QOE ASSESSMENT IN TELE-OPERATION WITH 3D VIDEO AND HAPTIC MEDIA

Ayano Tatematsu, Yutaka Ishibashi, Norishige Fukushima, and Shinji Sugawara

Graduate School of Engineering,
Nagoya Institute of Technology
Nagoya 466-8555, Japan
ayano_t@mcl.nitech.ac.jp, ishibasi@nitech.ac.jp, fukushima@nitech.ac.jp, shinji@nitech.ac.jp

ABSTRACT

In this paper, we investigate the influence of network delay on the quality of experience (QoE) for a tele-operation system with 3D (i.e., stereoscopic) video and haptic media. We assess the video output quality, operability of the haptic interface device, inter-stream synchronization quality (or interactivity), and comprehensive quality as QoE. We also evaluate the application-level quality of service (QoS). By multiple regression analysis, we demonstrate that it is possible to estimate QoE parameters from QoS parameters with a high degree of accuracy.

Index Terms— 3D video, haptic media, QoE, network delay

1. INTRODUCTION

Recently, 3D (i.e., stereoscopic) video attracts rising attention such as screening of 3D movie and experimental operation of 3D TV broadcasting. Many researches on 3D video have been done so far [1]-[3].

Also, haptic media come under the spotlight. A user is able to touch an object in a remote place and operate the object by using haptic media. It is expected that the efficiency of collaborative work over a network is greatly improved by using haptic media together with some other media [4]. An example of researches which deal with video and haptic media together is [5]. In [5], dynamic switching control of haptic transmission direction is proposed for a tele-operation system with non-stereoscopic video and haptic media. Also, automatic selection of switching time according to the contents of work is proposed, and the effectiveness is illustrated by quality of experience (QoE) [6] assessment (i.e., subjective assessment).

However, there are few papers that treat a combination of 3D video and haptic media. By using 3D video, a user is able to get high realistic sensation and the efficiency of work can be largely improved.

In this paper, we deal with a tele-operation system which transmits 3D video and haptic media of a real object. The system is referred to as the *tele-operation system with 3D video and haptic media* in this paper. By QoE assessment, we investigate the influence of the network delay. We also measure the application-level quality of service (QoS) by objective assessment. Then, we carry out QoS mapping [7] from application-level QoS parameters to QoE (also known as user-level QoS) parameters.

The rest of this paper is organized as follows. We explain the tele-operation system with 3D video and haptic media in Section 2. Section 3 describes the assessment method, and assessment results are presented in Section 4. Section 5 concludes the paper.

2. TELE-OPERATION SYSTEM WITH 3D VIDEO AND HAPTIC MEDIA

As shown in Fig. 1, the tele-operation system with 3D video and haptic media consists of a *master* terminal and a *slave* terminal. A 3D video camera (Minoru 3D webcam made by Promotion and Display Technology Ltd., Distance between cameras: about 6 cm, Field angle: about 42 degrees) is connected to the master terminal, and a 3D display (ZM-M220W made by Zalman Tech Co., Ltd.) is placed at the slave terminal. A 3D video is transmitted from the master terminal to the slave terminal. A user of the slave terminal is possible to view the video stereoscopically by wearing a pair of polarized glasses.

Also, we employ the PHANToM Omni [8] (just called PHANToM here) as a haptic interface device at each terminal. A clay cutter is attached to the PHANToM stylus at the master terminal by tapes (see Fig. 2). The distance between the 3D video camera and the clay cutter is about 80 cm.

