

Multidimensional Assessment and Principal Component Analysis of QoE in Interactive Multi-View Video and Audio IP Communications

Takayoshi Ichikawa, Toshiro Nunome and Shuji Tasaka
Department of Computer Science and Engineering,
Graduate School of Engineering,
Nagoya Institute of Technology
Nagoya 466-8555, Japan

Abstract—This paper compares interactive Multi-View Video and Audio (MVV-A) IP communications and interactive Single-View Video and Audio (SVV-A) ones in terms of QoE. We employ a task which makes use of characteristics of interactive MVV-A and perform a subjective experiment. We conduct multidimensional assessment and principal component analysis of QoE with 14 pairs of polar terms by the SD (Semantic Differential) method, which can assess an object from many viewpoints. We also assess the efficiency of the task objectively. As a result, we see that the users are more satisfied with MVV-A than SVV-A owing to viewpoint change. In addition, we show that the overall satisfaction improves when we select proper playout buffering time. From the result of the principal component analysis, we find that the first principal component is audiovisual quality and user's feeling, and the second principal component is viewpoint change.

Index Terms—multi-view video, QoE, multidimensional assessment, interactive IP transmission

I. INTRODUCTION

With the acceleration of transmission speed of IP networks, multimedia communications treating audio and video become popular. In particular, interactive communication services on the IP networks such as videophone have been widely spreaded.

The IP networks are generally best-effort. Audio and video packets can be lost during transmission and can be affected by network delay jitter. These impairments deteriorate the output quality of audio and video; then, *QoS (Quality of Service)* becomes lower.

Owing to the layered structure of the IP networks, the QoS also has a hierarchical structure. In the hierarchical structure, the top level QoS is called user-level QoS [1]; it is perceptual quality of the user. *ITU-T (International Telecommunication Union Telecommunication Standardization Sector)* refers to the user-level QoS as *QoE (Quality of Experience)* [2]. QoE is the most important because users are final recipients of the services.

In television broadcasting and video streaming sites in the Internet, users can watch a predetermined viewpoint given by the sender. All the users are not satisfied with the viewpoint. Therefore, *MVV (Multi-View Video)* [3], where the user can choose one video from multiple video streams of the same event, has been studied. However, previous studies on MVV mainly focus on encoding such as *MVC (Multi-view Video Coding)* [4], and then few studies assess QoE of MVV in IP communications.

In [5], Jimenez Rodriguez *et al.* assess the influence of playout buffering time and *GUI (Graphical User Interface)* for viewpoint change on QoE of *MVV-A (MVV and Audio)*, which is MVV accompanied by audio. In [6], Yamamoto *et al.* use the *SD (Semantic Differential)* [7] method to assess the influence of camera arrangements on QoE multidimensionally.

In [5] and [6], they assume asymmetrical real-time transmission, in which video and audio are transmitted from the server to the client. On the other hand, there are symmetrical interactive communications between client terminals like videophone. MVV-A can provide higher experience for the users in symmetrical interactive applications as in asymmetrical ones because the users can change the viewpoint which they want to watch. However, symmetrical interactive IP communications are not considered in [5] and [6].

In the symmetrical interactive applications with MVV-A, we must consider two types of interactivity. One is the viewpoint change response, and the other is interactivity of communication. However, no study assumes interactivity of human communications in MVV-A or MVV; the previous studies on MVV such as [3] and [8] consider the interactivity for viewpoint change only.

As for interactivity of human communications, various kinds of task can be assumed. The study of interactive communications using *SVV-A (Single-View Video and Audio)* [9], [10] employs calculation of numerals written on cards, picture comparison, counting numbers, and clapping once for each counting number as tasks for QoE assessment. However, these tasks do not assume MVV-A. For that reason, these tasks are not suitable for QoE assessment of MVV-A.

In this paper, we enhance the MVV-A system in [5] and [6] for symmetrical interactive applications; it is called *interactive MVV-A*. We devise a task to compare QoE of MVV-A with that of SVV-A in the interactive IP communication. We multidimensionally assess QoE by the SD method in order to clarify the difference between MVV-A and SVV-A in terms of QoE. Also, we assess the main factors affecting QoE of MVV-A by *PCA (Principal Component Analysis)*. We have chosen pairs of polar terms for QoE assessment so that they can exhibit principal components clearly. Through the assessment, we quantitatively show that MVV-A enhances QoE of the symmetrical interactive IP communication; it is the main objective of this paper.

