A Multidimensional QoE Monitoring System for Audiovisual and Haptic Interactive IP Communications

Eiichi Isomura*, Shuji Tasaka†, and Toshiro Nunome*
*Graduate School of Engineering, Nagoya Institute of Technology, Nagoya 466–8555, Japan
† Interdisciplinary Engineering Laboratory, Nagoya Institute of Technology, Nagoya 466–8555, Japan

***Abstract*—This paper deals with audiovisual and haptic interactive IP communications and designs a multidimensional QoE monitoring system with the QoE estimation in real time. For the estimation, we carry out regression analysis between QoE and application-level QoS and obtain equations for the estimation. We measured the application-level QoS and made a subjective experiment by assessing QoE multidimensionally with the SD method. To monitor QoE, we derive estimates for three QoE metrics. We apply linear and non-linear regression analysis to the experimental results. As a result, we find that the obtained equations can estimate QoE with high accuracy. By using the equations, we have built the QoE monitoring system.**

***Index Terms*—interactive IP communications, audiovisual and haptic, QoE, real time estimation, QoE monitoring**

I. Introduction

Haptic devices attract people's attention as new human interface devices, which are expected to spread through consumer environments [1], [2]. Combinations of haptic devices and audiovisual ones can enhance users' experience, in particular the one of the telepresence, which is one of the most promising applications in the near future multimedia IP consumer communications.

Audiovisual and haptic interactive IP communications in the real space can provide the users with high realistic sensation and high quality of interactivity; this enables the users to immerse into the work easily. For enhancement of physical intimacies in remote interaction, a haptic audiovisual teleconferencing system with a haptic jacket is presented in [3]. In [4], a haptic media is used for training and mastering skills of expert engineers with high efficiency. In any haptic system, users' perception is the most important.

It should be noted here that 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[1] is defined for each layer [5]. The quality of end-user experience (i.e., *QoE: Quality of Experience*), which is on the top of the QoS hierarchy, is the most important [6], [7]. Since QoS is not guaranteed in IP networks, it is important to monitor users' QoE in real time.

Subjective quality assessment is the most fundamental method of evaluating QoE. However, subjective experiment is time-consuming and expensive. For monitoring QoE, it is practically impossible that the user continuously assesses his/her perceptual quality while using the service. Therefore, some method for estimating the subjective quality from objective quality parameters is necessary.

We can find many methods to estimate the service quality for audio-video transmission. As for estimating voice quality for VoIP, ITU–T has standardized the E–model as Recommendation G.107 [8]. It is convenient for network planning purposes because it formulates the relationships among subjective quality, network and/or application parameters. ITU–T Recommendation G.1070 [9] has extended the E–model to interactive videophone applications over best-effort IP networks as a QoE/QoS planning tool.

To the best of the authors' knowledge, there is no publication and tool which monitor the QoE of the haptic (and collaborating haptic and audio-video) transmission in real time. Real-time QoE monitoring is based on QoE estimation in real time.

We can find several researches on QoE estimation of haptic media in the literature, e.g., [10] and [11], although they are not real-time estimation. Hamam *et al.* present a taxonomy for evaluating the QoE of a haptic virtual environment and proposes a fuzzy logic system to model the taxonomy [10]. The Fuzzy Inference System (the Mamdani system) is used to simulate and examine the proposed model. The Mamdani system has five inputs (i.e., media synchronization, fatigue, haptic rendering, degree of immersion, and user satisfaction) and an output (the overall QoE). Furthermore, a subjective experiment has been conducted, and the overall QoE measured has been compared with the output of the fuzzy logic system; it has given a satisfactory result. However, they do not consider the haptic communication in real space. Moreover, real-time QoE estimation is not studied. In [11], Iwata *et al.* investigate the effect of playout buffering control on QoE in a haptic application with sound and video transmission for beating the tambourine. In addition, they carried out *QoS mapping* [5] for multiple metrics of QoE (i.e., operability of haptic media, video output quality, sound output quality, synchronization quality and comprehensive quality) and obtained equations of estimated QoE with high accuracy. In their experimental system, however, the transmission of video and audio has been made in only one direction. Therefore, the system's operation does not suppose an equal interaction between the two users with video and audio, which is important in some haptic applications such as in [3]. Also, in the subjective experiment in [11], the video quality assessed is only the temporal quality since the video decoder performs skipping of damaged video

[1]In IP networks, six kinds of QoS are identified along the protocol stack: *physical-level, node(link)-level, network-level, end-to-end(transport)-level, application-level,* and *user-level*(QoE).

frames; the video spatial quality is not taken into consideration. Furthermore, no real-time QoE monitoring has been considered in [11].

