User-Level QoS Assessment with Psychometric Methods and QoS Mapping

Yoshihiro ITO and Shuji TASAKA
Department of Computer Science and Engineering
Graduate School of Engineering
Nagoya Institute of Technology,
Nagoya, 466-8555, Japan

ABSTRACT

This paper proposes a scheme for quantitative assessment of user-level QoS for audio-video transmission by means of two psychometric methods: the method of paired comparisons and the law of comparative judgment. We also discuss QoS mapping from application-level QoS to user-level QoS by the principal component analysis and multiple regression analysis. In the assessment, we simulate the transmission of an audio-video stream over a loaded network. In order to investigate the effect of the contents on QoS mapping, we treat two types of audio-video streams. By experiment, we demonstrate that our scheme can construct an interval scale as the user-level QoS parameter for each stream and represent it as a function of two application-level QoS parameters with high accuracy. We notice that the multiple regression line depends on the contents. We also propose the concept of control gain by media synchronization, which indicates how much media synchronization control subjectively lightens the average network load.

keywords: user-level QoS assessment, multimedia transmission, QoS mapping, psychometric methods, multiple regression analysis

1. INTRODUCTION

High-speed access networks of the Internet stimulate continuous multimedia applications such as VoD (Video on Demand). Unlike text-based data, continuous media have the temporal structure.

When continuous media are transmitted over the Internet, the temporal structure can be damaged by many factors like delay, delay jitter and packet loss. The preservation of the temporal structure requires Quality of Service (QoS) control over the Internet.

Owing to the layered architecture of the Internet, its QoS also has a layered structure. We can identify six levels of QoS in the Internet: physical-level, node-level, network-level, end-to-end-level, application-level and user-level [1].

Among the six levels of QoS, the user-level QoS is the most important since the users are the last recipients of the service. Therefore, it is desirable to control QoS at lower levels in order to make the user-level QoS satisfactory. For this purpose, a user-level QoS parameter needs to be measurable. Moreover, we have to clarify the QoS relationship between the user-level and each of the other five levels. Then, we can estimate the user-level QoS by measuring QoS at the lower levels, which is easy to measure. This QoS translation between the levels is referred to as QoS mapping.

The great majority of previous studies on QoS mapping treat the translation up to the end-to-end-level. In [2], the QoS mapping from the ATM layer to the IP layer is discussed. References [3] and [4] show methods that translate network-level QoS parameters into end-to-end-level ones.

With regard to the user-level QoS, references [1] and [5] deal with the relation between a user-level QoS parameter and application-level QoS parameters. In these studies, MOS (Mean Opinion Score) is used as the user-level QoS parameter. For measurement of MOS values, the rating-scale method [6], where an experimental subject classifies objects into some categories, is used. Each category is assigned an ordinal number. The ordinal number is regarded as a subjective score. A MOS value for an object is the averaged score over all subjects. MOS is commonly used as a subjective measure for a single medium: audio or video.

MOS is not necessarily adequate as a user-level QoS parameter of multimedia information for the following reasons. First, three or more categories in the rating-scale method make subjects’ burden heavy so that the reliability of the obtained data decreases. Second, MOS is an ordinal scale. In general, the four classes of scales can be considered. From lower to higher classes, we have the nominal scale, the ordinal scale, the interval scale and the ratio scale [6]. The mathematical operations that are applicable to the ordinal scale and the nominal scale are limited.

Many studies in the psychological field have reported methods of human subjectivity quantification. It is possible to apply the psychometric methods [6] to assessment of the user-level QoS. In the literature, however, we can find no study that utilizes the psychometric methods to assess user-level QoS. In this paper, we propose a novel scheme to assess user-level QoS by the psychometric methods. Moreover, we present a method of QoS mapping which translates application-level QoS parameters into a user-level QoS parameter with multiple regression analysis [7].

The rest of this paper is organized as follows. Section 2 presents how to assess the user-level QoS with psychometric methods. In Section 3, we describe application-level QoS parameters. Sections 4 and 5 give user-level QoS assessment and QoS mapping, respectively. Finally, Section 6 concludes the paper.

