Multidimensional QoE Assessment of Multi-View Video and Audio (MVV-A) IP Transmission: The Effects of User Interfaces and Contents

Erick Jimenez Rodriguez, Toshiro Nunome, and Shuji Tasaka
Department of Computer Science and Engineering, Graduate School of Engineering,
Nagoya Institute of Technology
Nagoya 466–8555, Japan
e-mail: {nunome, tasaka}@nitech.ac.jp

***Abstract*—This paper studies QoE (Quality of Experience) in Multi-View Video and Audio (MVV-A) transmission over IP networks where the users change the viewpoint by request. It jointly analyzes the effects of user interfaces for viewpoint change and contents in conjunction with network performance in a multidimensional way. In order to assess the QoE, we perform a subjective experiment which employs two user interfaces and two contents (a dog doll and a toy train). The assessors are instructed to follow the movement of the object in the content and evaluate the MVV-A system. As a result, we have found that even if the user feels comfortable with the user interface, QoE deteriorates when the load traffic is very high and the playout buffering time cannot absorb the delay jitter properly. The speed of the movement of the object in the content, and the similarity of the views of the cameras can affect users' perception of the viewpoint change response. We realized that in order to achieve high QoE, we need to employ a user interface that enables the users to change the viewpoint easily according to the object's movement.**

***Keywords*-real-time multi-view video, QoE, QoS, viewpoint change, IPTV**

I. Introduction

Internet Protocol Television (IPTV) [1], a system through which digital television service is delivered by using the architecture and networking methods of the Internet, is becoming very common. In the current IPTV services, however, as in the conventional television, the users cannot freely choose the view (e.g., the viewpoint, angle, and direction) which they want to watch. Instead, they are forced to look at the same view given by the sender, even if they would prefer to watch another view of the same content, program, or event.

In order to avoid this inconvenience, *omni-directional video* [2], *MVV (Multi-View Video)* [3], and *FTV (Free-viewpoint TeleVision)* [4] have been developed.

The omni-directional video captures images in all direction by an omni-directional camera. It is also referred to as *surround video* [5]. It can provide a panoramic view and can enhance viewing experience of the users. In [6], a telepresence system with an omni-directional video viewer on a web browser is proposed. By the system, the users can see the video contents in arbitrary direction.

In the omni-directional video, the users can change the view direction. However, they cannot change the viewpoint because the view is created by an omni-directional camera in a specific position.

In MVV, the users can choose one video from multiple video streams of the same content taken by multiple cameras from different positions. MVV systems can be applied to wide areas such as entertainment, sports, sightseeing, and education among others. MVV can be a base system of FTV, in which the users can select the viewpoint freely without the limitation of cameras' positions.

As IPTV, MVV can be transferred over the IP networks. In this paper, we focus on MVV over the IP networks in which the user's viewpoint is changed by his/her request.

The ultimate goal of the network services is to provide high *QoE (Quality of Experience)*. QoE represents the overall acceptability of an application or service, as perceived subjectively by the end-users [7]. ITU-T Rec. G.1080 [8] defines QoE requirements for IPTV services. However, this recommendation does not refer to the QoE in MVV systems.

There have been many studies regarding MVV systems. However, most of the studies focus on the coding techniques such as *MVC (Multi-view Video Coding)* and assess the effectiveness with the throughput and *PSNR (Peak Signal to Noise Ratio)*, which measures spatial quality of video.

In [9], Kurutepe *et al*. present a client-driven multi-view video streaming system by transmitting a small number of views selected according to his/her head position. In addition, lower quality versions of some other views are also prefetched for concealment if the current user's viewpoint differs from the predicted viewpoint. They evaluate the proposed system in terms of PSNR and the prediction error of head position.

Cheung *et al*. have addressed the problem of designing a frame structure for interactive multiview streaming [10]. They propose an algorithm to encode the video stream for interactive view switching with low transmission cost by means of an optimal selection of I-, P-, and "merge" frames. They assess the efficiency in terms of the transmission cost.

Note that both [9] and [10] do not discuss QoE of their MVV systems.

Even in network services, the user interface is an important factor that can affect QoE. The user can feel more satisfied with the system when employing user interfaces that are more intuitive and easier to use. Not only the type of the interface but also its usage can affect the QoE.

User interfaces by themselves are evaluated in many different areas in order to improve them by finding their possible problems in their functionality (e.g., [11] and [12]). However, the studies are limited only to the end-terminals, not taking into consideration the interface performance in conjunction with the network performance.

Some published works assess QoE of MVV systems by employing one user interface for viewpoint change (e.g., [13] and [14]). However, they do not perform a systematic QoE assessment when packet delay jitter and packet loss are present in the transmission and do not consider the effect of user interfaces for viewpoint change on QoE of MVV systems. Also, they do not treat audio; note that in the great majority of real applications, audio and MVV are transmitted together. In this paper, we refer to MVV accompanied by audio as *MVV-A*.

In [15], Ishikawa *et al*. propose a view generation method for realizing arbitrary view-directions at arbitrary viewpoints by means of multicast multiple omni-directional videos; the approach is close to FTV. They perform subjective evaluation to assess the validity of the proposed view generation method. However, the evaluation does not include the effect of network-level impairment, and they also do not treat audio.

In [16], the authors perform user behavior studies on MVV-A systems and analyze the effect of load traffic, packet delay, playout buffering time, and user interfaces on the QoE. However, they employed only one type of audio-video content and do not consider the effect of the user interfaces when using different types of contents. In addition, they do not consider criteria related to user's feelings, user's expectations, and user interface.

