PAPER Special Section on Quality of Communication Networks and Services

QoE Assessment of Multi-View Video and Audio IP Transmission

Erick JIMENEZ RODRIGUEZ[†a)], ***Student Member***, **Toshiro NUNOME**[†], ***Member***, ***and*** **Shuji TASAKA**[†], ***Fellow***

SUMMARY In this paper, we discuss QoE (Quality of Experience) requirements for MVV (Multi-View Video) and audio transmission over IP networks and study the effect of the playout buffering time, contents and viewpoint change interfaces on the QoE and user's behavior. Unlike previous works, which mainly discuss MVV transmission from aspects of video codecs, we study MVV and audio transmission under various IP traffic and delay conditions by experiment. We compare two schemes: a scheme that the user watches from a single viewpoint and the one that he/she can choose one viewpoint from many ones. As a result, we show that the users prefer the scheme where they can choose one viewpoint from many ones. We have found that when using proper buffering time, the users feel faster viewpoint changes; it improves their satisfaction compared to that when they watch on a single viewpoint. We have also noticed that the user pays more attention to the degradation of the video when watching on a single viewpoint. We have observed that the users tend to change the viewpoint more frequently in light traffic and low delay.

key words: *MVV, multi-view video, QoE, QoS, viewpoint change, IP network, IPTV*

1. Introduction

Television has realized a human's dream of seeing a distant world in real time. It has changed through time; for example, we have witnessed changes in video quality, audio quality, and screen size. Although many improvements have been made in television, the users can watch only the same viewpoint given by the sender even if they move their viewpoints in front of the display.

Because of this limitation in its functionality, *MVV (Multi-View Video)* [1], where the user can choose one video from multiple video streams of the same event, has been under development. In addition, *Free Viewpoint Television (FTV)* [2] and *3DTV* [3], which make use of MVV as their base system, have also been under investigation. In recent years, these technologies have been attractive owing to their enhanced viewing experience.

MVV can be used not only by broadcasting, but also there is a possibility of using an MVV system over IP networks. In this paper, we focus on MVV over the IP networks.

There are many challenges when implementing MVV systems. One of these challenges is how a large amount of data should be streamed on the network with limited capacity. Because of this reason, there are several works which focus on compression algorithms for MVV (e.g., [4] and [5]).

As for the method to measure the performance of an MVV system, most of the related works employ the throughput and *PSNR (Peak Signal to Noise Ratio)*, which measures spatial quality of video.

On the other hand, as the users are going to use the MVV system, it is also important to consider their opinion in order to provide better service. However, there has been still relatively few researches on *QoE (Quality of Experience)* assessment with MVV systems. QoE represents the overall acceptability of an application or service, as perceived subjectively by the end-users [6].

References [7] and [8] have performed a user study of interactive MVV systems. These references have assessed the effect of different features, such as viewpoint switching, frozen moment and viewpoint sweeping on the MVV system and the effect of the contents on the user's behavior. However, they do not consider audio; in real applications, audio and MVV are transmitted together.

At the same time, as the MVV system uses the IP network, problems such as packet loss and delay can arise. For this reason, it is also important to perform a systematic QoE assessment when delay and packet loss are present in the transmission. However, Refs. [7] and [8] do not perform systematic QoE assessment considering these two situations.

For IP transmission of traditional single-view television, ITU-T Rec. G.1080 [9] defines QoE requirements for IPTV services. However, no recommendation is available for MVV IP transmission. Therefore, we need to clarify QoE requirements for MVV and audio IP transmission.

In this paper, we first list up QoE requirements for MVV and audio IP transmission. Then, we perform subjective experiment to assess the QoE according to the criteria derived from the requirements. We compare the QoE when the assessors watch only from one viewpoint and that when they can choose one viewpoint from many ones. In this paper, we refer to the former scheme as "Fixed View" and the latter as "Selective View."

The rest of the paper is structured as follows. Section 2 describes QoE requirements. Section 3 outlines the system model. Section 4 discusses the conditions of the experiment we performed. We present the results of the experiment in Sect. 5, and Sect. 6 concludes this paper.

Manuscript received October 23, 2009.
Manuscript revised January 28, 2010.

[†]The authors are with the Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Nagoya-shi, 466-8555 Japan.
a) E-mail: erickjim@inl.nitech.ac.jp

DOI: 10.1587/transcom.E93.B.1373

2. QoE Requirements for MVV and Audio IP Transmission

In this section, we discuss QoE requirements for MVV and audio IP transmission. At first, we introduce *QoS (Quality of Service)* parameters. Later, we talk about the QoE metrics and finally show factors affecting QoE.

As practical examples of MVV applications, we can image a soccer match, a concert, etc. In these examples, cameras are placed around objects because the users want to focus on the objects. At the same time, the objects may move round a particular area, e.g., a stage, a field and a stadium.

Another important factor that should be considered is the audio. In general, sound is generated according to movements or actions of the objects. Therefore, in this study, we employ a dog doll that moves and barks while moving the head as a simple example of such a situation.

When showing the object to the users, we expect them to be interested in changing the viewpoint according to the object's movement. In the case when the dog doll is not facing the camera, the user will be interested in changing the viewpoint to see the face of the doll when it barks. Thus, the MVV system will satisfy the user with the ability of viewpoint change.

We focus on the viewpoint change function as the main feature of MVV systems. Unlike previous works, which mainly discuss MVV transmission from aspects of video codecs, we study the QoE of MVV and audio IP transmission from a network point of view. For this reason, we employ a simple MVV system as a first step to assess the QoE of MVV systems.

As our main objective is QoE assessment of MVV systems, we introduce several subjective QoE metrics. In addition, we also consider objective parameters that can be reflected on QoE. Among these parameters, we pick up the application-level QoS and the user's behavior. However, it is not clear how the application-level QoS and the user's behavior may be reflected on QoE. This relationship is not so simple. Thus, we perform assessment of the application-level QoS and the user's behavior to investigate how they are reflected on QoE.

One of the key elements involved in validating QoE in MVV service is how quickly users can change the viewpoint; this is referred to as *viewpoint change delay* at the application-level. There are some published works on the relationship between user perception and computer response times over a wide variety of application types (e.g., [10]). However, they are not directly applicable to the MVV services.

