Video Transmission and Presentation Methods for Multi-View Video and Audio IP Transmission

Toshiro Nunome and Takuya Ishida
Department of Computer Science and Engineering, Graduate School of Engineering,
Nagoya Institute of Technology, Nagoya 466–8555, Japan
nunome@nitech.ac.jp, i-takuya@inl.nitech.ac.jp

***Abstract*—In this paper, we compare three transmission and presentation methods in the same total bitrate for Multi-View Video and Audio (MVV-A) IP transmission in terms of QoE. We use an MVV-A system with four cameras, i.e., the system has four viewpoints. The three transmission and presentation methods are the one-view one-stream method, the one-view four-streams method and the four-views four-streams method. The one-view one-stream method transmits and displays only a viewpoint selected by a user with high bitrate. The one-view four-streams method transmits all the viewpoints with low bitrate and displays only a viewpoint selected by the user. The four-views four-streams method transmits and displays all the viewpoints with low bitrate. QoE is evaluated by subjective experiment with 11 adjective pairs. As a result, under the situation that the user can change the viewpoint quickly, the one-view one-stream method can satisfy the user. On the other hand, under the situation that the user cannot change the viewpoint quickly, the user feels that the usability of the four-views four-streams method is good.**

Keywords*—*MVV, audio-video IP transmission, QoE, multidimensional assessment, video display method

I. Introduction

As a new type multimedia service over the Internet, Multi-View Video (MVV) [1], in which users can watch video from various viewpoints, has been achieving much attention. We can consider various applications of MVV such as entertainment, sports, sightseeing, and education among others.

When we threat multiple viewpoints, we can consider not only viewing selected a viewpoint but also viewing plural viewpoints at once. If the user can watch only a viewpoint, he/she can miss an important moment. On the other hand, if the receiver terminal displays plural streams simultaneously, he/she can view from many aspects at once and then does not miss the important moment.

To provide high *QoE (Quality of Experience)* is the ultimate goal of the network services. QoE represents the overall acceptability of an application or service, as perceived subjectively by the end-users [2]. QoE-based management of the network services is one of the important issues in the current network systems.

References [3] and [4] considers various displaying methods for multiple videos. However, the papers have no discussion on QoE of IP transmission with the effect of delay, jitter, and loss.

In [5] and [6], QoE of MVV-A (MVV with Audio) IP transmission is assessed multidimensionally. The papers consider that the server transmits only one video stream selected by the user. In this case, the viewpoint change response will be quick as the playout buffering time decreases. However, the short buffering cannot absorb network delay jitter sufficiently, and then the output quality of audio and video degrades. In addition, the viewpoint change response is affected by the end-to-end delay between the server and the client.

On the other hand, Yamamoto *et al*. have assessed QoE of three simultaneous transmission methods of multi-view video [7]. However, the user in the paper watch only one viewpoint selected by him/her.

In practical MVV-A IP transmission, transmission methods for efficient network resource usage and display methods for the user are important factors affecting QoE. However, the study which jointly assess the effect of transmission methods and display methods on QoE have not seen yet.

In this paper, we investigate how the transmission methods and display methods enhance QoE of MVV-A IP transmission under limited network resource; we regulate total encoding bitrate of MVV-A. We consider an MVV-A system with four cameras, i.e., the system has four viewpoints. We then pick up three fundamental transmission and presentation methods: the one-view one-stream method, the one-view four-streams method, and the four-views four-streams method. The one-view one-stream method transmits and displays only a viewpoint selected by the user with high bitrate. The one-view four-streams method transmits all the viewpoints with low bitrate and displays only a viewpoint selected by the user. The four-views four-streams method transmits and displays all the viewpoints with low bitrate. The three methods exploit trade-off relationships among quality, response, and usability.

The remainder of this paper is organized as follows. Section II introduces the MVV-A system with multiple displays. Section III describes the experimental method. Section IV presents experimental results. Section V concludes this paper.

II. MVV-A system with multiple displays

MVV-A is a system in which the user can watch the video from various viewpoints while he/she chooses the viewpoints arbitrarily. It provides high flexibility of the service for the user.