The rest of this paper is structured as follows. Section II describes interactive MVV-A. Section III explains the environment of the experiment. Section IV outlines the method of QoE assessment. We show results of the experiment in Section V, and Section VI concludes this paper.

II. INTERACTIVE MULTI VIEW-VIDEO AND AUDIO

An MVV-A system [5], [6] consists of one server and at least one client that are connected to the IP networks. At the same time, several cameras are connected to the MVV-A server. The server captures the video of each camera. At the same time, the audio is captured by using at least one microphone. In this paper, we add symmetrical interactivity to this system.

In this paper, we consider interactive MVV-A between two users. Multiple cameras and a microphone are connected to

each media terminal in order to capture audio and video. Also, a headphone is connected to the terminal as an output device of audio. The two media terminals transmit audio and video streams with the capability of viewpoint change. Each user can select a viewpoint from the multiple viewpoints of the other end.

In MVV, the terminal transmits the video streams of multiple viewpoints or only a video stream of demanded viewpoint. In the case where a sender terminal sends only one viewpoint to a receiver terminal, the user must wait more time in order to see the new viewpoint than in the case of sending the video streams of multiple viewpoints. The receiver terminal must send a request for viewpoint change to the other end first. However, as the sender terminal is sending only one audio-video stream, the amount of data through the network is comparatively small. On the other hand, in the case of sending multiple viewpoints simultaneously, the viewpoint can be immediately changed at the receiver terminal. However, the amount of data through the network is comparatively large and can vary depending on the number of cameras connected to the other end.

In this paper, we extend the MVV-A system used in [5] and [6] to the symmetrical interactive communication. Hence, each terminal transmits only a video stream of demanded viewpoint to the other end. As a factor affecting QoE in this system, not only response of viewpoint change in MVV-A but also interactivity of communication is important. In the symmetrical interactive communication, two users carry out some tasks between the users. The task efficiency is important in QoE assessment.

III. EXPERIMENTAL SYSTEM

Figure 1 shows the configuration of the experimental system. Each *MT* (*Media Terminal*) transmits and receives media streams for the MVV-A application. *LS* (*Load Server*) is the server of the load traffic, and *LR* (*Load Receiver*) is the client. NISTNET, which is a PC, is laid out between the routers. This PC delays packets going through routers 1 and 2 by using NISTNET. Both router 1 and router 2 are Riverstone's RS3000. Each router and NISTNET are connected by a full-duplex Ethernet line of 10 Mb/s. All the other links are 100 Mb/s Ethernet.

Each *MT* is equipped with four video cameras and a headset with a microphone. For video encoding, we use H.264 with real-time encoding boards.

In the experiment, the cameras are placed with the circular arrangement as shown in Figure 2. Dotted lines in Figure 2 show the field of view of each camera. A center circle in Figure 2 is an area to roll a dice, which is used in the task of the experiment and is shown in Figure 3. We set the focus of the four cameras inside the center circle during the experiment.

In the experiment, we compare MVV-A and SVV-A. In MVV-A, we can select a viewpoint from all the four cameras. On the other hand, in SVV-A, the viewpoint is fixed; i.e., we cannot change it.

Figure 4 shows the user interface for viewpoint change. The user chooses a viewpoint by clicking a radio button with a mouse. The initial viewpoint is camera 1 in Figure 2 for both MVV-A and SVV-A.

When the user requests viewpoint change, the user's terminal sends information for viewpoint change of 4 bytes to the other terminal. If the other terminal receives the information, it changes the viewpoint and transmits the corresponding video to the user's terminal.

Each *MT* is placed in a different room. The light of the experimental rooms are bright enough to conduct the experiment. The rooms are air-conditioned.

Table I shows the specifications of the audio and video. We refer to the transmission unit at the application-level as an

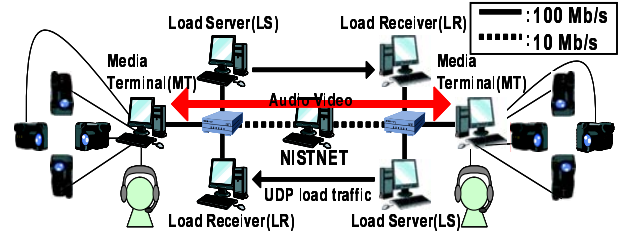


Fig. 1. System configuration

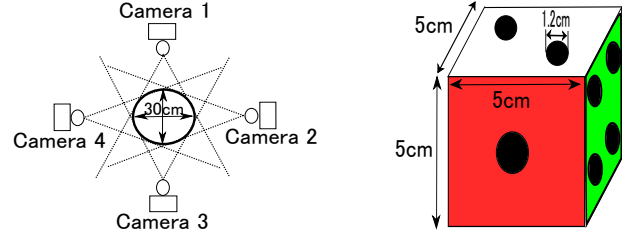


Fig. 2. Camera arrangement

Fig. 3. Dice with six colors



Fig. 4. GUI

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.

