Psychometric Analysis of the Effect of Buffering Control on User–Level QoS in an Interactive Audio–Visual Application

Yoshihiro Ito
yoshi@nitech.ac.jp

Shuji Tasaka
tasaka@nitech.ac.jp

Yoshihiko Fukuta
fukuta@inl.elcom.nitech.ac.jp

Department of Computer Science and Engineering,
Graduate School of Engineering,
Nagoya Institute of Technology
Gokiso-cho, Showa-ku Nagoya, 466-8555, Japan

ABSTRACT

This paper studies subjective tradeoff between fidelity and latency in an interactive audio–visual application over the Internet. In audio–video transmission, its temporal structure is disturbed by delay jitter of packets. The fidelity means how exactly the temporal structure is preserved. The degradation in the fidelity can be improved by playout buffer. Longer buffering time can suppress the disturbance caused by larger delay jitter, while it increases the latency, which is the difference between the time when media are generated at the sender and the time when the media are output at the receiver. The increase of the latency degrades the interactivity between users. The decrease of the fidelity and the increase of the latency degrade user–level QoS in interactive audio–visual applications. However, by changing the buffering time, we can improve either the fidelity or the latency. This means that the buffering control causes subjective tradeoff between the fidelity and the latency. Therefore, we must investigate a way of setting appropriate buffering time. In this paper, we study the effect of buffering time on user–level QoS in an interactive audio–visual application. In order to assess the user–level QoS quantitatively, we adopt psychometric methods. By experiment, we find an appropriate value of the initial buffering time in our experimental environment.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: measurement techniques, modeling techniques, performance attributes

General Terms

Performance

Keywords

interactive audio–visual application, buffering control, user–level QoS, psychometric methods

1. INTRODUCTION

Owing to the spread of high–speed networks and high performance terminals, many applications can transmit audio–video streams over the Internet. An audio–video stream has the temporal structure, which is disturbed by delay jitter of packets. The disturbance of the structure decreases *fidelity* of the audio–video stream. The fidelity in this paper indicates how exactly the temporal structure is preserved; this is referred to as *media synchronization* [1].

We can improve the fidelity with a playout buffer in the receiver. That is, packets which arrive at the receiver are stored in its buffer so that the delay jitter can be absorbed. We refer to the packet delay caused by the playout buffer as the *buffering time*. Absorption of larger delay jitter requires more buffer space. In reality, however, a receiver has finite buffer space. Consequently, we cannot make the buffering time infinity.

On the other hand, the utilization of the playout buffer increases *latency* because of the buffering time. The latency is the difference between the time when media are generated at the sender and the time when the media are output at the receiver. The increase of the latency causes degradation of interactivity between users in interactive audio–visual applications, such as TV conferences. Thus, larger buffering time does not always contribute toward improving quality in those applications.

It is important to consider how users subjectively assess the quality in audio–visual applications. As we mentioned earlier, the subjective quality in the interactive audio–visual applications is affected by both fidelity and latency. By increasing the buffering time, the fidelity improves, while the latency rises. That is, in the interactive audio–visual applications, there exists a subjective tradeoff between the fidelity and the latency by the buffering control. In this paper, therefore, we study a method of setting appropriate buffering time which makes subjective quality high in the interactive audio–visual applications. The subjective quality corresponds to *user–level QoS* (*Quality of Service*) in the context of the network architecture.

Since the Internet has a layered structure, QoS also has a layered structure. For example, Tasaka and Ishibashi [2] identified six levels of QoS: *physical–level, node–level, network–level, end–to–end–level, application–level* and user–level. The user–level QoS is subjective one. Since the fidelity relates to the quality of the media synchronization, it is application–level QoS. On the other hand, the latency concerns delay of media, which is application–level QoS.

We cannot control user–level QoS directly because it is judged by users. However, it is feasible to control QoS at lower–levels so that user–level QoS is kept high. In this paper, we consider improving user–level QoS in the interactive audio–visual applications by controlling application–level QoS: the fidelity and the latency.

We can find many researches which report the effect of buffering control on user–level QoS of the audio–video transmission. However, these researches treat only the fidelity. For example, Kouvelas *et al.* [3] showed that a reconstruction buffer must be added to a video system for lip synchronization. In [3], the effectiveness of lip synchronization is confirmed by subjective assessment. Steinmetz [4] assumed that audio and video are individually buffered to absorb delay jitter; he investigated the tolerance of skew, which is difference between audio delay and video one caused by buffering control.

Concerning both fidelity and latency, the authors reported the effect of delay jitter and mean delay on user–level QoS of the live audio–video transmission over packet networks [5]. In the literature, however, we can find no report that treats the subjective tradeoff between the fidelity and the latency caused by buffering control in interactive audio–visual applications.

This paper investigates the effect of buffering time on user–level QoS in an interactive audio–visual application. In order to assess the user–level QoS, we utilize the *psychometric methods* [6]. The psychometric methods have been proposed in the psychological field and are effective in assessment of human subjectivity. By experiments with the psychometric methods, we seek the most appropriate buffering time for the interactive audio–visual application.