The haptic media stream is transmitted from the master terminal to the slave terminal or between the terminals interactively. The former is called *one-way transmission*, and the latter *two-way transmission* in this paper. In one-way transmission, a user of the master terminal touches the object by using his/her PHANToM, and the master terminal conveys the haptic sensation experienced by him/her to the slave terminal. The haptic sensation is perceived by a user of the slave terminal. The user of the slave terminal only holds his/her PHANToM stylus, and he/she does not move the stylus independently of the master terminal's stylus. On the other hand, in two-way transmission, a user of the slave terminal remotely controls the PHANToM of the master terminal by manipulating his/her PHANToM, and he/she actively touches the object located at the master terminal while watching video of the object. A user of the master terminal only holds his/her PHANToM stylus, and he/she does not move the stylus independently. For the calculating method of the reaction force

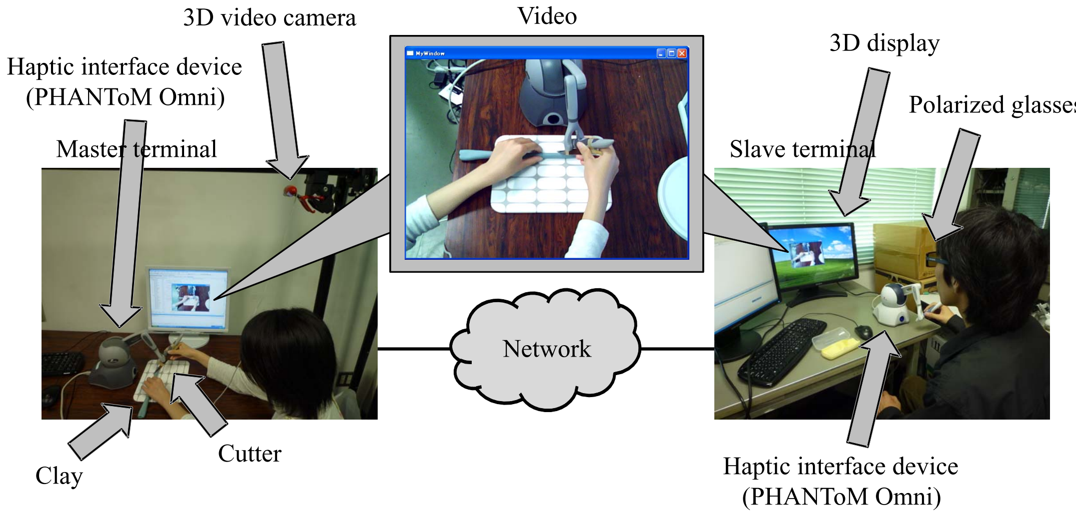

Fig. 1. Configuration of tele-operation system with 3D video and haptic media.

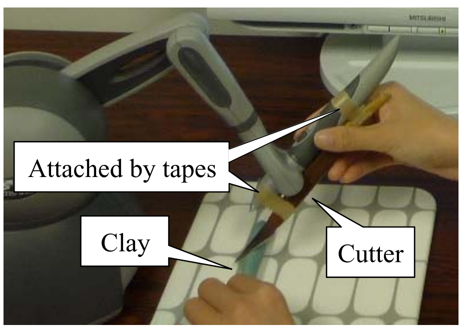

Fig. 2. PHANToM of master terminal.

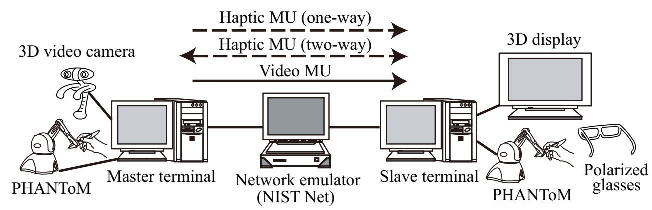

Fig. 3. Configuration of assessment system.

output against the PHANToM, the reader is referred to [9].

3. ASSESSMENT METHOD

3.1. Assessment System

We treat two assessments (*assessments (a) and (b)*). Assessment (a) investigates the influences of the network delay and delay jitter, and assessment (b) examines that of the inter-stream synchronization error.

(a) Influences of network delay and delay jitter

In this assessment, the master and slave terminals are connected to each other via a network emulator (NIST Net [10]) as shown in Fig. 3. NIST Net generates an additional delay according to the Pareto-normal distribution [10] for each packet transmitted between the master and slave terminals. 3D video and haptic media are transmitted independently as *media units* (*MUs*), each of which is the information unit for media output control. To investigate the influences of the network delay and delay jitter, we employ the *Skipping* algorithm [11] for the media output control at each terminal. The Skipping algorithm outputs MUs on receiving them. When multiple MUs are received at the same time, however, only the latest MU is output and the other MUs are discarded. Also, when the sequence number of a received MU is smaller than that of the last-output MU, the received MU is not output.

