Maximizing QoE of Interactive Services with Audio–Video Transmission over Bandwidth Guaranteed IP Networks

Shuji Tasaka and Norihito Misaki
Department of Computer Science and Engineering,
Graduate School of Engineering, Nagoya Institute of Technology, Nagoya 466–8555, Japan
Email: tasaka@nitech.ac.jp

***Abstract*— This paper identifies factors affecting QoE (Quality of Experience) of interactive audiovisual communications over bandwidth guaranteed IP networks and finds how to maximize QoE in that environment. Among the factors, we focus on the guaranteed bandwidth between the terminals, video encoding bit rate, the playout buffering time, and task. Under the condition that the guaranteed bandwidth for video is kept constant for a given amount of the guaranteed bandwidth for audio, we made experiment on QoE assessment for various values of the video encoding bit rate and playout buffering time. We suppose two kinds of task: one is audio-dominant, while the other is a task whose video affects QoE more than the former. As a result of the experiment, we see that an appropriate selection of the pair of the video encoding bit rate and playout buffering time maximizes QoE for the guaranteed bandwidth for video. Furthermore, we notice that the playout buffering time should be carefully chosen especially for the latter task.**

I. Introduction

The interactivity is an intrinsic function of communications; it has been a key requirement for the telecommunication system since its birth. In addition to the classical telephony, contemporary networked multimedia applications like video telephony, multimedia conferencing and networked gaming are based upon the function.

As typified by the telephony, such interactive services have been offered traditionally over bandwidth guaranteed channels in circuit–switching networks and dedicated systems. Although the services are also provided as best–effort ones over the current IP networks as well, we anticipate that they will be replaced with *QoS* (*Quality of Service*) guaranteed ones by the *Next Generation Network* (*NGN*) [1], which is an enhanced IP-based network with QoS guarantee; NGN has been under rapid deployment worldwide.

While the design and management methodologies for QoS of the interactive services over the legacy networks are available, the ones for NGN has not yet been established sufficiently; this is a very important open issue for the successful implementation of NGN.

The QoS to be considered in NGN is described as *network QoS* by ITU–T Recommendation Y. 1540 [2] in terms of QoS parameters including IPTD (IP packet transfer delay), IPDV (IP packet delay variation), IPLR (IP packet loss ratio), and IPER (IP packet error ratio). Furthermore, the objective values of the QoS parameters are specified by Rec. Y. 1541 [3], which gives some guidance for it by classifying typical applications into eight classes: Classes 0 through 7. Among the QoS classes, Classes 0 through 3 correspond to interactive applications. Note that the network QoS above is regarded as *network–level QoS* from a layered architectural point of view[1].

For further discussion on QoS of interactive communications, recall that the interactivity is a concept of end–to–end user communications, which contain the terminal equipments and user–to–user connection. Consequently, its QoS should involve not only the network–level but also the transport–level, application–level and user–level, which are outside the scope of ITU–T Recommendations Y. 1540 and Y. 1541.

Now let us consider what the ideal target quality is for the users, putting aside the tradition and convention of network design and management, since NGN as a new infrastructure should be realized in order to offer truly satisfactory services to the users. Then, we easily see that it is user–level QoS, which is referred to as *QoE* (*Quality of Experience*) in ITU–T; QoE represents *the overall acceptability of an application or service, as perceived subjectively by the end–user* [5].

It should be noted here that the objective values of network QoS specified by ITU–T Rec. Y. 1541 do not necessarily achieve QoE the users desire in efficient ways; the specified values may require too many resources to satisfy the users' demand or it may not realize the desired QoE, since QoE is beyond the scope of Y. 1541. Thus, in addition to the Y. 1541 guidance for the network QoS, we need another guidance for the quantification of QoE.

This paper presents a first step to the study on methodologies of the design and management for QoE enhancement/guarantee of interactive services over bandwidth guaranteed IP networks. We focus on interactive services with audio–video transmission since it forms an indispensable basis of networked multimedia applications.

QoE of interactive audiovisual communications over IP networks depends on various factors regarding the audio quality, video quality, media synchronization quality between audio and video, and responsiveness, which are affected in turn by network, application, and terminal equipment parameters

[1]In IP networks, six kinds of QoS can be identified along the protocol stack: *physical–level*, *node(link)–level*, *network–level*, *end–to–end(transport)–level*, *application–level*, and *user–level* [4].

as well as users' behavior. We need to clarify how these factors affect QoE and then establish some way of quantifying the impact of the factors so that they can maximize QoE. However, we can find no report on this subject in the literature. The purpose of this study is to give an answer to the above problem.