The rest of the paper is organized as follows. Section 2 describes application–level QoS parameters we use. Section 3 introduces a method of assessing user–level QoS with psychometric methods. Section 4 explains our experiment. We show our results and consideration in Sec. 5.

2. APPLICATION–LEVEL QOS PARAMETERS

In this paper, we consider the fidelity and latency as application–level QoS. In order to treat them quantitatively, we need to express them in terms of some application–level QoS parameters.

As described in the previous section, we assume that the fidelity indicates how exactly the temporal structure of media is preserved. The fidelity relates to media synchronization quality [2]. The media synchronization is defined for multimedia in general. In this paper, however, we treat only two types of media: audio and video.

The media synchronization for audio and video can be classified into *intra–stream synchronization* and *inter–stream synchronization* [1]. The former indicates the continuity of a single stream (audio or video), while the latter is synchro-

nization between an audio stream and the corresponding video one. We consider measures of media synchronization quality as application–level QoS parameters about the fidelity.

To represent media synchronization quality, the authors [5] used seven parameters for application–level QoS. We use the same application–level QoS parameters as those used in [5] as the parameters about the fidelity. First, we adopt the *coefficient of variation of output interval*, which is defined as the ratio of the standard deviation of the MU output interval of a stream to its average. MU stands for "media unit", which indicates an information unit for media synchronization. This parameter is denoted by C_a for audio and by C_v for video. Second, we use the *average MU rate* for audio R_a and that for video R_v; the average MU rate is defined as the average number of (either audio or video) MUs output in a second at the destination. Third, we treat the *mean square error of intra–stream synchronization*, which is defined as the average square of the difference between the output interval of MU at the destination and the generation one at the source. We denote it by E_a for audio and by E_v for video. These six parameters indicate the intra–stream synchronization quality.

The QoS parameter for the inter–stream synchronization is the *mean square error* E_{int}, which is defined as the average square of the difference between the output–time difference of the audio and corresponding video MUs and their timestamp difference. We use the seven application–level QoS parameters introduced so far as the fidelity measure.

For the purpose of examination of the application–level QoS from a latency point of view, we evaluate two application–level QoS parameters which are used in [5]. These are the *average MU delay* of audio and that of video. The average MU delay is the average time in seconds from the moment an MU is generated until the instant the MU is output. We denote it by D_a for audio and by D_v for video.

3. PSYCHOMETRIC METHODS FOR USER–LEVEL QOS ASSESSMENT

In order to investigate the effect of the buffering time on user–level QoS in the interactive audio–visual applications, we must assess the user–level QoS quantitatively. The most popular scheme for assessing user–level QoS is the one which utilizes *MOS* (*Mean Opinion Score*) as a user-level QoS parameter. However, MOS is not necessarily an exact measure since it is an *ordinal scale*. As exact measures, we can employ *interval* and *ratio* scales. By utilizing psychometric methods, we can assess user–level QoS with an interval scale. Four general classes of scales including ordinal, interval and ratio scales will be explained in the next subsection.

Some researches adopted the psychometric methods for the assessment [7], [8]. In [7], Ito and Tasaka assess user–level QoS of audio–video transmission by the method of paired comparisons and Thurstone's law of comparative judgment [6]. Reference [8] studies the mutually compensatory property of multimedia QoS by the *method of successive categories* [6]. The method of paired comparisons and Thurstone's law of comparative judgment can give more accurate values of the interval scale but takes longer experimental time than the method of successive categories. In this paper, we utilize the method of successive categories. The method of successive categories is introduced in this section.

3.1 Four General Classes of Measurement Scales

Before we explain the method of successive categories, let us consider four general classes of measurement scales which represent human subjectivity. In general, we can define four basic types of the measurement scales according to the mathematical operations that can be performed legitimately on the numbers obtained by the measurement; from lower to higher levels, we have *nominal*, ordinal, interval and ratio scales [6]. In the nominal scale, we use a number only as a label for a class or a category. The numbers assigned in the ordinal scale have the property of rank order. In the interval scale, numerically equal distances stand for empirically equal distances in some psychological aspect of objects. However, the origin and the unit of the interval scale are meaningless. In the ratio scale, the unique origin can be determined in addition to the property of the interval scale. Since almost all the statistical procedures can be applied to the interval scale and the ratio scale, it is desirable to represent user–level QoS by an interval scale or a ratio scale.

3.2 Method of Successive Categories

In the method of successive categories, a subjective score is measured by the *rating-scale method* [6]. In the method, experimental subjects (or observers) classify each stimulus into one of a certain number of categories. Note that a stimulus means an object, such as audio and video, for evaluation. Each category has a predefined number. For example, "excellent" is assigned 5, "good" 4, "fair" 3, "poor" 2 and "bad" 1. 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. Therefore, it is not desirable to use the assigned number for obtaining the user–level QoS parameter.

In order to obtain an interval scale as the user–level QoS parameter, we first measure the frequency of each category with which the stimulus was placed in the category by the rating-scale method. With the *law of categorical judgment* [6], we can translate the frequency obtained by the rating–scale method into an interval scale. We can apply almost all the statistical operations to the scale.