In order to study the user's behavior of an MVV system, we introduce the following parameters. Since the user of an MVV system can change the viewpoint as he/she like, the user's behavior of viewpoint change can vary according to the different conditions, such as load traffic and delay. For this reason, we employ *average number of viewpoint changes* and *average watching time on each camera* at the application-level.

Regarding the QoE metrics, it is important to evaluate the response of the viewpoint change as mentioned before. Therefore, in this paper, we employ "Quickness of the viewpoint change" as one of the QoE metrics.

As in traditional single-view video and audio IP transmission, video smoothness and media synchronization quality are also indispensable components of QoE. Because of this, we employ "Smoothness of the video" and "Synchronization of the video and audio" as two of the QoE metrics.

Since there are several factors that can affect the QoE of the MVV system, it is also important to consider all of them into one criterion that can depict how satisfactory the user is for the system. Therefore, we use "Overall evaluation" as one of the QoE metrics.

An important factor affecting QoE when we use an MVV system is the *playout buffering time* on the client. When the buffering time is short, the client cannot absorb the delay jitter. This will increase the packets that will not be in time for output. On the other hand, if the buffering time is very large, the viewpoint change delay will become larger; it degrades QoE. Because of this, the buffering time is important to be configured properly.

In addition, even in network services, the user interface is an important factor for the user's satisfaction. The user can feel more comfortable with a friendly and intuitive control interface.

At the same time, the usage of the MVV systems can change according to the content the user is watching. This is because the user can change the viewpoint in a different way according to the content.

In this paper, we examine the effect of the playout buffering time on the QoE of MVV and audio IP transmission by means of the criteria discussed above. Also, as the user interface and content can affect the QoE, we employ two user interfaces and two contents.

3. MVV System

An MVV system consists of one server and at least one client that are connected to an IP network. At the same time, several cameras are connected to the MVV server. The server captures the video of each camera. At the same time, the audio is captured by using at least one microphone.

The server can send one audio-video stream of a single viewpoint or it can send multiple audio-video streams of the viewpoints to the client. In the case of sending only one viewpoint to the clients, the users must wait more time in order to see the new viewpoint since the clients must send a request for viewpoint change to the server first. However, as the server is sending only one audio and video stream, the amount of data that is sent through the network is considerably low. On the other hand, in the case of sending multiple viewpoints of the cameras simultaneously, the viewpoint can be immediately changed at the clients. However, the amount of data that is sent through the network is consider-

ably high and can vary depending on the number of cameras connected to the server. For this reason, further investigation on this matter needs to be done.

In this paper, we focus on sending only an audio stream and a video stream of the selected view to the client at a time as a first step of research on MVV and audio IP transmission.

The client can choose one viewpoint from the cameras that are connected to the server. In order to do this, the user notifies the server of a desired viewpoint by using the viewpoint change interface. This request is sent to the server. When the server receives the request, it changes the viewpoint and start sending the audio-video stream of the new viewpoint. The transmission lasts until when the server receives another request for viewpoint change or when the session ends.

4. Experiment

In this section, we will explain the experiment's details. In the following subsections, we discuss the experimental system, experimental conditions, application-level QoS parameters, user's behavior, and QoE metrics.

4.1 Experimental System

Figure 1 shows the network topology used in the experiment. MS is the server of the MVV application, and MR is the client. Four cameras are connected to the server.

The server captures the video of each camera. At the same time, the audio is captured by a microphone. The server sends the audio and video of a viewpoint to the client by using UDP packets. The client receives these packets and outputs the audio and video decoded from them. The client can choose one viewpoint from the four cameras by sending a request with a UDP packet.

In this paper, we refer to the transmission unit at the application-level as a *Media Unit (MU)*; we define a video frame as a video MU and a constant number of audio samples as an audio MU. An audio MU is transmitted as a UDP packet. A video MU can be transmitted as multiple UDP packets. If all the packets of an MU are not correctly received in time for output, the MU is not output. We can apply some error concealment techniques; they may improve QoE. The effect of error concealment in MVV and audio IP transmission is one of our future studies.

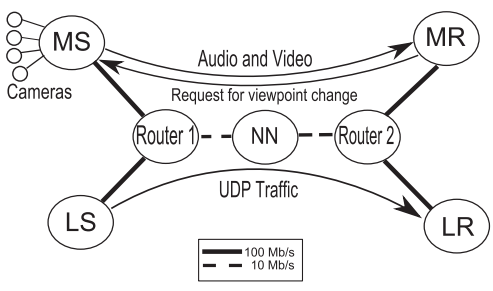

Fig. 1 Network topology.

On the other hand, LS is the server of the background traffic, and LR is the client. Both router 1 and router 2 are Riverstone's RS3000. At the same time, NN, which is a PC, is laid out between the routers. NN delays packets going through routers 1 and 2 by using NISTNET [11]. By adding this delay, we can see the effect of network delay on the QoE in the MVV system.

4.2 Experimental Conditions

We discuss the experimental conditions of our assessment in this subsection. At first, we introduce the contents. Then, we talk about the user interfaces and finally show the media specifications.

Figures 2 and 3 show the position of the cameras connected to MS with the two contents that were employed in this assessment. We refer to the content in Fig. 2 and that in Fig. 3 as "Content 1" and "Content 2," respectively.

We used two different dog dolls that move with battery for the two contents. The way one dog doll moves is different from the other one. They move inside the delimitation area and cannot go outside. Because of this, they were always inside the focus of the four cameras during the assessment.

In Content 1, the doll moves a few steps forward and barks with moving its tail while walking backwards. Later, it starts to walk forward again but in a different direction.

On the other hand, in Content 2, the doll walks for-

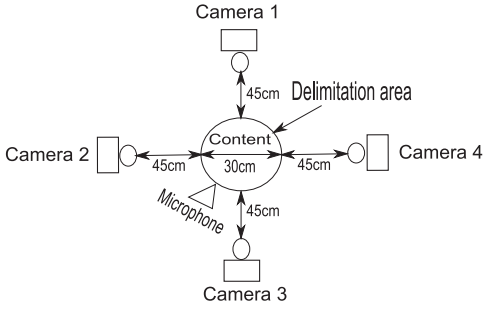

Fig. 2 Position of the cameras with Content 1.