Figure 1 shows an overview of the content employed in this paper. In the assessment, we ask assessors to follow movement of a toy train running on plastic rails.

For displaying methods, we employ the one-view method and the four-views method. In the one-view method, the receiver terminal displays only one viewpoint as shown in Fig. 2. The assessors select a viewpoint from the four viewpoints. This is the method employed in [5]-[7]. The user employs a user interface as shown in Fig. 4; it is shown as a small window

Fig. 1. Overview of content

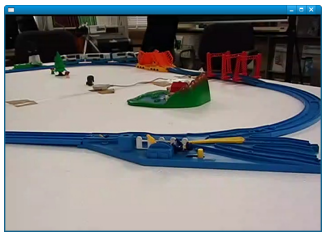

Fig. 2. One-view method

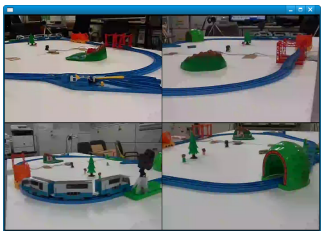

Fig. 3. Four-views method

on the display. The user can move this window to a desired position and can change the viewpoint by using the mouse.

In the four-views method, the receiver terminal displays all the four viewpoints with quarter image size of the one-view method for each stream. The display image of this method is shown in Fig. 3. In this method, the user does not need to change viewpoints.

III. Experimental Method

A. System

Figure 5 shows the configuration of the experimental system. Media Server is the server of MVV-A, and Media Receiver is the client. Load Server is the server of the load traffic, and Load Receiver is the client. Both Router 1 and Router 2 are Riverstone's RS3000. Between each router are connected by a full duplex Ethernet line of 10 Mb/s. All the other links are 100 Mb/s Ethernet.

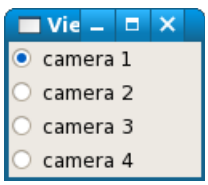

Fig. 4. User interface for viewpoint change

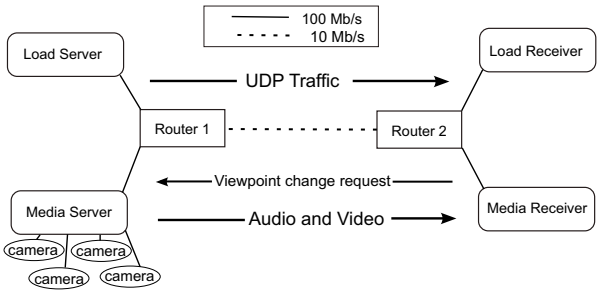

Fig. 5. Network configuration

Four SONY HDR-CX180 video cameras with the standard definition mode are connected to Media Server, which is equipped with real-time H.264 encoding boards. Media Server captures the video of each camera. At the same time, the audio is captured by a microphone. Media Server sends the audio and video to Media Receiver by using UDP packets. Media Receiver receives these packets and outputs the audio and video decoded from them. In the one-stream method, Media Receiver can choose one viewpoint from the four cameras by sending a request with SUBSCRIBE method of SIP (Session Initiation Protocol). In the four-streams method, Media Server transmits all the four video streams to the Media Receiver.

Load Server generates UDP datagrams of 1480 bytes each with exponentially distributed interval and sends them to corresponding Load Receiver.

B. Audio and video specification and experimental parameter

The specifications of audio and video are shown in Table I. We refer to the transmission unit at the application-level as an *MU (Media Unit)*. A video MU is a video frame and an audio MU is 320 audio samples. Each MU is transmitted as a UDP packet. We employ frame skipping as the output method of video. That is, when some packets consisting of an MU is lost, output of the MU is skipped.

We describe the three methods in Table II. The two types of video encoding bitrate are employed: 1 Mb/s and 4 Mb/s. For one-stream method, we use 4 Mb/s. On the other hand, for four-streams method, we use 1 Mb/s for each video stream. Thus, the total bitrate of the method is 1 Mb/s $\times$ 4 = 4 Mb/s. In the one-stream method, the user can see higher quality video than the four-streams method because of high encoding bitrate. However, in this method, when the user want to change the viewpoint, he/she experiences viewpoint change delay. On the other hand, in the four-streams method, the user does not suffer viewpoint change delay because all the streams are transmitted; however, the video quality is lower than the one-stream method.