A receiver terminal outputs an audio *MU* and a video *MU* after playout buffering. We set the buffering time to 60 ms, 100 ms, 150 ms, 300 ms, and 500 ms. We employ frame skipping as the output method of video. That is to say, when some packet for an *MU* is lost, output of the *MU* is skipped.

Each *LS* generates UDP packets of 1472 bytes each with exponentially distributed interval and sends them to corresponding *LR*. We assume two kinds of the average amount of UDP load traffic: 5.2 Mb/s and 7.5 Mb/s; they are based on [11], which reveals that the amount of daytime traffic is about 70 % of that of nighttime traffic. We have realized a situation in which congestion sometimes occurs between the two routers in Figure 1 on the nighttime traffic condition; considering this situation, we set the average amount to 7.5 Mb/s. The amount of daytime traffic is selected to be 5.2 Mb/s, which is about 70 % of 7.5 Mb/s.

The NISTNET software adds a constant delay, which can emulate a large scale network. We set three values of the delay: 0 ms, 75 ms, and 150 ms. We assume the value of 0 ms as communications delay inside a city, the values of 75 ms and 150 ms as the latency of international communications from Japan to U.S.A. and U.K., respectively. These values have been selected from [12], where the one-way delay from Japan to U.S.A. has a wide distribution from 60 ms to 150 ms, and the first peak of the distribution can be found at around 75 ms; as for the delay from Japan to U.K., the peak of the distribution is around 150 ms.

IV. ASSESSMENT METHOD

A. Task

In this paper, we design a new task to measure effectiveness of interactive MVV-A. Tasks which are performed in the study on interactive communications with SVV-A are not suitable for QoE assessment of interactive MVV-A. In the experiment, we employ dice rolling as a task in order to compare MVV-A

TABLE I
AUDIO AND VIDEO SPECIFICATIONS

	video	audio
coding method	H.264	ITU-T G.711 μ -law
average MU rate [MU/s]	30	25
average bit rate [kb/s]	2000	64
picture pattern	I	-
image size [pixels]	704 \times 480	-
playing time [s]	40	

TABLE II
COLOR PATTERNS OF DICE

spots \ dice	1	2	3	4	5	6
dice 1	red	white	yellow	green	black	blue
dice 2	blue	red	white	yellow	green	black
dice 3	yellow	black	green	white	blue	red
dice 4	black	green	red	blue	white	yellow
dice 5	green	yellow	blue	black	red	white

with SVV-A in terms of QoE. With MVV-A, users can see a dice from more viewpoints than with SVV-A; this advantage of MVV-A can enhance QoE.

We employ five dice for each subject. Each dice uses six colors (red, blue, yellow, green, black, and white). The color of each side of the dice is different as shown in Figure 3, and the combinations of spots and colors are different from each other. Table II shows color patterns of dice. If the color of side is black, the color of spots is white; otherwise, the color of spots is black.

The subjects perform the task in pairs. One of two subjects is called subject 1, the other subject is called subject 2 in this paper.

procedure: Subject 1 chooses a dice from among the five ones and a color from among the six ones. After confirming the number of the spots corresponding to the color, subject 1 tells the color to subject 2. Subject 2 answers “yes” if he/she understands the color which was chosen by subject 1, otherwise subject 2 answers “no”. If subject 2 answers “no”, subject 1 tells the color again. In the case where subject 1 hears the answer “yes”, subject 1 rolls the dice into the center circle in Figure 2. Subject 2 sees the side of the chosen color through the video and answers the number of spots to subject 1. Finally, subject 1 tells subject 2 whether the answer is correct or not. We assume a series of actions as a task.

