Methods of Selecting Threshold for the QoE–Based Video Output Scheme SCS

Shuji Tasaka and Akifumi Hirashima
Department of Computer Science and Engineering,
Graduate School of Engineering, Nagoya Institute of Technology, Nagoya 466–8555, Japan
Email: tasaka@nitech.ac.jp, tiery@inl.nitech.ac.jp

***Abstract*— SCS is a video output scheme which is designed to maximize QoE in audio-video IP transmission; it is featured by the applicability to end–terminals independent of the type of used networks. SCS switches between video error concealment and video frame skipping by comparing the percentage of error–concealed video slices in a frame to a threshold value, which is a key parameter for successful implementation of SCS. This paper proposes two practical methods of threshold selection for SCS: a *table lookup* method and a *user selection* method. We compare the two proposed methods with the pure error concealment and the pure video frame skipping. We conducted subjective experiments on the four methods for combinations of several contents and picture patterns and then evaluated QoE in terms of the psychological scale. As a result, we observe that the table lookup method and the user selection method can achieve higher QoE than the other two conventional methods.**

I. Introduction

Audio–video transmission is the most fundamental function of multimedia application services over IP networks, which have been increasingly important in our daily life. In order to realize high *QoS* (*Quality of Service*), a variety of techniques have been devised and developed for end–terminals in addition to network control technologies such as bandwidth guarantee mechanisms and QoS–aware routing. Typical techniques for the end–terminals are media encoding and decoding schemes, which include error-resilient coding, error-concealment [1], video frame skipping and playout buffering.

A practical advantage of the end–terminal techniques is the availability independent of the used network–types; the techniques are applicable to not only the current best-effort IP networks but also QoS guaranteed IP networks like the *Next Generation Network* (*NGN*) [2].

The great majority of previous studies on usage of the end–terminal techniques centers around maximization/enhancement of QoS at application and lower levels[1]. *PSNR* (*Peak Signal to Noise Ratio*) of pictures in a video stream, delay jitter of output video frames and audio/video throughput are examples of QoS parameters often employed.

Regarding the usage of the end–terminal techniques, we should recall here that the maximization/enhancement of QoS at application and lower levels is just an intermediate step to the final goal of the service offering, i.e., users' satisfaction, which means satisfactory *QoE* (*Quality of Experience*)[2]. Therefore, reconsideration on the usage of the techniques from a QoE point of view is worth studying for satisfactory service provision. *SCS* (*Switching between error Concealment and frame Skipping*) is an end-terminal technique of this kind [4], [5].

SCS is a video output scheme to enhance QoE by utilizing a tradeoff relation between spatial and temporal quality caused by *video error concealment* and *video frame skipping*. SCS switches between the two techniques according to the percentage of error–concealed video slices in a frame. It starts with the error–concealment mode; once the percentage exceeds a threshold value, it enters the video frame skipping mode, which continues until an intra–coded frame is decoded.

It has been demonstrated that SCS can achieve higher QoE than error–concealment only and video frame skipping only, if the threshold value is selected appropriately [4], [5]. In [4], the authors show the existence of the QoE tradeoff relation and the feasibility of QoE enhancement by the relation. Also, reference [5] has proposed a threshold selection method based on QoE real–time estimation, which selects the threshold value with the highest QoE estimate, and revealed its effectiveness. The QoE estimation method is straightforward in concept, but its implementation is not so easy since it necessitates the derivation of QoE estimation equations, which is still an open issue in the QoE research area. Thus, in addition to pursuing the QoE estimation method, we need some practical method of threshold selection, which should be easily implemented.

This paper proposes two practical methods of threshold selection: *table lookup* and *user selection*. We suppose that the threshold value is selected from among a finite set of values prepared in advance. The former method selects a threshold value by looking up a table whose entry is composed of an appropriate threshold value corresponding to each type of supposed media attributes like the content type and picture pattern. The latter method allows the user to select the threshold value by him/herself through a GUI (Graphical User Interface). We compile a lookup table and design a GUI for the user-selection. We also carry out subjective experiment to measure QoE of audio–video streams output at the receiver with the two proposed methods in addition to pure error–concealment and pure frame skipping. We then show the effectiveness of the two proposed methods by comparing the four methods in terms of QoE.