Building a real-time QoE monitoring system in audiovisual and haptic interactive IP communications with an equal interaction is the goal of this paper. In the authors' previous work [12], they conducted QoE assessment to evaluate how media degradations affect QoE for a task of *object movement*. By using the assessment results, the current paper estimates QoE by means of QoS mapping between QoE (user-level QoS) and application-level QoS. We perform the QoS mapping with *multiple regression analysis*. In the QoE monitoring, it should be noted that video, audio and haptic media have different properties in the allowable delay, the tolerance to packet loss and others. Therefore, user perception of quality deterioration depends on the media type. For this reason, we design a QoE monitoring system which can display three QoE metrics: video spatial quality, haptic operability and overall satisfaction; we adopt no metric of audio since the audio is used only for the directions for the task of object movement.

The remainder of this paper is structured as follows. An outline of the audiovisual and haptic interactive communication system is represented in Section II. A QoE monitoring model are described in Section III. Measurement methods of QoE and application-level QoS are explained in Section IV. QoE estimation methods and obtained equations are shown in Section V. Section VI illustrates the QoE monitoring system thus built. Finally, we conclude this paper and state future work in Section VII.

II. Audiovisual and Haptic Interactive Communication System

In this paper, we suppose that video, audio and haptic media are transmitted bidirectionally between two media terminals over a best-effort IP network. Fig. 1 illustrates functional components of the two media terminals. A haptic interface (*PHANToM Omni*), a video camera, an LCD monitor and a headset are connected to each media terminal. The media terminal has the ability to input and output the three media. Although the components of only media terminal 1 are shown in Fig. 1, media terminal 2 has the same configuration.

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.

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.

A reaction force of the haptic media which is presented to the user can be calculated by $\mathbf{F} = k\mathbf{V}$[13], where $\mathbf{F}$ is the reaction force [N], k is a spring constant ($= 0.1$ N/mm), and $\mathbf{V}$ is a displacement vector between the haptic interfaces [mm]. We illustrate the reaction force applied to the haptic interface in Fig. 2, where the position vector of its own terminal's PHANToM stylus is denoted by $\mathbf{A} = (x_a, y_a, z_a)$, and that of the other terminal's PHANToM stylus by $\mathbf{B} = (x_b, y_b, z_b)$. The displacement vector $\mathbf{V}$ is given by $\mathbf{V} = \mathbf{B} - \mathbf{A}$. As the end-to-end delay increases, $\mathbf{V}$ becomes larger. Therefore, the reaction force is stronger; this makes the operation of PHANToM tougher.

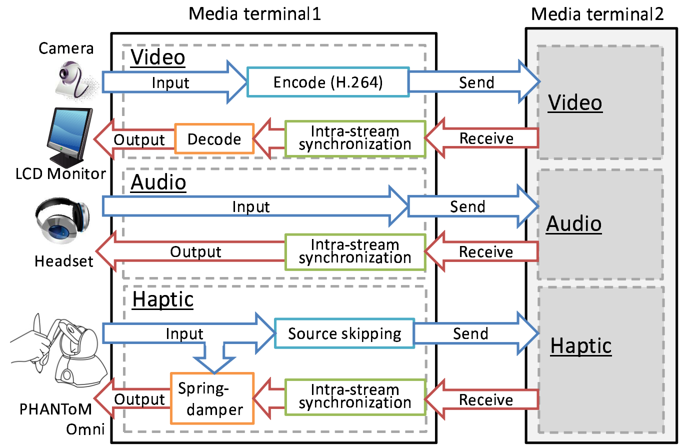


Fig. 1. Functional components of the media terminals.

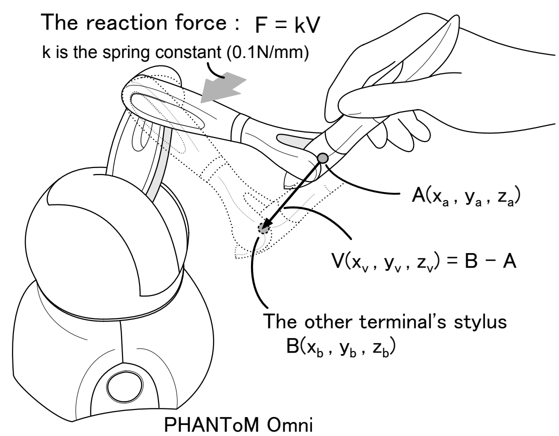


Fig. 2. Calculation of the reaction force.

