Relationships between usage situations and spatiotemporal QoE in audio and video IP transmission

Relationships between Usage Situations and Spatiotemporal QoE in Audio and Video IP Transmission

Toshiro Nunome[1] and Keita Abe[2]
[1]Department of Computer Science, Graduate School of Engineering,
[2]Department of Computer Science, Faculty of Engineering,
Nagoya Institute of Technology, Nagoya 466–8555, Japan
Email: nunome@nitech.ac.jp

***Abstract*— QoE (Quality of Experience) of audio and video IP transmission can change according to users' attributes and usage situations. This paper evaluates the effect of usage situations on QoE from a quality tradeoff point of view. As a usage situation, this paper employs distracted watching; the users watch video and audio while doing a calculation task. We perform a subjective experiment to compare two video output schemes: frame skipping and error concealment. We then find that the task can affect the tradeoff between spatial quality and temporal quality.**

Keywords-component; QoE, Temporal quality, Spatial quality, Tradeoff, Calculation task

I. INTRODUCTION

Multimedia communication services have been popularized owing to high-speed and broadband IP networks. However, general IP networks are best-effort and then cannot guarantee QoS (Quality of Service); packet losses and delays can occur. For users of the networks, QoE (Quality of Experience) [1] enhancement is important by mitigating the effect of delays and losses.

QoE of audio and video IP transmission can change according to users' attributes and usage situations.

To enhance QoE of audio and video IP transmission, Tasaka *et al.* have proposed SCS (Switching between error Concealment and frame Skipping) [2]. To exploit tradeoff between temporal quality and spatial quality, SCS switches two video output schemes: error concealment and frame skipping. Yokoi *et al.* have assessed the effect of users' attributes on QoE of threshold selection interfaces for SCS [3]. However, there is no assessment of usage situations on the spatial and temporal quality tradeoff.

On the other hand, Eguchi *et al.* have evaluated the effect of usage situations on Web page transition time [4]. However, they have not considered QoE of audio and video.

This paper evaluates the effect of usage situations on QoE from a quality tradeoff point of view. As a usage situation, this paper employs distracted watching; the users watch video and audio while doing a calculation task. We perform a subjective experiment to compare two video output schemes: frame skipping and error concealment.

The rest of the paper is structured as follows. Section II outlines the two video output schemes. Section III explains the task of distracted watching. Section IV describes methods of the experiment. We present results of the experiment in Section V, and Section VI concludes this paper.

II. VIDEO OUTPUT SCHEMES

This paper considers the relationships among video output schemes, distracted watching, and spatiotemporal QoE. As the temporal quality, we consider smoothness of output; freezing of output degrades it. The spatial quality is output image quality. It relates resolution, distortion, and imperfect interpolation for lost information.

The error concealment (Fig. 1) interpolates lost video slices with other information. The spatial quality of the error concealed video degrades compared to the original one. The degradation propagates to the succeeding frames in a unit of GOP (Group of Pictures).

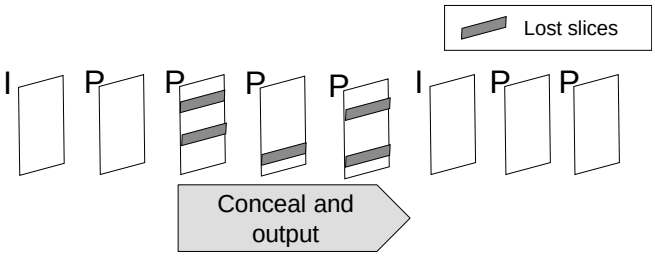

Figure 1. Error concealment

The frame skipping (Fig. 2) does not output video frames with lost slices. The spatial quality of the output video is kept original. However, the scheme degrades the temporal quality because of skipped frames.

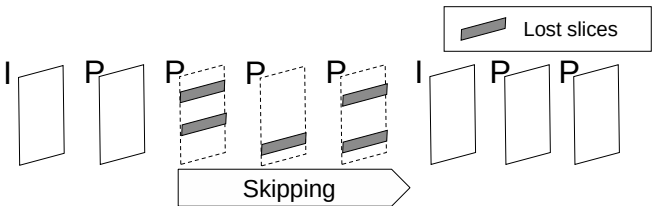

Figure 2. Frame skipping

III. TASK

In this paper, we ask the assessors to watch video and audio while doing calculations. Figure 3 shows the window for the calculation task. It appears to the right of the video window during the experimental run. The assessor uses a mouse to answer a question. He/She selects an answer from the three choices and then pushes the send button. The next question appears three seconds after pushing the send button. Each question is an addition of two-digit numbers. The questions are generated randomly. The fields "question" and "right" mean the number of answered questions and the number of correct answers, respectively.