Also, in the experiment, the subjects measure the time for a task. We employ a button on the right side of GUI shown in Figure 4. When subject 1 rolls the dice and confirms that the dice stops, he/she pushes the button to start measurement of the time. After subject 1 hears the answer from subject 2, he/she pushes the button again to finish measurement. The notation of the button is changed from “START” to “STOP” when subject 1 pushes the button to start measurement. When subject 1 pushes the button again to finish measurement, the notation of the button returns from “STOP” to “START”. Subject 1 retrieves the dice when subject 2 chooses one of his/her own dice and a color out of the six. After that, the subjects change their roles and repeat the task for 40 seconds.

In MVV-A, when subject 2 cannot answer the number of spots even after more than about ten seconds, subject 2 tells subject 1 “I cannot understand”. In SVV-A, when subject 2 is asked the color of side which he/she cannot see from the viewpoint, or when subject 2 cannot answer the number of spots even after more than about five seconds, subject 2 tells subject 1 “I cannot understand”.

B. Objective assessment

We perform objective assessment to measure efficiency of a task. For objective assessment, we employ three parameters. *The average task completion time* is the average time until a subject answers the number of spots of dice after throwing the dice. *The average number of performed tasks* is the average number of tasks that the subjects performed for 40 seconds. *The ratio of correct answers* is the ratio of the number of correct answers to the number of tasks that the subjects performed for 40 seconds.

C. QoE assessment

QoE is assessed multidimensionally with the SD method. The SD method can assess an object for evaluation, which is called 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., warm and cool. In this method, we can assess QoE in detail. Table III shows the polar terms used in this paper. The polar terms are classified into six categories. In Table III, “v” means video, “a” audio, “r” response, “s” synchronization, “t” task, and “p” psychology. Note that this experiment was performed in Japanese. This paper has translated the used Japanese terms into English. Therefore, the meaning of adjectives written in English here may slightly differ from those of Japanese one.

Also, for each selected pair of polar terms, a subjective score of an object for evaluation is measured by the rating scale method with five grades. The best grade (score 5) represents the positive adjective (left or upper side one in each pair in Table III). The worst grade (score 1) means the negative adjective (right or lower side one). The middle grade (score 3) is neutral. The scores 4 and 2 show slightly positive and slightly negative, respectively. The subjects assess quality expressed by pairs of polar terms with an assessment GUI, which is displayed on the screen, whenever an experiment is finished.

The rating scale method is also used to measure MOS (Mean Opinion Score), which is widely utilized for assessment of a single medium. In the rating scale method, assessors classify each stimulus into one of a certain number of categories. Each category has a predefined number, i.e., a score. However, 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* [13], 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 [14] proposed a method of testing the goodness of fit for a scale calculated with Thurstone’s law of comparative judgment [13], 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.

In this paper, we employ two kinds of average bit rate of load traffic, five kinds of playout buffering time, three kinds of fixed additional delay, and two types, i.e., MVV-A and SVV-A. In total, we consider 60 stimuli obtained by these combinations. Each subject assessed the 60 stimuli in two terms. In a term, he/she evaluated 30 stimuli and additional four dummies. Before the experiment, subjects practiced and confirmed a workflow of the task. The total assessment time

TABLE III
PAIRS OF POLAR TERMS

ID	polar terms
v1	The video is smooth - The video is rough
v2	The video is sharp - The video is blurred
v3	The video is easy to see - The video is hard to see
a1	The audio is natural - The audio is artificial
a2	The audio is clear - The audio is dull
r1	The viewpoint change response is rapid - The viewpoint change response is slow
s1	The audio and video are in synchronization - The audio and video are out of synchronization
t1	The task is easy - The task is difficult
t2	The communication is natural - The communication is unnatural
p1	I feel relaxed - I feel frustrated
p2	I feel free - I feel constrained
p3	I feel fulfilled - I feel impatient
p4	I feel easy - I feel troublesome
p5	Excellent - Bad

of a pair is about 120 minutes including the practice, task, and assessment. For our experiment, we employed 20 male students in their twenties as subjects.

V. EXPERIMENTAL RESULTS

A. Objective assessment

Figure 5 shows the average task completion time as a function of the playout buffering time. Figure 6 indicates the average number of performed tasks. Figure 7 depicts the ratio of correct answers. These figures show the results in fixed additional delay 0 ms.

In Figures 5, 6 and 7, squares show the results of MVV-A, and diamonds for SVV-A. As for the UDP load traffic, black symbols represent 5.2 Mb/s, while gray symbols show 7.5 Mb/s.

In Figure 5, we find that the average task completion time in MVV-A is longer than that in SVV-A. When the subject cannot see spots for the answer in MVV-A, he performs viewpoint change; then, it takes more time for a task in MVV-A than that in SVV-A.

