Adaptation of the QoE-Based Video Output Scheme SCS to Time-Varying Traffic in IP Networks

Tomohiro Yokoi, Shuji Tasaka and Toshiro Nunome
Department of Computer Science and Engineering, Graduate School of Engineering,
Nagoya Institute of Technology
Nagoya 466–8555, Japan

***Abstract*—This paper studies the QoE-Based Video Output Scheme SCS (Switching between error Concealment and frame Skipping) for audio-video IP transmission in terms of QoE; it adapts the threshold-setting method of SCS to time-varying traffic. The conventional error concealment method corresponds to a constant threshold value of 100%. We then employ three adaptive methods for threshold-setting; looking up into a reference table, setting by estimating QoE, and selection by the user. We conducted subjective experiments on the four methods for combinations of two contents, three picture patterns, and two patterns of variation in the amount of load traffic. We assessed QoE by the method of successive categories. As a result, we observe that the adaptive methods can achieve higher QoE than the conventional error concealment method for time-varying traffic.**

***Index Terms*—Audio-Video IP Transmission, QoE, SCS, Error Concealment, Frame Skip, Time-Varying Traffic**

I. Introduction

As IP networks are becoming faster, many people use various services on them. In particular, the services with audio-video IP transmission grow rapidly.

IP networks basically offer best-effort services. *QoS (Quality of Service)* for audio-video transmission can deteriorate because of packet loss, delay and delay jitter, which also degrade *QoE (Quality of Experience)* [1].

In order to enhance QoE for audio-video IP transmission, Tasaka and Yoshimi proposes a video output scheme *SCS (Switching between error Concealment and frame Skipping)* [2]. SCS defines the error concealment ratio (R_c)[%] as the ratio of the number of lost video slices to the total number of slices in a frame, and a threshold value T_h[%]. When R_c is larger than T_h in a video frame, the video frame is skipped, i.e., *frame skipping*. When an I frame with $R_c \leq T_h$ comes out, the output scheme is switched to error concealment.

Video error concealment is a method to compensate lost video slices due to packet drop with other information of the video stream. However, spatial quality of the video is degraded because the method cannot perfectly compensate the lost information. In addition, there is a problem that the degradation propagates to the succeeding frame in a unit of GOP (Group of Pictures).

Frame skipping is a method which does not output video frames with lost slices. The method keeps the spatial quality of the output video original, while it degrades the temporal quality because of the skipping.

To increase the effectiveness of SCS, we need to adopt some appropriate threshold setting methods. In [3], Tasaka and Hirashima propose the *table lookup method* and the *user selection method* and show that they achieve higher QoE than the conventional error concealment method by experiment, which is performed with stationary traffic. However, load traffic in best-effort IP networks is not normally stationary. Therefore, we need to study the effectiveness of threshold-setting method for SCS with nonstationary traffic, i.e., in time-varying traffic environments.

In this paper, we deal with four threshold setting methods as in [3]: (1) 100% method, (2) table lookup method, (3) QoE estimation method and (4) user selection method. We then examine which method can achieve the highest QoE for time-varying traffic. The methods (2) and (4) in this paper are modified from those in [3] so that they can adapt to time-varying traffic.

The remainder of the paper is organized as follows. Section II introduces the time-varying traffic used in the experiment. Section III demonstrates a method to assess QoE with the *psychological scale*. Section IV describes the experimental method. Section V presents experimental results. Section VI concludes the paper.

II. Time-Varying Traffic

We should use realistic time-varying traffic for the experiment. Because of popularity of Web traffic, we adopt two time-varying Web traffic models in this paper.

Time-varying traffic based on self-similarity

It is known that network traffic has *self-similarity* [4]. It is a feature focusing on the changing pattern of traffic for a period of time; even if the time-scale is changed, the pattern is similar to the original one.

In this paper, we apply the feature to the pattern of HTTP hits per second in Asia (9/9/2010 at 3 pm. Japan time) produced by Akamai Technologies [5]. Traffic 1 in Fig. 1 shows the result. The pattern of Traffic 1 has been created by changing the timescale from 24 hours to 60 seconds.