In order to enhance haptic QoE, we adopt a method of *source skipping*, which reduces the transmission rate of the haptic media from 1000 MU/s (i.e., the original rate) to 500 MU/s by alternating sending and skipping a haptic MU.

Each terminal transmits the three media as three separate UDP streams.

The received MUs are delivered to *intra-stream synchronization control* for maintaining the temporal structure [14]. In this paper, we deal with *Skipping* and *playout buffering* as the intra-stream synchronization control. Skipping outputs only the latest MU out of a group of received MUs[2]. Thus, Skipping performs the quickest response. Playout buffering stores the MUs in a receive-buffer until the *target output time* of each MU, which is determined by the MU birth time and playout buffering time. When the MU arrives after the target output time, it is either output or discarded. In this paper, if the MU is received after its 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 the playout buffering time enough to absorb delay jitter can maintain the output quality. However, as the playout buffering time increases, the output delay becomes

[2]Note that this type of Skipping is different from the source skipping mentioned earlier; the two should be distinguished.

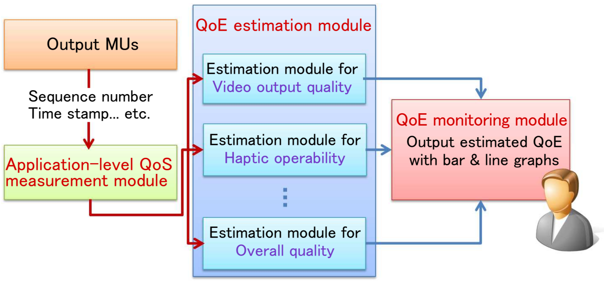


Fig. 3. QoE monitoring model.

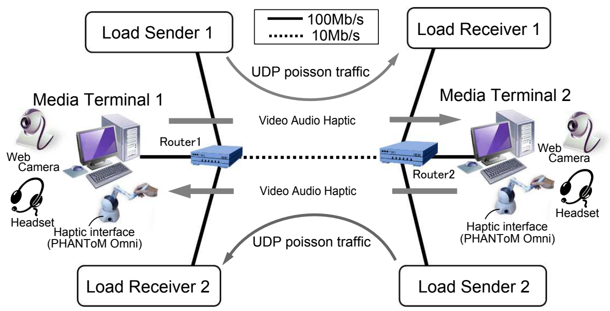


Fig. 4. Experimental system configuration.

longer. Thus, there is a trade-off relation between the output quality and responsiveness.

III. QoE MONITORING MODEL

We take a QoE monitoring model, which estimates and displays the QoE in real time, as shown in Fig. 3. The QoE monitoring model works as follows. First, the application-level QoS measurement module acquires information from each of output MUs: the birth time, target output time and sequence number. Second, the application-level QoS measurement module calculates values of application-level QoS parameters (e.g., MU loss ratio and MU output delay) with the inputs at regular intervals and outputs the results to the QoE estimation module. Third, QoE is estimated from the application-level QoS parameters. The QoE estimation module is composed of several estimation modules, each of which corresponds to a QoE metric. Each module has an individual equation for the QoE estimation. Finally, the QoE monitoring module displays the estimated QoE multidimensionally with bar and line graphs. The graphs are updated at the regular intervals. Thus, the users can monitor the QoE in real time.

IV. MEASUREMENT OF QoE AND APPLICATION-LEVEL QoS

Fig. 4 shows the experimental system, which consists of six PC's (two Media Terminals, two Load Senders and two Load Receivers) and two routers (Cisco 2811). The routers are connected by a full-duplex Ethernet link of 10 Mb/s. All the other links are 100 Mb/s full-duplex Ethernet.

Tables I and 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 [15]. We take two values of the haptic MU rate [MU/s]: 1000 (without the source skipping) and 500 (with the source skipping).

TABLE I
SPECIFICATIONS OF VIDEO AND AUDIO

	video	audio
encoding scheme	H.264 (x264) 800 $\times$ 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 MEDIA

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

Taking into consideration the features of the three media properties, we adopt the following three intra-stream synchronization control schemes in the experiment [12]:

Scheme 1. Media adaptive buffering (MAB)

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

Scheme 2. Skipping & buffering (S&B)

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

Scheme 3. Buffering (BUF)

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

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 set to 10 ms because of its smoothness and operability.

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].

We have used the task of object movement in [12]. We aimed 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 a 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. Fig. 5 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 area. In the task, the role of the subjects is two kinds: the instructor and the manipulator. One subject plays a role of the instructor,

[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; load traffic of over 6.0 Mb/s degrades the operability of the haptic interface owing to queuing delay.