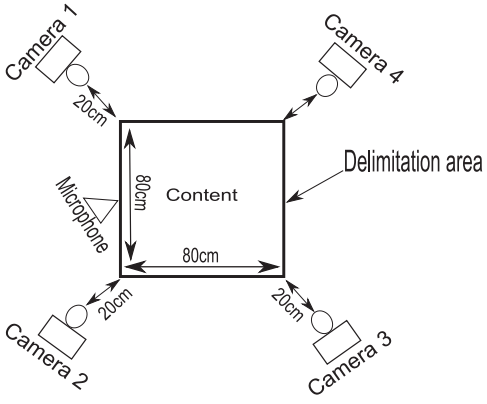

Fig. 3 Position of the cameras with Content 2.

ward and stops. Then, it barks for a little moment and jumps backward in the air. Later, it walks forward changing the direction.

As the doll of Content 2 jumps in the air, the delimitation area for Content 2 is larger than that for Content 1 in order for the doll not to fall down or go outside the delimitation area.

The user interface may affect the user's behavior and also may affect the QoE of the system as well. However, there are no standardized recommendations of interfaces for MVV systems. Simple and intuitive interfaces can be suitable for the user to change the viewpoints. Therefore, we employed simple two interfaces that can be used with a mouse as a first step of research in order to analyze the effect on the user's behavior and the QoE.

We used two different user interfaces only with Content 1. They are shown as a small window on the display. The user can move this window to a desired position and can change the viewpoint by using the mouse. With the first one, the user can change the viewpoint by selecting the number of the camera, as shown in Fig. 4. The second one lets the user change the viewpoint by using the direction of the camera position, as shown in Fig. 5. In this paper, we refer to the first interface as "Interface 1" and to the second one as "Interface 2." For Content 2, we applied only "Interface 2."

In addition, Fig. 6 shows a deployment of our system with Content 1. Figure 7 shows an screenshot of the MVV client application. The specifications of the audio and video are shown in Table 1.

We employed a simple scheme of playout buffering control to absorb network delay jitter and set the buffering time to 60 ms, 100 ms, and 140 ms.

While the server sends the video and audio to the client, LS generated UDP packets of 1472 bytes each with exponentially distributed interval and sends them to LR. The average bit rate was set to 7.2 Mb/s, 7.4 Mb/s, and 7.6 Mb/s. The delay in the computer NN was 0 ms or 100 ms.

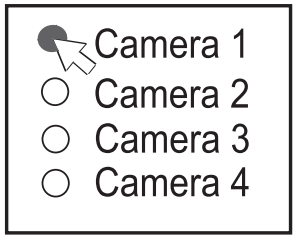

Fig. 4 User interface with camera numbers (Interface 1).

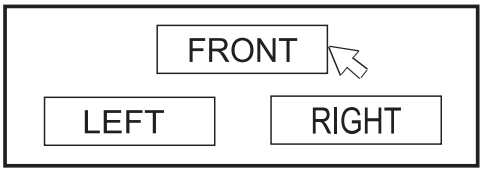

Fig. 5 User interface with camera directions (Interface 2).

For our experiment, we employed 17 male students of between 22 and 27 years old as assessors.

4.3 Application-Level QoS Parameters

We employ the *viewpoint change delay* and *MU loss ratio* as application-level QoS parameters.

The viewpoint change delay is defined as the time in seconds from the moment the destination sends a request for viewpoint change until the instant a new viewpoint is output at the destination.

The MU loss ratio is defined as the ratio of the number of MUs not output to the total number of MUs transmitted.

4.4 User's Behavior

In order to assess the user's behavior, we use the *average number of viewpoint changes* and the *average watching time on each camera*.

The average number of viewpoint changes is defined as the average of the number of viewpoint changes during an

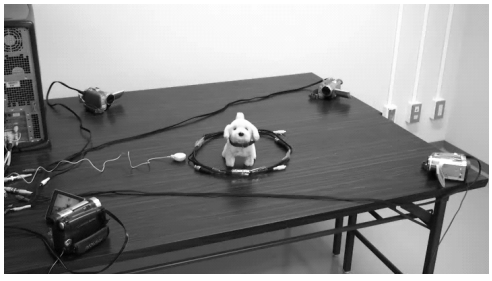

Fig. 6 A deployment of our system with Content 1.

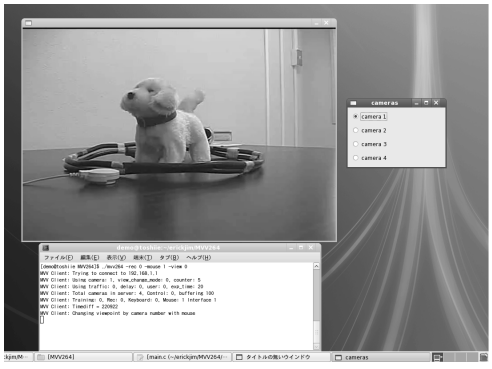

Fig. 7 Screenshot of the MVV client application with Content 1 and Interface 1.

Table 1 Media specifications.

	Audio	Video
Coding scheme	G.711 μ-law	H.264
Image size [pixels]	-	704×576
Picture pattern	-	I
Coding bit rate [kb/s]	64	2000
Average MU rate [MU/s]	25	25
Media duration [s]	20	

experiment.

The average watching time on each camera is defined as the average time in seconds when a camera's video is displayed.

4.5 QoE Metrics

In this paper, we do not use MOS (Mean Opinion Score), which is widely used in subjective assessments. This is because MOS is an ordinal scale where the integer assigned to the categories only has a greater-than-less-than relation between them. Instead, we express QoE in terms of the *interval scale*.

The interval scale can be calculated by one of the psychometric methods [12]. For the calculation of the interval scale, this paper adopts the *method of successive categories*, which is composed of two steps: the *rating scale method* and the *law of categorical judgment*. 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; an assessor classifies the stimuli into a certain number of categories (e.g., five) each assigned an integer (typically 5 through 1 in order of highly perceived quality).

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 [13]. Once the goodness of fit has been confirmed, we use the interval scale as the QoE parameter, which is therefore called the *psychological scale* [14].