This paper proposes an integrated assessment method of QoE in MVV-A IP transmission by taking into consideration user interfaces for viewpoint change and contents in addition to network performance such as packet loss and delay jitter. For that purpose, we resort to a multidimensional assessment method: the *SD (Semantic Differential)* method [17]. Since the criteria employed in [16] are not enough to examine the joint effects of the user interfaces and contents, we introduce new criteria in this paper. In the assessment, the users are instructed to follow the movement of the object in two types of contents, i.e., a dog doll and a toy train.

In addition to several subjective QoE metrics, we introduce objective measures that can be reflected on QoE. Among these measures, we pick up application-level QoS parameters and measures for the user's behavior. We then perform *PCA (Principal Component Analysis)* with the subjective QoE metrics and these objective measures in order to investigate relationships among them and clarify main factors affecting QoE.

The rest of the paper is structured as follows. Section 2 discusses factors affecting QoE of MVV-A IP transmission. Section 3 outlines methods of the experiment we performed. We present results of the experiment in Section 4, and Section 5 concludes this paper.

II. Factors affecting QoE of MVV-A IP transmission

We need to clarify QoE requirements for MVV-A IP transmission. In this section, we discuss the QoE metrics and show factors affecting QoE.

As practical examples of MVV applications, let us imagine a soccer match and a concert, for instance. In these examples, cameras are placed around objects because the users want to focus on the objects. The objects may move round a particular area, e.g., a stage, a field and a stadium. Also, in general, when the objects move or perform an action, sound is generated according to these movements or actions. Therefore, not only video but also audio should be considered, namely, MVV-A.

When showing the object to the users, we expect them to be interested in changing the viewpoint according to the object's movement. Thus, the MVV-A system will satisfy the user with the ability of viewpoint change.

As mentioned earlier, the user interface for viewpoint change is an important factor which influences the user's satisfaction. The *user's expectation* for the user interface can also affect the user's satisfaction; unless the user interface comes up to his/her expectations, QoE can deteriorate compared to the case in which the user has no expectation.

In addition to the above mentioned factors, we also consider the *playout buffering time* as a factor that can affect QoE of MVV-A systems. When the MVV-A client uses short buffering time, it cannot absorb the delay jitter properly. 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 interactivity of the MVV-A system. Thus, there is a tradeoff relationship between the viewpoint change response and the output quality; this implies that the buffering time should be set properly.

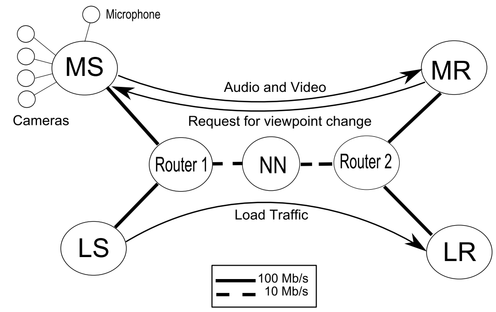


Figure 1. Experimental system.

When analyzing the QoE of any system, multiple factors should be evaluated for a better understanding of how these factors are related to each other and how they can affect the QoE of that system. For this reason, we employ multidimensional assessment considering multiple factors that can affect QoE.

We classify these factors into the following categories:

- Audio-video (AV): As in traditional single-view video and audio IP transmission, audio and video smoothness are indispensable factors affecting QoE.
- Interactivity (I): One of the key elements involved in validating QoE in the MVV-A service is how quickly the users can change the viewpoint.
- User Interface (UI): It is important to evaluate the user's satisfaction with the user interface under a variety of conditions of the playout buffering time, load traffic and network delay.
- MVV-A system (MVV): Opinions that the user has about the MVV-A system as a whole should be considered.
- Content (CO): Depending on the content, its actions and movements can vary. Since the user can change the viewpoint according to the content, the usage of the MVV-A systems may also vary. This can affect the QoE as well.
- User's feelings (UF): The user's feelings, such as impatience and irritancy, can change according to the response of the MVV-A system. They should be taken into consideration in order to see how they are reflected on the QoE.
- User's overall evaluation (O): Since there are multiple factors that can affect the QoE of the MVV-A system, it is important to consider all of them in terms of a single criterion that can depict how satisfactory the user is for the system.

In this paper, we examine the effect of the user interface in MVV-A IP transmission in terms of the criteria discussed above by designing two user interfaces. Also, as the content can affect the QoE, we employ two contents.

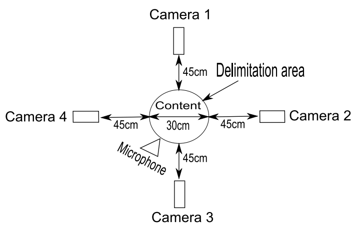
III. Experimental method

In this section, we will explain the details of our experiment. In the following subsections, we discuss the experimental system, experimental conditions, and QoE metrics.

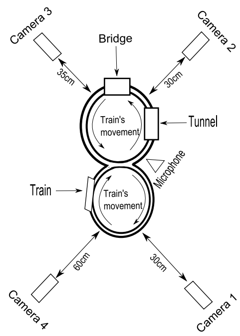
A. Experimental system

Figure 1 shows the experimental system. MS is the server of the MVV-A application, and MR is the client. Four cameras and a microphone are connected to the server.

The server captures the video of each camera. At the same time, the audio is captured by the microphone. The server sends the audio and video of a selected viewpoint to the client as two separate UDP packet flows. The client receives



(a) Dog



(b) Train

Figure 2. Positions of the cameras.

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.

