Mosteller, "Remarks on the method of paired comparisons: III. a test of significance for paired comparisons when equal standard deviations and equal correlations are assumed," *Psychometrika*, vol. 16, no. 2, pp.207–218, June 1951.