The remainder of the paper is organized as follows. Section II identifies factors affecting QoE and gives a brief description of related work. Section III defines the bandwidth guaranteed IP network we suppose in this paper. Section IV introduces a method of QoE assessment for our interactive communications. Section V demonstrates an experimental system, tasks to be assessed, and a method of measuring QoE. Section VI presents experimental results and examines how the factors affect QoE. Section VII concludes the paper.

II. Factors Affecting QoE and Related Work

In this section, we first list up possible factors affecting QoE of interactive audiovisual IP communications and focus on dominant ones. We restrict the factors to design parameters, i.e., parameters controllable by the system designers, operators and end–users, since the purpose of this paper is to maximize QoE by the design and management. We then describe related work to this issue.

The possible factors from the above mentioned point of view are: (1) the guaranteed bandwidth between the terminals, (2) audio/video encoding schemes including the bandwidth compression methods, image size, bit rates, frame rate and picture pattern, (3) audio/video decoding schemes including error–concealment methods and video frame skipping, (4) the end–to–end delay consisting of the propagation delay, the delays at routers (i.e., queueing and processing delay) and the terminal delay, which includes jitter-buffer delay (i.e., *playout buffering time*) for jitter absorption in addition to queueing and processing delay, and (5) tasks for the application.

Since there are a variety of design parameters, as a first step toward the study, this paper focuses on the following ones, setting the others to typical ones:

1) the guaranteed bandwidth between the terminals
2) the video encoding bit rate
3) the playout buffering time
4) the task

Note that the first parameter is a factor dominating the network QoS, and the second one corresponds to the sender–part at the application–level of the terminal, while the third one is the receiver–part. The task represents the user's behavior.

It is clear that the guaranteed bandwidth between the terminal is crucial to QoE since the former dominates the end–to–end throughput and delay. In practical service offering by the NGN carriers, the user is often required to select one service class from among a finite number of the set, each of which usually guarantees a certain amount of the bandwidth. In this paper, we suppose this situation.

It is also important how the encoding bit rate of video is set for a given value of the guaranteed bandwidth[2]. The video encoding bit rate fluctuates; therefore, the encoding up to the guaranteed bandwidth is not necessarily the best strategy since video packets exceeding the bandwidth are lost. A bit rate lower than the maximum at the guaranteed bandwidth may achieve the best quality of video.

Also, among the components of the end–to–end delay, the playout buffering time plays an important role in interactive audiovisual packet communications. In bandwidth guaranteed IP networks, the delays at routers are basically low, and environments of not so large propagation delay are usually supposed for ensuring quick response; therefore, it is often the case that the playout buffering time is the dominant component.

The playout buffering time produces a tradeoff relationship between *fidelity* and *latency* [6], [7]. As the buffering time increases, the fidelity of output audio and/or video improves because packets arriving later can be output (otherwise discarded) by absorbing larger delay jitter[3]. At the same time, however, the longer buffering time increases the latency, which degrades the responsiveness. Shorter buffering time brings about the opposite effects on the output media quality and responsiveness. Thus, we have the tradeoff relationship. Note that the fidelity and latency in this context correspond to application–level QoS.

Furthermore, the task strongly affects QoE of interactive audiovisual communications since it is the work context in which the real–time communication system is employed. For instance, how does it draw the attention of the user to audio and/or video signals? (e.g., audio is dominant since the task is speech-oriented, or video is dominant because he/she gets main information from video), and how often conversation switching occurs? ITU–T Rec. P.920 [8] defines interactive test methods for audiovisual communications. It presents five examples of the task according to the degree to which free conversation can occur and the importance of video: the Name–Guessing task, the Story–Comparison task, the Picture–Comparison task, the building blocks task, and the object–description task. In addition, the protocols for the tasks are specified.

Among the factors mentioned earlier, many of previous studies on networked multimedia applications focus on the playout buffering time from an application–level QoS point of view; moreover, the great majority of the studies treat either audio only or video only (e.g., see [7], [9], [10] and [11]), while few deal with both media (e.g., see [12]). It should be noted that these studies adopt end–to–end *one–way delay* metrics under the condition that a maximum acceptable value of the delay is specified; this way of the modeling cannot reflect the interactivity sufficiently.

[2]This is also the case with audio; however, for simplicity of discussion in this paper, we restrict ourselves to video only by keeping the audio bit rate constant.

[3]Since the buffering time is set to a finite value, packets arriving late can be either discarded or output with jitter; in both cases, the temporal quality of the output stream deteriorates.