Time-varying traffic with burstiness

Recently, Internet traffic tends to increase the percentage of video distribution [6]. It causes bursts of the traffic by a large amount of data transmissions in a short term. In order to reflect this situation, we have designed Traffic 2 in Fig. 1; keeping the accumulated numbers of Web clients during the interval between 5 and 55 seconds the same as those of Traffic 1, we have increased the number of peak traffic in Traffic 2, namely from two to three.

III. QoE Measure

As the QoE measure, we adopt the *psychological scale* [7]. This is because the psychological scale, which is *the interval scale* in psychometric theory, can represent human subjectivity more accurately than *MOS (Mean Opinion Score)*, which is the most popular QoE measure.

In this method, each subject gives a score (e.g., an integer between 5 and 1) to the audio-video stream output at the media receiver; the method is called *the rating-scale method*. Then, we apply *the law of categorical judgement* to the scores to obtain the interval scale of an evaluation object.

TABLE I
VIDEO SPECIFICATIONS

Coding method	H.264(JM17.2)
Image size [pixel]	320 × 240
Number of slices in a frame	15
Average MU rate [MU/s]	30
Picture pattern	I, IPPPP, IPPPPPPPPPPPPPP (I+14P's)
Playing time [s]	60

TABLE II
AUDIO SPECIFICATIONS

Coding method	Linear PCM 24kHz 16bit 1ch
Average MU rate [MU/s]	50
Average bit rate [kb/s]	384
Playing time [s]	60

In the case of the interval scale, we further conduct Mosteller's test [8] to confirm the goodness of fit for the obtained scale. Once the goodness of fit has been confirmed, we use the interval scale as the psychological scale.

IV. EXPERIMENT

A. Experimental network

Fig. 2 illustrates the experimental network. The network consists of two routers and four PC's which are used as a media sender (MS), a media receiver (MR), a Web server (WS) and a Web client (WC). The two routers are Cisco 2811 by Cisco Systems. The links between terminals and routers are full duplex Ethernet channels of 100 Mb/s. The link between two routers is a full duplex Ethernet channel of 10 Mb/s. The MS transmits an audio stream and the corresponding video stream to the MR with RTP/UDP. The WS sends the load traffic to the WC. The MR adapts playout buffering of 1 second to absorb delay jitter. We do not perform inter-stream synchronization control between audio and video.

Tables I and II show the specification of video used in the experiment and that of audio, respectively. *MU (Media Unit)* in these tables means the information unit for transfer between the application layers. We define a video MU as a video frame and an audio MU as audio samples for 20 ms. We utilize H.264/MPEG-4 AVC reference software JM17.2 [9] for video encoding; we do not adopt FMO (Flexible Macroblock Ordering).

In the experiment, we select two content types: *sport* and *music video*, referring to the final report of VQEG [10]. We apply the QoE estimation equations derived in [11] to these contents, which are regarded as non-training data. Sport is the same content as sport 2 used in [3]; it is a soccer game. In the music video, a woman is singing a song with a slow tempo; it is the same as music video 1 in [3].

Table III shows the average bit rate and TI value of each video. The TI value is an index of the degree of video

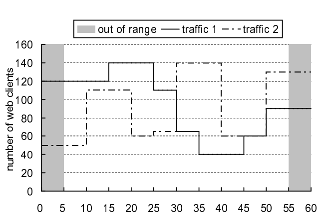

Fig. 1. Traffic patterns

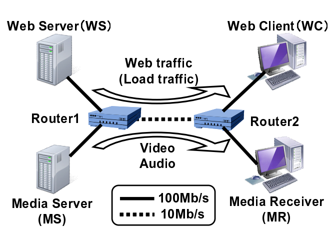

Fig. 2. System configuration

TABLE III
VIDEO BIT RATE AND TI VALUE FOR EACH CONTENT

Content type	GOP	Average bit rate [kb/s]	TI value
sport	I	3045.73	16.405
	IPPPP	1657.77	
	I+14P's	1436.26	
music video	I	1812.20	3.287
	IPPPP	526.09	
	I+14P's	325.68	

motion. It is specified by ITU-T Rec. P.910 [12]. As the value increases, the video motion becomes larger.