In the assessment, we deal with two cases for two-way transmission. In one case, the average additional delay is mainly changed. In the other case, the standard deviation of the additional delay is chiefly changed. In the former case, the average additional delays of 3D video and haptic media are changed from 0 ms to 200 ms at intervals of 50 ms. Then, the standard deviation of the additional delay is set to 0 ms or 20 ms for the two media streams. However, when the average additional delay is 0 ms, the standard deviation is set to 0 ms. In the latter case, the standard deviation of the additional delay is changed from 0 ms to 50 ms at intervals of 10 ms. Then, the average additional delay is set to 50 ms or 100 ms. For one-way transmission, we treat only the latter case, in which the average additional delay is set to 500 ms[1].

[1] In one-way transmission, the average additional delay does not affect the media output quality. To change the standard deviation of the additional delay largely, we set the average additional delay to 500 ms.

(b) Influence of inter-stream synchronization error

In this assessment, the master and slave terminals are connected to each other directly by a 100 BASE-T Ethernet cable. Then, each terminal outputs each received MU after adding a constant delay to the MU.

In one-way transmission, the constant delay of haptic media is set to 500 ms, and that of 3D video is changed from 100 ms to 900 ms. At this time, the inter-stream synchronization error between the 3D video and haptic media is from -400 ms to $+400$ ms, where negative values of the inter-stream synchronization error mean that the 3D video is ahead of the haptic media.

Also, in two-way transmission, the constant delay of haptic media is set to 10 ms or 100 ms. When the constant delay is 10 ms, the inter-stream synchronization error is changed from $+40$ ms to $+400$ ms; when the constant delay is 100 ms, that is changed from -80 ms to $+400$ ms.

In the two assessments, the size of a haptic MU is 40 bytes, and the average size of a video MU is 50 kbytes. The transmission rate of 3D video is 30 MU/s, and that of haptic media is 1000 MU/s. The average bit rate of 3D video is 12 Mbps, and the bit rate of haptic media is 320 kbps. Each video MU contains right and left video images. MUs are transmitted by User Datagram Protocol (UDP).

3.2. QoE Assessment

In QoE assessment, each subject operates the PHANToM of the slave terminal and one of the authors always manipulates the PHANToM of the master terminal. In one-way transmission, the author cuts a stick of clay several times by a clay cutter attached to the PHANToM of the master terminal. On the other hand, in two-way transmission, a subject moves his/her PHANToM as if he/she actually cut the clay.

Before the assessment, each subject was asked to do such a task on the condition that there was no additional delay. Then, he/she was asked to base his/her judgment about the video output quality, operability of PHANToM, inter-stream synchronization quality (in two-way transmission, the quality means the interactivity), and comprehensive quality in terms of wording used to define the subjective scale (5: imperceptible, 4: perceptible, but not annoying, 3: slightly annoying, 2: annoying, 1: very annoying). Each subject gave a score from 1 through 5 to each test to obtain the *mean opinion score* (*MOS*) [12], which is one of QoE parameters. The video output quality was assessed about the smoothness of the video output. The distance between each subject and 3D display was about 110 cm. The operability of PHANToM was graded on the smoothness of reaction force output and the easiness of operation. The inter-stream synchronization quality was rated on the synchronism between movement of subject's hand and that of PHANToM in the video. The comprehensive quality is a synthesis of the operability and the other two types of quality (i.e., the video output quality and inter-stream synchronization quality). The subjects were 20 persons whose ages were between 21 and 25.

3.3. Application-Level QoS Assessment

In this subsection, we define application-level QoS parameters for QoS mapping to QoE parameters (i.e., MOS). We adopt the *average MU rate*, *coefficient of variation of output interval*, *average MU delay*, *coefficient of variation of MU delay*, *average force*, *average speed*, *average error of inter-stream synchronization*, and *mean square error of inter-stream synchronization* [13].