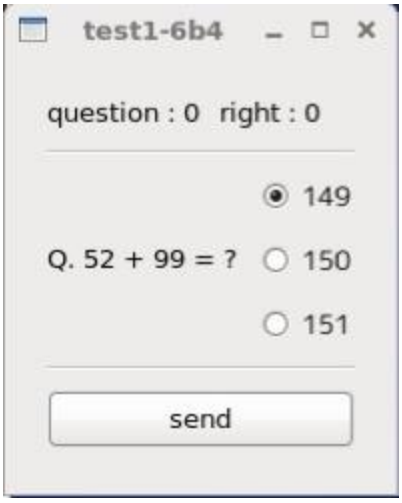

Figure 3. Task window

IV. EXPERIMENTAL METHOD

Figure 4 shows the experimental system in this paper. It consists of four PCs (Media Server, Media Client, Web Server, and Web Client) and two routers (Riverstone RS3000). All the links in the network are 100 Mb/s full-duplex Ethernet. Media Server transmits video and audio streams to Media Client through RTP/UDP. The OS of both Media Server and Media Client is CentOS 6.3.

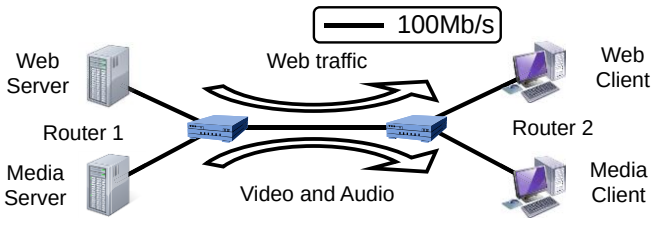

Figure 4. Experimental system

Tables I and II show the specifications of video and audio, respectively. Here, an *MU (Media Unit)* means a unit for media synchronization. In video, an MU is a video frame. In audio, an MU consists of a constant number of audio samples.

TABLE I
SPECIFICATION OF VIDEO

coding method	H.264/AVC
image size [pixels]	640 × 360
number of slices per frame	23
picture pattern	IPPPP
average MU rate [MU/s]	30
average encoding bitrate [kb/s]	1500
duration [s]	20

TABLE II
SPECIFICATION OF AUDIO

coding method	MPEG4 AAC
sampling rate [kHz]	48
channels	2
average MU rate [MU/s]	46.875
average encoding bitrate [kb/s]	128
duration [s]	20

We employ *music* (two women sing a song) and *sport* (a football game) as the contents. The former is an audio-dominant content, and the latter is a video-dominant one.

For video error concealment, we use *Frame Copy* and the interpolation from neighboring macroblocks of *FFmpeg* [5]. We do not adopt *FMO (Flexible Macroblock Ordering)*.

Media Receiver employs simple playout buffering control for absorbing network delay jitter. The playout buffering time is set to 500 ms.

As the interference traffic of audio and video, Web Server transmits Web traffic to Web Client according to requests generated by *WebStone 2.5* [6], which is a Web server benchmark tool. For the number of client processes, we employ 20, 50, and 80. As the number of client processes increases, the amount of interference traffic increases.

In this paper, we employ two contents, three load conditions (i.e., the number of Web clients), two video output schemes (frame skipping and error concealment), and with or without the task. In total, we consider 52 stimuli obtained by these combinations and two dummy stimuli for each content. The assessors are 37 students in their twenties; 19 male students who major in computer science and 18 female students who do not major in computer science.

In the assessment, we employ six pairs of polar terms. Table III shows the pairs of polar terms in the subjective experiment. For each pair, a subjective score is measured by the *rating scale method* [7]. 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 (score 5 is

the best, and score 1 is the worst). Finally, we calculate the mean opinion score (MOS), which is an average of the rating scale scores for all the users.

TABLE III
PAIRS OF POLAR TERMS

category	pair of polar terms
Video temporal	The video is smooth - The video is rough
Video spatial	The video is sharp - The video is blurred
Audio	The audio is natural - The audio is artificial
Task	The task is easy - The task is difficult
Synchronization	The audio and video are in synchronization - The audio and video are out of synchronization
Overall satisfaction	Excellent - Bad

V. EXPERIMENTAL RESULTS

Figures 5 through 8 show the assessment result of the adjective pair "Excellent - Bad (Overall satisfaction)", that of the adjective pair "The audio is natural - The audio is artificial", that of the adjective pair "The video is sharp - The video is blurred", and that of the pair "The video is smooth - The video is rough", respectively. The abscissa means the combination of the number of Web client processes, with or without the task, and the content.