We can find few reports on interactive audiovisual applications over IP networks from a QoE point of view. ITU–T Rec. G.1070 [13] describes a computational model for interactive videophone applications over best–effort IP networks as a QoE/QoS planning tool; it incorporates network, application and terminal quality parameters such as video bit rate, video frame rate, speech/video packet–loss rate, and speech/video delay (end–to–end one–way delay). The playout buffering time is regarded just as a part of the speech/video delay and therefore not treated in any explicit way. Also, the task assumed in the model is only "free conversation", which does not exhibit strong interactivity. Consequently, G.1070 does not mention the tradeoff relationship between fidelity and latency.

To the best of the authors' knowledge, only [6] and [14] investigate the subject from a QoE point of view by explicitly showing the tradeoff relationship. However, the two papers suppose best–effort IP networks; they do not examine the effects of the guaranteed bandwidth between terminals and the video encoding bit rate. Therefore, they can give no guidance for QoE enhancement/guarantee of interactive services over bandwidth guaranteed IP networks.

The current study is a first–step trial of obtaining the guidance. Under the condition that the total guaranteed bandwidth for audio and video is kept constant, we set various values of the video encoding bit rate and playout buffering time and examine their effects on QoE for two kinds of task.

III. Bandwidth Guaranteed IP Network

In this paper, we suppose a simplified version of NGN as the bandwidth guaranteed IP network; it has the *transport functions* and *RACF* (*Resource and Admission Control Functions*) [15]. The RACF interacts with the transport functions for QoS related transport resource control including bandwidth reservation and allocation.

For the bandwidth control of the transport functions, this paper assumes a network composed of bandwidth–controllable routers. The RACF instructs the routers on the guarantee of required bandwidth, which is commonly realized with packet schedulers. The packet scheduling algorithm for bandwidth guarantee is typified by *WFQ* (*Weighted Fair Queueing*) and *CBWFQ* (*Class–Based Weighted Fair Queueing*) [16]; they are installed into many of commercially available routers.

In this paper, we examine only the *data transfer phase* of interactive audiovisual communications and focus on a single pair of users for simplicity of discussion. This implies that static bandwidth allocation at the routers is sufficient for our purpose; the session establishment and termination along with bandwidth reservation is outside the scope of this paper[4].

IV. Method of QoE Assessment

We require some method for assessing QoE of interactive audiovisual communications, especially how to structure conversational tasks for interactive applications. ITU-T Rec. P.920 gives a guideline for it; the tasks should be structured so as to represent the applications as regards of the *rate of information exchange* and the *degree of audio and video signal utilization*. On the basis of this guideline, P.920 illustrates the five tasks mentioned in Sec. II. Referring to the tasks, this paper develops two tasks suitable for our experiments, which are described in the next section.

As in the authors' previous studies on QoE [17], [18], the QoE metric adopted in this paper is the *psychological scale*, which is an *interval scale* in the *psychometric methods* [19], [20]. Note that the QoE metric 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*, which is composed of the *rating–scale method* and the *law of categorical judgment* [17]. The rating–scale method specifies how the subjective measurement is made on *stimuli*, which are audio-video streams output at the receiver in our case; a *subject* (i.e., an *assessor*) classifies the stimuli into a certain number of categories (e.g., five) each assigned an integer score (typically 5 through 1 in order of highly perceived quality). Simply taking an average of the scores for a stimulus over all subjects gives MOS. Instead, we apply the law of categorical judgment to the measurement results by the rating–scale method in order to obtain the interval scale.

Since the law of categorical judgment is based on several assumptions, we have to confirm the goodness of fit for the obtained scale. For a test of goodness of fit, we conduct *Mosteller's test* [19]. Once the goodness of fit has been confirmed, we use the interval scale as the psychological scale.

V. Experimental Method

This section explains an experimental method for examining the effects of the four design parameters on QoE of interactive audiovisual IP communications. We suppose a pair of subjects communicating interactively by means of audio–video transmission which competes with random data traffic. We first present an experimental system and then define two tasks to be assessed along with a method of measuring QoE.

A. Experimental system

Figure 1 shows the configuration of the experimental system; it consists of six PC's (Terminal 1, Terminal 2, Load Sender 1, Load Sender 2, Load Receiver 1, and Load Receiver 2), two bandwidth controllable routers (Routers 2 and 3: Cisco System's 7301), and two ordinary routers (Routers 1 and 4: RiverStone's RS3000).

Each Terminal transmits an audio stream and the corresponding video stream as two separate UDP streams to the other Terminal. A real–time H.264 video encoding board (DSP Research Inc.) equipped with a video camera has been installed into each Terminal along with a microphone and headphones. The nominal error ratio of the average encoding

[4]This is the reason why we do not suppose the *SCF* (*Service Control Functions*) explicitly in this paper.

bit rate of the board is less than 10 %. The operating system for the Terminals is Linux (Fedora 8).