The remainder of the paper is organized as follows. Section II gives an outline of the SCS and proposes a table lookup method and a user selection method. Section III introduces a QoE metric. Section IV describes an experimental network, contents to be assessed, methods of threshold selection and a method of measuring QoE. Section V presents experimental results and show the effectiveness of the two proposed methods. Section VI concludes the paper.

II. SCS

This section first describes the principle of SCS and then presents methods of selecting threshold for SCS.

A. Principle

The operation of SCS depends only upon the receiver, not on any network function.

[1]In IP networks, six kinds of QoS are identified along the protocol stack: *physical–level*, *link–level*, *network–level*, *transport(end–to–end)–level*, *application–level*, and *user–level*.

[2]QoE represents the overall acceptability of an application or service, as perceived subjectively by the end–user [3]. It corresponds to user–level QoS.

SCS is based upon two basic video output techniques that cope with packet loss and error: error concealment and frame skipping [4]. It switches from error concealment to frame skipping once the percentage of error–concealed slices in a frame exceeds a threshold value T_h; the frame skipping continues until an intra–coded frame is decoded.

SCS utilizes the QoE tradeoff relation caused by the two techniques; furthermore, it takes into consideration the cross-modal interaction between audio and video[3]; although SCS is a video output scheme, it can reflect the effect of audio on the overall QoE.

A brief comment on why this type of switching is effective in improving QoE follows.

The video error concealment intends to conceal the visual effects of packet loss and error by either interpolating a missing block from its neighboring blocks or replacing it with some appropriate block in a previously decoded frame (see [8] and [9], for instance). However, since the concealment is not necessarily perfect, it causes residual errors, which can propagate to the succeeding frames. Therefore, concealed video frames are output with degraded picture quality compared to the original pictures. This implies that the spatial quality deteriorates. At the expense of this, however, the output frame rate is kept high, i.e, high temporal quality.

The video frame skipping here does not decode a frame unless all packets of the frame are correctly received. This means frame skipping until an intra-coded frame is decoded. Therefore, the spatial quality of the output frames is kept original, whereas the output frame rate decreases, i.e., low temporal quality.

Thus, the two techniques bring about the opposite effects on the temporal and spatial quality of output video, which leads to the QoE tradeoff relation. Because of the relation, we can expect that an appropriate mixture of the two techniques enhances QoE compared to the adoption of either of the two techniques.

SCS is a simple implementation example of the utilization of the relation; it adopts a threshold value as the criterion of the switching. The threshold selection also enables the reflection of the cross–modality between audio and video in addition to the video tradeoff relation between spatial and temporal quality. This is because we can adjust the weight of video information in the output audiovisual stream through the threshold value, depending on whether the content is audio–dominant like music video or video–dominant like sport.

Note that the case of $T_h = 100\%$ is equivalent to the pure error concealment technique, whereas $T_h = 0\%$ implies the simple frame skipping without error concealment.

B. Threshold selection methods

This paper examines four methods of threshold selection: (1) pure video frame skipping, which corresponds to $T_h = 0\%$ and therefore is referred to as the 0% *method*, (2) pure error concealment, which is called the 100% *method*, (3) the *table lookup method*, and (4) the *user selection method*.

Now, let us define the table lookup method and the user selection method in more detail. For simplicity of discussion, this paper supposes a finite set of threshold values to be selected (e.g., 100%, 40%, 20% and 0% as in [4] and [5]).

1) Table lookup method: This method looks up a table showing appropriate threshold values corresponding to the attribute of used media; for example, a threshold value is given for every combination of the content type, image size, picture pattern for permitted sets of video and audio encoding schemes in the target system. We suppose that the information of the media attribute is delivered from the sender to the receiver at the session setup and/or through packet headers.

[3]The interaction between the two media has been well-known for a long time [6]. ITU–T Recommendation J.148 sets a framework for this issue [7].

This method needs to compile the table in advance. For that purpose, we have to carry out subjective experiment on measuring QoE for all possible combinations of media attributes. Once the table becomes available, the operation in service is quite simple, just looking up on the table; this is an advantage of the method.

A disadvantage of the method is that it keeps a constant threshold value regardless of network conditions and the user's taste.