For video error concealment, we apply *Frame Copy* of JM16.2 to a P frame and the interpolation from neighboring macroblocks to an I frame.

Load traffic used in this experiment is generated by Web server benchmark tool *WebStone 2.5* [13]. In order to create time-varying traffic, we restart WebStone after a rewrite of the number of Web clients at the change point of traffic volume. We adopt the two time-varying patterns in Fig. 1. In these patterns, a certain number of clients continues for more than 5 seconds. This is because WebStone would finish before the expected amount of load traffic is generated if we changed the number of Web clients in a shorter period than 5 seconds.

In the experiment, we remove the first 5 seconds from the scope of the assessment because a subject may not have prepared an assessment. In addition, the last 5 seconds are out of range because of losing logs at the end of playing.

B. Threshold setting methods

We explain the four threshold setting methods used in the experiment.

(1) 100% method

This method always sets a constant threshold value of 100%. It means pure error concealment.

(2) table lookup method

This method looks up a table showing appropriate threshold values corresponding to the attribute of media; content type, picture pattern and *slice arriving ratio*. The slice arriving ratio is defined as the ratio of the number of received slices to the number of sent slices in units of %. In the experiment, the MR calculates the slice arriving ratio to estimate the degree of network congestion and refers to the reference table, which is shown in Table IV, in 1 second intervals. Table IV has been constructed by referring to the results of [3] and [11].

In picture pattern I, frame skipping does not degrade the temporal quality so much. Therefore, we set the threshold value of 0% regardless of the content type and the slice arriving ratio.

In picture pattern IPPPP, we set low threshold values (i.e., 0% and 20%) in sport, when the slice arriving ratio is over 90%; otherwise, we select a threshold value of 100% in order not to degrade the temporal quality. Compared to sport, we set higher threshold values in music video, which is an audio-dominant content, so that video freezing due to frame skipping cannot largely impair lip sync between audio and video.

In picture pattern I+14P's, we adopt the threshold value of 100% because frame skipping degrades temporal quality greatly.

(3) QoE estimation method

The QoE estimation method sets the threshold value using real time estimation of QoE. In the beginning, we prepare a certain number of threshold values as proposed in [11]: 100%, 40%, 20% and 0%. The MR estimates the QoE for each threshold value in real time, and selects the threshold value which achieves the maximum QoE.

TABLE IV
REFERENCE TABLE FOR THE TABLE LOOKUP METHOD

Features of contents		Slice arriving ratio A [%]		
Content type	GOP	$A \geq 96$	$90 \leq A < 96$	$A < 90$
sport	I	0%	0%	0%
	IPPPP	0%	20%	100%
	I+14P's	100%	100%	100%
music video	I	0%	0%	0%
	IPPPP	20%	40%	100%
	I+14P's	100%	100%	100%

This method can set an appropriate threshold value in response to time-varying traffic. However, low accuracy of QoE estimation does not achieve high QoE. This method is hard to implement because we need to map application level QoS to QoE with high accuracy.

In the experiment, we utilize QoE estimation equations used in [11]. The interval of estimation is set to 1 second, which is the same as the interval of referring to the table in the table lookup method.

The QoE estimation equations are as follows.

$$\begin{aligned} \hat{U}_S &= 4.492 - 1.363 R_c + 2.939 \times 10^{-1} R_c^2 \\ &\quad -2.296 \times 10^{-2} R_c^3 - 1.317 \times 10^{-1} L_v \\ &\quad +1.699 \times 10^{-3} L_v^2 - 8.607 \times 10^{-6} L_v^3 \end{aligned} \tag{1}$$

$$\begin{aligned} \hat{U}_M &= 4.826 - 2.372 R_c + 7.479 \times 10^{-1} R_c^2 \\ &\quad -8.025 \times 10^{-2} R_c^3 - 1.074 \times 10^{-1} L_v \\ &\quad +2.106 \times 10^{-3} L_v^2 - 2.499 \times 10^{-5} L_v^3 \end{aligned} \tag{2}$$

where R_c is the error concealment ratio of video, and L_v the MU loss ratio, which is defined as the ratio of the number of lost video MU to the number of transmitted video MU in units of %. Equation (1) has been derived for sport 2 in [11], and Eq. (2) for music video 1 in [11]. The two contents in [11] are regarded as training data for the QoE estimation, while sport and music video in our experiment are non-training data.