3.3 The Law of Categorical Judgment

The law of categorical judgment makes the following assumptions. Let the number of the categories be $m + 1$. When stimulus $j(j = 1, \cdots, n)$ is presented to a subject, a psychological value designated by s_j occurs on a *psychological continuum*, which is an interval scale, in him/her. For the $m + 1$ categories, their *boundaries* have values on the interval scale. We denote the upper boundary of category $g(g = 1, \cdots, m + 1)$ by c_g and define $c_0 \overset{\Delta}{=} -\infty$ and $c_{m+1} \overset{\Delta}{=} +\infty$. The subject classifies n stimuli into the $m + 1$ categories $(n > m + 1)$ by comparing s_j with c_g. If $c_{g-1} \le s_j < c_g$, then stimulus j is classified into category g. The categories can be arranged in a rank order, in the sense that each stimulus in category g is judged to have a psychological value which is "less than" the one for any stimulus in category $g + 1$. This statement holds for all values of g from 1 to m. The variable c_g is normally distributed with mean t_g and standard deviation d_g. Also, the variable s_j is normally distributed with mean R_j and standard deviation σ_j. Then, we can consider R_j as 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 [9] proposed a method of testing the goodness of fit for a scale calculated with Thurstone's law of comparative judgment [6], which is one of psychometric methods. The method can be applied to a scale obtained by the law of categorical judgment. In this paper, we use Mosteller's method to test the goodness of fit.

4. EXPERIMENTS

4.1 Experimental Environment

In order to clarify the subjective tradeoff between the fidelity and the latency, we set up an experimental environment shown in Fig. 1. In the experimental environment, a
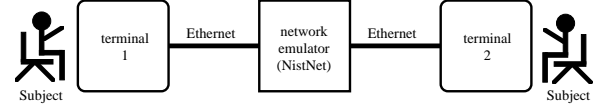

Figure 1: Experimental environment.

pair of subjects sit in front of their terminals, and each terminal transmits a pair of audio–video streams of the subject to each other over a network emulator which produces delay jitter. The subjects assess the output audio–video stream subjectively. Each terminal can perform the buffering control. By adjusting the buffering time at the terminals, we can realize the tradeoff between the fidelity and the latency. If we increase the buffering time, the fidelity improves, but the latency rises. Conversely, if we reduce the buffering time, the latency decreases, but the fidelity deteriorates.

The audio–video streams are encapsulated into UDP datagrams. The media specifications of the audio–video streams are shown in Table 1 .

Table 1: Specifications of audio–video streams.

	audio	video
coding scheme	unsigned 8bit PCM	MPEG1
image size[pixels]	-	240×180
picture pattern	-	IPPP
average MU size[byte]	400	3138
average MU rate[MU/s]	20	20
average MU interval[ms]	50.0	50.0
average bit rate[kb/s]	64	502

In the subjective assessment of the audio–video streams, the subjects perform a task. The task is a simple game. One subject indicates a direction, right, left, up or down, with his/her forefinger. At the same time, the other one turns his/her face toward a direction. If the direction toward which the latter turned his/her face is the same as the one

indicated by the former's forefinger, the winner is the former. They change their roles alternately.

The subjects assess their subjectivity of the audio–video stream with the rating–scale method. In this method, we use five categories (i.e., $m = 4$) of impairment: "imperceptible" assigned integer 5, "perceptible, but not annoying" 4, "slightly annoying" 3, "annoying" 2, and "very annoying" 1.

We use NistNet [10] as the network emulator. NistNet can delay packets according to a specified probability distribution.

We delay packets according to the Pareto–normal distribution to emulate packet delay of the Internet. The Pareto-normal distribution is the normal distribution with the Pareto tail. Fujimoto *et al.* [11] showed that the Pareto–normal distribution is an appropriate model of packet delay distributions in the Internet. Therefore, we adopted this model.

In this experiment, as the first step of our research, we set the mean of delay and the standard deviation of delay to 50 ms and 20 ms, respectively.

4.2 Scheme of Buffering Control

In order to change the buffering time at the terminals in our experimental environment, we utilize a simple buffering control scheme, which is introduced below.

For the description of the scheme, we define the following notations for stream j ($j = 1$ for audio, and $j = 2$ for video). Firstly, we let $T_n^{(j)}$ ($n = 1, 2, \cdots$) denote the timestamp of the n–th MU in stream j, which is attached when it generates, and define $\sigma_{n,m}^{(j)} \triangleq T_m^{(j)} - T_n^{(j)}$ ($n \leq m$; $m = 1, 2, \cdots$). Secondly, let $J_{\max}$ be an estimate of the maximum delay jitters. We call $J_{\max}$ the *initial buffering time*. Thirdly, let $A_n^{(j)}$ and $D_n^{(j)}$ represent the arrival time and output time, respectively, of the n–th MU in stream j at the destination.