Next, in Figure 5, we notice that when the playout buffering time is 60 ms and 100 ms, the average task completion time on the UDP load traffic 7.5 Mb/s is longer than that on the traffic 5.2 Mb/s. This is because the MU loss due to congestion causes pause of audio and video.

We find in Figure 6 that the average number of performed tasks in SVV-A is the same level as or larger than that in MVV-A. This is because it takes longer time for MVV-A to answer owing to viewpoint change.

In Figure 7, we notice that the ratio of correct answers in MVV-A is higher than that in SVV-A regardless of the playout buffering time and the UDP load traffic. In this figure, MVV-A has the ratio of correct answers of about 90 %, and the ratio in SVV-A is about 50 %. The number of sides which the subject can watch from a viewpoint is two or three. For this reason, the ratio of correct answers can be approximately from $\frac{1}{3}$ to $\frac{1}{2}$ in SVV-A.

B. Psychological scale

From among the pairs of polar terms shown in Table III, we focus on r1, t1, p2, and p5 because they are related to the difference of MVV-A and SVV-A. 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

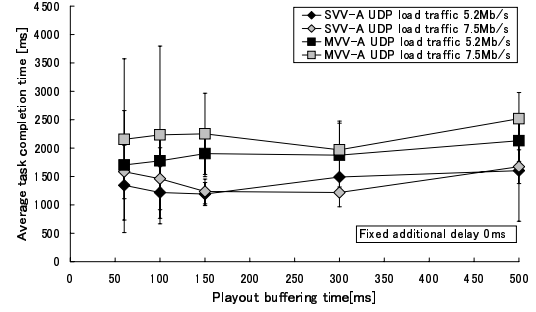


Fig. 5. Average task completion time (fixed additional delay 0 ms)

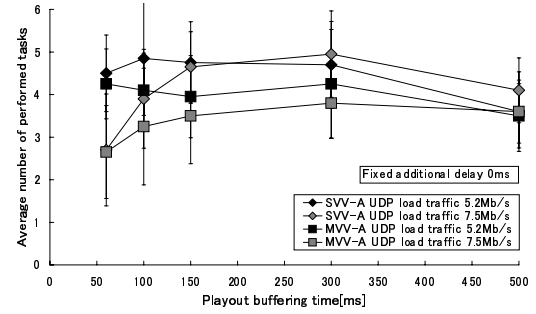


Fig. 6. Average number of performed tasks (fixed additional delay 0 ms)

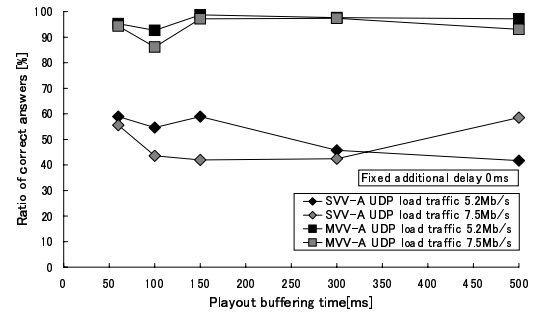


Fig. 7. Ratio of correct answers (fixed additional delay 0 ms)

a significance level of 0.05. 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.

Figure 8 shows the psychological scale for viewpoint change response (r1). Figure 9 indicates that for difficulty of task (t1). Figure 10 depicts that for freedom (p2). Figure 11 shows that for overall satisfaction (p5).

Each of Figures 8 through 11 shows the results for the three fixed additional delay in NISTNET: 0 ms, 75 ms, and 150 ms. These figures depict the assessment results of MVV-A and those of SVV-A for the two UDP load traffic values: 5.2 Mb/s or 7.5 Mb/s. SVV-A is excepted in Figure 8. The abscissa is the playout buffering time: 60 ms, 100 ms, 150 ms, 300 ms, and 500 ms.