As discussed in Sect. 2, every time the assessor sees an audio-video stream, he evaluates it according to the following criteria:

- Smoothness of the video
- Synchronization of the video and audio
- Quickness of the viewpoint change
- Overall evaluation

Each criterion is evaluated to be one of five levels between 1 (the worst case) and 5 (the best case). With "Fixed View," all the criteria except "Quickness of the viewpoint change" are evaluated because the user does not change the viewpoint in this scheme. All the questions about the criteria are written in Japanese.

After one audio-video stream of each scheme has been shown, the user evaluates the quality with an additional criterion in order to express his opinion on which scheme was better under the given condition of delay and traffic. We refer to this criterion as "Fixed View vs. Selective View."

5. Assessment Results

In this section, we will present the experimental results of the application-level QoS assessment, the user's behavior, and the QoE assessment.

5.1 Application-Level QoS

Figures 8 and 9 show measured values of the application-level QoS parameters as a function of the load traffic for the two values of the additional delay at NN and the three values of the buffering time. The legends we use in the results are as follows. "Delay" corresponds to the additional delay in NN and "Buffer" is the playout buffering time that was used in the MVV client. For example, "Delay 0 ms - Buffer 60 ms" means that the additional delay and the buffering time were set to 0 ms and 60 ms, respectively.

Figure 8 depicts the viewpoint change delay for video with "Selective View" of Content 1 with Interface 1 versus the amount of load traffic. Figure 9 shows the MU loss ratio for video with "Selective View" of Content 1 with Interface 1.

In Fig. 8, we can see that with a small buffering time, we can expect small viewpoint change delay with low traffic and no additional delay. Also we notice that the viewpoint change delay increases as UDP load traffic and the buffering time increase.

At the same time, we can notice in Fig. 9 that the MU loss ratio increases as the UDP load traffic increases. This is because when the load traffic increases, the available bandwidth decreases to a point where there is not enough bandwidth to send both load traffic and the audio-video streams.

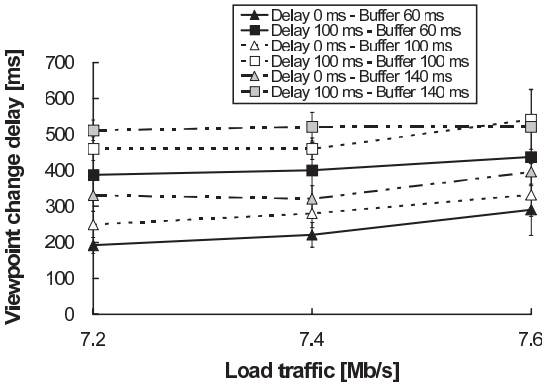

Fig. 8 Viewpoint change delay of Content 1 with Interface 1.

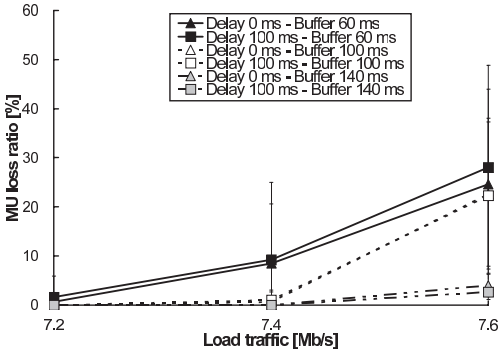

Fig. 9 MU loss ratio for video of "Selective View" of Content 1 with Interface 1.

This causes the packets to be delayed or discarded.

We can also find in Fig. 9 that with the buffering time of 140 ms, the MU loss ratio is considerably low compared to the cases when using buffering time of 60 ms and 100 ms. That is, the MU loss ratio increases as the buffering time decreases. The reason is as follows. As the delay jitter increases when the UDP load traffic increases, the buffering time needs to be longer in order to absorb the jitter that is present during the transmission. If the buffering time is not long enough to absorb the delay jitter, the number of skipped packets increases because they are not in time for output.

In preliminary experiments where the buffering time is larger than 140 ms, we noticed that the MU loss ratio is comparable to that with 140 ms. For this reason, we have not employed larger values than 140 ms in this paper.

We assessed the MU loss ratio and viewpoint change delay for the two contents and the two interfaces. As a result, we found that the contents and the interfaces scarcely affect the application-level QoS.

We have also assessed the MU loss ratio for "Fixed View" and that for "Selective View"; we then noticed that the MU loss ratio of the video for "Fixed View" is almost the same as that with "Selective View." Also, we observed that the two schemes have almost the same MU loss ratio of audio and that the maximum MU loss ratio is almost 1% when the UDP load traffic is 7.6Mb/s. For this reason, it is difficult for the assessor to notice the degradation of the audio.

5.2 User's Behavior

Figure 10 shows the average number of viewpoint changes of Content 1 with Interface 1. The legends we use are the same as in Sect. 5.1.

We can notice that the user changed the viewpoint in a similar way regardless of the delay and buffering time. However, we can find that under the heavy loaded condition, the number of viewpoint changes starts to decrease. This is related with the MU loss ratio; as the MU loss ratio increases, the time to wait for the new viewpoint to be displayed increases. As this time increases, the user may not change the viewpoint in the same way that he has been changing it before. For this reason, the load traffic can affect the user's behavior.

Figure 11 depicts the average number of viewpoint changes for buffering time of 60 ms without additional delay of Content 1 with Interface 1, Content 1 with Interface 2, and Content 2 with Interface 2. The legend we employ shows the content number and interface type used. For example, "Content 1 with Interface 1" means that Content 1 was used with Interface 1 for that result.

As we can notice in Fig. 11, the user changed the viewpoint more frequently when watching Content 1 with Interface 2 than when watching Content 2 with Interface 2. Therefore, the user watched for each viewpoint for Content 2 longer than that for Content 1. This is because the doll of Content 2 kept still for a while before it jumps, so the user waited for the dog of Content 2 to start to move. On the other hand, the dog of Content 1 always moved without stopping, so the user often changed the viewpoint. For this reason, the content's type of movement is related with the user's tendency to change the viewpoint.

In addition, we observe in Fig. 11 that the user changed the viewpoint more times when using Interface 2 than when using Interface 1. This implies that it is easier for the user to change the viewpoint using the direction of the position of the camera (Interface 2) rather than using the camera number (Interface 1).