Load Sender 1 and Load Sender 2 are senders of UDP load traffic; they generate UDP datagrams of 1472 bytes each at exponentially distributed intervals. Load Receiver 1 and Load Receiver 2 are the corresponding receivers.

The links between the routers and ones between a router and a PC are all full duplex Ethernet channels. The transmission rate of the link between Router 2 and Router 3 is 10 Mb/s, while the others are 100 Mb/s. Note that the link of 10 Mb/s becomes a bottleneck.

The bandwidth control is exerted between Router 2 and Router 3; *LLQ* (*Low Latency Queueing*) is adopted as the packet scheduling algorithm. In LLQ, we can set a *PQ* (*Priority Queueing*) class and CBWFQ classes [16]: Each class has a dedicated buffer. Packets in the PQ class are served with high priority until its buffer becomes empty; then, the server goes down to the CBWFQ classes. The PQ class is usually utilized for VoIP, which requires low delay; in our experiment, this class is also assigned to the audio streams. The video streams and the UDP load traffic are treated as two separate CBWFQ classes. The guaranteed bandwidth for audio is kept constant at 90 kb/s, while that for video is set to either 1, 2, 3, or 4 Mb/s[5]. The remaining bandwidth is allocated to the UDP load traffic, whose average transmission rate is set to the allocated bandwidth so that the link between Router 2 and Router 3 can be congested.

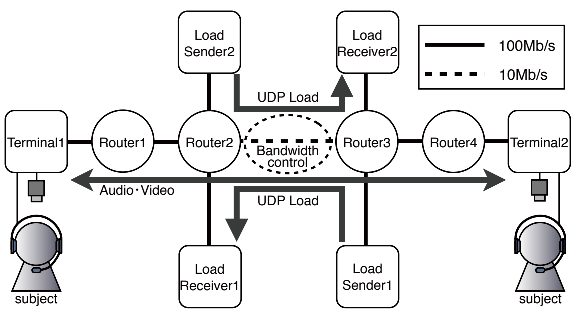

Fig. 1. Configuration of the experimental system

Tables I and II give the specification of audio and that of video, respectively. We refer to the transmission unit at the application layer as a *media unit* (*MU*); in this paper, we define a constant number of audio samples as an audio MU and a video frame as a video MU; a timestamp is attached to each MU when it is generated. An audio MU composes a single IP datagram, while a video MU is usually divided into two or more IP datagrams. The encoding bit rate of audio is kept constant at 64 kb/s in all experimental runs. On the other hand, the video encoding bit rate is set to various values according to the guaranteed bandwidth (GB) for video as shown in Table II; it takes a value between the guaranteed bandwidth minus 0.3 Mb/s and the guaranteed bandwidth in steps of 0.1 Mb/s.

[5]Strictly speaking, this is the minimum guaranteed bandwidth because of the CBWFQ scheduling.

We have adopted only "I" as the picture pattern for fast encoding and for simplicity of discussion as a first step of this study.

TABLE I
SPECIFICATION OF AUDIO

encoding scheme	ITU-T G.711 μ–law
bit rate [kb/s]	64 (constant)
average MU interval [ms]	40
MU rate [MU/s]	25

TABLE II
SPECIFICATION OF VIDEO

encoding scheme	H.264
image size [pixel]	704 × 480
picture pattern	I
encoding bit rate [Mb/s]	when GB=1: 0.7,0.8,0.9,1.0 when GB=2: 1.7,1.8,1.9,2.0 when GB=3: 2.7,2.8,2.9,3.0 when GB=4: 3.7,3.8,3.9,4.0
average MU interval [ms]	40
average MU rate [MU/s]	25
monitor	17 inch LCD
monitor resolution [pixel]	1280 × 1024

Unless all IP datagrams of a video MU are received correctly, the MU is not assembled; in other words, no error concealment is made, and therefore, the MU is skipped (i.e., pure video frame skipping). This is because the pure video frame skipping is usually the best output strategy of video with picture pattern I from a QoE point of view [18].

Each Terminal carries out playout buffering control on received MU's in each stream (i.e., audio or video) independently; this means no inter–stream media synchronization control is exerted in our experiment. We have chosen this way of the control so that end–to–end delay characteristics can explicitly affect the quality of inter–stream media synchronization, which influences QoE. We take nine values of the playout buffering time: 40 ms, 80 ms, 120 ms, 160 ms, 200 ms, 300 ms, 400 ms, 500 ms, and 1000 ms. The same value is chosen at both Terminals in an experimental run.