(4) user selection method

The user selects his/her favorite threshold value from among the four values through an interface.

The advantage of this method is to be able to adapt to the individual user's inclination. On the other hand, it has a disadvantage of placing the burden on the user. The method is easy to implement because the user controls the threshold value.

In the experiment, the user sets the threshold value in real time through a GUI of Fig. 3, where A, B, C and D correspond to the threshold values of 100%, 40%, 20% and 0%, respectively.

C. QoE assessment experiment

We recorded the audio-video streams that the MR output; the recorded streams are regarded as objects to be evaluated for QoE measurement. Thus, we totally have 48 objects to be evaluated because of the two contents, three picture patterns for each content, two patterns of time-varying load traffic, and four threshold setting methods.

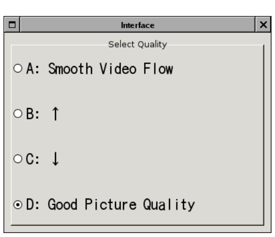

Fig. 3. GUI for the user selection method

The procedure of this experiment is as follows.

Step 1: The original audio and video streams are presented to the subject in order for him/her to obtain the base of QoE comparison with recorded streams.

Step 2 (the user selection method only): The recorded stream to be compared is played to the subject; while playing, the subject continuously selects his/her favorite threshold values without assessing the quality.

Step 3: The subject continuously evaluates audiovisual quality by using *DCR (Degradation Category Rating)* with the following *five-level impairment scale*: "imperceptible" assigned score 5, "perceptible, but not annoying" 4, "slightly annoying" 3, "annoying" 2, "very annoying" 1. The evaluation is conducted utilizing a key from 1 to 5 of a keypad in real time. The subject updates the score by pushing a key when his/her evaluation changes. At this time, the MR samples the subject's score at intervals of 0.5 seconds. Note that only the user selection method needs a "two-pass" procedure (i.e., Steps 2 and 3) for the evaluation.

Step 4: The subject makes *comprehensive evaluation*: evaluation for the 60 seconds of the audio and video output with DCR just after the play of audio and video stream finishes.

Step 5: Repeat from Step 2 to Step 4 until the subject finishes all evaluation.

The number of the subjects is 55: 26 Japanese males and 29 Japanese females who were university students. Their age ranged from teens through twenties. It took about 80 minutes including training and break time for a subject to evaluate all the audio and the video.

V. EXPERIMENTAL RESULTS

In this section, we present the psychological scale calculated from the result of the subjective experiment.

A. Psychological scale

In the continuous evaluation, we define the *stimulus* as an audio-video stream for half a second in the 50 second observation interval (i.e., from 5 seconds through 55 seconds). Thus, the number of the stimuli by the continuous evaluation is 4848: 48 objects to be evaluated and 101 evaluation sampling in the 50 second interval. In the comprehensive evaluation, an audio-video stream for the 60 seconds is regarded as a stimulus. The number of the stimuli for the comprehensive evaluation is 48 because of the 48 objects to be evaluated.

We applied Mosteller's test to the interval scale obtained by *the method of successive category*. As a result, we have found that the test with a significance level of 0.05 can reject the hypothesis that the observed value equals the calculated one. Then, we removed the stimuli which give large errors in Mosteller's test. Removing 514 stimuli in the continuous evaluation and 4 stimuli in the comprehensive evaluation, we saw that the hypothesis cannot be rejected. Consequently, we utilize the interval scale as the psychological scale. Since we can select an arbitrary origin in an interval scale, we set the minimum value of the psychological scales to unity for each evaluation.