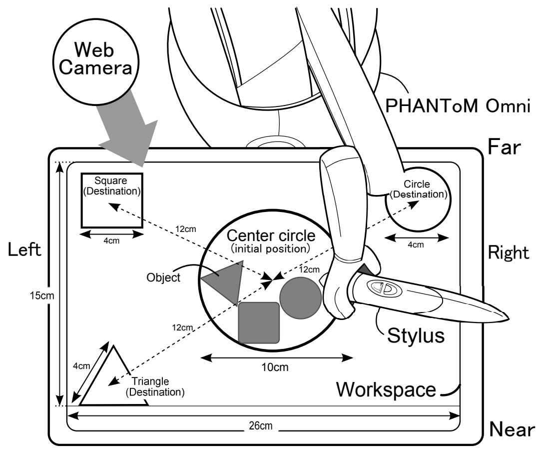


Fig. 5. Workspace and layout at each terminal.

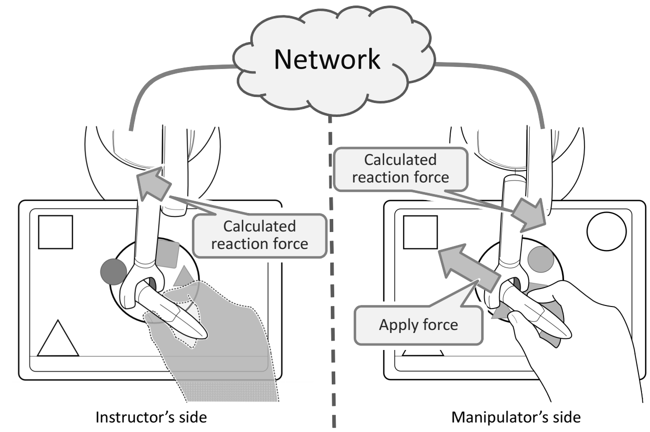
TABLE III
PAIRS OF POLAR TERMS

class	item	polar terms
Video:v1	Spatial quality	Video is corrupt–clear
Video:v2	Temporal quality	Video is jerky–smooth
Video:v3	Usefulness	Image is hard to grasp–easy
Haptic:h1	Operability	Manipulation is heavy–light
Haptic:h2	Smoothness	Movement is awkward–smooth
Haptic:h3	Stability	Manipulation is unstable–stable
Audio:a1	Naturalness	Artificial–Natural
Inter-stream sync:s1	Video and haptic	Out of sync–In sync
Interactive:i1	Response	Response is slow–rapid
Interactive:i2	Communicability	Hard to communicate–Easy
Interactive:i3	System comfort	Uncomfortable–Comfortable
Interactive:i4	Work difficulty	Difficult to work–Easy
Overall	Overall satisfaction	Unsatisfied–Satisfied

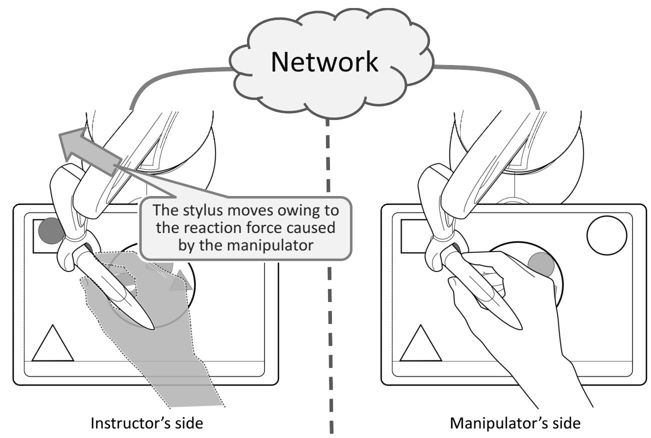
and the other is the manipulator. First, the instructor selects an object in the center circle and its destination on his/her own side; then he/she tells the instruction to the manipulator using the headset microphone. The manipulator replies to the instruction and then manipulates PHANToM to move the specified object to the destination on the instructor's side, while watching the video and grasping the positional relation between PHANToM and the object on the instructor's side. Fig. 6 shows the situation in which the circle object on the instructor's side is moved to the square destination on the same side by the manipulator's remote operation. 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 instructor only holds his/her PHANToM stylus and surrenders him/herself to the manipulator's movement.

For multidimensional QoE assessment, we use the *SD (Sematic Differential)* method [16]. 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* (i.e., the totality of the three media output, operation and responsiveness, etc.) by the *rating scale method* [17] 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 meanings of score 4 and score 2 are the meanings modified by "slightly" of score 5 and score 1, respectively. The middle grade (score 3) means neutral.