In this paper, playout buffering control is used for absorbing delay jitter in Media Receiver. In the MVV-A system, playout buffering control brings trade-off between the viewpoint change response and output quality [5]. In order to investigate the effect of the playout buffering time, we employ five values: 60, 100, 150, 300, and 500 ms.

We assume two values of the average amount of UDP load traffic: 3.8 Mb/s and 5.4 Mb/s. They are selected on

TABLE I. SPECIFICATIONS OF AUDIO AND VIDEO

	video	audio
coding method	H.264	G.711 μ-law
picture pattern	I	-
coding bitrate	4 [Mb/s] (one-stream method) 1 [Mb/s] (four-streams method)	64 [kb/s]
picture size	704 $\times$ 480 (one-view method) 352 $\times$ 240 (four-views method)	-
MU rate	30 [MU/s]	25 [MU/s]
duration	20 [sec]	

TABLE II. TRANSMISSION AND PRESENTATION METHODS

method	bitrate of transmitted video stream	picture size
one-view one-stream	4 [Mb/s] $\times$ 1	704 $\times$ 480
one-view four-streams	1 [Mb/s] $\times$ 4	704 $\times$ 480
four-views four-streams	1 [Mb/s] $\times$ 4	four 352 $\times$ 240

the basis of [8], which reveals that the amount of daytime traffic is about 70 % of that of nighttime traffic in access ISP networks. We have realized a situation in which congestion sometimes occurs between the two routers in Fig. 5 on the nighttime traffic condition; as considering this situation, we set the average amount to 5.4 Mb/s. The amount of daytime traffic is selected to be 3.8 Mb/s, which is about 70 % of 5.4 Mb/s.

C. QoE assessment methods

In this paper, we assess QoE of MVV-A in a similar way as the methodology in [6].

We ask the users watch the running toy train in the assessment. We employ two kinds of average load, five kinds of playout buffering time, and the three transmission and presentation methods. In total, we consider 30 stimuli obtained by these combinations. Before the experiment, assessors have confirmed a work flow of the task through practices. The total assessment time for an assessor is about 40 minutes including the practices and experimental runs. We employed 20 male students in their twenties as assessors.

In the experiment, we perform multidimensional QoE assessment with the *SD (Semantic Differential)* method [9]; it is a technique for evaluating an object from many aspects by means of many pairs of polar terms. The pairs of polar terms in the subjective experiment are shown in Table III. The pairs are classified into five categories; there are four pairs for video, a pair for audio, four pairs for psychology, a pair for response, and a pair for overall satisfaction. Abbreviated names from v1 to o1 are attached to the pairs of polar terms.

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* [10]. 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.

TABLE III. PAIRS OF POLAR TERMS

category	pair of polar terms
Video	v1: The video is smooth - rough v2: The video is comfortable - jarring v3: The video is sharp - blurred v4: The video is powerful - poor
Audio	a1: The audio is natural - artificial
Psychology	p1: I feel free - restricted p2: I feel comfortable - uncomfortable p3: I feel powerful - well-behaved p4: I feel simple - difficult
Response	r1: The viewpoint change response is fast - slow
Overall satisfaction	o1: Excellent - Bad

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 we assess 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 stimulus is placed in the category. With the law of categorical judgment [10], we can translate the frequency obtained by the rating scale method into an interval scale. Since the law of categorical judgment is a suite of assumptions, we must test goodness of fit between the obtained interval scale and the measurement result. Mosteller [11] proposed a method of testing the goodness of fit for a scale calculated with Thurstone's law of comparative judgment [10], which is one of psychometric methods. The method can be applied to a scale obtained by the law of categorical judgment. This paper uses Mosteller's method to test the goodness of fit. Once the goodness of fit has been confirmed, we refer to the interval scale as the psychological scale; it is a QoE metric.