The average MU rate is defined as the average number of MUs output in a second at each terminal. The average MU rates of 3D video and haptic media at the slave terminal are denoted by $R_{\mathrm{v,s}}$ and $R_{\mathrm{h,s}}$, respectively, in this paper. Then, the average MU rate of haptic media at the master terminal is denoted by $R_{\mathrm{h,m}}$. Note that "v" and "h" of subscripts mean 3D video and haptic media, respectively. Also, we abbreviate "master" and "slave" to "m" and "s." The coefficient of variation (the standard deviation divided by the average) of output interval indicates the smoothness of output of a media stream and is denoted by $C_{\mathrm{v,s}}$ for 3D video, and $C_{\mathrm{h,s}}$ and $C_{\mathrm{h,m}}$ for haptic media. The average MU delay is defined as the average of the difference between the output time and generation time of an MU. We denote it by $D_{\mathrm{v,s}}$ for 3D video, and $D_{\mathrm{h,s}}$, $D_{\mathrm{h,m}}$ for haptic media. The coefficient of variation of MU delay ($C_{\mathrm{d_v,m}}$, $C_{\mathrm{d_h,m}}$, $C_{\mathrm{d_h,s}}$) is calculated by dividing the standard deviation of MU delay by the average of MU delay. The average force F is the average of force presented to a subject. The average speed S is the average of speed at which the PHANToM moves at the slave terminal. The average error of inter-stream synchronization E is defined as the average of the difference between the output time of each slave MU and its *derived output time* [13] (i.e., the output time of the corresponding master MU plus the difference between the timestamps of the two MUs). The mean square error of inter-stream synchronization E_{sqr} is the average square of the difference between the output time of each slave MU and its derived output time.

4. ASSESSMENT RESULTS

4.1. Influences of Network Delay and Jitter

We show a part of QoE assessment results in Figs. 4 through 10, where the 95 % confidence intervals are also plotted. Since the other parts of the QoE assessment results had similar tendencies to these figures, the results are omitted here. Figures 4 and 5 plot the MOS values of the video output quality and comprehensive quality, respectively, as a function of the standard deviation of the additional delay in one-way transmission. In Figs. 6 through 10, we plot the MOS values of the operability of PHANToM, inter-stream synchronization quality, and comprehensive quality in two-way transmission. The estimated values of MOS shown by solid lines and broken lines in Figs. 4 through 10 will be described in Subsection 4.3.

In one-way transmission, from Figs. 4 and 5, we see that as the standard deviation becomes larger, the MOS values of the video output quality and comprehensive quality decrease.

In two-way transmission, we observe in Fig. 6 that the MOS value of the operability of PHANToM decreases as the average additional delay becomes larger. Figure 7 reveals that the MOS value tends to increase slightly as the standard deviation becomes larger. This is because the average MU delay decreases owing to Skipping (note that Skipping discards obsolete MUs), and MU loss does not lead to deterioration of the operability (this means that the MU rate of the haptic me-

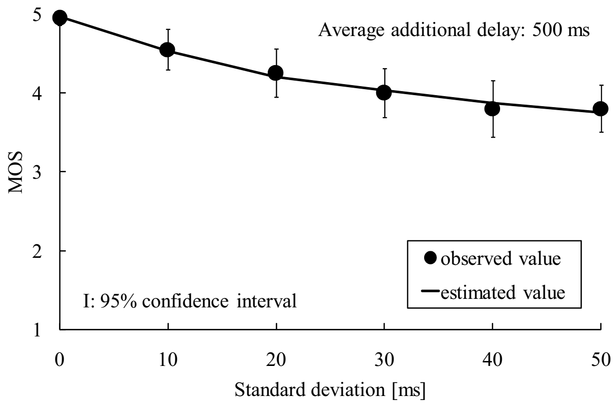

Fig. 4. MOS of video output quality versus standard deviation of additional delay in one-way transmission.