On the other hand, LS is the server of the load traffic, and LR is the client. Both router 1 and router 2 are Riverstone's RS3000. NN, which is a PC, is laid out between the routers. NN delays packets going through routers 1 and 2 by using NISTNET [18]. By adding this delay, we can see the effect of network delay on the QoE in the MVV-A system.

B. Experimental conditions

We discuss experimental conditions of our assessment in this subsection. At first, we introduce the contents and the user interfaces, and we then mention the traffic specifications and assessors.

1) Contents: In this assessment, we employ two types of contents in order to analyze the effect of the following factors:

- The speed of movement of the object in the content
- The similarity of the views of each camera
- The direction of the object's movement
- The distance between the cameras and the content area

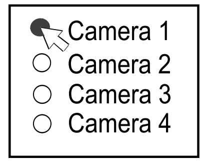
Figure 2 shows the positions of the cameras and the microphone connected to MS for two contents that were employed. We refer to the content in Fig. 2(a) and that in Fig. 2(b) as "Dog" and "Train", respectively. We can assume various camera arrangements. In this paper, as one of the basic arrangements, we employ a circular arrangement surrounding the contents.

As the object in Dog, we employ a dog doll which moves with battery. When the switch of the doll is turned on, the doll walks a few steps forward and barks with moving its tail while walking backwards. Later, the doll starts to walk forward again, but in a different direction; it moves in the counterclockwise direction.

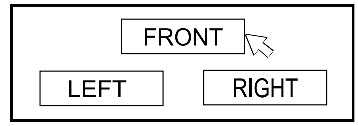
In Train, we use a toy train which moves with battery. When the train is turned on, it moves continuously on the rail. The arrows inside the two circles in the center of Fig. 2(b) show the direction of movement of the train. Since the train moves on the rail, the user sees the train moving in a different direction depending on the camera from which the train is being watched.

As seen in Fig. 2(b), we made the upper half of the rail different from the lower one in order to analyze how different views of the same content can affect the QoE. We included a tunnel and a bridge only in the upper half.

2) User interface: We used two different user interfaces. Each interface is 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. 3(a). The second one lets the user change the viewpoint by using the camera's direction, according to the camera that is currently being watched, as shown in Fig. 3(b). In this paper, we refer to the first interface as "Interface 1" and to the second one as "Interface 2".



(a) Interface 1: with camera numbers



(b) Interface 2: with camera directions

Figure 3. User interfaces.

Table I
SPECIFICATIONS OF AUDIO AND VIDEO.

	Audio	Video
Coding scheme	G.711 μ-law	H.264
Image size [pixels]	-	704 $\times$ 480
Picture pattern	-	I
Coding bit rate [kb/s]	64	2000
Average MU rate [MU/s]	25	25
Media duration [s]	20	

Prior to this experiment, the assessors received instructions when watching the contents. In Dog, the assessors were instructed to follow the dog doll's face. In Train, the instruction was to follow the train. We gave these instructions to the assessors so that we can measure the level of fulfillment of each interface with which the assessors accomplish the instruction.

3) Traffic specifications: 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. The specifications of the audio and video are shown in Table I.

We employed a simple scheme of playout buffering control at the client to absorb network delay jitter and set the buffering time to 60 ms, 100 ms, and 140 ms. If all the packets of an MU are not correctly received in time for output, the MU is not output.

While the server (MS) sends the video and audio to the client (MR), LS generates 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 set to 0 ms, 100 ms or 300 ms.

We refer to an object for evaluation as a *stimulus*, which is an audio-video stream output at the receiver in each experimental run. For each combination of the content and the user interface, the assessor evaluates 30 stimuli including three dummies in a random order. In addition, the order of the content and the user interface differs for each assessor.

4) Assessors: We employed 21 assessors: 15 Japanese males, a Malaysian male, a Chinese male, and 4 Japanese females. Their age ranges from around twenty through forty. Note that the non-Japanese assessors are sufficiently familiar with Japanese language for the subjective assessment.

C. Application-level QoS parameters

In this paper, we employ the following application-level QoS parameters, where the subscripts "a" and "v" mean audio and video, respectively.

- MU loss ratio (L_a and L_v [%]): the ratio of the number of MUs not output at the recipient to the number of MUs transmitted by the sender.

- average MU delay (D_a and D_v [ms]): the average time in seconds from the moment an MU is generated until the instant the MU is output.

In addition, we employ *viewpoint change delay* (D_c [ms]) as a specific application-level QoS parameter for MVV-A. It is defined as the time in seconds from the moment the client sends a request for viewpoint change until the instant a new viewpoint is output at the client.

D. User's behavior

There exist various measures representing the user's behavior. In this paper, as a measure, we express the user's behavior in terms of *number of viewpoint changes*, which is defined as the number of times that the user (assessor) changed the viewpoint in each experimental run, i.e., 20 seconds. In this assessment, we use the *average number of viewpoint changes* (N_{avg}) and the *variance of the number of viewpoint changes* (N_{var}).

E. Subjective QoE metrics

In this paper, we employ the SD (Semantic Differential) method for subjective QoE assessment. This method was proposed by Osgood as a method of measuring meaning. This method can assess a stimulus from many points of view with many *pairs of polar terms*. A pair of polar terms consists of one adjective and its opposite one; e.g., quiet and noisy.