IV. EXPERIMENTAL RESULTS

A. Application-level QoS

Figure 6 show the MU loss ratio of audio and that of video, respectively. It is the ratio of the number of MUs not output at the recipient to the number of MUs transmitted by the sender for the displayed video streams. These figures show the MU loss ratio versus the playout buffering time for each amount of load traffic. In the following discussion, we call the amount of load traffic 3.8 Mb/s as lightly loaded condition and the amount of load traffic 5.4 Mb/s as heavily loaded condition.

We see in Fig. 6 that under the lightly loaded condition, the MU loss merely occurs. On the other hand, under the heavily loaded condition, we notice that the MU loss occurs for the playout buffering time 60 ms and 100 ms; the MU loss ratio decreases as the playout buffering time increases. This is because the small buffering time cannot absorb network delay jitter under the condition.

Figure 7 depicts the average viewpoint change delay for the one-view methods. It is defined as the time in seconds from the moment the user inputs a request for viewpoint change by the user interface until the instant a new viewpoint is output

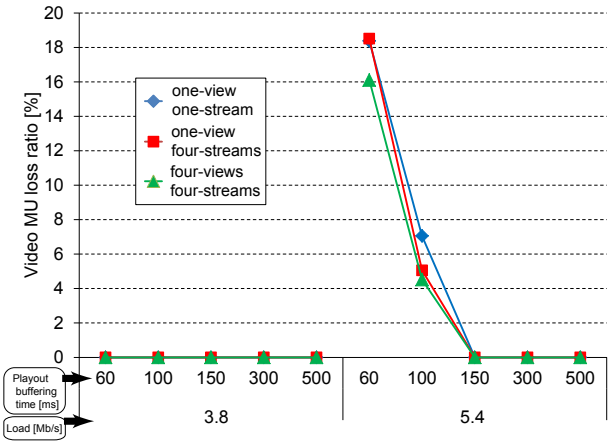

Fig. 6. Video MU loss ratio

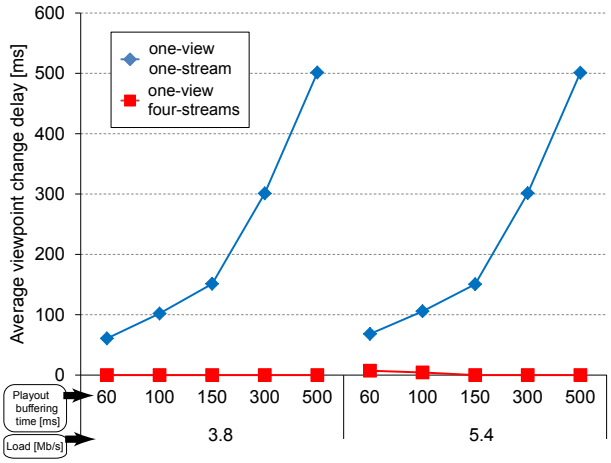

Fig. 7. Viewpoint change delay

at the client. This figure does not plot the results of the four-views four-streams method because the users do not change the viewpoint.

In Fig. 7, the viewpoint change delay is approximately the summation of the round trip delay between the server and the client and the playout buffering time. That is, the playout buffering time degrades the response of the viewpoint change.

On the other hand, in the one-view four-streams method, the viewpoint change delay is very small. This is because all the viewpoints are received simultaneously, and then the receiver can change the viewpoint without request to the server.

B. Psychological scale values

We calculated the interval scale for each criterion. We then carried out the Mosteller's test. As a result, we have found that the hypothesis that the observed value equals the calculated one can be rejected with significance level of 0.05 in some criteria. Therefore, we removed the stimuli which have large errors until the hypothesis cannot be rejected. In this paper, we use obtained values by these processes as the psychological scale.

Since we can select an arbitrary origin in an interval scale, for each criterion, we set the minimum value of the psychological scale to unity.