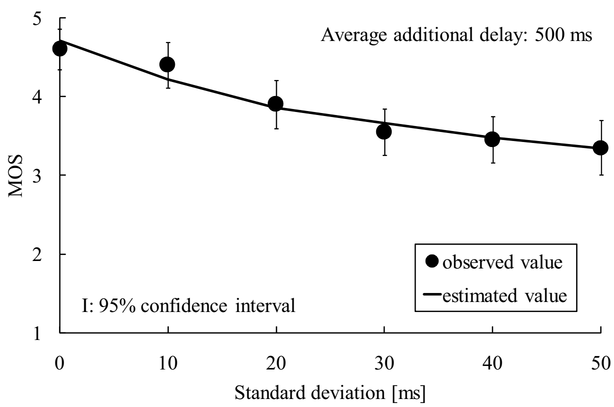

Fig. 5. MOS of comprehensive quality versus standard deviation of additional delay in one-way transmission.

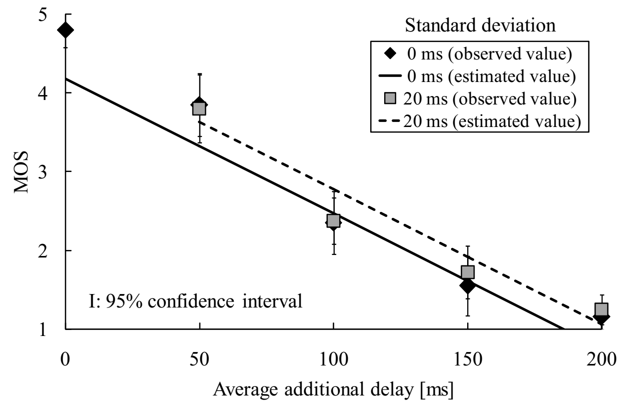

Fig. 6. MOS of operability of PHANToM versus average additional delay in two-way transmission.

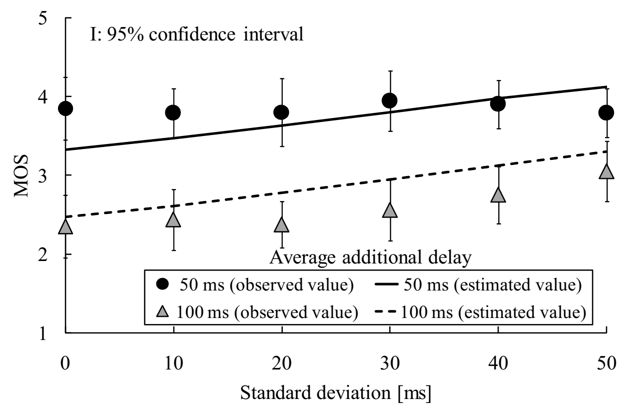

Fig. 7. MOS of operability of PHANToM versus standard deviation of additional delay in two-way transmission.

dia is large enough); as a result, the operation of PHANToM becomes easier. From Fig. 8, we find that the MOS value of the inter-stream synchronization quality decreases by growing of the average additional delay, and we also see that the MOS value is hardly affected by the standard deviation of the additional delay. In Figs. 9 and 10, we find that as the average additional delay becomes larger, the MOS values of the comprehensive quality decrease, and the MOS values hardly depend on the standard deviation.

4.2. Influence of Inter-Stream Synchronization Error

Figures 11 and 12 show the MOS values of the inter-stream synchronization quality in one-way transmission and two-way transmission, respectively, as a function of the inter-stream synchronization error.

From Fig. 11, we see that the MOS value is the highest when the inter-stream synchronization error is between about -80 ms and 0 ms. In Fig. 12, the MOS value is the highest when the inter-stream synchronization error is about 40 ms and the additional delay of haptic media is 10 ms. Also, when the inter-stream synchronization error is between about -40 ms and 0 ms and the additional delay of haptic media is 100 ms, the MOS value is the highest.

Therefore, the MOS value of the inter-stream synchronization quality is the highest when the 3D video and haptic media are output at the same time or the 3D video is output slightly ahead of haptic media. These results are similar to those in [14], which investigates the influence of inter-stream synchronization error between non-stereoscopic video and haptic media. However, we got many opinions that it is easier to do the task with 3D video than with non-stereoscopic video.