An MU in each stream is output or discarded as follows. The MU received first at the application layer is output the playout buffering time later. Each of the following MU's is output at the *target output time* if it is received before that time; otherwise, it is discarded. The target output time of an MU is given by adding the timestamp difference between the MU and the first MU to the output time of the first MU.

B. Tasks and QoE measurement method

We have designed two tasks for conversational tests, referring to ITU–T Rec. P.920. In the design, we have intended to assure a pair of subjects that approximately the same amount of information can be exchanged between the subjects in a predetermined interval for an experimental run.

We specify the two tasks as follows.

- *task 1*: One subject selects a number randomly from 1 through 5 and reads the numbers from 1 to the selected number aloud. Immediately after the reading, the other subject reads the same numbers aloud. This interaction is repeated during a predetermined interval.

- *task 2*: One subject selects a number randomly from 1 through 5 and reads the numbers from 1 to the selected number aloud, *clapping once for each number*. Immediately after the reading, the other subject reads the same numbers aloud, clapping once for each number. This interaction is repeated during a predetermined interval.

Note that task 1 is audio–dominant since the subjects exhibit only low motion, whereas task 2 has a visual impact on QoE because of clapping.

After an experimental run, each subject assesses the stimuli, which are audio–video streams output at his/her own Terminal during the run, by the rating–scale method, where the *Absolute Category Rating* with the following five–level quality scale is used: "excellent" assigned score 5, "good" 4, "fair" 3, "poor" 2 and "bad" 1. These categories are referred to as Category 5 through Category 1 according to the score.

The subjects are Japanese male and female students in their twenties; the number of the subjects is 24. Each pair of subjects assessed 288 stimuli because of 16 video encoding bit rates, nine values of the playout buffering time, and two tasks. The interval of an experimental run (i.e., the stimulus presentation time) was set to 30 seconds; in addition, we needed approximately 30 seconds for system setup and voting for each experimental run. Thus, it took about 360 minutes including break time for a subject to assess all the stimuli [6].

VI. Experimental Results

In this section, we first calculate the psychological scale for QoE assessment and then examine the effects of the video guaranteed bandwidth, video encoding bit rate, playout buffering time and task on QoE.

The psychological scale was calculated as an interval scale as in [17]. We applied the law of categorical judgment to all the classification results of the 288 stimuli together so that we can compare the interval scales for the stimuli on the same basis. We carried out Mosteller's test for a test of the goodness of fit of the interval scale. We then found that the test with a significance level of 0.05 can reject the hypothesis that the observed value equals the calculated one. Therefore, we checked stimuli which give large errors of Mosteller's test to find 38 ones. Removing the 38 stimuli, we saw that the hypothesis cannot be rejected. Consequently, for the 250 $(= 288 - 38)$ stimuli, we can consider the interval scale as the psychological scale.

Since we can select an arbitrary origin in an interval scale [17], we set the minimum value of the psychological scales for the 250 stimuli to the origin. Under this condition, we also calculated the lower boundaries of the categories and found 3.628 for Category 5, 2.705 for Category 4, 1.896 for Category 3, and 0.986 for Category 2.

[6]During each experimental run, we also measured application–level QoS and end–to–end–level QoS parameters.

Figures 2 through 5 plot the psychological scales for task 1 at the four values of the video encoding bit rate as a function of the playout buffering time for the video guaranteed bandwidth 1 Mb/s, 2 Mb/s, 3 Mb/s and 4 Mb/s, respectively. Figures 6 through 9 represent the psychological scales for task 2 in the same way. In these figures, the lower boundaries of the categories are also plotted as straight dotted lines parallel to the abscissa. It should be noted that the results of the stimuli removed by Mosteller's test are not shown in the figures.

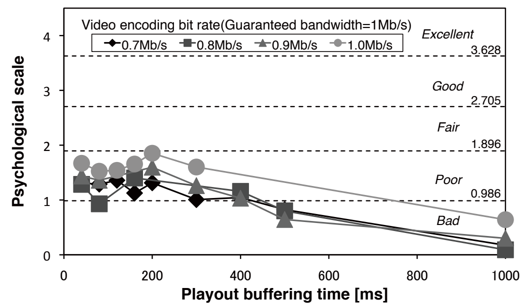

Fig. 2. Psychological scale versus playout buffering time (task 1, video guaranteed bandwidth: 1 Mb/s).