First, we determine the output time of the first MU in each stream, which is also used to obtain the time–origin for output control at the destination. Defining $A_1 \triangleq \max\left(A_1^{(1)}, A_1^{(2)}\right)$ and $T_1 \triangleq \min\left(T_1^{(1)}, T_1^{(2)}\right)$, we set the output time of the first MU in stream j ($j = 1$ and 2) to

$$D_1^{(j)} = A_1 + T_1^{(j)} - T_1 + J_{\max} \qquad (1)$$

Next, we define the *ideal target output time* $x_n^{(j)}$ of the n–th MU in stream j as

$$x_1^{(j)} = D_1^{(j)} \qquad (2)$$

$$x_n^{(j)} = x_1^{(j)} + \sigma_{1,n}^{(j)}(n = 2, 3, \cdots) \qquad (3)$$

We calculate the output time of each MU with the ideal target output time.

If $A_n^{(j)} \leq x_n^{(j)}$, $D_n^{(j)}$ is set to $x_n^{(j)}$. Otherwise, the n–th MU in stream j is dropped.

In our experiment, we vary the initial buffering time $J_{\max}$ to change the buffering time. We chose ten values from 10 ms to 100 ms as those of the initial buffering time, i.e., $J_{\max} = 10j$ ms ($j = 1, 2, \cdots, 10$). For each initial buffering time, we consider the stream which is output at the receiver as a stimulus. We refer to the stimulus for $J_{\max} = 10j$ as stimulus j.

4.3 Subjects

We used 16 subjects in the subjective assessment. The subjects were non–experts in the sense that they were not directly concerned with audio and video quality as a part of their normal work. They are male, and their ages were between 20 and 25. It took about thirty minutes for a subject to finish all assessment.

5. EXPERIMENTAL RESULTS

5.1 Results of Application–Level QoS Assessment

Figs. 2 through 6 plot measured application–level QoS parameters, which were introduced in Sec. 2. The figures also show a 95% confidence interval on each measured value.

Figure 2 displays the average MU rate for audio and that for video versus the initial buffering time. From Fig. 2, we
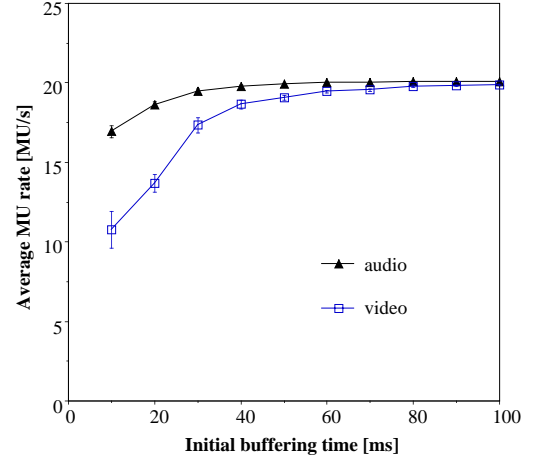

Figure 2: Average MU rate versus initial buffering time.

find that increase of the initial buffering time augments the average MU rate for audio and that for video.

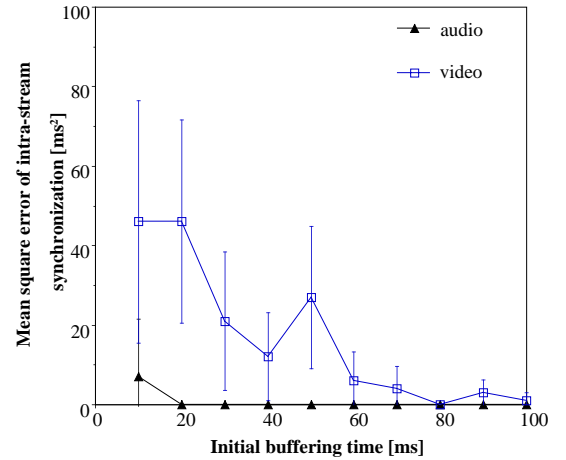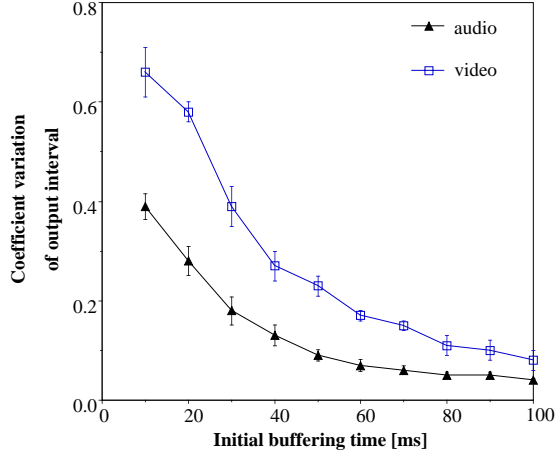Figure 3 denotes the mean square error of intra–stream synchronization versus the initial buffering time. It shows

Figure 3: Mean square error of intra–stream synchronization versus initial buffering time.

that the mean square error of intra–stream synchronization of audio is close to zero when the initial buffering time is 20 ms or more. On the other hand, the mean square error of intra–stream synchronization of video fluctuates, but it tends to decrease as the initial buffering time increases. This is because an audio MU can be transmitted with a single IP packet while a video MU consists of a few IP packets. Note that even if only one of the packets which compose a video MU arrives late, the output of the MU must be delayed.