In this subsection, we picked up adjective pairs which represent characteristics of the three methods. Figure 8 shows the psychological scale for "The video is comfortable - jarring [v2]". Figure 9 depicts the psychological scale for "The video is powerful - poor [v4]". Figures 10 and 11 present the psychological scales for "I feel comfortable - uncomfortable [p4]" and "The viewpoint change response is fast - slow [r1]", respectively. The psychological scale of "I feel simple - difficult [p4]" is found in Fig. 12. In these figures, removed stimuli by the Mosteller's test are not shown.

We notice in Fig. 8 that the one-view one-stream method has the highest psychological scale values of [v2] and that the four-views four-streams method has the second highest. Thus, the user feels comfortable to see the high quality one-view video than the low-quality four-views video. In addition, for the same low bitrate video, the user prefers four-views than one-view.

We also notice in Figs. 8 and 9 that the psychological scale for [v4] has the same tendency as that for [v2]. Thus, we can confirm that the user feels higher presence with the high quality video.

In Fig. 10, we see that the psychological scale values of [p2] decrease as the playout buffering time increases under the lightly loaded condition for the one-view one-stream method. This is because the viewpoint change response degrades as we notice in Fig. 11; the increase of viewpoint change delay causes the user's uncomfortable.

On the other hand, in Fig. 10, we find that under the heavily loaded condition, the short buffering time cause the degradation of the psychological scale values for all the methods. Thus, the degradation of video output quality also degrades the user's comfortable.

As we compare the four-views four-streams method and the one-view one-stream method in Fig. 10, the four-views four-streams method achieves higher psychological scale for large buffering time. This is because the user does not suffer viewpoint change delay in the four-streams method, while the viewpoint change delay for the one-stream method increases as the playout buffering time increases.

We see in Fig. 12 that the four-views four-streams method has the highest psychological scale for [p4] among the three methods. This is because the user can watch all the viewpoints without operation of the viewpoint change interface in the four-views four-streams method.

C. Overall satisfaction

Table IV shows the correlation coefficient between the psychological scale of the overall satisfaction and that of each adjective pair except for [r1] in descending order. We notice in this table that [p2], [p4] and [v2], which are related to comfort and simplicity, are highly correlate with the overall satisfaction.

Figure 13 depicts the psychological scale of [o1]. At first, we focus on the results under the lightly loaded condition. The one-view one-stream method degrades the psychological

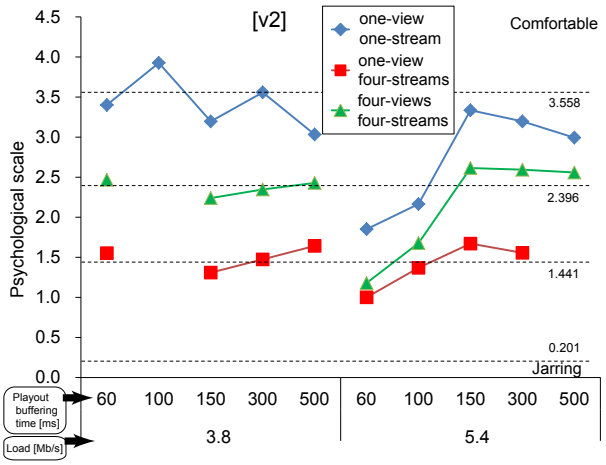

Fig. 8. v2: The video is comfortable - jarring

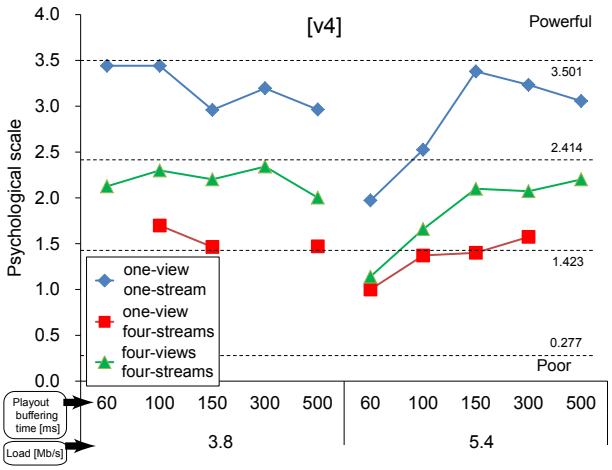

Fig. 9. v4: The video is powerful - poor

scale values as the playout buffering time increases. This is because the viewpoint change response degrades as the buffering time increases. When the playout buffering time is smaller than 300 ms, the one-view one-stream method has higher psychological scale than the four-views four-streams method; otherwise, the four-views four-streams method is the best. That is, the user prefers good picture quality under quick viewpoint change response. However, the user wants to see all the viewpoint at once when the response becomes slow. In addition, the one-view four-streams method is not preferred by the user.