In the SD method, how to select pairs of polar terms used for the assessment is important. In order to select the polar terms, we performed preliminary tests analyzing different criteria regarding the audio-video streams (AV), the interactivity (I), the user interface (UI), the MVV-A system (MVV), the content (CO), user's feelings (UF), and the overall satisfaction (O). When we could not find any appropriate adjective in order to evaluate a particular criterion, we adopted a verb instead. After the tests, we chose the following criteria which the assessor evaluates for each stimulus:

- Smoothness of the video (AV1): the video is smooth – the video is interrupted
- Smoothness of the audio (AV2): the audio is smooth – the audio is interrupted
- Viewpoint change response (I1): the viewpoint change response is fast – the viewpoint change response is slow
- Usefulness of the interface (UI1): the interface is useful – the interface is not useful
- Reliability of the MVV-A system (MVV1): the system is reliable – the system is not reliable
- Convenience of the MVV-A system (MVV2): the system is convenient – the system is inconvenient
- Willingness of using the MVV-A system (MVV3): I want to use the system – I do not want to use the system
- Fulfillment of the MVV-A system (MVV4): the system fulfills my expectations – the system does not fulfill my expectations
- Possibility of following the content's movement (CO1): I can follow the content's movement – I cannot follow the content's movement
- Level of impatience (UF1): I feel relaxed – I feel impatient
- Level of irritancy (UF2): I am not irritated – I am irritated
- Overall satisfaction (O1): excellent – bad

Note that the experiment was performed with the Japanese language. This paper has translated the used Japanese terms into English. Therefore, the meanings of adjectives or verbs written in English here may slightly differ from those of Japanese ones.

For each criterion, a subjective score is measured by the *rating scale method* [19]. In the method, an assessor classifies the stimuli into a certain number of categories; here, each criterion is evaluated to be one of five grades. The best grade (score 5) represents the positive adjective (the left-hand side one in each pair), while the worst grade (score 1) means the negative adjective. The middle grade (score 3) is neutral. The rating scale method is also used to measure *MOS (Mean Opinion Score)*, which is widely utilized for assessment of a single medium. However, in MOS, the numbers assigned to the categories only have a greater-than-less-than relation between them; that is, the assigned number is nothing but an ordinal scale. When assessing the subjectivity quantitatively, it is desirable to use at least an *interval scale*.

In order to obtain an interval scale from the result of the rating scale method, we first measure the frequency of each category with which the object for evaluation is placed in the category. With the *law of categorical judgment* [19], we can translate the frequency obtained by the rating scale method into an interval scale. However, 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 [20]. Once the goodness of fit has been confirmed, we use the interval scale as the QoE metric, which is therefore called the *psychological scale* [21].

In order to understand the user's opinion regarding the user interface and the content, we adopted additional criteria that may not be affected by the network conditions, such as packet loss and packet delay jitter. At the end of the experiment, the assessors were asked the following additional questions:

- Intuitiveness of the interface (UI2): the interface is intuitive – the interface is not intuitive
- Ease of using the interface (UI3): the interface is easy to use – the interface is difficult to use
- Ease of watching the content (CO2): I feel easy to watch the content – I feel difficult to watch the content

These questions were also evaluated to be one of five grades; the highest score is 5. We then calculated a MOS value of each question. This is because only four combinations of the user interfaces and contents are not enough to calculate the interval scale value for each criterion.

IV. Assessment results

In this section, we will present experimental results of both objective measures and subjective QoE. At first, we show the assessment results of application-level QoS and user's behavior. We then discuss the results of subjective QoE assessment and PCA.

A. Application-level QoS

Figures 4 and 5 show measured values of the application-level QoS parameters as a function of the load traffic for the three values of the additional delay at NN and the three values of the buffering time. Figure 4 depicts the viewpoint change delay for video of Dog with Interface 1. Figure 5 shows the MU loss ratio for video of Dog with Interface 1.

In Fig. 4, we see that with a short buffering time, we can expect small viewpoint change delay when there is low traffic and no additional delay. We also notice that the viewpoint change delay increases as the UDP load traffic and the buffering time increase. Also, for each of the three values of the buffering time, as the value of additional delay in NN increases, the viewpoint change delay also increases.

At the same time, we notice in Fig. 5 that the MU loss ratio increases as the UDP load traffic increases. This is because when the load traffic increases, the available

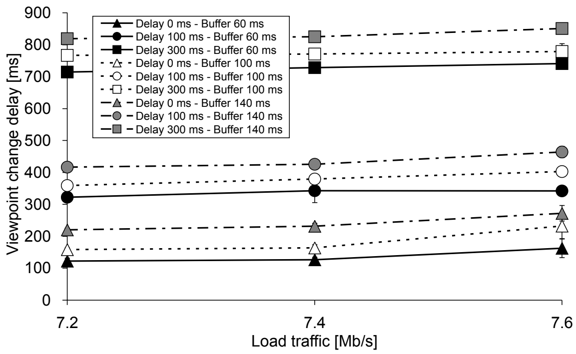


Figure 4. Viewpoint change delay for video of Dog with Interface 1.

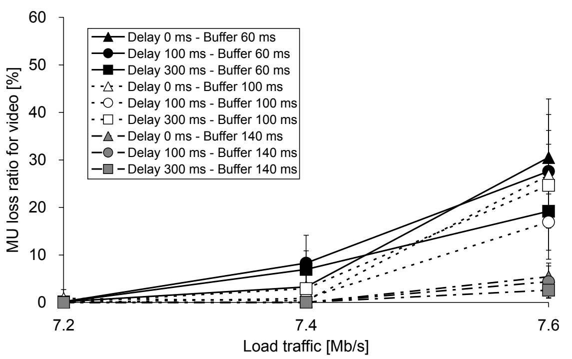


Figure 5. MU loss ratio for video of Dog with Interface 1.