2) User selection method: This method allows the user to select the most favorite threshold value through a GUI which displays the possible alternatives as we will see later in Fig. 2. At the beginning of the media transfer, the user can try all the threshold values until he/she finds his/her favorite. The default value can be set to the value in the table lookup method, for instance.

This method can achieve the highest QoE for individual users in principle, since it can manage each user's inclination. On the other hand, however, this obliges the user to interact.

In this paper, we assume that the user does not change the threshold value during the session for simplicity of discussion.

III. QoE Metric

As the QoE metric, this paper adopts the *psychological scale* [10] as in [4] and [5]. This is because the psychological scale, which is the *interval scale* in psychometric theory [11], [12], can represent human subjectivity more accurately than *MOS* (*Mean Opinion Score*), which is the most popular QoE metric.

The calculation of the interval scale is based partly on the same method as that of MOS: the *rating–scale method*. In this method, each assessor (i.e., subject) gives a score to the audio-video stream output at the receiver, which is referred to as *stimulus* in psychometric theory; the score is an integer within 5 through 1 in order of highly perceived quality, for instance. Simply taking an average of the scores over all the subjects gives MOS, while the interval scale is obtained by applying the *law of categorical judgment* to the scores.

In the case of the interval scale, we further conduct *Mosteller's test* [11], [13] 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. See [10] for more detail.

IV. Experimental Methodology

In order to show the effectiveness of the table lookup method and the user selection method, we conducted experiment on the assessment of QoE for the four methods of threshold selection. This section describes a methodology for the experiment, including an experimental network, contents, implementation of the two proposed methods of threshold selection and how to measure QoE.

A. Experimental network

Figure 1 shows the configuration of the experimental network; it consists of two routers, Router 1 and Router 2, both of which are RiverStone's RS3000, and four PC's, which are used as a media sender (MS), a media recipient (MR), a Web server (WS), and a Web client (WC). The link between the routers and ones between a router and a PC are all full duplex Ethernet channels of 100 Mb/s.

The MS transmits an audio stream and the corresponding video stream, which are transferred as two separate streams with RTP/UDP, to the MR. The information unit for transfer between the application layers is referred to as the *MU* (*Media Unit*). A video MU is defined as a video frame and an audio MU as a constant number of audio samples. The MR exerts playout buffering control of 1 second to absorb delay jitters of received MU's. For video encoding and decoding, we utilize the H.264/MPEG–4 AVC reference software JM15.0 [14].

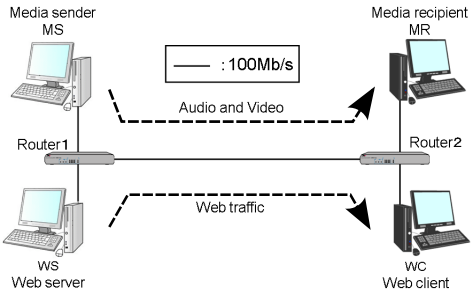

Fig. 1. Configuration of the experimental network

Table I gives specifications of audio and video used in the experiment. The audio has been encoded with linear PCM using 16 bits per sample at a sampling rate of 24 kHz. An audio MU consists of 480 samples, which compose a single UDP datagram. A video MU (i.e., frame) is divided into 15 UDP datagrams each of which corresponds to a slice. We adopt no *FMO* (*Flexible Macroblock Ordering*) of slice group map types. We have set three kinds of picture patterns, which are I followed by $n-1$ P's (n=1, 5, and 15), to examine their effects on QoE achieved by the threshold selection methods.

TABLE I

SPECIFICATIONS OF AUDIO AND VIDEO

audio coding scheme	Linear PCM 24kHz 16bit 1ch
audio MU size [byte]	960
audio average bit rate [kb/s]	384 (=50MU/s)
video coding scheme	H.264 (JM15.0)
image size [pixel]	320×240
number of slices in a picture	15 (20 macroblocks/slice) (no FMO)
video average MU rate [MU/s]	30
picture pattern (GOP)	I IPPPP IPPPPPPPPPPPPPP (I+14P's)
recording time [s]	10

As the error concealment technique in this paper, we employ the one implemented in JM15.0. A missing block in an I frame is interpolated from its four neighboring blocks. For P–frames, we use a temporal approach of *Frame Copy*; it simply replaces the missing block with the spatially corresponding one of the previously output frame[4].