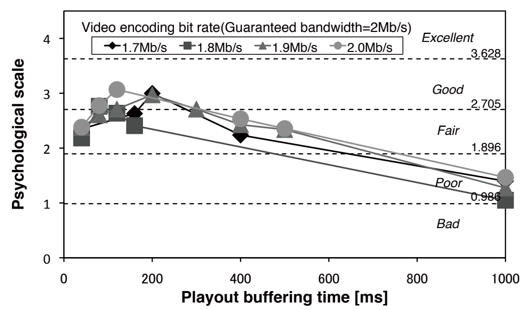

Fig. 3. Psychological scale versus playout buffering time (task 1, video guaranteed bandwidth: 2 Mb/s).

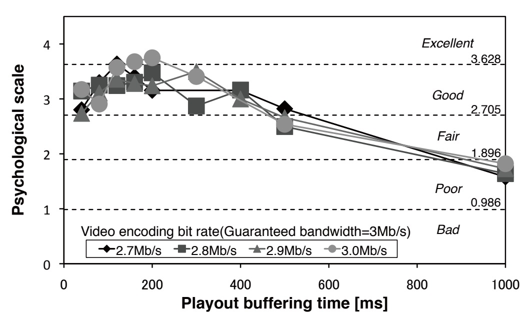

Fig. 4. Psychological scale versus playout buffering time (task 1, video guaranteed bandwidth: 3 Mb/s).

In Figs. 2 through 9, we make the following observations:

(a) For each pair of the video guaranteed bandwidth and video encoding bit rate, there exists a playout buffering time which maximizes the psychological scale value; we call this value the *individual maximum*.

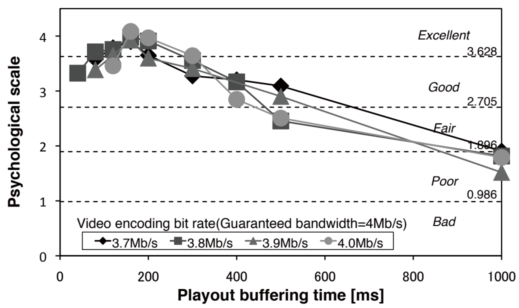

Fig. 5. Psychological scale versus playout buffering time (task 1, video guaranteed bandwidth: 4 Mb/s).

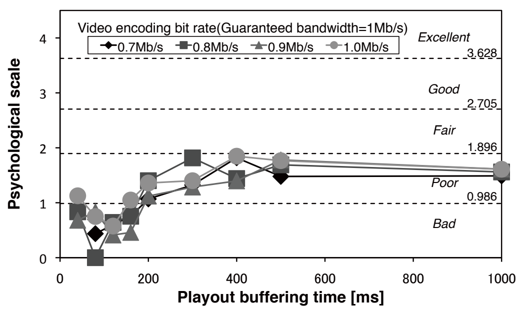

Fig. 6. Psychological scale versus playout buffering time (task 2, video guaranteed bandwidth: 1 Mb/s).

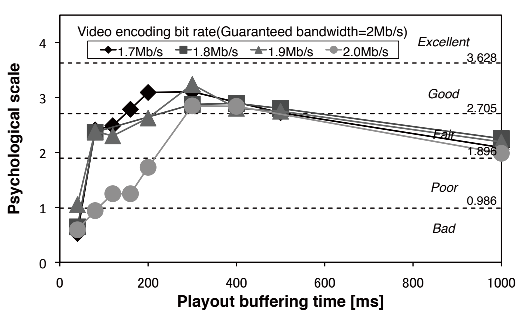

Fig. 7. Psychological scale versus playout buffering time (task 2, video guaranteed bandwidth: 2 Mb/s).

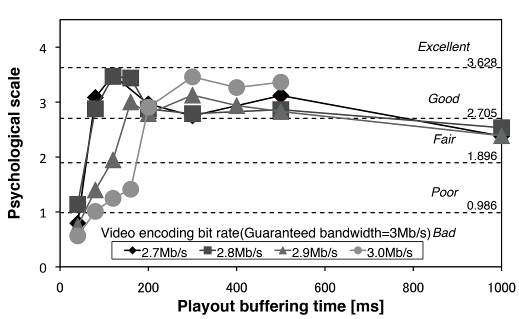

Fig. 8. Psychological scale versus playout buffering time (task 2, video guaranteed bandwidth: 3 Mb/s).

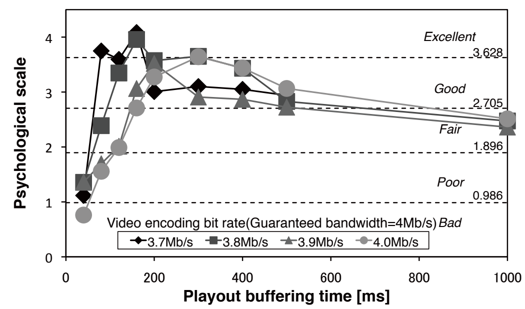

Fig. 9. Psychological scale versus playout buffering time (task 2, video guaranteed bandwidth: 4 Mb/s).