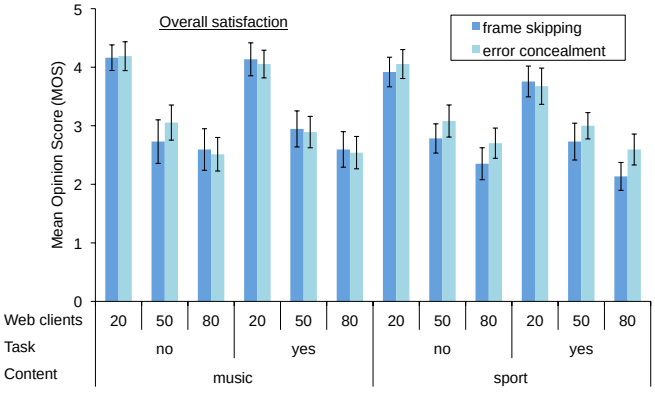

Figure 5. Overall satisfaction (excellent – bad)

We see in Fig. 5 that the frame skipping scheme tends to have larger MOS values than the error concealment scheme for music with the task. For sport, when the number of Web clients is 20, the task affects the tradeoff; in general, the error concealment is effective for sport because of its video-dominant character and large movement.

In Fig. 6, we notice that the tendency of the output schemes on audio quality is almost the same as that of the overall satisfaction. This is because of cross-modality of audio and video although the output quality of audio is not affected by the video output schemes.

We find in Fig. 7 that the task affects the MOS values of video spatial quality for the frame skipping scheme in sport. Besides, we can observe in Fig. 8 that the task decreases the difference between the video output schemes, especially in music. Thus, we can think that the task makes the assessors insensitive to video quality.

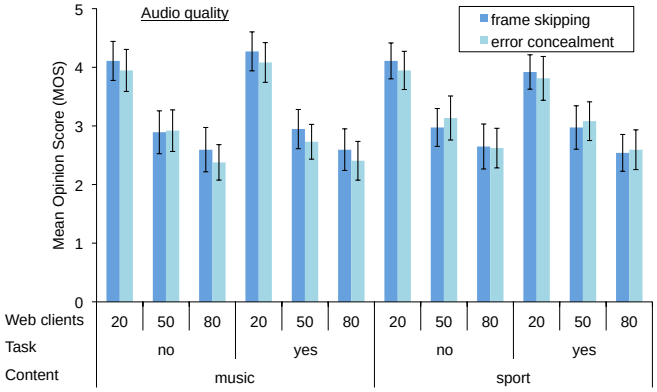

Figure 6. Audio quality (natural – artificial)

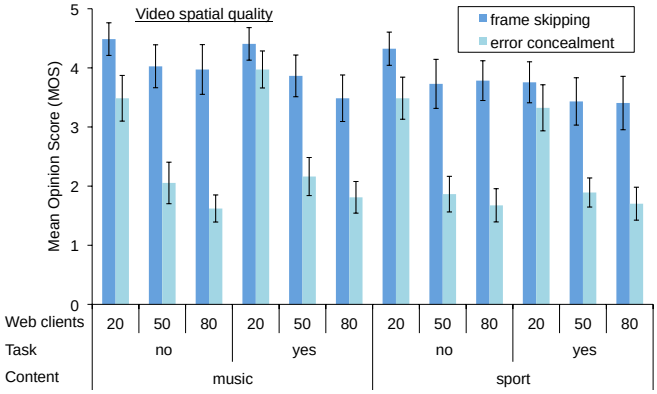

Figure 7. Video spatial quality (sharp – blurred)

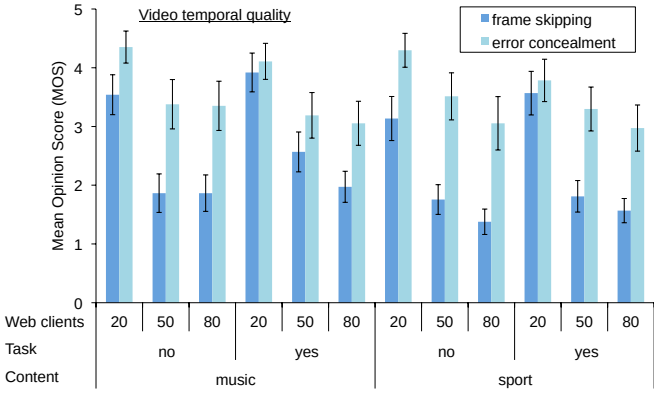

Figure 8. Video temporal quality (smooth – rough)

VI. Conclusions

In this paper, we evaluated the effect of usage situations on QoE from a quality tradeoff point of view. As a usage situation, we employed distracted watching; the users watch video and audio while doing a calculation task. From the QoE assessment results, we found that the task can affect the tradeoff between spatial quality and temporal quality. Also, the assessors become insensitive to video quality by the task.

For future study, we need to evaluate more diverse situations and investigate the effect of the situations on QoE.