(a) The manipulator operates his/her own PHANToM stylus while watching the video. The instructor only holds his/her PHANToM stylus and surrenders him/herself to the manipulator's movement.



(b) The stylus of the instructor's PHANToM moves owing to the reaction force caused by the manipulator. The circle object of the instructor's side is pushed by the stylus.

Fig. 6. The procedure for the task of object movement.

In this paper, we express QoE multidimensionally in terms of the scores given by the subjects for each pair of polar terms; the scores are converted into the *interval scale*[4] by the *method of successive categories*. Moreover, we confirm the goodness of fit for the obtained scale by using Mosteller's test [18]. Once the goodness of fit has been confirmed, we use the interval scale as the QoE metric, which is called the *psychological scale* [19].

The subjects in the experiment were male and female students in their teens or twenties; the number of the subjects is 56 (28 males and 28 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

[4] A simple arithmetic average of the scores gives *MOS (Mean Opinion Score)*. We do not use MOS because of higher accuracy of the interval scale [19].

TABLE IV
APPLICATION-LEVEL QOS PARAMETERS AND THE NOTATIONS

parameter	video	audio	haptic
video slice arrival ratio[%]	S	–	–
MU output rate[MU/s]	R_v	R_a	R_h
average MU output delay[ms]	D_v	D_a	D_h
Coefficient of variation of output interval	C_v	C_a	C_h
MU loss ratio[%]	L_v	L_a	L_h
MSE of inter-stream synchronization [ms^2]	E_{va}, E_{vh}, E_{ah}		

about 60 minutes for a subject to assess all the stimuli.

During each experiment run, Media Terminals measured application-level QoS parameters which are listed in Table IV. 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 *average MU output delay* is defined as the average of the difference between the output time and generation time of an MU. The *coefficient of variation of MU output interval* is defined as the ratio of the standard deviation of the MU output interval to its average; therefore, it represents the smoothness of the output stream. The *MU loss ratio* is the ratio of the number of MUs not output at the recipient to the number of MUs transmitted by the sender. The *Mean Square Error (MSE) of inter-stream synchronization* is defined as the average square of the difference between the output time difference of one media and another media and their time stamp difference.

We easily see that all the application-level QoS parameters above are automatically measurable; therefore, we can use them to estimate the psychological scale in real time.

V. ESTIMATION OF QOE

A. QoS Mapping

As the metrics for QoE monitoring, we employ the overall satisfaction and two metrics selected from among the 12 QoE measures in Table III (v1 through i4); the choice was made by using *principal component analysis*[5].

Fig. 7 plots the principal component loading values. In Fig. 7, many adjective pairs of the interactive class and haptic class have large positive values of the first principal component loading. Moreover, the video class and audio class have large positive values of the second principal component loading.

On the basis of the results shown in Fig. 7, as the metrics for QoE monitoring, we selected haptic operability h_op (= h1) from the first principal component, video spatial quality v_sp (= v1) from the second principal component, and overall satisfaction ov.

In order to estimate QoE, we resort to multiple regression analysis by defining the psychological scale for each metric as the *dependent variable*. As the *independent variables*, we employ the application-level QoS parameters. We perform both linear analysis and nonlinear one based on exponential functions.

Regression analysis requires us to select appropriate independent variables with low cross-correlations from among the introduced variables. Therefore, we computed principal component loadings of the application-level QoS parameters (i.e., the independent variables) and classified them into five classes as shown in Table V. We then pick up a variable from each class and calculate a multiple regression line for every combination of the variables.

[5]Principal component analysis helps us find the correlations between the values. We compute the principal component loadings of each variable up to the principal component that provides a large cumulative contribution rate (e.g., over 90%). On the basis of the principal component loadings, we classify the values into a certain number of classes.

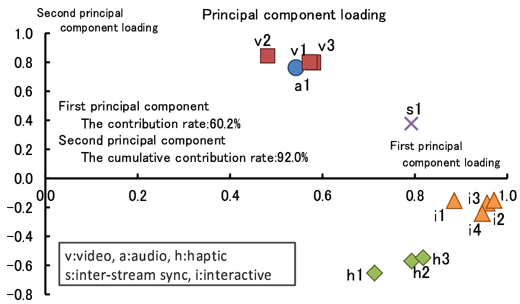


Fig. 7. First and second principal component loading.