(b) For a given amount of the video guaranteed bandwidth, the individual maximum value depends on the video encoding bit rate; that is, one rate among the four takes the largest of the individual maximum values, which we call the *total maximum* value for the video guaranteed bandwidth.

(c) As the video guaranteed bandwidth increases, the convex curves of the psychological scale value become steeper, and the total maximum value becomes higher.

(d) For a given amount of the video guaranteed bandwidth, task 1 and task 2 have approximately the same total maximum.

(e) In order to achieve "good" (Category 4) or "excellent" (Category 5), we need the video guaranteed bandwidth of 3 Mb/s or 4 Mb/s as seen from Figs. 4, 5, 8 and 9. In these cases, the side of the convex curve corresponding to the total maximum for task 2 is steeper than that for task 1.

We can easily understand the reasons for Observations (a) through (d), while Observation (e) needs some consideration. This is due to the difference in the task property; as already mentioned in Subsection V-B, task 1 is audio–dominant in the sense that the video shows only low motion, whereas the video motion of task 2 is high because of clapping, which gives a visual impact on QoE. However, note that the difference in video motion between the two tasks scarcely produces the difference in the actual video encoding bit rate for a specified encoding bit rate; this is due to the picture pattern we have selected in the experiment, namely, picture pattern I.

Figures 10 and 11 show the video MU loss ratio[7] for task 1 and task 2, respectively, when the video guaranteed bandwidth is 4 Mb/s. From these figures, we can confirm that the video MU loss ratios for the two tasks hardly differ from each other.

Then, why have we observed different behavior of the psychological scale as a function of the playout buffering time in the two tasks? This is particularly noticeable when QoE is high (i.e., "good" and "excellent"); so, let us focus on Figs. 5 and 9, for instance.

Comparing the two figures, we first find that the individual maximum values in task 1 do not depend on the video encoding bit rate so much, while those in task 2 differ from

[7]The MU loss ratio is defined as the ratio of the number of MUs not output at the receiver to the number of MUs transmitted by the sender. The audio MU loss ratio in this experiment is zero in all cases.

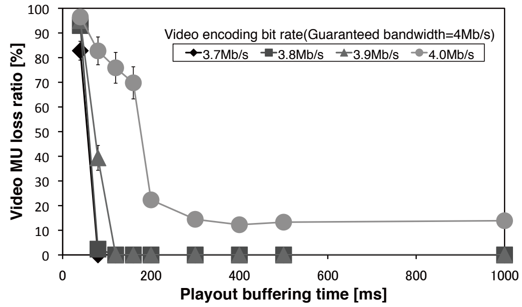

Fig. 10. Video MU loss ratio versus playout buffering time (task 1, video guaranteed bandwidth: 4 Mb/s).

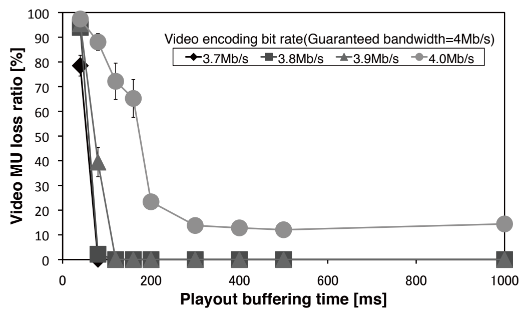

Fig. 11. Video MU loss ratio versus playout buffering time (task 2, video guaranteed bandwidth: 4 Mb/s).

each other apparently. The reason in task 1 is the audio–dominant property. Even if the video MU loss ratio is rather high, especially when the video encoding bit rate is 4 Mb/s (see Fig. 10), the subjects scarcely cared it since their attention was mainly drawn to the audio. On the other hand, the psychological scale value for task 2 is affected by the video MU loss ratio. As seen from Figs. 9 and 11, the total maximum is achieved around the minimum value of the playout buffering time which realizes no MU loss. Note that in the case of the 4 Mb/s, which exhibits nonzero floor values of the video MU loss ratio, the individual maximum is achieved at the minimum playout buffering time which takes the floor value.

From the observations so far, we see that in order to provide high QoE, the video encoding bit rate and playout buffering time should be carefully chosen; in particular, when the task is such that it draws the attention of the user to both audio and video, QoE is sensitive to the playout buffering time.