In Figure 8, we notice that for each fixed additional delay, the psychological scale value has a peak around the playout buffering time of 100 ms on the UDP load traffic of 5.2 Mb/s. When the UDP load traffic is 7.5 Mb/s, the user feels that the viewpoint change response is fast with the playout buffering time of 150 ms or 300 ms. This is because these values of the playout buffering time can decrease MU loss, and then the viewpoint change delay is not felt long. On the other hand, the playout buffering time of 60 ms cannot absorb the delay

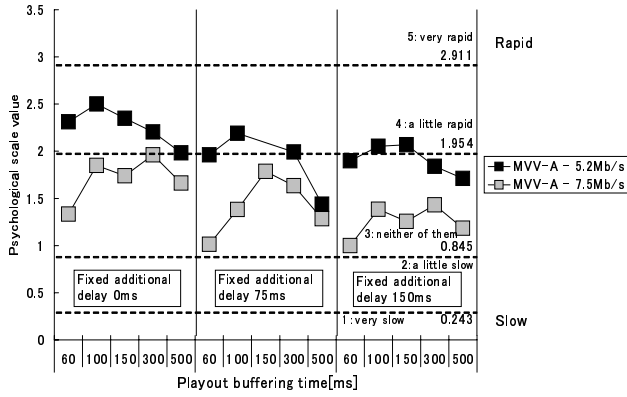


Fig. 8. Psychological scale for viewpoint change response (r1)

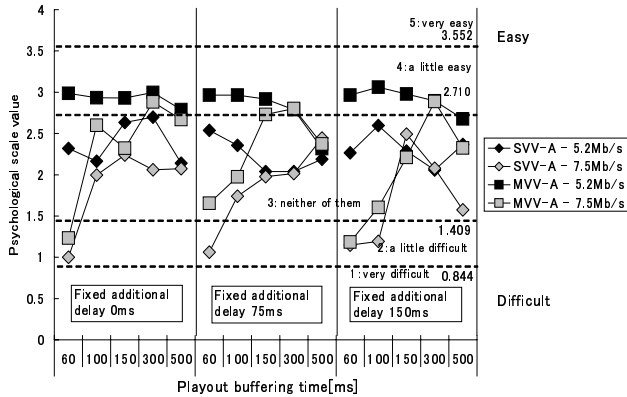


Fig. 9. Psychological scale for difficulty of task (t1)

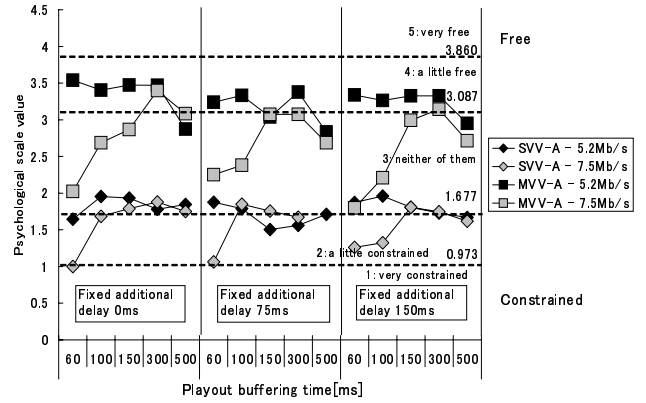


Fig. 10. Psychological scale for freedom (p2)

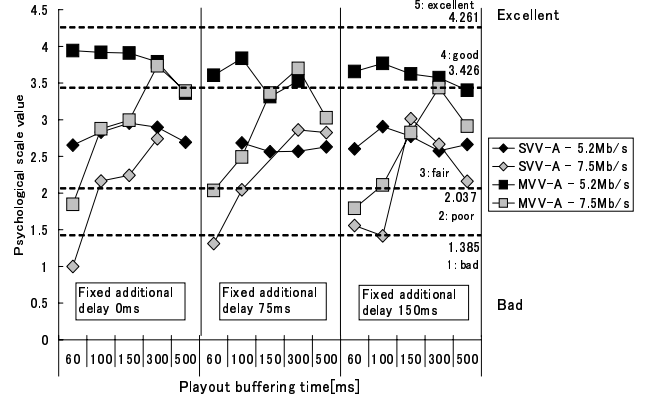


Fig. 11. Psychological scale for overall satisfaction (p5)

jitter; then, the number of packets which are not in time for the scheduled output time increases. Therefore, the new viewpoint cannot be shown immediately, and then the subject feels that the change is slow.

Also, we find in Figure 8 that as the fixed additional delay increases, the psychological scale values for viewpoint change response decrease. This is because the viewpoint change delay becomes large as the additional delay increases.

In Figure 9, we notice that when the UDP load traffic is 7.5 Mb/s in MVV-A, the psychological scale value is the highest with the playout buffering time of 300 ms. This is because the buffering time is not so large as to affect the perceptual output delay of audio and video and is long enough to absorb the jitter. On the other hand, the value decreases for the playout buffering time larger than 300 ms because the output delay of audio and video becomes long.