TABLE V
CLASSIFICATION OF APPLICATION-LEVEL QOS PARAMETERS

class	parameters
A	$S, R_v, R_a, E_{va}, L_v, L_a, C_v, C_a$
B	D_v, D_a
C	R_h, L_h, C_h
D	D_h
E	E_{vh}, E_{ah}

From among the multiple regression lines thus calculated, we finally select one that achieves the largest value of the *contribution rate adjusted for degrees of freedom* (R^{*2}), which indicates goodness of fit of estimates to the corresponding measured values.

We first performed linear multiple regression analysis of all combinations of the application-level QoS parameters under the condition that one parameter is selected from one class. As a result, we found the following multiple regression lines; in the equations, the estimate of the psychological scale is represented by $\hat{X}^L_{ME}$, where the subscript ME (v_sp, h_op, or ov) means the kind of the QoE metric, and the superscript L indicates linear analysis; R^* denotes the *multiple correlation coefficient adjusted for degrees of freedom*.

$$\hat{X}^L_{v\_sp} = -20.265 + 0.235S \quad (R^* = 0.887) \quad (1)$$

$$\hat{X}^L_{h\_op} = 3.054 - 1.533 \times 10^{-2} D_h \quad (R^* = 0.943) \quad (2)$$

$$\hat{X}^L_{ov} = 0.104S - 1.230 \times 10^{-2} D_h - 0.439C_h - 6.754 \quad (R^* = 0.912) \quad (3)$$

We next performed nonlinear multiple regression analysis based on exponential functions for the video spatial quality and overall satisfaction. Note that Equation (2) can estimate the haptic operability with high accuracy. Thus, we did not derive any nonlinear equation for the haptic operability. Below, we show the obtained multiple regression lines, where $\hat{X}^N_{ME}$ represents the nonlinear estimate of the psychological scale for the kind of the metric ME (v_sp or ov).

$$\hat{X}^N_{v\_sp} = 2.920 \times 10^{-5} e^{0.116S} \quad (R^* = 0.933) \quad (4)$$

$$\hat{X}^N_{ov} = 2.767 e^{1.186 \times 10^{-2} S} + 4.257 e^{-4.100 \times 10^{-3} D_h} - 0.575C_h - 9.500 \quad (R^* = 0.925) \quad (5)$$

Comparing the linear and nonlinear equations thus obtained for the video spatial quality and overall satisfaction, we see that the nonlinear equations provide larger values of R^* than the linear ones. In the following discussion, we utilize Equations (2), (4) and (5).

B. Accuracy of estimation

We now examine the accuracy of the estimate with the multiple regression lines by comparing them to the corresponding measured values. Because of space limitations, we show the results only for limited experimental conditions.

Fig. 8 plots estimated values of the video spatial quality along with the measured ones in the case where the haptic MU rate is 1000 MU/s as a function of playout buffering time which is set for video, audio, and haptic media of BUF (scheme 3). In Fig. 8, we see that Equation (4) can estimate the video spatial quality with good accuracy.

Fig. 9 presents the results of the haptic operability in the case where the haptic MU rate is 500 MU/s. In Fig. 9, we find that the estimated values of the haptic operability can approximate the measured ones with high accuracy. In BUF, the end-to-end delay of the haptic media increases as the playout buffering time becomes longer; therefore, the haptic manipulation becomes tougher.

Fig. 10 displays the result of the overall satisfaction in the case where the haptic MU rate is 500 MU/s. The estimated values are close to the measured ones. When the playout buffering time is 20 ms, the estimated and measured QoE are degraded because of video slice loss. In BUF, the result is similar to that of the haptic operability because of the haptic end-to-end delay.

In addition, we conducted another experiment with 16 subjects who are not included by the 56 subjects. For the experiment, the task is the object movement. By using the results, we confirmed that the equations can estimate QoE with high accuracy. Due to limitations of space, the results are not shown in this paper.

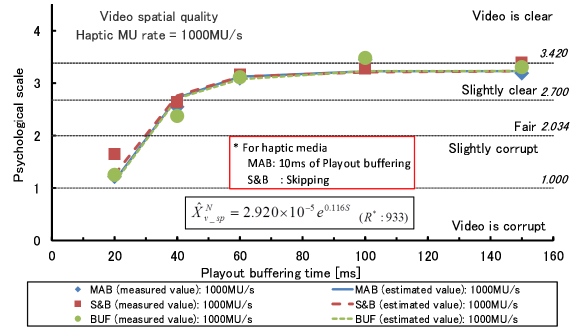


Fig. 8. Video spatial quality (haptic MU rate = 1000 MU/s).