The coefficients of variation of output interval of audio and that of video are indicated in Fig. 4. From this figure,

Figure 4: **Coefficient of variation of output interval versus initial buffering time.**

we see that the value of the coefficient decreases as the initial buffering time increases. However, the rate of the decrease also reduces according to the increment of the initial buffering time. This is because more packets can arrive at the receiver during the interval of the initial buffering time as it has a longer value. Figure 4 also shows that the coefficient of variation of output interval for video takes larger values than that for audio. This was caused by the difference in MU size between audio and video.
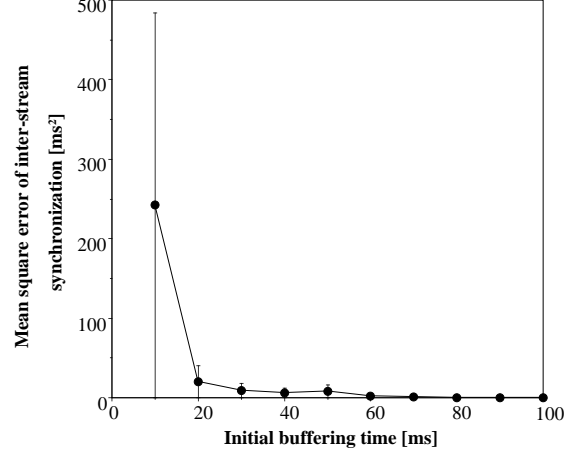
Figure 5 indicates the mean square error of inter–stream synchronization versus the initial buffering time. From Fig. 5, we find that the mean square error of inter–stream synchronization diminishes as the initial buffering time grows. This is because more buffering time can absorb larger delay jitter.

From Figs. 2 through 5, we can confirm that the fidelity of the audio–video streams improves with the increase of the initial buffering time.

Figure 6 plots the average MU delay of audio and that of video. This figure indicates that the average MU delay, that is, the latency, increases linearly as the initial buffering time rises. Thus, we can confirm that the increment of the initial buffering time improves the fidelity but increases the latency.

5.2 Calculation of User–Level QoS Parameter

Table 2 indicates the measurement result by the rating–scale method. Each entry in this table represents the number of subjects who classified the stimulus into the entry. From this table, we see that when the initial buffering time

Figure 5: **Mean square error of inter–stream synchronization versus initial buffering time.**
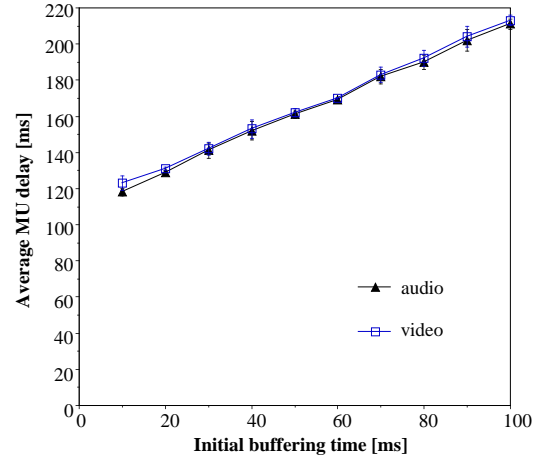
Figure 6: **Average MU delay versus initial buffering time.**

Table 2: **Measurement result by the rating–scale method.**

sti-mulus j	J_{max} [ms]	category				
		1	2	3	4	5
1	10	14	1	0	1	0
2	20	6	7	3	0	0
3	30	2	4	5	5	0
4	40	0	7	4	3	2
5	50	0	1	4	7	4
6	60	0	0	5	7	4
7	70	0	0	5	4	7
8	80	0	1	4	4	7
9	90	0	1	5	5	5
10	100	0	1	6	6	3

is 40 ms, for example, three subjects classified the stimulus into category 4.

Table 3 shows the cumulative proportion calculated from Table 2. The cumulative proportion for a category is defined as the proportion of subjects who classified the stimulus into the category or one below it to the total number of subjects. Therefore, the cumulative proportion of category 5 is 1.000.

Table 3: Cumulative proportion.

sti-mulus j	$J_{\max}$ [ms]	category				
		1	2	3	4	5
1	10	0.875	0.938	0.938	1.000	1.000
2	20	0.375	0.813	1.000	1.000	1.000
3	30	0.125	0.375	0.688	1.000	1.000
4	40	0.000	0.438	0.688	0.875	1.000
5	50	0.000	0.063	0.313	0.750	1.000
6	60	0.000	0.000	0.313	0.750	1.000
7	70	0.000	0.000	0.313	0.563	1.000
8	80	0.000	0.063	0.313	0.563	1.000
9	90	0.000	0.063	0.375	0.688	1.000
10	100	0.000	0.063	0.438	0.813	1.000