We can observe in Figure 9 that when the UDP load traffic is 7.5 Mb/s for the playout buffering time of 60 ms, the psychological scale values of both MVV-A and SVV-A are small. Since MU loss occurs frequently on this traffic with the short buffering time, output quality of audio and video deteriorates. Because of this, the user feels that the task is difficult.

In Figure 10, we find that the psychological scale value in MVV-A is higher than that in SVV-A for all the experimental parameters considered here. Because the subjects can freely select viewpoints from four cameras in MVV-A, they easily feel freedom in MVV-A.

Also, in Figure 10, for the playout buffering time shorter than 300 ms with the UDP load traffic of 7.5 Mb/s, the psychological scale value for MVV-A decreases as the playout buffering time decreases. This is because the subject feels constrained since the short buffering time cannot absorb the

jitter enough, and the output quality of audio and video degrades by MU loss.

In Figure 11, we notice that under most of the experimental conditions considered here, the psychological scale value in MVV-A is larger than that in SVV-A. In MVV-A, the subject can distinguish all sides of the dice owing to the viewpoint change. Thus, the overall satisfaction of the subject can be enhanced. When the playout buffering time is 300 ms, which can absorb the delay jitter enough, the psychological scale value is the maximum in MVV-A with the UDP load traffic of 7.5 Mb/s.

Also, we find that the overall satisfaction in MVV-A decreases as the fixed additional delay increases. The reason is that the fixed additional delay increases the viewpoint change delay.

In Figure 11, for the UDP load traffic of 5.2 Mb/s in MVV-A, we see that the playout buffering time of 60 ms and that of 100 ms have higher psychological scale values than the other values. This is because the viewpoint change response becomes rapid with these values of the playout buffering time under this load condition.

C. Principal component analysis

We conducted principal component analysis for psychological scale values except for the overall satisfaction (p5). As a result, the contribution rate of the first principal component is 78.713 %. The cumulative contribution rate of the first two principal components is 93.229 %. We then employ the two principal components in this paper.

Figure 12 shows the principal component loading values. The abscissa is the first principal component, and the ordinate is the second principal component. We notice in Figure 12 that the first principal component has a positive correlation with

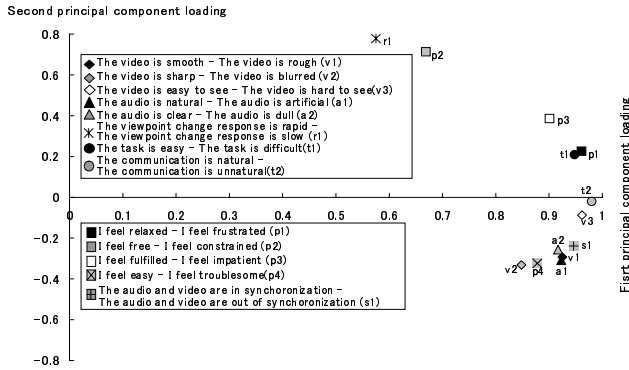


Fig. 12. Principal component loading values

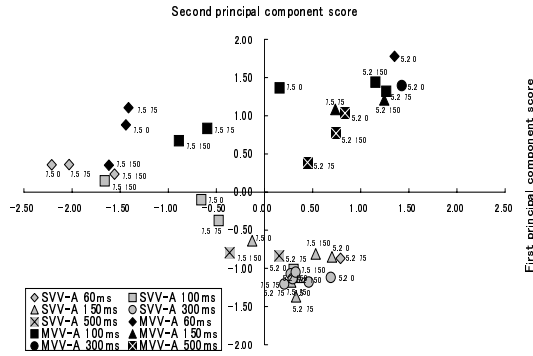


Fig. 13. Principal component score

all the pairs of polar terms. The first principal component is strongly associated with the naturalness of the task ($t2$) and the easiness to see the video ($v3$). Thus, we can interpret the first principal component as audiovisual quality and user's feeling.

Also, in Figure 12, the second principal component is strongly associated with the viewpoint change response ($r1$) and the freedom ($p2$). Thus, we can notice that the second principal component is viewpoint change.

Figure 13 shows the principal component score. The abscissa is the first principal component, and the ordinate is the second principal component. In Figure 13, black symbols show results of MVV-A, and gray symbols show results of SVV-A. As for the playout buffering time, diamonds are for 60 ms, squares show 100 ms, triangles represent 150 ms, circles are for 300 ms, and crosses show 500 ms. The figure accompanying each symbol show the combination of the UDP load traffic [Mb/s] and fixed additional delay [ms]. Also, we do not show the stimuli excluded by the Mosteller's test.