In the experiment, we cause audio/video packet loss by transmitting interference traffic to the audio–video streams. We transfer HTTP messages from the Web server (WS) to the Web client (WC) as in [15]; it is generated according to the configuration of WebStone 2.5 [16]. WebStone generates Web client processes on the WC PC; those client processes retrieve specified files from the WS PC continuously. In the experiment, the number of the Web client processes was set to 20, 30, 40, 50, 75 and 100. As the number of the processes increases, the amount of the Web traffic becomes larger, and therefore packet loss occurs more frequently.

In the experiment on the SCS, we set the threshold value T_h to 100%, 40%, 20% and 0% as in [4] and [5].

B. Contents

Referring to the VQEG multimedia test plan [17], we have selected three types of contents: *sport, music video* and *animation*. Sport has been selected as a video–dominant content type, where video plays a more important role than audio, while music video is considered audio–dominant. Animation has different features from sport and music video, especially in video with respect to the picture property and frame rate; the animation is usually made at a lower frame rate (say 24 fps or less) than the others.

Table II is a list of contents used in the experiment along with a brief comment on scenes in each content[5]. For each of the three content types, we have prepared two contents; a content labeled 1 (say sport 1) has low motion video, and the one labeled 2 (say sport 2) exhibits high motion.

TABLE II

CONTENTS USED IN THE EXPERIMENT

Content	Scenes
sport 1	Scenes of a singles game of tennis by two female players. The audio includes strokes and a commentator's voice. There is no scene change.
sport 2	Scenes of a soccer game. Players pass the ball several times and then score a goal. The audio includes a commentator's voice and spectators' cheers. No scene change.
music video 1	Views of a young lady from the shoulders up singing a slow tempo song. She hardly moves. No scene change.
music video 2	Three scenes of a rock band with five people. The camera focuses on the vocalist, zooming in and panning rapidly in each scene. The music is high tempo.
animation 1	Two scenes in which two characters are talking with soft background music. One scene is the characters in the distance, and the other is a close up.
animation 2	A male character riding a dragon is falling from the sky. The audio is mainly the character's voice with loud background music. There are four scene changes.

Table III shows the video average bit rate and the *TI* (*Temporal perceptual Information*) value[6] for the three picture patterns (GOP) in each content. The TI values in this table have been calculated by eliminating the effect of scene changes. We notice that the second content in each type has a larger TI value than the first. Note that the TI value is not used in the operation of SCS; it is shown just for information of the degree of video motion.

C. The table lookup method for threshold selection

This subsection gives an implementation example of the table lookup method, utilizing the experimental results reported in [4]. We can summarize the results in [4] as follows.

1) Threshold values that achieve high QoE depend on the content type, picture pattern, and degree of video motion.
2) Pure frame skipping ($T_h = 0\%$) is the most effective for picture pattern I in all the contents.
3) When the picture pattern includes P's in video–dominant contents, nonzero T_h values work well; as the number of P's increases, a larger value of T_h achieves higher QoE.
4) When the picture pattern is IPPPP in audio–dominant contents with low motion video and animation, $T_h = 0\%$ is still better than the nonzero values.

[4]In JM15.0, another temporal approach is available, i.e., *Motion Copy*. It utilizes the information of the motion vector in the replacement. The application of Motion Copy is left as future work.

[5]These contents are different ones from those in [4] and [5], though the content types are the same.

[6]The TI value indicates the amount of temporal changes of a video sequence [18]. A higher value implies higher motion.

TABLE III

VIDEO BIT RATE AND TI VALUE FOR EACH CONTENT

content	GOP	Average bit rate [kb/s]	TI value
sport 1	I	2560.866	5.325
	IPPPP	801.429	
	I+14P's	523.278	
sport 2	I	3078.036	16.498
	IPPPP	1544.885	
	I+14P's	1311.782	
music video 1	I	1748.034	1.631
	IPPPP	472.004	
	I+14P's	278.821	
music video 2	I	3554.912	23.618
	IPPPP	2135.814	
	I+14P's	1924.333	
animation 1	I	1668.740	1.839
	IPPPP	445.900	
	I+14P's	243.734	
animation 2	I	1818.880	15.660
	IPPPP	1018.413	
	I+14P's	900.554	

5) Even in audio–dominant contents and animation, nonzero values of T_h become effective when the number of P's increases to IPPPPPPPPPPPPPP (I+14P's).
6) For pictures with many P's and/or when the video motion is high, $T_h = 100\%$ is the best.