4.3. Relations between QoE Parameters and Application-Level QoS Parameters

By multiple regression analysis, we have carried out QoS mapping to estimate the QoE parameters (i.e., MOS) from the application-level QoS parameters (results of the application-level QoS assessment are omitted here owing to limited space). In the analysis, the application-level QoS parameters are the predictor variables (i.e., the independent variables), and the QoE parameters are the criterion variables (i.e., the dependent variables).

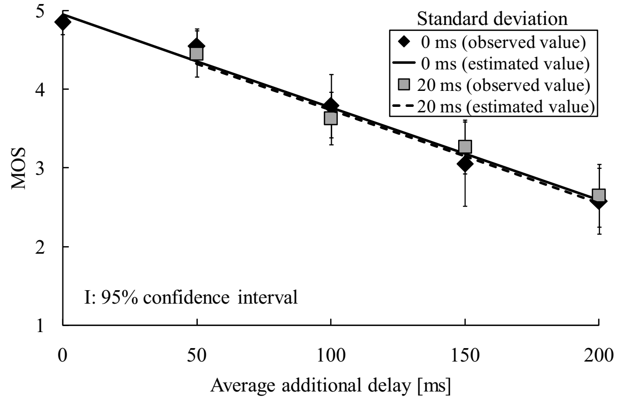

Fig. 8. MOS of inter-stream synchronization quality versus average additional delay in two-way transmission.

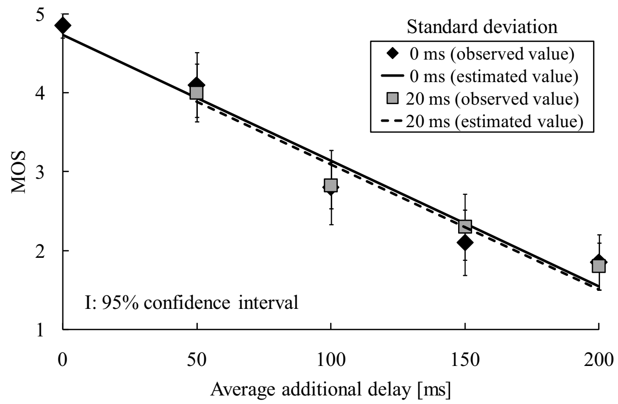

Fig. 9. MOS of comprehensive quality versus average additional delay in two-way transmission.

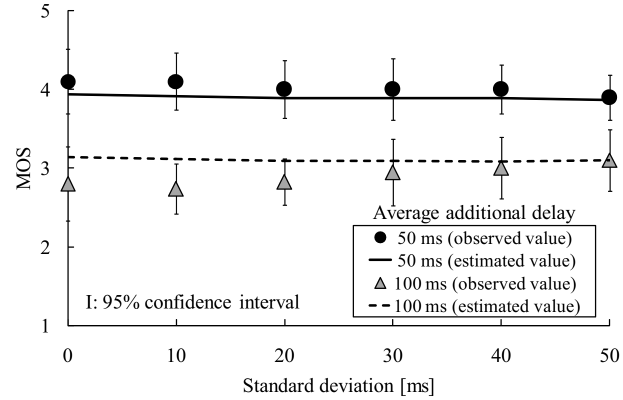

Fig. 10. MOS of comprehensive quality versus standard deviation of additional delay in two-way transmission.

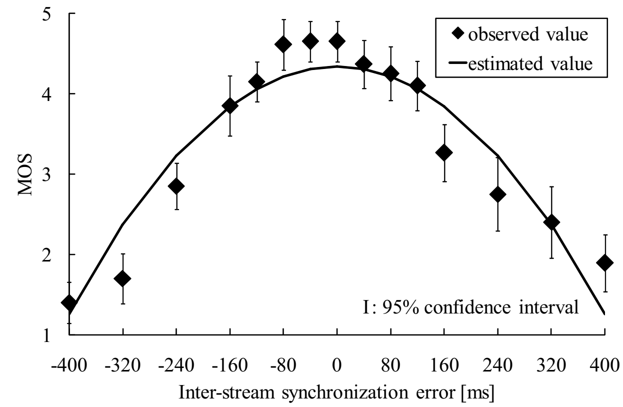

Fig. 11. MOS of inter-stream synchronization quality versus inter-stream synchronization error in one-way transmission.