Figure 12 shows the average watching time on each camera for buffering time of 60 ms and load traffic of 7.2 Mb/s without additional delay of Content 1 with Interface 1, Content 1 with Interface 2, and Content 2 with Interface 2.

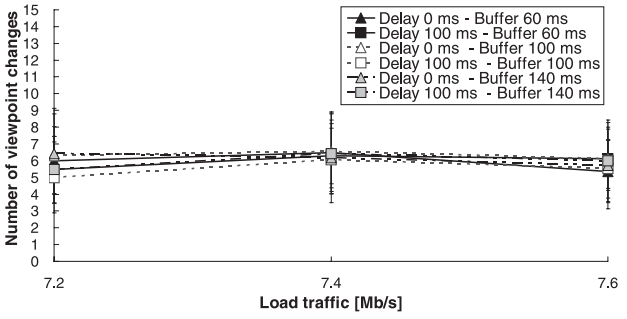

Fig. 10 Average number of viewpoint changes of Content 1 with Interface 1.

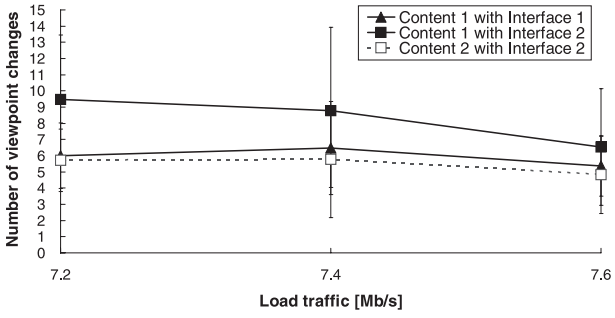

Fig. 11 Average number of viewpoint changes for buffering time of 60 ms without additional delay.

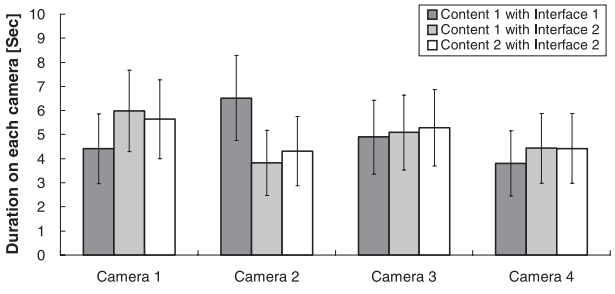

Fig. 12 Average watching time on each camera for load traffic of 7.2 Mb/s without additional delay.

In Interface 1, the duration decreases as the camera number increases except for camera 1. In Interface 2, the difference among each camera is smaller than that in Interface 1. This behavior is related to the default camera of the interfaces and the way the viewpoint is changed. The reason is as follows.

Camera 1 is the default camera for the two interfaces, and therefore the users will always see camera 1 first. In the case of using Interface 1, the user needs to select a camera number to change the viewpoint. As camera 2 is the next to camera 1, the user may want to change the camera to the next one. He will watch the viewpoint for less time as long as the camera of that viewpoint gets farther from the default camera with Interface 1.

On the other hand, when using Interface 2, the user can know the position of the cameras. Thus, he will concentrate more on watching the movement of the doll and will change the viewpoint to the direction he wants. By doing this, the time that the user watched on every camera does not have a clear relationship with the default camera, while this is related to the content movement.

In addition, we also show the variances for the parameters of the user's behavior. Figure 13 represents the variance of the number of viewpoint changes of Content 1 with Interface 1. Figure 14 depicts the variance of the number of viewpoint changes for buffering time of 60 ms without additional delay of Content 1 with Interface 1, Content 1 with Interface 2, and Content 2 with Interface 2. At the same time, Fig. 15 shows the variance of the watching time on each camera.

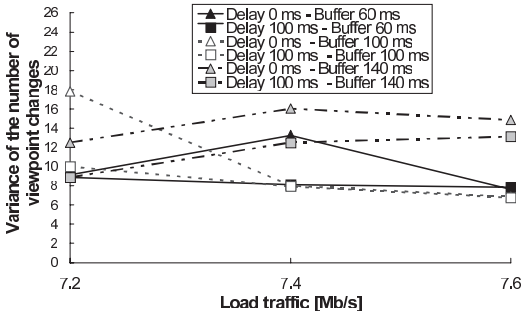

Fig. 13 Variance of the number of viewpoint changes of Content 1 with Interface 1.

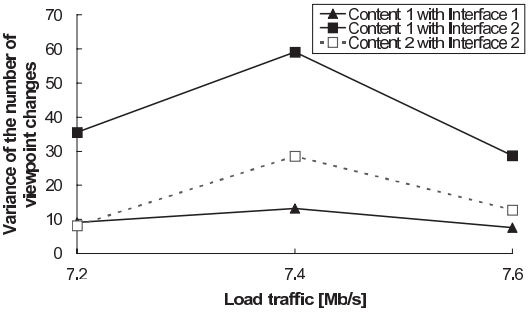

Fig. 14 Variance of the number of viewpoint changes for buffering time of 60 ms without additional delay.

In Figs. 13 and 14 we notice that the variance is large when using Interface 2. This is related to the way of changing the viewpoint.

With Interface 1, the user changes the viewpoint by the camera number. Thus, in order to change the viewpoint, the user needs to move the mouse to a different camera number from the one already selected and then select it by clicking the mouse. On the other hand, with Interface 2, the viewpoint is changed by the direction from the currently selected camera. Since the cameras are placed in a circular shape, in order to change the viewpoint, the user may press the same button without moving the mouse. As in our experiment we employed four cameras, for example, when the left or right buttons are pressed four times sequentially, the application displays the same viewpoint that was being displayed initially.

Since the user can change the viewpoint without moving the mouse when using Interface 2, some users changed the viewpoint many times, while others did not change the viewpoint so many times. The difference of the behavior makes the large variance.

5.3 QoE

We calculated the interval scale for each criterion. Then, we carried out the Mosteller's test. As a result, we have found that the test with a significance level of 0.01 cannot reject the hypothesis that the observed value equals the calculated one. We then use the interval scale as the QoE parameter.