On the basis of the above results, we have made Table IV as an example of the lookup table for possible combinations of the picture pattern and the content type; we have used this table in the experiment.

The reason for the selection of the T_h values in the table are obvious from the above summary. However, $T_h = 40\%$ for music video with picture pattern IPPPP might need some explanation. Although this content is audio–dominant, frequent video freezing due to frame skipping makes out of lip–sync noticeable; this leads to low QoE. $T_h = 40\%$ can suppress frequent video freezing, while it does not display largely degraded pictures.

TABLE IV

LOOKUP TABLE FOR T_h

Picture pattern	Content type	T_h [%]
I	sport	0%
	music video	
	animation	
IPPPP	sport	20%
	music video	40%
	animation	0%
IPPPPPPPPPPPPPP	sport	100%
	music video	
	animation	

D. The user selection method for threshold selection

Figure 2 illustrates the GUI designed for the user selection method; radio buttons A, B, C and D correspond to 100%, 40%, 20% and 0% respectively. The initial (i.e., default) value is set to the value specified in the table lookup method. The user can select any button by clicking it.

Once the output of the audio-video stream begins, the user can try all the buttons until he/she chooses the most favorite one. In this paper, we assume that the threshold value remains the same during the session for simplicity of the experiment, though the change would be possible. Allowing the changes during the session provides the adaptability to time-varying network conditions; a study on this scheme is left as future work.

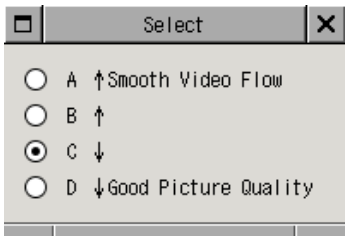

Fig. 2. GUI for the user selection method

E. QoE assessment

In order to obtain the psychological scale, we first carried out subjective experiment by the rating–scale method. We employed the *Degradation Category Rating* (*DCR*) [18] with the following *five–level impairment scale*: "imperceptible" assigned score 5, "perceptible, but not annoying" 4, "slightly annoying" 3, "annoying" 2, and "very annoying" 1, which are referred to as *Category 5* through *Category 1*, respectively.

We define an experimental run as the transmission of a content with a picture pattern at a constant level of the average Web traffic (i.e., when the number of Web client processes is kept constant). During each experimental run, we recorded the audio–video streams that the media recipient output; the recorded streams are regarded as stimuli for QoE measurement. Thus, we totally have 432 stimuli because of six contents, three picture patterns for each content, four values of T_h, and six levels of the average Web traffic.

We put the 432 stimuli in a random order and presented them to subjects, using a PC with headphones and a 17–inch LCD display. The distance between the display and each subject was set to that in the case where he/she usually uses a PC (i.e., approximately 50 cm through 1 m).

The number of the subjects is 45: 27 Japanese males and 18 Japanese females, who were university students and homemakers. Their age ranged from around twenty through forty. They were non–experts of audio and video quality assessment. It took about 4.5 hours including break time for a subject to assess all the stimuli.

V. EXPERIMENTAL RESULTS

In this section, we first present the psychological scale values calculated from the result of the subjective experiment on the four methods of threshold selection. We then examine the values to show the effectiveness of the proposed methods.

A. Psychological scale

As already mentioned in Sec. III, we calculate the interval scale by the rating–scale method and the law of categorical judgment, which are collectively called the *method of successive categories*. In order to compare the interval scales for the six contents on the same basis, we applied the law of categorical judgment to all the results obtained by the rating–scale method for the six contents together, i.e., the 432 stimuli.

Then, by carrying out Mosteller's test for the goodness of fit of the interval scale, 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. Removing 28 stimuli which give a large error of Mosteller's test, we saw that the hypothesis cannot be rejected. Consequently, we can consider the interval scale for the 404 ($= 432 - 28$) stimuli as the psychological scale.