Regarding the continuous evaluation, we present the result of sport only because the result of music video is close to it. Figs. 4 through 11 display QoE for each combination of picture pattern and load traffic pattern.

Figs. 4 through 9 show the psychological scale in the continuous evaluation. The abscissa in the figures represents the playing time, and the ordinate shows the psychological

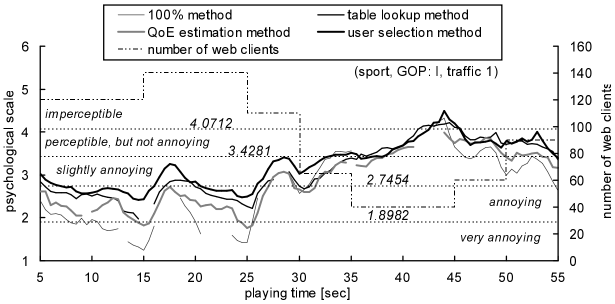

Fig. 4. Psychological scale in the continuous evaluation (content: sport, picture pattern: I, load traffic pattern: traffic 1)

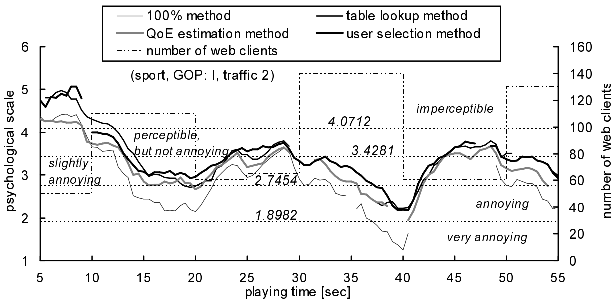

Fig. 5. Psychological scale in the continuous evaluation (content: sport, picture pattern: I, load traffic pattern: traffic 2)

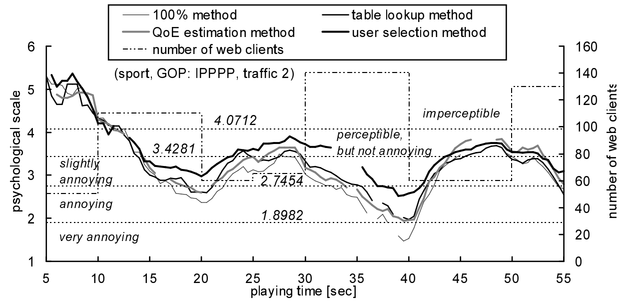

Fig. 7. Psychological scale in the continuous evaluation (content: sport, picture pattern: IPPPP, load traffic pattern: traffic 2)

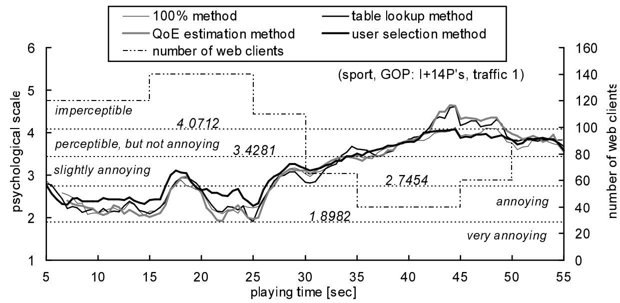

Fig. 8. Psychological scale in the continuous evaluation (content: sport, picture pattern: I+14P's, load traffic pattern: traffic 1)

scale and the number of Web clients. Straight broken lines parallel to the abscissa represent the lower boundaries of the categories. Note that the results removed by the Mosteller's test are not plotted in the figures.

Figs. 10 and 11 present the result of the comprehensive evaluation and the average of the psychological scale over time in the continuous evaluation.

B. Consideration on QoE for each picture pattern

Picture pattern I: As we see in Figs. 4 and 5, the table lookup method, the QoE estimation method and the user selection method achieve higher QoE than the 100% method. These adaptive methods can also provide high QoE compared with the 100% method in Figs. 10 and 11. In picture pattern I, frame skipping does not degrade the temporal quality greatly. Therefore, the three threshold setting methods, which use frame skipping, lead to high QoE.