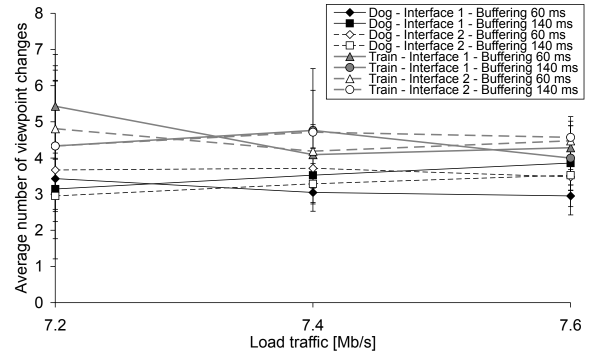


Figure 6. Average number of viewpoint changes without the additional delay.

Table II
RESULTS OF THE ADDITIONAL QUESTIONS.

Criterion	Content	Interface	MOS
Intuitiveness of the interface (UI2)	Dog	1	3.476
	Dog	2	4.190
	Train	1	2.667
	Train	2	4.238
Ease of using the interface (UI3)	Dog	1	3.476
	Dog	2	4.333
	Train	1	2.952
	Train	2	3.905
Ease of watching the content (CO2)	Dog	1	4.143
	Dog	2	4.288
	Train	1	3.571
	Train	2	3.619

bandwidth decreases to a point where there is not enough bandwidth to send both load traffic and the audio-video streams. This causes the packets to be delayed or discarded.

We also find in Fig. 5 that with the buffering time of 60 ms and that of 100 ms, the MU loss ratio is considerably high compared to the case when using buffering time of 140 ms. If the buffering time is not long enough to absorb the delay jitter, the number of skipped MUs 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 viewpoint change delay and MU loss ratio for the two contents and two interfaces. As a result, we found that the types of contents and user interfaces scarcely affect the application-level QoS.

In addition, we observed that the maximum MU loss ratio of audio is almost 1 % when the UDP load traffic is 7.6 Mb/s. For this reason, it is difficult for the assessor to have noticed the degradation of the audio.

B. User's behavior

Figure 6 shows the average number of viewpoint changes for each combination of content, user interface, and playout buffering time without additional delay. We notice in this figure that the user changed the viewpoint more times in Train than in Dog. This is because the train's speed was faster than that of the dog doll; in order for the user to continue watching the train, the user needed to change the viewpoint more frequently.

C. Additional questions

Table II provides the results of the additional questions of the assessment. The first column shows the criterion, and the fourth column presents the MOS value.

In Table II, we notice that the MOS value with Interface 2 is larger than that with Interface 1 for "Intuitiveness of the interface (UI2)" and "Ease of using the interface (UI3)" in each content. We confirmed this by the Student's t-test. Thus, it is easier for the users to understand and use Interface 2 than Interface 1 for the two contents. With Interface 2, the viewpoint is changed according to the direction of the camera that is being watched. With Interface 1, the viewpoint is changed according to the camera number. Since it is not necessary for the user with Interface 2 to remember the position of the cameras, the user found that Interface 2 is easier to understand and use than Interface 1.

As for "Ease of watching the content (CO2)", in Table II, we can see that the users feel easy to watch Dog more than Train. We also confirmed this by the Student's t-test. Since the doll dog moves slower, it is easier for the user to follow than the train.

D. Subjective QoE assessment

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 can reject the hypothesis that the observed value equals the calculated one in some criteria. Thus, we checked stimuli which give large errors of Mosteller's test and removed them in order not to reject the hypothesis. We then use the interval scale as the psychological scale.

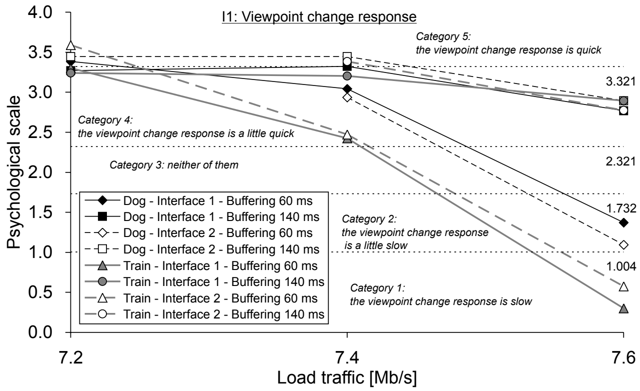


Figure 7. Psychological scale value for the criterion "Viewpoint change response" without the additional delay.

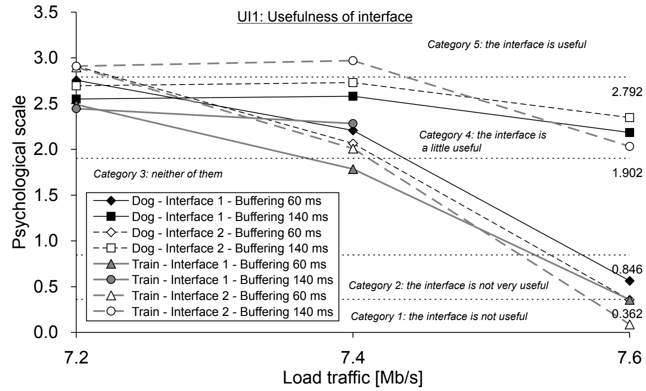


Figure 8. Psychological scale value for the criterion "Usefulness of the interface" without the additional delay.