Since we can select an arbitrary origin in an interval scale [10], let us set the minimum value of the psychological scales for the 404 stimuli to unity (i.e., 1). We then calculated the lower boundaries of the categories to be 4.878 for Category 5, 3.871 for Category 4, 2.914 for Category 3, and 1.861 for Category 2.

Figures 3 through 8 plot the psychological scale versus the number of Web client processes for the four methods of threshold selection. Straight broken lines parallel to the abscissa in the figures represent the lower boundaries of the categories. Figures 3 through 5 present the results of sport 2 for picture patterns I, IPPPP and I+14P's, respectively. Figure 6 corresponds to animation 1 with picture pattern IPPPP, and Figs. 7 and 8 to music video 2 with picture pattern IPPPP and I+14P's, respectively. Note that the results removed by the Mosteller's test are not shown in the figures.

B. Comparison of the threshold selection methods

In Figs. 3 through 8, we first notice that the psychological scale value for each method tends to be smaller as the number of Web client processes increases. This is because the amount of interference traffic increases, which degrades the output quality of the audio–video streams. When the number of Web client processes is 20, however, we have hardly observed quality degradation such as packet loss and delay jitter; this implies that the difference in the psychological scale value among the four method in this case is due to the fluctuation of the measurement. Therefore, in the following discussion, we will focus on the results for 30 and more Web client processes and examine them for each of the three picture patterns in turn.

1) Picture pattern I: Figure 3 shows that the 100% method exhibits the lowest value of the psychological scale among the four methods. That is, pure error concealment is ineffective in improving QoE for picture pattern I, whereas the other methods work well. As Table IV indicates, the table lookup method always selects $T_h = 0\%$ for this picture pattern. Regarding the user selection method, we have observed that the relative frequency of $T_h = 0\%$ in the subjective experiment is approximately 0.76. The above observations are consistent with those in [4] and [5].

2) Picture pattern IPPPP: Figures 4, 6 and 7 deal with this picture pattern. As seen from Fig. 4, the 0% method for sport 2 tends to provide the lowest QoE among the four methods. On the other hand, Fig. 6 reveals that the 100% method is clearly the worst for animation 1 and that the others including the 0% method are approximately the same. The two proposed methods work well for both content types. In the case of music video 2, the four methods are comparable.

Low QoE of the 0% method for sport 2 comes from the vulnerability of sport 2 to frame skipping owing to the high motion video. The table lookup method, which works better than the 0% method, employs $T_h = 20\%$.

The result of animation 1 is due to its features of picture property and frame rate. That is, residual errors in animation's pictures caused by incomplete error concealment are noticeable compared to the other content types, and therefore frame skipping is effective; this is validated by lower frame rates required by animation. Note that the table lookup method selects $T_h = 0\%$ in this case; also, the measurement of the user selection method in this case indicated that $T_h = 0\%$ was selected with a relative frequency of 0.77.

We can understand the result of music video by noting the audio dominance of this content type.

3) Picture pattern IPPPPPPPPPPPPP: From Figs. 5 and 8, we confirm that the 0% method exhibits the lowest QoE; this is due to much lower output frame rates caused by skipping many P frames. Note that even in music video, which is an audio dominant content, the 0% method cannot achieve high QoE, though it tends to work better than the case of sport 2. For this picture pattern, the table lookup method set $T_h = 100\%$, and the user selection method chose $T_h = 100\%$ with a relative frequency of 0.68 for sport 2 and 0.61 for music video 2.

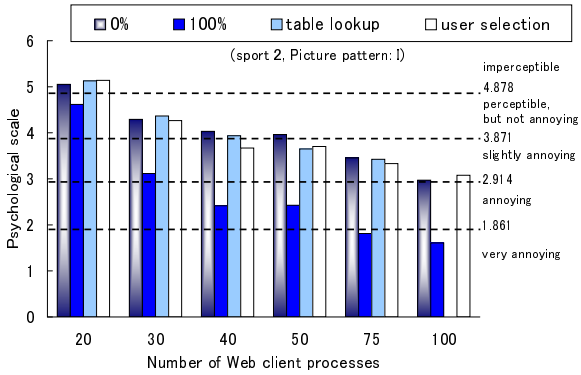

Fig. 3. Psychological scale versus number of Web client processes (sport 2, Picture pattern: I).