Picture pattern IPPPP: In Figs. 6, 7, 10 and 11, we see that the table lookup method, the QoE estimation method and the user selection method exhibit QoE approximately equal to or higher than that of the 100% method. In particular, the user selection method achieves the highest QoE in the four methods. However, the difference in QoE between the threshold setting methods is small compared with the case of picture pattern I. This is because the degradation of the temporal quality becomes larger as the number of P frames increases.

Picture pattern I+14P's: The behavior of the 100% method and the table lookup method for this picture pattern are the same; therefore, we notice that QoE of these methods are close to each other (see Figs. 8 and 9).

The QoE estimation method may select the threshold value other than 100%; therefore, we do not notice any improvement of QoE over the 100% method at peak traffic because of much degradation of temporal quality.

In the experiment, we found that in the user selection method, more than half of the subjects chose the threshold value other than 100%. This result was observed regardless of the content type and the load traffic pattern. Frame skipping in the video stream with many P frames greatly degrades the

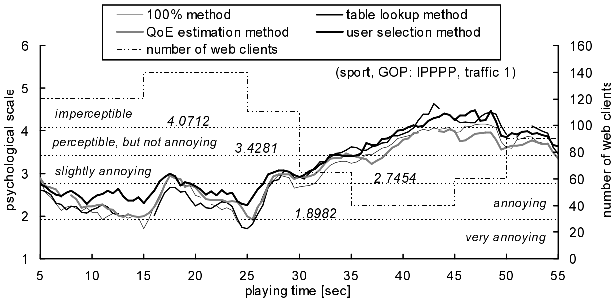

Fig. 6. Psychological scale in the continuous evaluation (content: sport, picture pattern: IPPPP, load traffic pattern: traffic 1)

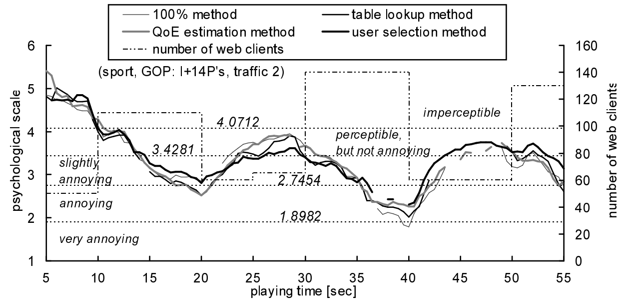

Fig. 9. Psychological scale in the continuous evaluation (content: sport, picture pattern: I+14P's, load traffic pattern: traffic 2)

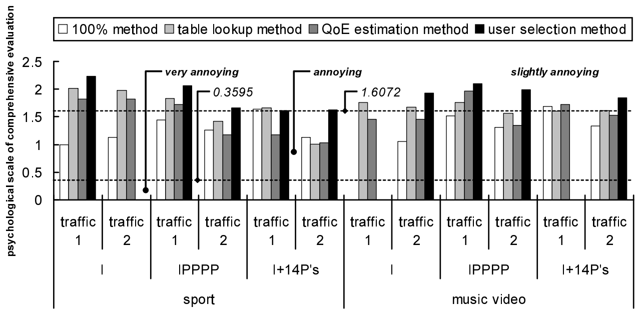

Fig. 10. Psychological scale in the comprehensive evaluation

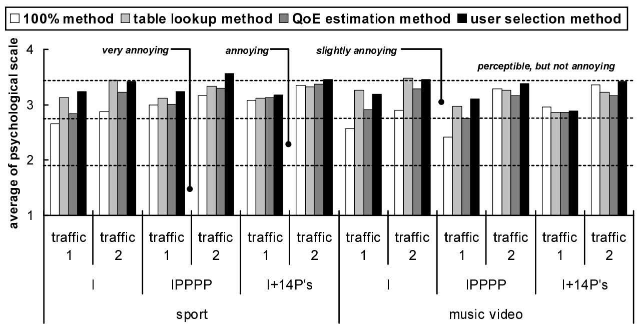

Fig. 11. Average psychological scale in the continuous evaluation

temporal quality. However, Figs. 10 and 11 show that the user selection method achieves QoE close to or higher than that of the 100% method. This is because the user selection method can adapt to individual users' inclination and satisfies the subjects by enabling them to control the video quality by themselves.