In Figure 13, in both MVV-A and SVV-A, the first principal component score on the UDP load traffic of 5.2 Mb/s is higher than that on the UDP load traffic of 7.5 Mb/s. With heavy load traffic and short playout buffering time, output quality of audio and video deteriorates, and then the subject feels that the task is hard to accomplish. Therefore, the first principal component score becomes small.

On the other hand, when the playout buffering time is 300 ms, the first principal component score is positive even in the UDP load traffic of 7.5 Mb/s. This is because the playout buffering time can absorb delay jitter enough.

From the above observation, we can confirm that the first principal component is audiovisual quality and user's feeling.

In Figure 13, we notice that all the stimuli related to MVV-A have positive scores of the second principal component, and many stimuli related to SVV-A have negative scores. Also, the second principal component score of MVV-A becomes high

for the playout buffering time of 60 ms and that of 100 ms with the UDP load traffic of 5.2 Mb/s since the viewpoint change delay is short. Thus, we find that the second principal component is the viewpoint change as we have already seen, and then the viewpoint change is one of the important factors affecting QoE in interactive MVV-A.

VI. CONCLUSIONS

In this paper, we conducted multidimensional assessment of QoE in interactive MVV-A. We employed dice rolling as a task to compare the difference between MVV-A and SVV-A from a viewpoint of QoE. As a result, we find that the psychological scale values in MVV-A are higher than those in SVV-A for pairs of polar terms for the psychology and the task.

The average number of performed tasks in MVV-A is less than that in SVV-A, and the average task completion time in MVV-A gets longer than that in SVV-A. However, the subject can confirm spots of all sides by changing viewpoint in MVV-A. Thus, the subject feels easy to do the task with MVV-A.

From these, we find that interactive MVV-A improves QoE than interactive SVV-A.

Also, as a result of principal component analysis, we notice that the first principal component is audiovisual quality and user's feeling, and the second principal component is viewpoint change.

As future work, we will examine the influence of the interactivity on QoE of interactive MVV-A in detail for different types of task. In addition, we will consider the influence of video quality.

VII. ACKNOWLEDGMENT

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

REFERENCES

- [1] S. Tasaka and Y. Ishibashi, "Mutually compensatory property of multimedia QoS," *IEEE Commun. ICC 2002*, vol. 2, pp. 1105-1111, Apr. 2002.
- [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] I. Ahmad, "Multi-view video: Get ready for next-generation television," *IEEE Distributed Systems Online 2007*, vol. 8, no. 3, pp. 6, Mar. 2007.
- [4] P. Merkle, A. Smolic, K. Muller and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 11, Nov. 2007.
- [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] M. Yamamoto, T. Nunome and S. Tasaka, "The effects of camera arrangements and contents on QoE in multi-view video and audio IP transmission," *Proc. IEEE TENCON 2010*, pp. 1450-1455, Nov. 2010.
- [7] C. E. Osgood, "The nature and measurement of meaning," *Psychological Bulletin*, vol. 49, no. 3, pp. 197-237, May 1952.
- [8] J. Lou, H. Cai, and J. Li, "A real-time interactive multi-view video system," *Proc. ACM Multimedia 2005*, pp. 161-170, Nov. 2005.
- [9] D. Sugiya, S. Tasaka and T. Nunome, "Multidimensional assessment of QoE in interactive audiovisual IP communications: Effect of task types," *Technical Report of IEICE*, vol. 110, no. 287, CQ2010-63, pp. 85-90, Nov. 2010, in Japanese.
- [10] S. Tasaka and N. Masaki, "Customization of interactive services for QoE enhancement in audio-video transmission over bandwidth guaranteed IP networks," *Proc. IEEE ICME 2010*, pp. 546-551, July 2010.
- [11] K. Cho, K. Fukuda, H. Esaki and A. Kato, "Observing slow crustal movement in residential user traffic," *ACM CoNEXT2008*, Dec. 2008.
- [12] T. Tejima, A. Tan, S. Watanabe and K. Yoshida, "Large-scale measurements of network quality using online game," *IEICE Trans. on Commun.*, 192-B(10), pp. 1566-1578, Oct. 2009, in Japanese.
- [13] J. P. Guilford, *Psychometric methods*, McGraw-Hill, N. Y., 1954.
- [14] 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.