Next, we focus on the heavily loaded condition. For the playout buffering time equal to or smaller than 150 ms, as the buffering time increases, the psychological scale values for all the methods increase. This is because the short buffering time cannot absorb delay jitter enough and then the skipping of output MUs occurs. Under the condition, the one-view one-stream method is the best, while the four-views four-streams method has the highest psychological scale for the buffering time 500 ms.

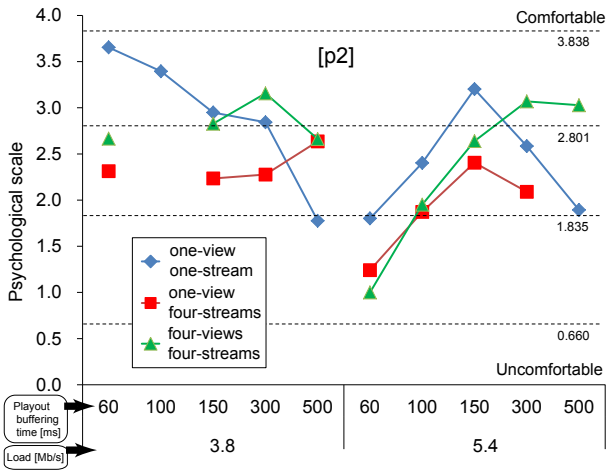

Fig. 10. p2: I feel comfortable - uncomfortable

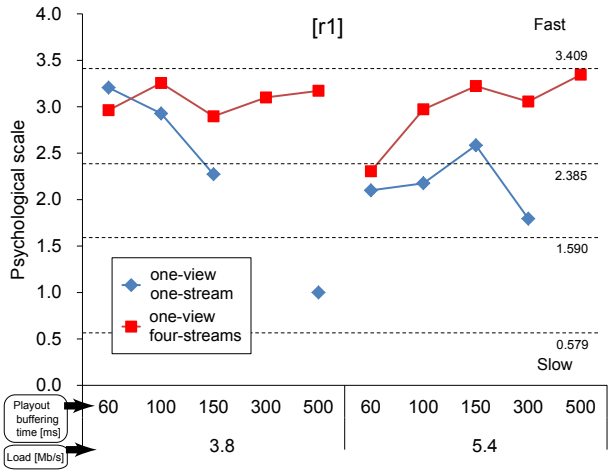

Fig. 11. r1: The viewpoint change response is fast - slow

V. Conclusions

In this paper, we compared the three transmission and presentation methods of MVV-A in terms of QoE; they are the one-view one-stream method, the one-view four-streams method, and the four-views four-streams method. As a result, we notice that the user prefers the one-view method with high bitrate to the multiple-views with low bitrate when the viewpoint change response is enough quick. On the other hand, when the server transmits the plural low bitrate video streams simultaneously, displaying multiple viewpoints is ef-

TABLE IV. Correlation coefficients

adjective pairs	coefficient
p2: I feel comfortable - uncomfortable	0.959
p4: I feel simple - difficult	0.832
v2: The video is comfortable - jarring	0.813
v4: The video is powerful - poor	0.751
a1: The audio is natural - artificial	0.751
v1: The video is smooth - rough	0.723
v3: The video is sharp - blurred	0.638
p1: I feel free - restricted	0.636
p3: I feel powerful - well-behaved	0.488