C. Effects of the content type on QoE

Fig. 11 shows that the difference in QoE between 100% method and each of the other three methods tends to be larger in sport than music video. Since music video is an audio-dominant content, controlling video quality does not affect QoE largely. On the other hand, sport is a video-dominant content; therefore, improvement of video output quality by SCS enhances QoE.

D. Effects of the traffic pattern on QoE

In Fig. 11, the average of the psychological scale over time for traffic 2 is higher than that for traffic 1. This is because there are more periods of low traffic load in traffic 2 and they improved QoE of the subjects. However, the psychological scale in the comprehensive evaluation in Fig. 10 shows that QoE of traffic 1 tends to be higher than that of traffic 2. This is due to *the recency effect* [14], which is a feature that information given before the decision strongly affects the result. That is, as seen in Fig. 1, the amount of load traffic in the second half of traffic 1 is small, whereas traffic 2 has high load traffic in the second half. Therefore, traffic 1 produces better result than traffic 2 in the comprehensive evaluation.

E. Comparison of the threshold setting methods

Fig. 11 shows that the 100% method can achieve high QoE when the GOP length is long. However, this method degrades QoE because of deterioration of spatial quality when the GOP length is short. This property is noticeable in Figs. 4 and 5. The table lookup method and the QoE estimation method exhibit stable and high QoE regardless of content type, picture pattern and load traffic pattern. The user selection method achieves the highest QoE in many cases.

Taking into consideration the burden on the user, we can say that the table lookup method is the most practical as the threshold setting method of SCS for time-varying traffic. However, we can also effectively utilize the user selection method, which always achieves high QoE.

VI. Conclusions

In this paper, we studied the four threshold setting methods of SCS for time-varying traffic. As a result of the experiment, the table lookup method, QoE estimation method and user selection method can achieve higher QoE than the 100% method.

In addition, the table lookup method, which is easy to implement, accomplishes higher QoE than the QoE estimation method, which is hard to implement.

As future work, we will study a method which enables the user to select a threshold value if he/she wants; otherwise, sets a value by the table lookup method.

Acknowledgment

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

References

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

[2] S. Tasaka and H. Yoshimi, "Enhancement of QoE in audio-video IP transmission by utilizing tradeoff between spatial and temporal quality for video packet loss," in *Conf. Rec. IEEE GLOBECOM*, Dec. 2008.

[3] S. Tasaka and A. Hirashima, "Methods of selecting threshold for the QoE-based video output scheme SCS," in *Conf. Rec. IEEE GLOBECOM*, Dec. 2010.

[4] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking, Vol. 2, No. 1, pp. 1-15*, Feb. 1994.

[5] Akamai Technologies, "Visualizing Akamai," *http://www.akamai.com/html/technology/dataviz3.html*

[6] Cisco Systems, Inc. "Hyperconnectivity and the approaching zettabyte era," Cisco Visual Networking Index, June 2010.

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

[8] 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.

[9] "H.264/MPEG-4 AVC reference software JM," *http://iphome.hhi.de/suehring/tml/index.htm*

[10] The Video Quality Experts Group, "The video quality experts group web site," *http://www.its.bldrdoc.gov/vqeg/index.php*

[11] S.Tasaka, H.Yoshimi, A.Hirashima and T.Nunome, "The effectiveness of a QoE-based video output scheme for audio-video IP transmission," in *Proc. ACM Multimedia2008*, pp.259-268, Oct. 2008.

[12] ITU-T Rec. P.910, "Subjective video quality assessment methods for multimedia applications," Apr. 2008.

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

[14] B. B. Murdock, Jr., "The serial position effect of free recall," *Journal of Experimental Psychology, vol. 64, No. 5, pp. 482-488*, 1961.