Appendix A. Application-level QoS assessment results

In the experiment, we also assess the application-level QoS. The application-level QoS is closely related to QoE because they adjoin at the layered network model. In this paper, we treat the audio MU loss ratio, the video MU loss ratio, and the error concealment ratio for video as the application-level QoS parameters. The error concealment ratio represents the percentage of slices error-concealed (i.e., lost slices) in a frame; it shows the image quality of video stream. The MU loss ratio is the ratio of the number of MUs not output at the recipient to the number of MUs transmitted by the sender.

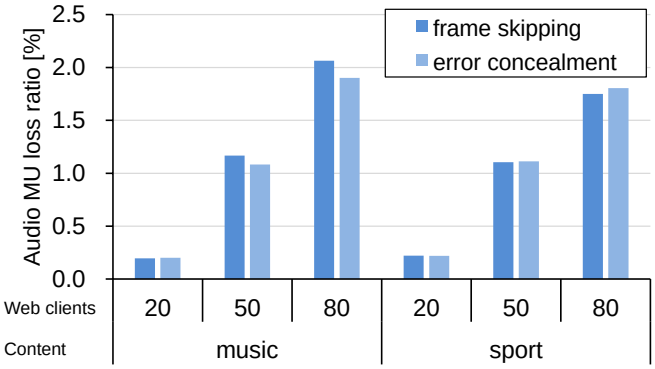

Figure 9. Audio MU loss ratio

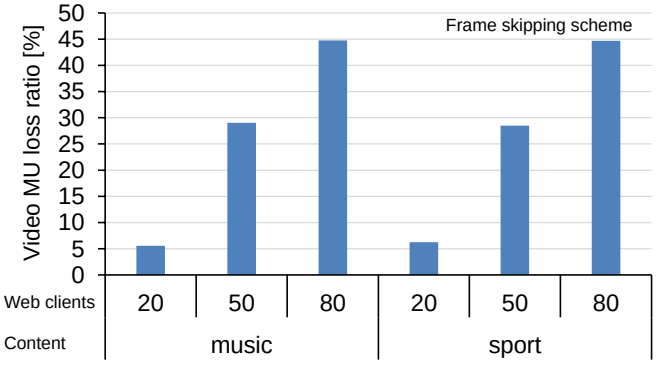

Figure 10. Video MU loss ratio

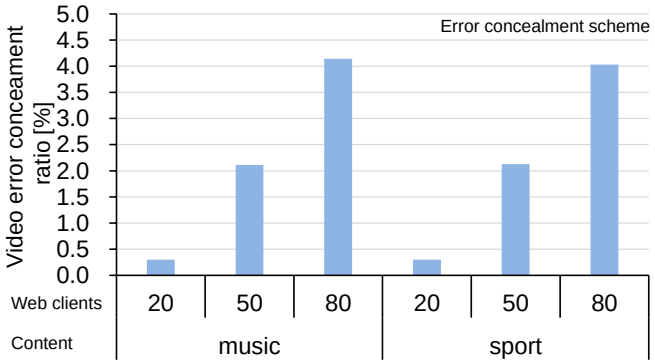

Figure 11. Video error concealment ratio

Figure 9 depicts the MU loss ratio of audio. The video MU loss ratio is shown in Fig. 10. Figure 11 represents the error concealment ratio of video. Each parameter value is the average of 37 experimental runs (i.e., 37 assessors).

In Figure 10, we only show the results of the frame skipping scheme. The reason is that the MU loss ratio is 0 in the error concealment scheme because there is no MU of which the whole slices are lost. On the other hand, Fig. 11 only depicts the results of the error concealment scheme because there is no degradation due to concealment in the frame skipping scheme.

We notice in Figures 9 through 11 that the application-level QoS parameters increase as the number of Web client processes increases. In addition, the difference between the two contents is very small. This is because the difference scarcely affects the application-level QoS.

References

[1] ITU-T Rec. P.10/G.100, "Amendment 5: New definitions for inclusion in Recommendation", July 2016.

[2] S. Tasaka, H. Yoshimi, A. Hirashima, and T. Nunome, "The effectiveness of a QoE-based video output scheme for audio-video IP transmission," *Proc. ACM Multimedia 2008*, pp. 259-268, Oct. 2008.

[3] T. Yokoi, S. Tasaka and T. Nunome, "The effect of subject attributes on QoE of threshold selection interfaces in the QoE-based video output scheme SCS," *Tech. Rep. IEICE*, CQ2012-35, July 2012. (in Japanese)

[4] M. Eguchi, T. Miyoshi, K. Yamori, and T. Yamazaki, "Structuring relational models between users' situation factor and their QoE evaluation," *Tech. Rep. IEICE*, CQ2010-52, Nov. 2010. (in Japanese)

[5] "FFmpeg, " *http://www.ffmpeg.org/*

[6] Mindcraft Inc, "WebStone benchmark information," *http://www.mindcraft.com/webstone/*

[7] J. P. Guilford, *Psychometric Methods*, McGraw-Hill, N. Y., 1954.