Since we can select an arbitrary origin in an interval scale, for each criterion, we set the minimum value of the psychological scales to the origin. We obtained the boundary of each category. The upper boundaries of Category 1 (C1) to Category 4 (C4) are shown in Table 2.

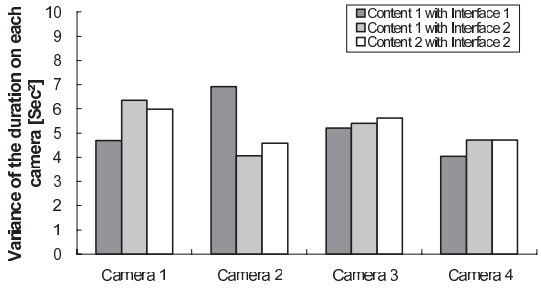

Fig. 15 Variance of the watching time on each camera for load traffic of 7.2 Mb/s without additional delay.

Table 2 Boundaries of the categories for each criteria.

	C1	C2	C3	C4
Smoothness of the video	0.836	1.450	1.839	3.032
Synchronization of the video and audio	0.670	1.296	1.714	2.940
Quickness of the viewpoint change	0.097	0.726	1.389	2.680
Overall evaluation	0.790	1.496	2.059	3.415
Fixed View vs. Selective View	−0.807	−0.537	0.210	1.589

Figure 16 shows the psychological scale value for the criterion of "Smoothness of the video" of Content 1 with Interface 1 and no additional delay. Figure 17 depicts the psychological scale value for the criterion of "Synchronization of the video and audio" in the same way as Fig. 16. The legends we use in the results are as follows. "Buffer" corresponds to the playout buffering time that was used in the MVV client. "Fixed View" or "Selective View" represents the scheme that was used for that result. For example, "Buffer 60 ms - Fixed View" means that the buffering time was set to 60 ms and that "Fixed View" was employed for that result.

Regarding both "Smoothness of the video" and "Synchronization of the video and audio," except the case with buffering time of 140 ms in the criterion "Synchronization of the video and audio," the evaluation for the two criteria of "Selective View" is better than that of "Fixed View" when the UDP load traffic is 7.6Mb/s. This is because the users noticed the deterioration in the video when watching on only one viewpoint easier than when watching on several viewpoints.

At the same time, we can see that the buffering time affects the evaluation of the two criteria. As the buffering time decreases and as UDP traffic increases, the MU loss ratio increases. When this happens, the psychological scale value of the two criteria decreases.

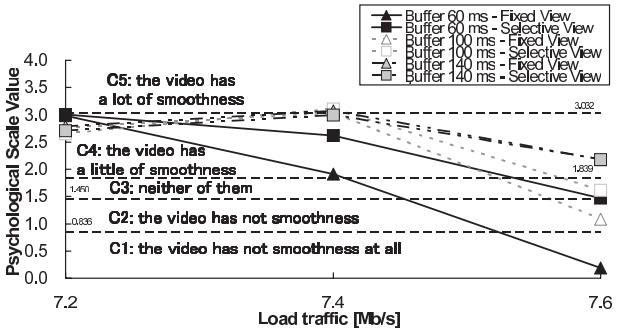

Fig. 16 Psychological scale value for the criterion "Smoothness of the video" for the two schemes of Content 1 with Interface 1 without additional delay.

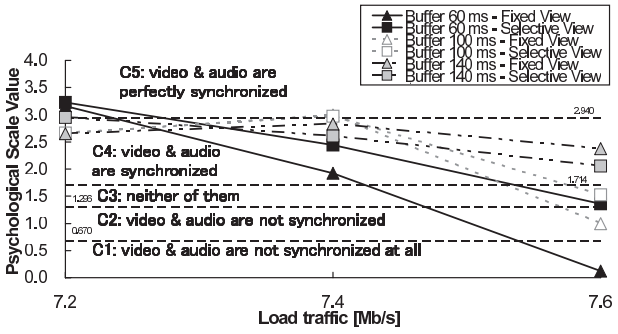

Fig. 17 Psychological scale value for the criterion "Synchronization of the video and audio" for the two schemes of Content 1 with Interface 1 without additional delay.

Figure 18 shows the psychological scale value for the criterion of "Quickness of the viewpoint change" with "Selective View" of Content 1 with Interface 1. The legends are the same as in Sect. 5.1.

We can notice that for the amount of load traffic equal to 7.2 Mb/s, the user felt fast viewpoint changes with small buffering time. On the other hand, when the load traffic is 7.6 Mb/s, the viewpoint change with long buffering time is faster than that with short buffering time. The reason is as follows.

As the load traffic increases, the delay jitter increases. With short buffering time and high delay jitter, a few MUs of the new viewpoint can be discarded owing to their delayed arrival. When this happens, the video freezes until the new viewpoint is displayed. Therefore, the user may feel slow viewpoint changes. Even if the user felt fast viewpoint changes in the same experimental run, once he experienced video freezing after having changed the viewpoint, his perception can be that the viewpoint change was slow. Also, in the case where the user changed the viewpoint in the moment when the video is frozen, he can feel he waited for more time than the time he actually waited. These two situations can make the user feel slow viewpoint changes.

Figure 19 shows the psychological scale value of "Quickness of the viewpoint change" with "Selective View" for buffering time of 60 ms and 140 ms without additional delay for Content 1 with Interface 1, Content 1 with Interface 2, and Content 2 with Interface 2. The legend shows the content, the interface and the playout buffering time. For

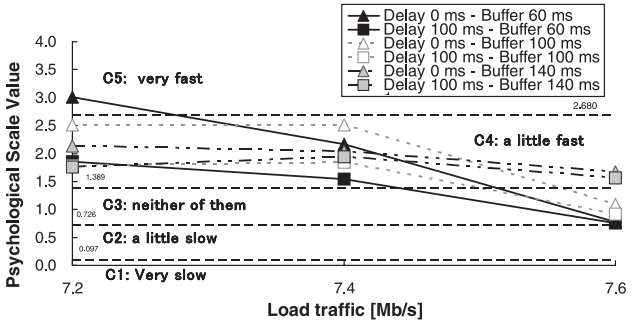

Fig. 18 Psychological scale value for the criterion "Quickness of the viewpoint change" for "Selective View" of Content 1 with Interface 1.