From Table 3, we calculate an interval scale with the law of categorical judgment. In the law, we can consider four conditions: conditions A, B, C and D, which differ in assumptions, approximations, and degree of simplification [12]. In this paper, we first try the simplest condition D. If we cannot confirm the goodness of fit for results thus obtained, we adopt another condition. Let the probability that s_j is less than c_g be p_{jg}. We regard the cumulative proportion as the observed value of the probability. Under condition D, the law of categorical judgment can be represented by

$$t_g - R_j = Z_{jg} \tag{4}$$

where Z_{jg} is the *normal deviate*, which is normally distributed with zero mean and unit variance, corresponding to the proportion p_{jg} as shown in Fig. 7.
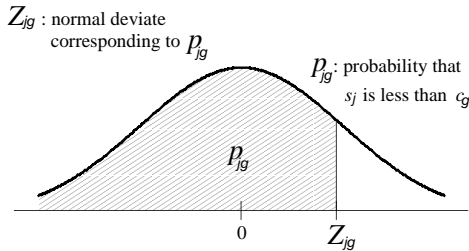
Figure 7: Normal deviate Z_{jg} corresponding to probability p_{jg}.

In the law of categorical judgment, we assume that there exist the true values for the quantities introduced in the previous section; we represent the corresponding observed values and estimated ones from observed data by attaching the prime ($'$) and the double prime ($''$), respectively, to the true values.

First, we regard the cumulative proportion shown in Table 3 as the probability p'_{jg}, which provides an estimated value of the normal deviate Z''_{jg}. Table 4 denotes the estimates of normal deviates Z''_{jg} calculated from Table 3. If an observed probability p'_{jg} is 0 or 1, Z''_{jg} becomes negative or positive infinity. In this case, Z''_{jg} is considered as a *missing entry*. Because each p'_{j5} is 1.000 in Table 3, Z''_{j5} is positive infinity. That is, we must consider all Z''_{j5}'s as missing entries. Therefore, we omit the column of category 5 from Table 4.

Table 4: Estimate of normal deviate Z''_{jg}.

sti-mulus j	$J_{\max}$ [ms]	category g			
		1	2	3	4
1	10	1.150	1.534	1.534	-
2	20	-0.319	0.887	-	-
3	30	-1.150	-0.319	0.489	-
4	40	-	-0.157	0.489	1.150
5	50	-	-1.534	-0.489	0.674
6	60	-	-	-0.489	0.674
7	70	-	-	-0.489	0.157
8	80	-	-1.534	-0.489	0.157
9	90	-	-1.534	-0.319	0.489
10	100	-	-1.534	-0.157	0.887

Using Z''_{jg}, we calculate the estimated values t''_g and R''_j for t_g and R_j, respectively. If some missing entries exist except for the column of category 5, we have an alternative of finding t''_g first or R''_j first. We select the former here; that is, we calculate t''_g first and then use the obtained values to determine R''_j.

In order to determine t''_g, we first estimate the average width of each category. Then, we regard one boundary as the origin and calculate the boundaries of the others. An estimated value $t''_{g+1} - t''_g$ can be calculated by

$$t''_{g+1} - t''_g = \frac{1}{q_g} \sum_{j}^{q_g} (Z''_{j,g+1} - Z''_{jg}) \tag{5}$$

where $\sum_{j}^{q_g}$ means the summation for j for which both Z''_{jg} and $Z''_{j,g+1}$ are available, and q_g is the number of the available data for a given g. Table 5 indicates $Z''_{j,g+1} - Z''_{jg}$ for $g = 1, 2$ and 3.

From Table 5, we can calculate $t''_{g+1} - t''_g$. By regarding the mean t''_1 of the upper boundary of category 1 as the origin of the obtained interval scale, we can obtain the mean of the upper boundary of each category. Thus, we have $t''_1 = 0.000$, $t''_2 = 0.807$, $t''_3 = 1.684$ and $t''_4 = 2.560$.

Next, R''_j can be obtained by

$$R''_j = \frac{1}{q_j} \sum_{g}^{q_j} (t''_g - Z''_{jg}) \tag{6}$$

where $\sum_{g}^{q_j}$ means the summation for g for which Z''_{jg} is available, and q_j is the number of the available data for a given j.

Table 6 displays R''_j in the rightmost column. The obtained R''_j is an interval scale, that is, the user–level QoS parameter.
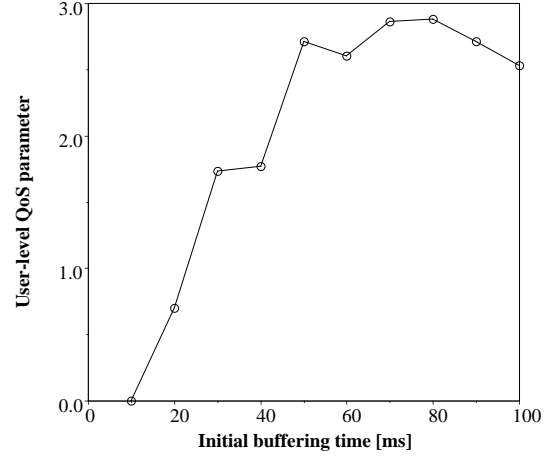
To verify the obtained interval scale, we have performed Mosteller's test. As a result of Mosteller's test, the null

hypothesis that the obtained interval scale fits the observed data cannot be rejected at significance level 0.05. That is, if the hypothesis is right, the probability that the hypothesis is rejected by mistake is less than 0.05. Therefore, we consider that the obtained scale is appropriate for the user–level QoS parameter.