VII. Conclusions

We examined how QoE is affected by the video guaranteed bandwidth, video encoding bit rate, playout buffering time and kind of task under the condition that the audio bit rate is kept constant. As a result, we first observed that a certain pair of the video encoding bit rate and playout buffering time maximizes QoE for a given amount of the video guaranteed bandwidth, which determines the maximum value of the QoE metric. We then found that the playout buffering time should be carefully chosen particularly when the content of the task depends on both audio and video.

Since this paper has focused only upon the data transfer phase, we should extend our consideration to session setup and termination phases as well. This leads to an extension of QoE guarantee architecture *GPSQ* (*Guarantee of Psychologically Scaled Quality*), which the authors proposed previously [21], for interactive audiovisual applications.

As future work, we should investigate cases with P and B picture patterns and the optimum bandwidth allocation between video and audio. It is important to study different tasks from those in this paper and the effects of packet scheduling algorithms. How video error concealment affects QoE in interactive audiovisual IP communications is interesting.

Acknowledgment

The authors thank Dr. Toshiro Nunome for his valuable discussions. This work was supported by the Grant-In-Aid for Scientific Research of Japan Society for the Promotion of Science under Grant 21360183 and by The Telecommunications Advancement Foundation.

References

[1] C.–S. Lee and D. Knight, “Realization of the next–generation network,” *IEEE Commun. Mag.*, vol.43 No.10 pp.34–41,Oct. 2005.
[2] ITU-T Rec. Y.1540, “Internet protocol data communication service – IP packet transfer and availability performance parameters,” Dec. 2002.
[3] ITU-T Rec. Y.1541, “Network performance objectives for IP–based services,” Feb. 2006.
[4] S. Tasaka and Y. Ishibashi, “Mutually compensatory property of multimedia QoS,” in *Conf. Rec. IEEE ICC2002*, pp. 1105–1111, Apr./May 2002.
[5] ITU-T Rec. G.100/P.10 Amendment 1, “Amendment 1: new appendix I definition of Quality of Experience (QoE),” Jan. 2007.
[6] Y. Ito and S. Tasaka, “Tradeoff relationship between fidelity and latency in interactive audio–video applications over IP networks,” *IEICE Trans. on Commun.*, vol. E90–B, no. 5, pp. 1112–1121, May 2007.
[7] N. Laoutaris and I. Stavrakakis, “Intrastream synchronization for continuous media streams: A survey of playout schedulers,” *IEEE Network Magazine*, vol.16, no.3, pp.30–40, May 2002.
[8] ITU-T Rec. P.920, “Interactive test methods for audiovisual communications,” May 2000.
[9] R. Ramjee, J. Kurose, D. Towsley and H. Schulzrinne, “Adaptive playout mechanisms for packetized audio applications in wide–area networks,” in *Proc. IEEE INFOCOM1994*, pp.680–688, June 1994.
[10] M. Kalman, E. Steinbach and B. Girod, “Adaptive media playout for low–delay video streaming over error–prone channels,” *IEEE Trans. Circuits Syst. Video Technol.*, vol.14, no.6, pp.841–851, June 2004.
[11] K. Yamagishi and T. Hayashi, “Opinion model for estimating video quality of videophone services ,” in *Conf. Rec. IEEE GLOBECOM2006*, Nov. 2006.
[12] S. Tasaka, T. Nunome and Y. Ishibashi, “Live media synchronization quality of a retransmission-based error recovery scheme,” in *Conf. Rec. IEEE ICC2000*, pp. 1535–1541, June 2000.
[13] ITU-T Rec. G.1070, “Opinion model for video-telephony applications,” April 2007.
[14] Y. Ito, S. Tasaka and Y. Fukuta, “Psychometric analysis of the effect of buffering control on user–level QoS in an interactive audio–visual application,” in *Proc. ACM NRBC ’04* , pp. 2–10, Oct. 2004.
[15] ITU-T Rec. Y.2111, “Resource and admission control functions in next generation networks,” Sept. 2006.
[16] S. Vegesna, “IP Quality of Service,” Cisco Press,2001.
[17] 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.
[18] S. Tasaka, H. Yoshimi, A. Hirashima and T. Nunome, “The effectiveness of a QoE–based video output scheme for audio–video IP transmission,” in *Proc. ACM Multimedia2008*, pp.259–268, Oct. 2008.
[19] J. P. Guilford, *Psychometric methods*, McGraw-Hill, N. Y., 1954.
[20] J. C. Nunnally and I. H. Bernstein, *Psychometric theory, Third edition*, McGraw-Hill, N. Y., 1994.
[21] S. Tasaka, Y. Ito, H. Yamada and J. Sako, “A method of user–level QoS guarantee by session control in audio–video transmission over IP networks,” in *Conf. Rec. IEEE GLOBECOM2006*, Nov. 2006.