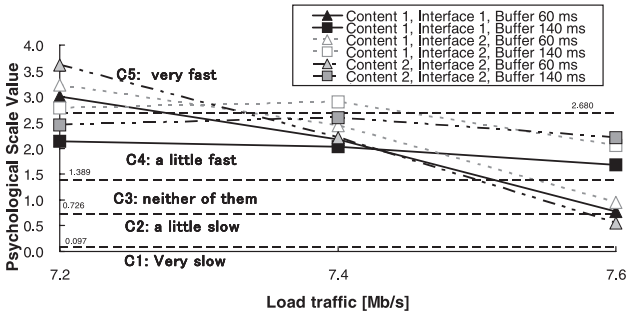

Fig. 19 Psychological scale value for the criterion "Quickness of the viewpoint change" for "Selective View" without additional delay.

example, "Content 1, Interface 1, Buffer 60 ms" means that Content 1 with Interface 1 and the playout buffering time of 60 ms was used for that result.

We can notice that the evaluation of quickness of the viewpoint change is better for Content 2 with Interface 2 than for Content 1 with Interface 2 at the load traffic of 7.6 Mb/s and the buffering time of 140 ms. At the same time, at the load traffic of 7.2 Mb/s and with buffering time of 60 ms, the user felt faster viewpoint changes when watching Content 2 with Interface 2 than when watching Content 1 with Interface 2. This is related to the movement of the doll dog. The doll of Content 2 kept still before jumping in the air. Therefore, the user felt a faster viewpoint change as the doll kept still at the moment the user changed the viewpoint. On the other hand, the doll of Content 1 continuously moved without stopping. Thus, the user may have changed the viewpoint at the moment the doll was moving, making him feel slower viewpoint changes.

At the same time, we can notice that the psychological scale value of quickness of the viewpoint change when the user watches Content 1 with Interface 2 is better than when the user watches Content 1 with Interface 1. As Interface 2 is easier to use, the user can change the viewpoint faster. For this reason, the evaluation improves.

Figures 20 and 21 show the psychological scale value of "Overall evaluation" with "Fixed View" and that with "Selective View," respectively, for buffering time of 60 ms and 140 ms without additional delay for Content 1 with Interface 1, Content 1 with Interface 2, and Content 2 with Interface 2. We can observe in Figs. 20 and 21 that the two schemes have a similar tendency; as the load traffic increases, "Overall evaluation" deteriorates especially in small buffering time. However, we can notice that the QoE of "Selective View" is higher than that of "Fixed View" of the two contents. There are two reasons for this. The first is because the user easily noticed the degradation of the video when watching only one viewpoint than when watching several viewpoints. The second is that the effective viewpoint change can enhance QoE.

We can also notice that "Overall evaluation" with the buffering time 140 ms is better than that with the buffering time 60 ms when the amount of load traffic is 7.6 Mb/s. This is because the buffering time 140 ms is large enough to absorb the delay jitter in the experiment. On the other hand, if we use smaller values for the buffering time, the delay jitter is not properly absorbed; it increases skipped MUs.

In addition, we can observe in Fig. 21 that the evaluation of "Overall evaluation" of "Selective View" with Content 1 is higher than with Content 2 when the UDP load traffic is 7.6 Mb/s for the buffering time of 60 ms. As explained before, this is because the user kept the viewpoints with Content 2 for longer time, making the degradation of the video easier to be noticed.

Regarding the interface, we can observe that when the amount of load traffic is 7.4 or 7.6 Mb/s, in Content 1, Interface 2 can provide higher values of "Overall evaluation" than Interface 1 for buffering time 140 ms. This is because the evaluation of "Quickness of the viewpoint change" is better with Interface 2.

Figure 22 shows the evaluation of the criterion "Fixed View vs. Selective View" for the buffering time of 60 ms and 140 ms without additional delay for Content 1 with Interface 1, Content 1 with Interface 2, and Content 2 with Interface 2.

In Fig. 22, we notice that the user seems to be more satisfied with "Selective View" when the buffering time is long enough to absorb the delay jitter. At the same time, as the load traffic increases, the QoE will gradually deteriorates until the point where the user is equally satisfied with the two schemes.

As for the interface, we can notice in Fig. 22 that, in

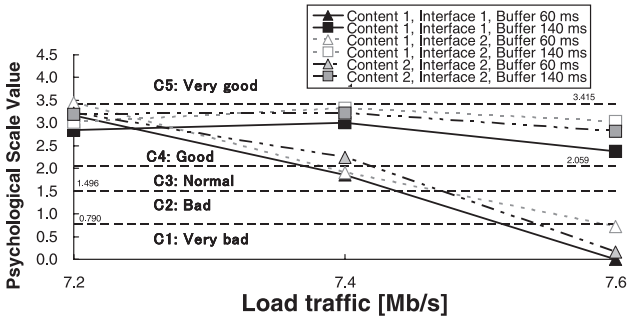

Fig. 20 Psychological scale value for the criterion "Overall evaluation" for "Fixed View" without additional delay.

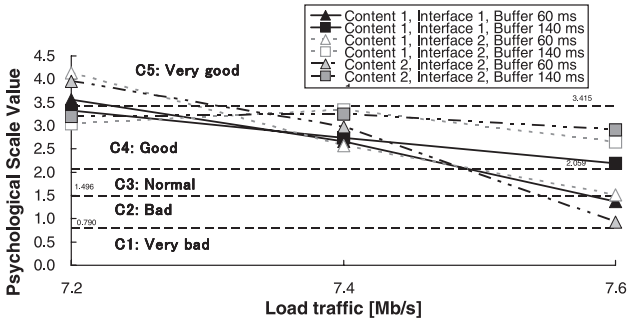

Fig. 21 Psychological scale value for the criterion "Overall evaluation" for "Selective View" without additional delay.

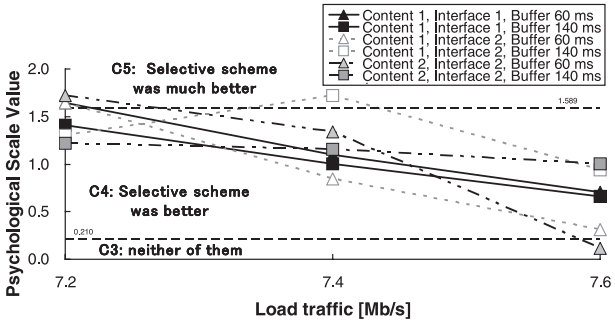

Fig. 22 Psychological scale value for the criterion "Fixed View vs. Selective View" without additional delay.