We plot the calculated user–level QoS parameter in Fig. 8. Note that, in an interval scale, we can select an arbitrary origin and any unit of scale. For convenience, then, we set the smallest value of the user–level QoS parameter as the origin. That is, we have added 0.576 to each value of the user–level QoS parameter in Table 6.

Table 5: Estimate of category width.

sti-mulus j	$J_{\max}$ [ms]	$Z''_{j2} - Z''_{j1}$	$Z''_{j3} - Z''_{j2}$	$Z''_{j4} - Z''_{j3}$
1	10	0.384	0.000	-
2	20	1.206	-	-
3	30	0.832	0.807	-
4	40	-	0.646	0.662
5	50	-	1.045	1.163
6	60	-	-	1.163
7	70	-	-	0.646
8	80	-	1.045	0.646
9	90	-	1.215	0.807
10	100	-	1.377	1.044
	Ave.	0.807	0.877	0.876

Figure 8: User–level QoS parameter versus initial buffering time.

From Fig. 8, we see that the value of user–level QoS parameter grows as the initial buffering time increases until it reaches about 70 ms. However, when the initial buffering time exceeds 70 ms, the user–level QoS parameter does not gain. Moreover, when the initial buffering time becomes more than 80 ms, the value of the user–level QoS parameter begins to decrease. From Figs. 2 through 6, we have found that application–level QoS parameters about the fidelity rise but those about the latency increase with the increment of the initial buffering time. Therefore, we confirm the subjective tradeoff between the fidelity and the latency caused by the buffering control.

5.3 QoS Mapping

To investigate the subjective tradeoff between the fidelity and the latency by increasing the initial buffering time quantitatively, we perform QoS mapping between application–level QoS and user–level QoS. In [5], the authors perform QoS mapping between application–level QoS and user–level one by *multiple regression analysis*. In this paper, we also utilize multiple regression analysis as a QoS mapping scheme.

We consider the user–level QoS parameter and application–level QoS parameters as the criterion variable and predictor variables, respectively. We select two application–level QoS parameters as predictor variables; one is an application–level QoS parameter about the fidelity and the other is one about the latency. Therefore, we first select one parameter from among R_a, R_v, E_a, E_v, C_a, C_v and E_{int}. Then, we adopt either D_a or D_v. In this case, we must examine 7×2 com-

Table 6: Interval scale value R''_j of the stimulus j.

sti-mulus j	$J_{\max}$ [ms]	$t''_1 - Z''_{j1}$	$t''_2 - Z''_{j2}$	$t''_3 - Z''_{j3}$	$t''_4 - Z''_{j4}$	R''_j
1	10	-1.150	-0.727	0.150	-	-0.576
2	20	0.319	-0.080	-	-	0.119
3	30	1.150	1.126	1.195	-	1.157
4	40	-	0.964	1.195	1.409	1.190
5	50	-	2.341	2.173	1.885	2.133
6	60	-	-	2.173	1.885	2.029
7	70	-	-	2.173	2.402	2.287
8	80	-	2.341	2.173	2.402	2.305
9	90	-	2.341	2.002	2.071	2.138
10	100	-	2.341	1.841	1.673	1.952

binations of application–level QoS parameters as predictor variables. We select one combination of the application–level QoS parameters whose contribution rate adjusted for degrees of freedom is the highest. The contribution rate adjusted for degrees of freedom indicates goodness of fit of the obtained multiple regression line.

Table 7 shows contribution rates adjusted for degrees of freedom for the 14 combinations of the application–level QoS parameters. From Table 7, we find that two combinations,

Table 7: Contribution rates adjusted for degrees of freedom.

	R_a	E_a	C_a	R_v	E_v	C_v	E_{int}
D_a	0.945	0.877	0.976	0.962	0.853	0.964	0.884
D_v	0.945	0.873	0.976	0.962	0.848	0.965	0.880

(D_a, C_a) and (D_v, C_a) have the highest value of the contribution rate. In this paper, we select a parameter regarding audio and one concerning video. Therefore, we choose (D_v, C_a) as predictor variables. As a result of multiple regression analysis, the multiple regression line becomes

$$S = 3.859 - 3.496 \times 10^{-3} D_v - 9.145 C_a \qquad (7)$$

where S is an estimate of the user–level QoS parameter. This equation can be rewritten as

$$S = 3.859 - 9.145(C_a + 3.823 \times 10^{-4} D_v) \qquad (8)$$

From Eq. (8), we find the following. When we increase the initial buffering time by $\Delta J_{\max}$, for example, we suppose that the values of C_a and D_v increase by ΔC_a and ΔD_v, respectively. The negative value of ΔC_a means that the fidelity improves by the increment of the buffering time, while the positive value indicates that the fidelity deteriorates. In the same way, the negative value of ΔD_v indicates the decrease of the latency, while the positive one stands for the increase of the latency. Let H denote $\Delta C_a + 3.823 \times 10^{-4} \Delta D_v$. If H takes a negative value, the user–level QoS parameter will gain by increasing the initial buffering time. Otherwise, we should not increase the initial buffering time.