We show equations of estimated MOS and the *contribution rates adjusted for degrees of freedom* (called *adjusted* R^2 here), which show goodness of fit with the estimated equations, in Tables 1 and 2. In the tables, $\hat{V}_{\mathrm{video}}$, $\hat{V}_{\mathrm{ope}}$, $\hat{V}_{\mathrm{sync}}$ and $\hat{V}_{\mathrm{total}}$ denote the estimated MOS values of the video output quality, operability of PHANToM, inter-stream synchronization quality, and comprehensive quality, respectively. All the values of adjusted R^2 in Tables 1 and 2 are high. To confirm that the equations can estimate the observed values with high accuracy, the estimated values of MOS are shown in Figs. 4 through 12. From these figures, we find the close agreement between the observed values and estimated ones. Therefore, we can estimate MOS with high accuracy by using the application-level QoS parameters.

5. CONCLUSIONS

In this paper, we investigated the influences of the network delay, delay jitter, and inter-stream synchronization error between 3D video and haptic media on QoE for the teleoperation system with 3D video and haptic media. As a result, we found that as the network delay jitter becomes larger, the QoE deteriorates in one-way transmission. On the other hand, in two-way transmission, we saw that the QoE mainly depends on the average network delay and hardly relates to the delay jitter. Also, we noticed that the inter-stream synchronization quality is the highest when the 3D video and haptic media are output at the same time or the 3D video is output slightly ahead of haptic media. In addition, we demonstrated that it is possible to estimate QoE parameters from application-level QoS parameters with high accuracy.

As the next step of our research, we will study media synchronization control which takes advantage of the results obtained in this paper for 3D video and haptic media.

6. REFERENCES

[1] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview imaging and 3DTV," Special Issue of IEEE Signal Processing Magazine, vol. 24, no. 6, pp. 10-21, Nov. 2007.

[2] W. Matusik and H. Pfister, "3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of

Table 1. Equations of estimated MOS and adjusted R^2 in assessment (a).

Transmission	Item	Equations of estimated MOS	Adjusted R^2
One-way	Video output quality	$\hat{V}_{\mathrm{video}} = 5.012 - 3.103 \times 10\, C_{\mathrm{d_v,s}}$	0.986
	Operability of PHANToM	$\hat{V}_{\mathrm{ope}} = 4.973 - 1.594 \times 10^2 C_{\mathrm{d_h,s}}$	0.942
	Synchronization quality	$\hat{V}_{\mathrm{sync}} = 4.567 - 3.787 \times 10^{-5} E_{\mathrm{sqr}}$	0.983
	Comprehensive quality	$\hat{V}_{\mathrm{total}} = 4.759 - 3.488 \times 10^{-1} C_{\mathrm{d_v,s}}$	0.945
Two-way	Video output quality	$\hat{V}_{\mathrm{video}} = 3.053 - 5.471 \times 10^{-2} R_{\mathrm{v,s}} - 2.576 \times 10^{-3} D_{\mathrm{v,s}}$	0.916
	Operability of PHANToM	$\hat{V}_{\mathrm{ope}} = 4.192 - 8.550 \times 10^{-3} (D_{\mathrm{h,s}} + D_{\mathrm{h,m}})$	0.903
	Synchronization quality	$\hat{V}_{\mathrm{sync}} = 5.044 - 3.929 \times 10^{-3} (D_{\mathrm{v,s}} + D_{\mathrm{h,s}} + D_{\mathrm{h,m}})$	0.970
	Comprehensive quality	$\hat{V}_{\mathrm{total}} = 4.863 - 5.305 \times 10^{-3} (D_{\mathrm{v,s}} + D_{\mathrm{h,s}} + D_{\mathrm{h,m}})$	0.942

Table 2. Equations of estimated MOS and adjusted R^2 in assessment (b).