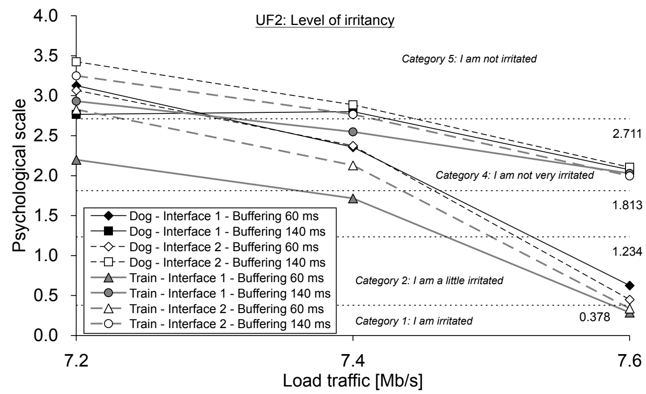


Figure 9. Psychological scale value for the criterion "Level of irritancy" without the additional delay.

Since we can select an arbitrary origin in an interval scale, for each criterion, we set the minimum value of the psychological scale to the origin. The upper boundaries of Category 1 to Category 4 are plotted as straight broken lines parallel to the abscissa in the figures to be shown. Note that the lower bound of Category 1 is $-\infty$, and the upper bound of Category 5 is ∞. In addition, the removed stimuli by the Mosteller's test are not shown in the figures.

In this subsection, we discuss the results of the "Viewpoint change response (I1)", "Usefulness of the interface (UI1)", "Level of irritancy (UF2)", and "Overall satisfaction (O1)".

We have assessed the effect of the additional delay in terms of the criteria. Then, we have found that the additional delay only affects the evaluation result of the criterion "Viewpoint change response"; the result without the additional delay is better than that with the additional delay. In this subsection, we show the results without the additional delay.

1) Viewpoint change response: Figure 7 shows the psychological scale value for the criterion of "Viewpoint change response" for each combination of the content, user interface, and playout buffering time without the additional delay.

First, we focus on the user interface. Let us examine the case where the buffering time is 140 ms; note that the packet delay jitter is absorbed properly in this case. In Fig. 7, we can notice that the psychological scale value of the "Viewpoint change response" is very similar for the two interfaces.

Secondly, we examine the effect of the buffering time. It clearly appears when the load traffic is 7.6 Mb/s; in this case, we notice that the viewpoint change response with buffering time of 60 ms is slower than that with the buffering time of 140 ms. The reason is as follows. When the traffic is high and the buffering time cannot absorb the delay jitter properly, the MU loss ratio becomes high, as seen from Fig. 5. The video freezing due to MU loss makes the user feel slow viewpoint changes; as for the detailed discussion, see [16].

Thirdly, we consider the content. We find in Fig. 7 that the viewpoint change response of Train tends to be slower than that of Dog. This is related to the content's speed of movement and the similarity of the views of each camera. The views of each camera with Train are not so similar as those with Dog, and the train moves faster than the dog doll. Thus, the user feels slower viewpoint changes with Train.

2) Usefulness of the interface: Figure 8 depicts the psychological scale value of "Usefulness of the interface" for each combination of the content, user interface, and playout buffering time without the additional delay.

As seen from Fig. 8, the user feels that Interface 2 is more useful than Interface 1 when the buffering time is enough to absorb delay jitter. However, when the buffering time is 60 ms under heavily loaded condition, the psychological scale value of Interface 2 is smaller than that of Interface 1. This is due to the user's expectation. Since Interface 2 is easier to use, the user has more expectations for Interface 2 than those for Interface 1, as shown in Table II. However, since the buffering time is not enough to absorb the delay jitter properly when the traffic is high, the MVV-A system will not fulfill those expectations. Therefore, the psychological scale value of Interface 2 decreases even more.

3) Level of irritancy: Figure 9 depicts the psychological scale value of "Level of irritancy" for each combination of the content, user interface, and playout buffering time without the additional delay.

In Figs. 8 and 9, we can see that "Level of irritancy" has the same tendency as "Usefulness of the interface"; the user tends to feel less irritated with Interface 2 than Interface 1. This is because Interface 2 is intuitive and easy to use.

4) Overall satisfaction: Figure 10 shows the psychological scale value of "Overall satisfaction" for each combination of the content, user interface, and playout buffering time without the additional delay. We observe in this figure that "Overall satisfaction" has the same tendency as "Level of irritancy" shown in Fig. 9, although the difference of contents and user interfaces in "Overall satisfaction" is smaller than that in "Level of irritancy".

We also calculated the correlation coefficients between the psychological scale value of "Overall satisfaction" and

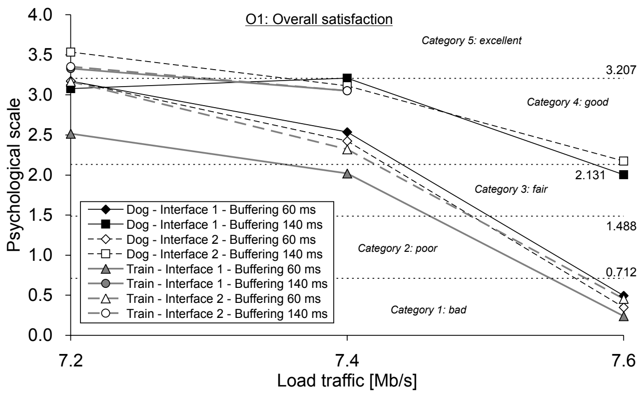


Figure 10. Psychological scale value for the criterion "Overall satisfaction" without the additional delay.