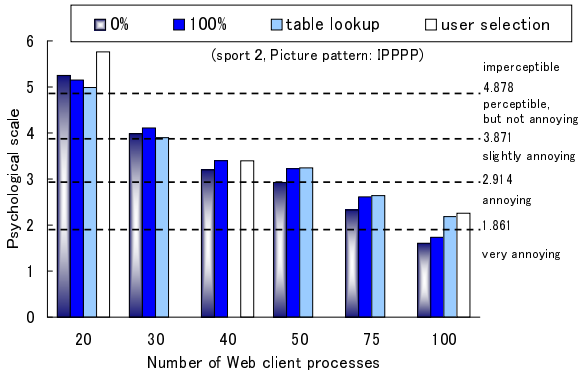

Fig. 4. Psychological scale versus number of Web client processes (sport 2, Picture pattern: IPPPP).

C. Discussion on the threshold selection methods

In the previous subsection, we observed that the table lookup method and the user selection method work well and therefore they are practically usable.

It should be noted here that the results in this paper have been obtained under many conditions. In particular, we have assumed that the threshold value is not changed during the session for simplicity of discussion. This restriction is inappropriate for time–varying traffic. The two proposed methods should be extended so that they can cope with this situation.

A solution to this problem for the table lookup method is to prepare multiple threshold values for each type of supposed media attributes and some mechanism to switch the threshold value according to the amount of traffic, which can be monitored at the receiver.

The user selection method can deal with time-varying traffic more flexibly in principle; however, it imposes a heavier burden on the users. We need some mechanism to alleviate the burden; for example, when the amount of network traffic changes drastically, the threshold value is set to the one specified by the table lookup method; then, the user can adjust the value through the GUI if necessary.

Another approach to time-varying traffic is the QoE estimation method, though it was not studied in this paper; the method can adapt to this situation automatically.

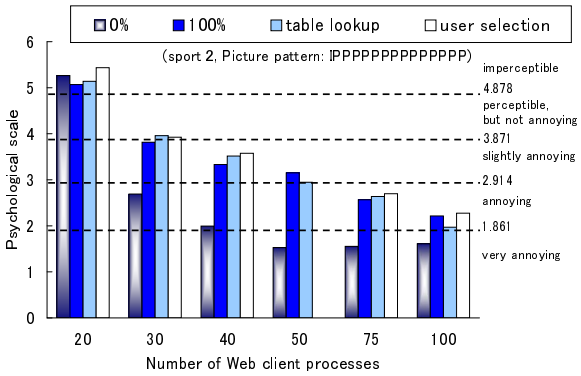

Fig. 5. Psychological scale versus number of Web client processes (sport 2, Picture pattern: IPPPPPPPPPPPPP).

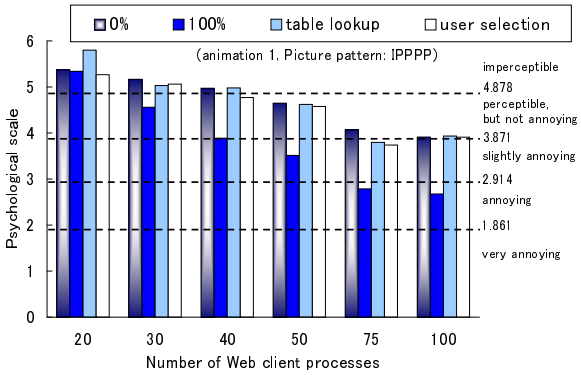

Fig. 6. Psychological scale versus number of Web client processes (animation 1, Picture pattern: IPPPP).

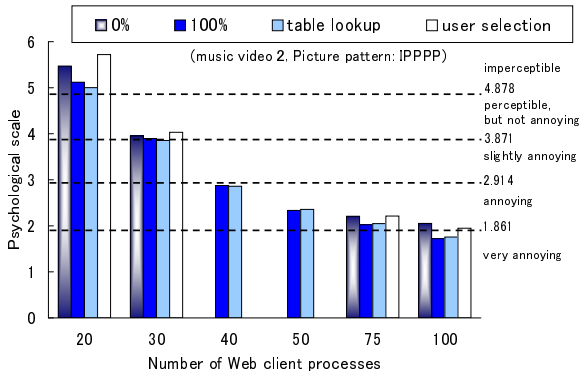

Fig. 7. Psychological scale versus number of Web client processes (music video 2, Picture pattern: IPPPP).