Transmission	Item	Equations of estimated MOS	Adjusted R^2
One-way	Synchronization quality	$\hat{V}_{\mathrm{sync}} = 4.335 - 1.922 \times 10^{-5} E^2$	0.885
Two-way	Synchronization quality	$\hat{V}_{\mathrm{sync}} = 7.500 - 1.295 \times 10^{-5} E^2 - 5.509 \times 10^{-2} S$	0.903

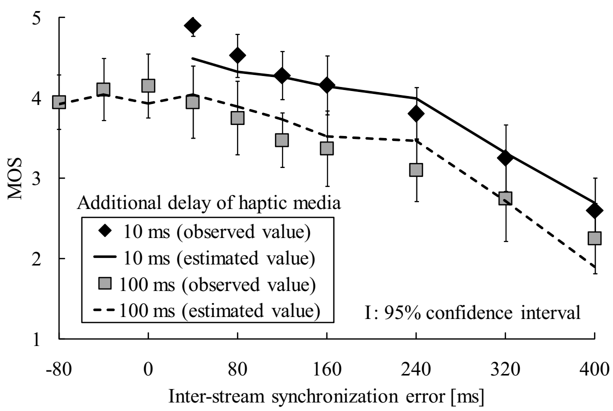

Fig. 12. MOS of inter-stream synchronization quality versus inter-stream synchronization error in two-way transmission.

dynamic scenes," ACM Trans. on Graphics (in Proc. ACM SIGGPRAPH 2004), vol. 23, no. 3, pp. 814-824, Aug. 2004.

[3] G. B. Akar, A. M. Tekalp, C. Fehn, and M. R. Civanlar, "Transport methods in 3DTV - A survey," IEEE Trans. on Circuits and Systems for Video Technology, Special Issue on Multi-view Video Coding and 3DTV, vol. 17, no. 11, pp. 1622-1630, Nov. 2007.

[4] M. A. Srinivasan and C. Basdogan, "Haptics in virtual environments: Taxonomy, research status, and challenges," Computers and Graphics, vol. 21, no. 4, pp. 393-404, Apr. 1997.

[5] T. Watanabe, Y. Ishibashi, N. Fukushima, and S. Sugawara, "Dynamic switching control of haptic transmission direction in remote control system," in Proc. the Haptics Symposium (Haptics'10), pp. 207-213, Mar. 2010.

[6] ITU-T Rec. G.100/P.10 Amendment 1, "New appendix I - Definition of quality of experience (QoE)," Jan. 2007.

[7] Y. Ito and S. Tasaka, "Quantitative evaluation and mapping of user-level QoS focusing on media synchronization in audio-video transmission," (in Japanese), IEICE Trans. on Commun., vol. J86-B, no. 3, pp. 485-498, Mar. 2003.

[8] J. K. Salisbury and M. A. Srinivasan, "Phantom-based haptic interaction with virtual objects," IEEE Computer Graphics and Applications, vol. 17, no. 5, pp. 6-10, Sep./Oct. 1997.

[9] T. Watanabe, Y. Ishibashi, N. Fukushima, and S. Sugawara, "Dynamic switching control of haptic transmission direction in remote control system," in Proc. ACM International Conference on Advances in Computer Entertainment Technology (ACE'10), Nov. 2010.

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

[11] Y. Ishibashi, S. Tasaka, and T. Hasegawa, "The virtual-time rendering algorithm for haptic media synchronization in networked virtual environments," in Proc. the 16th International Workshop on Communication Quality & Reliability (CQR'02), pp. 213-217, May 2002.

[12] ITU-R BT. 500-11, "Methodology for the subjective assessment of the quality of television pictures," 2002.

[13] Y. Ishibashi and S. Tasaka, "A synchronization mechanism for continuous media in multimedia communications," in Proc. IEEE INFOCOM'95, pp. 1010-1019, Apr. 1995.

[14] S. Kameyama and Y. Ishibashi, "Influences of inter-stream synchronization error on haptic media and video transmission," (in Japanese), in Proc. IEICE Society Conference, B-11-24, Sep. 2006.