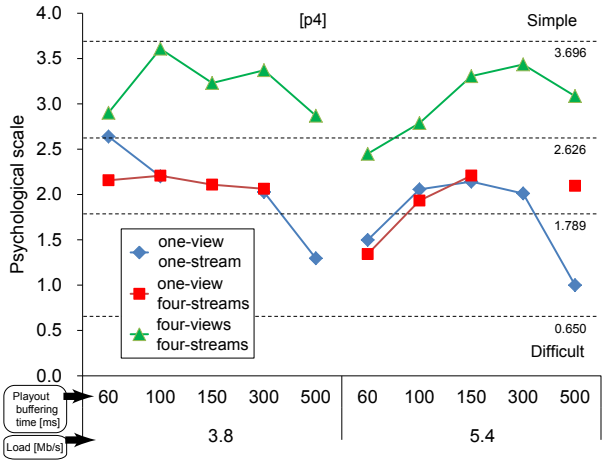

Fig. 12. p4: I feel simple - difficult

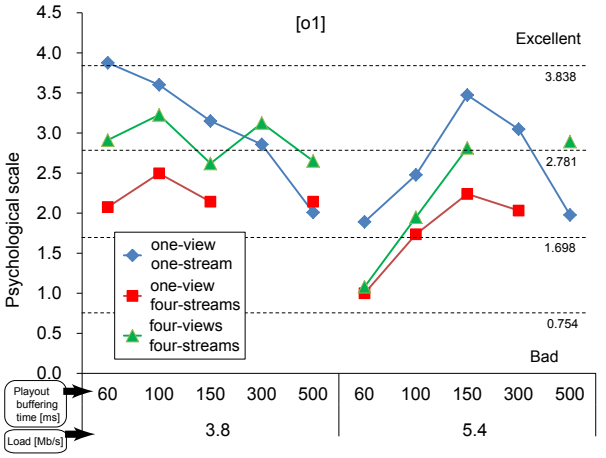

Fig. 13. o1: Excellent - Bad

fective than displaying a viewpoint with the viewpoint change function. Thus, the user's satisfaction can be enhanced by using appropriate transmit and display method according to the network condition.

In future work, we need to confirm the results in other contents. In addition, we will devise other display methods for further QoE enhancement. We will also consider the effect of user attributes on QoE.

ACKNOWLEDGMENT

We thank Professor Emeritus Shuji Tasaka for his valuable discussion. This work was supported by The Telecommunications Advancement Foundation.

REFERENCES

[1] I. Ahmad, "Multi-View Video: Get Ready for Next-Generation Television," *Proc. IEEE Distributed Systems Online*, vol. 8, no. 3, art. no. 0703-o3006, Mar. 2007.

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

[3] D. Ikeda and T. Naemura, "A study on an interface for simultaneous browsing of multiple videos," *Tech. Rep. IEICE*, HIP2004-81, Dec. 2004 (in Japanese).

[4] T. Porat and J. Silbiger and M. Rotem-Hovev, "Switch and deliver: Display layouts for MOMV (Multiple Operator Multiple Video feed) environment," *Proc. IEEE CogSIMA 2011*, pp. 264-267, Feb. 2011.

[5] E. Jimenez Rodriguez, T. Nunome and S. Tasaka, "QoE assessment of multi-view video and audio IP transmission," *IEICE Trans. Commun.*, vol. E92-B, no. 6, pp. 1373-1383, June. 2010.

[6] E. Jimenez Rodriguez, T. Nunome, and S. Tasaka "Multidimensional QoE assessment of multi-view video and audio (MVV-A) IP transmission: The effects of user interfaces and contents," Proc. Advanced Information Networking and Applications Workshops (WAINA), pp. 91-98, Mar. 2012.

[7] M. Yamamoto, T. Nunome and S. Tasaka, "QoE assessment of simultaneous transmission methods of multi-view video and audio IP transmission," *Tech. Rep. IEICE*, CQ2012-15, Apr. 2012 (in Japanese).

[8] K. Cho, K. Fukuda, H. Esaki and A. Kato, "Observing slow crustal movement in residential user traffic,"*Proc. ACM CoNEXT2008*, Dec. 2008.

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

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

[11] 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, June 1951.