Content 1, the psychological scale value of "Fixed View vs. Selective View" with Interface 2 is better than that with Interface 1 for the buffering time of 140 ms and the load traffic of 7.4 Mb/s or 7.6 Mb/s. This is also related to the evaluation of the criterion "Overall evaluation." As the evaluation of "Overall evaluation" is better with Interface 2 than that with Interface 1, the evaluation of the criterion "Fixed View vs. Selective View" improves for Content 1 with Interface 2.

6. Conclusions

We made experiments on MVV and audio IP transmission under various traffic and delay conditions. We compared the two schemes: "Fixed View" and "Selective View." We assessed the effects of the IP traffic and the delay on QoE for the two schemes. We also employed two contents.

From the application-level QoS evaluation results, we saw that the ability of viewpoint change does not affect the application-level QoS parameters such as the MU loss ratio in our MVV application.

From the user study results of our subjective evaluation, we observed that users tend to change the viewpoint frequently in light traffic and low delay. Also, we found that when there is no additional delay, the number of viewpoint changes slightly decreases as the load traffic increases.

From the QoE evaluation results, we found that the user prefers "Selective View." However, as the traffic and delay gradually increases, the QoE will gradually deteriorate until the point where the user is equally satisfied with the two schemes. Also, we found out that both content and user interface can affect the QoE of the MVV system.

Also, under heavily loaded conditions, the user feels fast viewpoint changes when the buffering time is appropriate for absorbing the network delay jitter. The appropriate buffering time also improves the overall evaluation. Also, we noticed that the user pays more attention to the degradation of the video when watching only one viewpoint. For this reason, "Overall evaluation" of "Selective View" is higher than that of "Fixed View."

For future work, we will study the effect of the contents on QoE and the user's behavior with MVV systems in more detail. At the same time, we will use the audio of each camera instead of using a microphone.

Acknowledgment

This work was supported by the Grant-In-Aid for Scientific Research of Japan Society for the Promotion of Science under Grant 21360183.

References

[1] I. Ahmad, "Multiview video: Get ready for next-generation television," Proc. IEEE Distributed Systems Online, vol.8, no.3, art. no.0703-o3006, March 2007.
[2] M. Tanimoto, "Free viewpoint television — FTV," Proc. Picture Coding Symposium 2004, Dec. 2004.
[3] W. Matusik and H. Pfister, "3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes," ACM Trans. Graph., vol.23, no.3, pp.814–824, Aug. 2004.
[4] X. Guo, Y. Lu, F. Wu, W. Gao, and S. Li, "Distributed multi-view video coding," Proc. SPIE Visual Communications and Image Processing 2006, vol.6077, pp.290–297, Jan. 2006.
[5] E. Martinian, A. Behrens, J. Xin, A. Vetro, and H. Sun, "Extensions of H.264/AVC for multiview video compression," Proc. IEEE ICIP 2006, pp.2981–2984, Oct. 2006.
[6] ITU-T Rec., P.10/G.100 Amendment 1, "Amendment 1: New appendix I — Definition of quality of experience (QoE)," Jan. 2007.
[7] J. Lou, H. Cai, and J. Li, "A real-time interactive multiview video system," Proc. ACM Multimedia 2005, pp.161–170, Nov. 2005.
[8] L. Zuo, J. Lou, H. Cai, and J. Li, "Multicast of real-time multi-view video," Proc. IEEE ICME 2006, pp.1225–1228, July 2006.
[9] ITU-T Rec. G.1080, "Quality of experience requirements for IPTV services," Dec. 2008.
[10] J. Hoxmeier and C. DiCesare, "System response time and user satisfaction: An experimental study of browser-based applications," Proc. AMCIS 2000, Aug. 2000.
[11] "NISTNET," http://snad.ncsl.nist.gov/nistnet/
[12] J.C. Nunnally and I.H. Bernstein, Psychometric Theory, Third ed., McGraw-Hill, N.Y., 1994.
[13] F. Mosteller, "Remarks on the method of paired comparisons: III a test of significance for paired comparisons when equal standard deviations and equal correlations are assumed," Psychometrika, vol.16, no.2, pp.207–218, June 1951.
[14] S. Tasaka and Y. Ito, "Psychometric analysis of the mutually compensatory property of multimedia QoS," Conf. Rec. IEEE ICC2003, pp.1880–1886, May 2003.

Erick Jimenez Rodriguez received the B.E. degree in computer science engineering from Costa Rica Institute of Technology in 2005. He was with IBM Costa Rica from 2005 to 2007. He is now a student in the Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Nagoya, Japan. His current research interests include Multi-View Video, QoS and QoE assessment.

Toshiro Nunome received the B.S., M.S. and Ph.D. degrees from Nagoya Institute of Technology, Nagoya, Japan, in 1998, 2000, and 2006, respectively. From 2000 to 2001, he was with Pioneer Corp. In 2002, he joined Nagoya Institute of Technology as a Research Associate. Now, he is an Assistant Professor in the Department of Computer Science and Engineering, Graduate School of Engineering. His research interests include media synchronization, multicast communications and wireless networks. Dr. Nunome is a member of the IEEE.

Shuji Tasaka received the B.S. degree in electrical engineering from Nagoya Institute of Technology, Nagoya, Japan, in 1971, and the M.S. and Ph.D. degrees in electronic engineering from the University of Tokyo, Tokyo, Japan, in 1973 and 1976, respectively. Since April 1976, he has been with Nagoya Institute of Technology, where he is now a Professor in the Department of Computer Science and Engineering, Graduate School of Engineering. In the 1984–1985 academic year, he was a Visiting Scholar in the Department of Electrical Engineering at the University of California, Los Angeles. His current research interests include wireless networks, QoE/QoS, and multimedia communication protocols. He is the author of a book entitled Performance Analysis of Multiple Access Protocols (MIT Press, Cambridge, MA, 1986). Dr. Tasaka is a member of the IEEE, ACM and Information Processing Society of Japan.