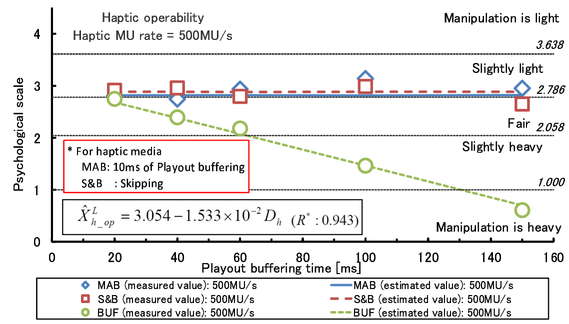


Fig. 9. Haptic operability (haptic MU rate = 500 MU/s).

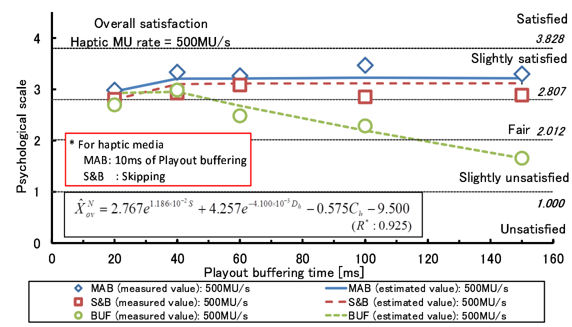


Fig. 10. Overall satisfaction (haptic MU rate = 500 MU/s).

VI. QoE Monitoring System

We implemented a QoE monitoring system in the experimental system with the derived QoE estimation equations. Fig. 11 shows a snapshot of the implemented QoE monitoring. In Fig. 11, the system depicts graphs of estimated values of the three QoE metrics as a function of time: the video spatial quality, haptic operability and overall satisfaction. For estimation of the video spatial quality, Equation (4) is used. Similarly, Equation (2) is used for the haptic operability, and Equation (5) for the overall satisfaction. The application-level QoS and estimated QoE are calculated and updated at intervals of a second in the system.

In the QoE monitoring system, the current estimated QoE is represented with a bar graph on the left hand side; below the bar graph, the system shows the quality category (Categories 1 through 5), to which the estimated value belongs. There are five kinds of category notation: Excellent, Good, Fair, Poor and Bad. For example, in the case of the haptic operability of Fig. 11, Excellent means that the operation of the PHANToM is light, while Bad means the heavy operation. In addition, the meaning of Fair is neutral, Good and Poor express Categories 4 and 2, respectively.

The right-hand side of the system displays the estimated QoE history with the line graph; the right edge of the graph corresponds to the current estimated value. Horizontal dashed lines in the graph represent the category boundaries, and the notations on the left edge of the graph (i.e., Cate.1 through Cate.5) means the category number. The system can render the estimated QoE history for two minutes with the line graph.

The monitoring system shows more than one QoE metrics simultaneously; this implies monitoring of multidimensional QoE. Moreover, the value of application-level QoS parameters, which are measured in real time, can be displayed (see the bottom of Fig. 11).

In the QoE monitoring of this paper, the most important QoE metric is the overall satisfaction. However, dominant factors affecting the overall satisfaction can depend on individual users. In order to achieve the highest QoE for the individual users, we need some method of customization [20]. The customization allows users to select favorite values of important parameters (e.g., video encoding bit rate and playout buffering time). We are planning to implement this type of customization into our audiovisual and haptic interactive communication system in order to accommodate the individual users' preferences.

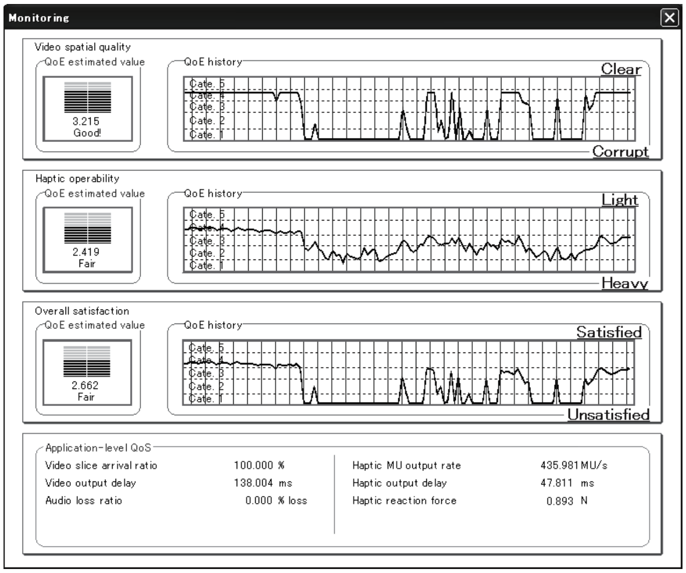


Fig. 11. A snapshot of the QoE monitoring.