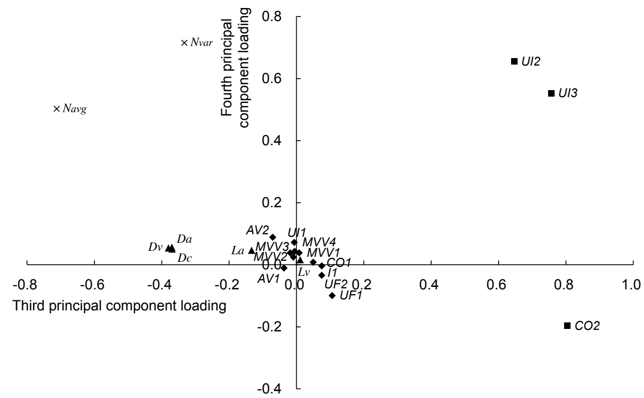


Figure 12. Third and fourth principal component loading.

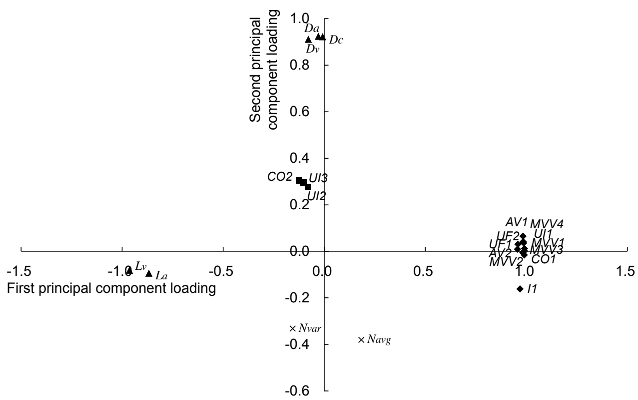


Figure 11. First and second principal component loading.

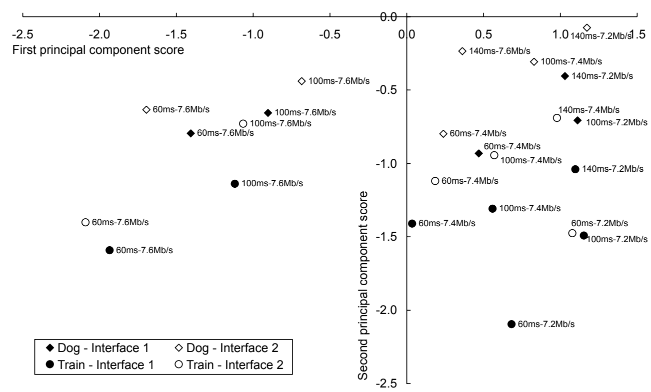


Figure 13. First and second principal component scores.

that of the other adjective pairs. We then found that the subjective QoE metrics discussed in this subsubsection have high correlation coefficients with the "Overall satisfaction"; the values are larger than 0.9.

E. Principal component analysis

In order to investigate main factors affecting QoE, we performed principal component analysis for application-level QoS parameters, measures for user's behavior, and the subjective QoE metrics except for "Overall satisfaction" together. We then picked up four components which have eigenvalues larger than unity. The contribution rate of the first principal component is 58.697 %. The cumulative contribution rate of the first two principal components is 73.464 %, and the rate of the first three components is 86.429 %. The cumulative contribution rate of the first four components is 93.944 %.

Figures 11 and 12 plot the principal component loading values. In Fig. 11, the abscissa indicates the first principal component loading, and the ordinate represents the second one. The abscissa and ordinate of Fig. 12 are the third principal component loading and the fourth one, respectively.

We find in Fig. 11 that the subjective QoE metrics (i.e., AV1, AV2, I1, UI1, MVV1, MVV2, MVV3, MVV4, CO1, UF1, and UF2) have high positive values of the first principal component loading. This is because there exists high correlation among the metrics. In addition, L_a and L_v, which mean the MU loss ratio of audio and that of video, respectively, have negative loading values of the first component. When L_a and L_v are large, the output quality of audio and video degrades. Thus, the first principal component can be considered as the output quality of MVV-A.

We can observe in Fig. 11 that the application-level QoS parameters in terms of delay (D_a, D_v, and D_c), which are regarded as objective QoE measures, have high loading values of the second principal component. On the other hand, the subjective QoE metrics have loading values around zero except for I1 in the second component. That is, the application-level delay does not strongly correlate with the subjective criteria in the assessment. The second component can be interpreted as the interactivity.

We notice in Fig. 12 that CO2, UI2 and UI3 have high loading values of the third principal component. In addition, among the subjective QoE metrics, UF1 and UF2 have slightly higher loading values than the others.

In Fig. 12, we also find that the fourth principal component loading of N_{var} is high. In addition, N_{avg}, UI2 and UI3 also have high loading values.

Figures 13 and 14 show the principal component scores for stimuli without additional delay. The abscissa and ordinate of Fig. 13 are the first principal component score and the second one, respectively. In Fig. 14, the abscissa is the third principal component score, and the ordinate is the fourth one. The figures accompanying each symbol represent the pair of playout buffering time and load traffic.

We see in Fig. 13 that all the plots have negative scores of the second component. This is because the second component means the interactivity, and this figure only shows

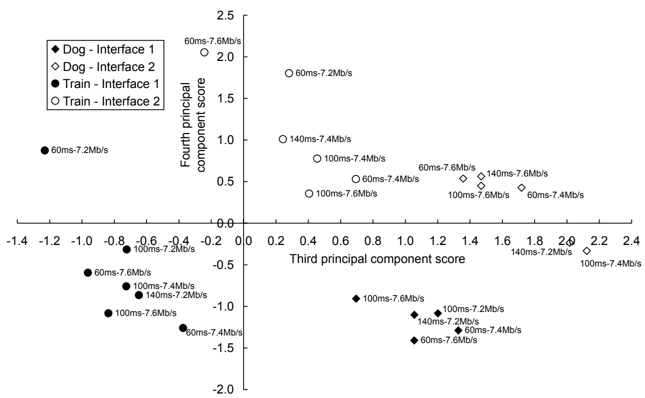


Figure 14. Third and fourth principal component scores.

the plots for the stimuli without additional delay.