We set $\Delta J_{\max}$ to 10 ms, for instance. Figure 9 shows calculated values of H versus the initial buffering time in this case. From Fig. 9, we see that H takes a negative value at the initial buffering time of 20 ms and tends to gain as the initial buffering time increases. Moreover, this figure shows that the values of H are close to 0 when the initial buffering time is over 70 ms. That is, the user–level QoS does not gain even if we increase the initial buffering time over 70 ms. This accords with the conclusion in the previous subsection. Therefore, once we get a multiple regression line for the QoS mapping, we can adaptively set the initial buffering time which makes user–level QoS the highest according to network status without assessing the user–level QoS.

Note that Fig. 9 shows a different result from that of Fig. 8 when the initial buffering time is large. From Fig. 9, we see that the values of H do not always take positive values when the initial buffering time is over 80 ms. That is, this figure implies that the user–level QoS does not always deteriorate greatly by the increase of the latency. This is because the QoS mapping is imperfect. Development of more appropriate QoS mapping schemes is left as our future work.
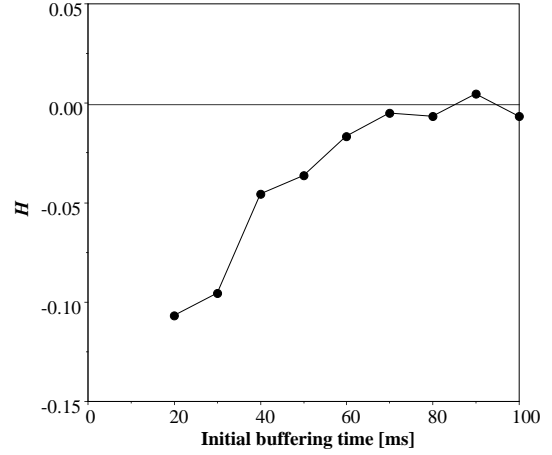
Figure 9: H versus initial buffering time.

6. CONCLUSIONS

In this paper, we assessed the effect of the buffering control on user–level QoS of an interactive audio–visual application. By experiment, we changed the initial buffering time while assessing the user–level QoS with the method of successive categories. As a result, we found that the appropriate value of the initial buffering time is about 80 ms in this experimental environment. That is, the initial buffering time less than 80 ms aggravates the user–level QoS because of degradation in the fidelity; the initial buffering time more than 80 ms causes subjective degradation owing to increase of the latency.

We have some issues to be investigated as our future work. First, in our experiment, we used one specific task. However, how the fidelity is subjectively traded for the latency depends on the kind of task. Some tasks may be tolerant of the latency. Therefore, we will treat other tasks.

Second, we used only one experimental environment. We will try other environments. Especially, we will perform experiments in actual networks, such as the Internet.

Finally, by considering the subjective tradeoff between the fidelity and the latency, we can perform the buffering control adaptively according to the network status so that the user–level QoS is kept high. We will devise some scheme based on this idea and implement it.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] G. Blakowski and R. Steinmetz. A media synchronization survey: Reference model, specification and case studies. *IEEE J. Sel. Areas in Commun.*, 14(1):5–35, January 1996.

[2] S. Tasaka and Y. Ishibashi. Mutually compensatory property of multimedia QoS. *Conf. Rec. IEEE ICC2002*, pages 1105–1111, April/May 2002.

[3] I. Kouvelas V. Hardman and A. Watson. Lip synchronisation for use over the Internet: Analysis and implementation. *Conf. Rec. IEEE GLOBECOM'96*, pages 893–898, November 1996.

[4] R. Steinmetz. Human perception of jitter and media synchronization. *IEEE J. Sel. Areas in Commun.*, 14(1):61–72, January 1996.

[5] Y. Ito, S. Tasaka, and Y. Fukuta. Psychometric analysis of the effect of end–to–end delay on user–level QoS in live audio-video transmission. *Conf. Rec. IEEE ICC2004*, pages 2214–2220, June 2004.

[6] J. P. Guilford. *Psychometric methods*. McGraw-Hill, New York, 1954.

[7] Y. Ito and S. Tasaka. Quantitative assessment of user–level QoS and its mapping. *IEEE Trans. Multimedia, to appear.*

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

[9] 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*, 16(2):207–218, June 1951.

[10] "URL http://www.itl.nist.gov/div892/itg/carson/nistnet/"

[11] K. Fujimoto, S. Ata, and M. Murata. Statistical analysis of packet delays in the internet and its application to playout control for streaming applications. *IEICE Trans. Commun.*, E84(6):1504–1512, June 2001.

[12] W. S. Torgerson. *Theory and methods of scaling*. J. Wiley, New York, 1958.