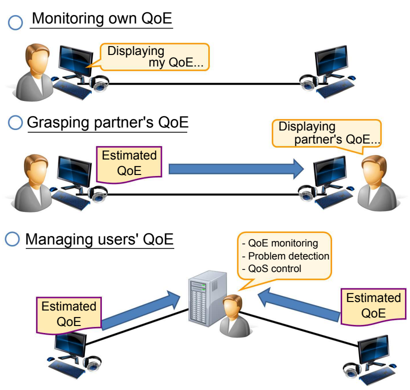


Fig. 12. Utilization of the monitoring system.

Several methods of utilization of the QoE monitoring system are conceivable as shown in Fig. 12. In the current implementation, the terminal which receives and outputs media displays the monitoring window. By sending the information of QoE estimations to another terminal, the terminal becomes a QoE monitoring terminal. In other words, if the estimated values of QoE are sent to the other side of the interactive service, the other user can be informed of the partner's QoE. As the third example, we can set up a management terminal separately from the media terminals, and both media terminals send the QoE estimation results to the management terminal. Then, the service provider can monitor the users' QoE in real time on the management terminal in order to detect problems and perform QoS control for QoE enhancement.

VII. Conclusions

We designed and implemented a QoE monitoring system for audiovisual and haptic interactive IP communications. For real time monitoring, we carried out QoS mapping using multiple regression analysis to estimate QoE from application-level QoS parameters. Three QoE metrics (video spatial quality, haptic operability and overall satisfaction) were picked up for the estimation and monitoring. As a result of the analysis, we found that the derived equations can estimate QoE with high accuracy. By using the equations, we built the QoE monitoring system, which displays multidimensional QoE in real time.

As future work, we will study different tasks from that in this paper. In addition, QoS control with QoE monitoring is important. It is also interesting to investigate the goodness of fit between estimated values and measured values in the case of time varying load traffic environments. Furthermore, some methods of QoE customization should be developed as the next step of this study.

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 Challenging Exploratory Research of Japan Society for the Promotion of Science under Grant 23656253.

References

[1] F. Danieau, J. Fleureau, A. Cabec, P. Kerbiriou, P. Guillotel, N. Mollet, M. Christie, and A. Lecuyer, "A framework for enhancing video viewing experience with haptic effects of motion," in *Proc. IEEE Haptics Symposium 2012*, pp. 541–546, Mar. 2012.

[2] M. Cruz and D. Grant, "High definition haptics for consumer electronics," in *Proc. IEEE ICCE2011*, pp. 57–58, Jan. 2011.

[3] M. Eid, J. Cha, and A. El Saddik, "HugMe: a haptic videoconferencing system for interpersonal communication," in *Proc. IEEE VECIMS2008*, pp. 5–9, July 2008.

[4] T. Ishii and S. Katsura, "Switching bilateral control system for haptic training," in *Proc. SICE2011*, pp. 633–638, Sept. 2011.

[5] S. Tasaka and Y. Ishibashi, "Mutually compensatory property of multimedia QoS," in *Conf. Rec. IEEE ICC2002*, vol.2, pp.1105–1111, Apr./May 2002.

[6] 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.

[7] ETSI TR 102 643, "Human Factors (HF); Quality of Experience (QoE) requirement for real-time communication services," Jan. 2010.

[8] ITU–T Rec. G.107, "The E–model: a computational model for use in transmission planning," Apr. 2009.

[9] ITU–T Rec. G.1070, "Opinion model for video-telephony applications," Apr. 2007.

[10] A. Hamam, M. Eid, A. El Saddik, and N. D. Georganas, "A fuzzy logic system for evaluating Quality of Experience of haptic-based applications," in *Proc. EuroHaptics2008*, pp.129–138, June 2008.

[11] 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. ACM International Conference on Advances in Computer Entertainment Technology (ACE'10)*, Nov. 2010.

[12] E. Isomura, S. Tasaka, and T. Nunome, "QoE enhancement in audiovisual and haptic interactive IP communications by media adaptive intra-stream synchronization," in *Proc. IEEE TENCON2011*, pp.1085–1089, Nov. 2011.

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

[14] 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.

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

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

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

[18] 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.

[19] 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.

[20] S. Tasaka and N. Misaki, "Customization of interactive services for QoE enhancement in audio-video transmission over bandwidth guaranteed IP networks," in *Proc. IEEE ICME2010*, pp.546–551, July 2010.