In Fig. 14, we find that Dog has higher principal component scores for the third principal component than Train. We also notice that many plots for Interface 2 have positive scores of the fourth principal component. On the other hand, most of the fourth principal component scores for Interface 1 are negative values. Thus, the third and fourth components imply the characteristics of content and the usage of interface, respectively.

From the above observation, we notice that the main factors affecting QoE of MVV-A are the output quality, the interactivity, the characteristics of content, and the usage of interface.

V. Conclusions

We made experiments on MVV-A IP transmission with two user interfaces (Interface 1 and Interface 2) and two contents (Dog and Train). We assessed the effects of the IP traffic and the delay on QoE in a multidimensional way.

From the assessment results, we found that the users had high expectations when using Interface 2 since it is easier to use. However, when the load traffic is high and the playout buffering time cannot absorb the delay jitter properly, the MVV-A system cannot fulfill those expectations. When this happens, the usefulness of Interface 2 decreases more than that of Interface 1.

In addition, the speed of the movement of the object in the content and the similarity of the views of each camera can affect the QoE especially in the viewpoint change response.

From the result of PCA, we find the four main factors: the output quality, the interactivity, the characteristics of content, and the usage of interface.

The user interface and content affect the QoE of MVV-A system. In order to achieve high QoE, we need to employ a user interface that enables the user to change the viewpoint easily, fulfilling his/her expectations, i.e., following the object's movement of the content.

As future work, we will study QoE and the user's behavior with MVV-A systems for other kinds of contents and user interfaces, e.g., more intuitive interfaces with buttons corresponding to the cameras' physical locations. At the same time, we will use the audio of each camera instead of using a microphone.

Acknowledgments

The authors thank Makoto Yamamoto for his support in the experiment. This work was supported by the Grant-In-Aid for Scientific Research of Japan Society for the Promotion of Science under Grant 21360183 and the Grant-In-Aid for Young Scientists under Grant 23760332.

References

[1] ITU-T Rec. H.720, "Overview of IPTV terminal devices and end systems," Oct. 2008.

[2] T. E. Boult, R. J. Micheals, M. Eckmann, X.Gao, C. Power, and S. Sablak, "Omnidirectional video applications," *Proc. 8th International Symposium on Intelligent Robotic Systems*, July 2000.

[3] I. Ahmad, "Multiview video: Get ready for next-generation television," *Proc. IEEE Distributed Systems Online*, vol. 8, no. 3, art. no. 0703-o3006, Mar. 2007.

[4] M. Tanimoto, "FTV (Free viewpoint TV) and creation of ray-based image engineering," *ECTI TRANSACTIONS on Electrical Eng., Electronics, and Communications,* vol. 7, no. 2, Aug. 2009.

[5] F. Nielsen, "Surround video: A multihead camera approach," *The Visual Computer*, (21) 1-2, pp. 92-103, Feb. 2005.

[6] T. Ishikawa, K. Yamazawa, T. Sato, S. Ikeda, Y. Nakamura, K. Fujikawa, H. Sunahara, and N. Yokoya, "Networked telepresence system using Web browsers and omni-directional video streams," *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XII*, vol. 5664, Jan. 2005.

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

[8] ITU-T Rec. G.1080, "Quality of experience requirements for IPTV services," Dec. 2008.

[9] E. Kurutepe, M. R. Civanlar, and A. M. Tekalp, "Client-driven selective streaming of multiview video for interactive 3DTV," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1558-1565, Nov. 2007.

[10] G. Cheung, A. Ortega, and N. -M. Cheung, "Interactive streaming of stored multiview video using redundant frame structures," *IEEE Trans. on Image Processing*, vol. 20, no. 3, pp. 744-761, Mar. 2011.

[11] M. Ivory and M. Hearst, "The state of the art in automating usability evaluation of user interfaces," *ACM Computing Surveys*, vol. 33, no. 4, pp. 470-516, Dec. 2001.

[12] A. Lecerof and F. Paterno, "Automatic support for usability evaluation", *IEEE Transactions on Software Engineering*, vol. 24, no. 10, pp. 863-888, Oct. 1998.

[13] J. Lou, H. Cai, and J. Li, "A real-time interactive multiview video system," *Proc. ACM Multimedia 2005*, pp. 161-170, Nov. 2005.

[14] 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.

[15] T. Ishikawa, K. Yamazawa, and N. Yokoya, "Real-time networked telepresence by multi-casted omni-directional videos," *IPSJ SIG Tech. Rep.*, 2008-CVIM-163, May 2008.

[16] E. Jimenez Rodriguez, T. Nunome, and S. Tasaka, "QoE assessment of multi-view video and audio IP transmission," *IEICE Trans. on Commun.*, vol. E93-B, no. 6, pp. 1373-1383, June 2010.

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

[18] "NISTNET," http://snad.ncsl.nist.gov/nistnet/.

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

[20] 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.

[21] S. Tasaka and Y. Ito, "Psychometric analysis of the mutually compensatory property of multimedia QoS," *Conf. Rec. IEEE ICC 2003*, May 2003.