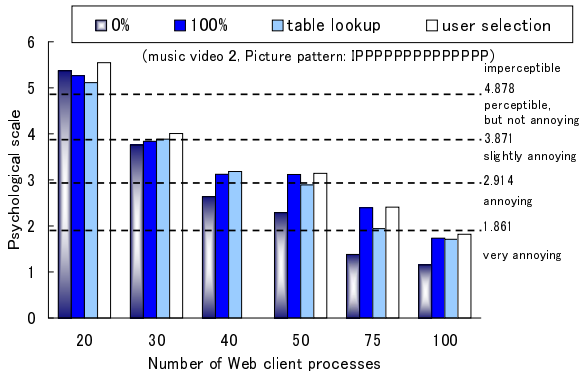

Fig. 8. Psychological scale versus number of Web client processes (music video 2, Picture pattern: IPPPPPPPPPPPPP).

VI. Conclusions

This paper proposed two practical methods of threshold selection for SCS, the table lookup and the user selection, and demonstrated their effectiveness. We saw that the two methods outperform the two conventional end–terminal techniques for coping with packet loss and error, pure error concealment (the 100% method) and pure frame skipping (the 0% method), in many cases and perform comparably in the other cases.

Future work includes extension of the two proposed methods to allow the change of the threshold value during the session and QoE assessment of the extended methods in time–varying traffic environments. Also, the QoE estimation method is an important approach to this issue. Investigation on SCS using HD video is one of our research subjects.

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] P. A. Chou and M. van der Schaar ed., *Multimedia over IP and wireless networks: compression, networking, and systems*, Academic Press, 2007.

[2] C.–S. Lee and D. Knight, "Realization of the next–generation network," *IEEE Commun. Mag.*, vol.43 No.10 pp.34–41,Oct. 2005.

[3] ITU-T Rec. G.100/P.10 Amendment 1, "Amendment 1: new appendix I definition of Quality of Experience (QoE)," Jan. 2007.

[4] 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 GLOBECOM2008*, Dec. 2008.

[5] 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 Multimedia*, pp.259–268, Oct. 2008.

[6] S. Tasaka, J. Sako and Y. Ito, "Enhancement of user–level QoS in audio–video IP transmission by utilizing the mutually compensatory property," in *Conf. Rec. IEEE GLOBECOM2006*, Nov. 2006.

[7] ITU-T Rec. J.148, "Requirements for an objective perceptual multimedia quality model," May 2003.

[8] L. Atzori, F. G. B. De Natale, and C. Perra, "A spatio-temporal concealment technique using boundary matching algorithm and mesh–based warping (BMA–MBW)," *IEEE Trans. Multimedia*, vol.3,no.3,pp.326–338, Sep. 2001.

[9] S. Belfiore, M. Grangetto, E. Magli, and G. Olmo, "spatio-temporal video concealment with perceptually optimized mode selection," in *Proc. IEEE ICASSP*, Apr. 2003.

[10] S. Tasaka and Y. Ito, "Psychometric analysis of the mutually compensatory property of multimedia QoS," in *Conf. Rec. IEEE ICC2003*, pp. 1880–1886, May 2003.

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

[12] J. C. Nunnally and I. H. Bernstein, *Psychometric theory, Third edition*, McGraw-Hill, N. Y., 1994.

[13] 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.

[14] "H.264/MPEG-4 AVC reference software JM15.0," http://iphome.hhi.de/suehring/tml/index.htm.

[15] S. Tasaka, Y. Ito, H. Yamada and J. Sako, "A method of user–level QoS guarantee by session control in audio–video transmission over IP networks," in *Conf. Rec. IEEE GLOBECOM2006*, Nov. 2006.

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

[17] The Video Quality Experts Group, "Multimedia group test plan, draft version 1.19," http://www.its.bldrdoc.gov/vqeg/.

[18] ITU-T Rec. P.911, "Subjective audiovisual quality assessment methods for multimedia applications," Dec. 1998.