References
2. APPLICATION OF PSYCHOMETRIC METHODS AND MULTIPLE REGRESSION ANALYSIS

2.1 User-level QoS assessment

We propose to utilize the method of paired comparisons [6] as a scheme of user-level QoS assessment for the following reasons. In the method of paired comparisons, which is one of the psychometric methods, all stimuli are presented to an experimental subject in all possible pairs. A stimulus means an object of a research such as an audio-video stream. Then, the subject judges which one of the pair is superior to the other from some perspective. By repeating this judgement for many subjects, we obtain the proportion of times each stimulus is decided better than every other stimulus.

Since the subject only has to judge which one of the pair is superior to the other, the use of the method of paired comparisons imposes less burden on the subject for a judgment than that of other methods like the rating-scale method.

By using the method of paired comparisons, we can obtain the proportion of times each stimulus is judged superior to the other. The proportion is an ordinal scale. We must translate the ordinal scale into an interval scale or a ratio scale. Thurstone’s law of comparative judgment [6], [8] is the most popular scheme of translation from an ordinal scale obtained by the method of paired comparisons into an interval scale. It should be noted that the ordinal scale has only the ordinal relation between the numbers obtained by the measurement (i.e., a greater-than-less-than relation); however, in the interval scale, the size of the differences between the numbers as well as their ordinal relation has meaning.

In the law of comparative judgment, when stimulus \( i \) is presented to a subject, a psychological value designated by \( s_i \) is assumed to occur on a scale called the psychological continuum. The variable \( s_i \) is assumed to be a normally distributed variable with mean \( S_i \) and standard deviation \( \sigma_i \). When two stimuli \( i \) and \( j \) are presented, we make a judgement on the superiority of the stimuli according to \( s_j - s_i \). We can estimate a distribution of \( (s_j - s_i) \) from proportions obtained with the method of the paired comparisons. Consequently, the mean \( S_i \) of \( s_i \) can be calculated from the estimated distribution of \( (s_j - s_i) \) by the method in [9]. We can regard \( S_i \) as an interval scale.

We have to confirm the goodness of fit for obtained results since Thurstone’s law of comparative judgment is just an assumption. In [10], a test of goodness of fit is proposed for Thurstone’s law of comparative judgment. In this test, a proportion of the superiority for stimulus \( i \) is estimated from a calculated value of \( S_i \), which is denoted by \( S'_i \). Then, the estimated proportions are compared with the actual proportions measured by the method of paired comparisons.

If we can confirm the goodness of fit as a result of the test proposed in [10], we refer to this interval scale as the psychological scale. We use the psychological scale as the user-level QoS parameter for audio-video transmission.

2.2 QoS mapping

Now, we discuss a method of mapping from QoS at lower levels to the user-level QoS. To estimate a variable from other variables, we often utilize multiple regression analysis. As a preliminary study, we focus on QoS mapping schemes from application-level QoS to the user-level QoS by multiple regression analysis. That is, we consider that application-level QoS parameters are the predictor variables and the user-level QoS parameter is the criterion variable.

To do this, we must select appropriate application-level QoS parameters as predictor variables. In this paper, we adopt nine measures from a media synchronization quality point of view [11]. From among the nine measures, we choose necessary predictor variables by the principal component analysis.

3. APPLICATION-LEVEL QoS PARAMETERS

3.1 Media synchronization

With regard to application-level QoS for continuous media transmission, it is important to keep the temporal structure; this is referred to as media synchronization [12]. Therefore, in this paper, we consider media synchronization quality as application-level QoS as in [1].

There are two types of media synchronization: intra-stream synchronization and inter-stream synchronization [12]. The former keeps continuity of a single stream. The latter means synchronization among multiple media streams.

The temporal structure can be disturbed by various causes: packet loss and delay jitters during the transmission through communication networks, for example. In order to transmit multimedia without degradation of its media synchronization quality, we need media synchronization control.

Many schemes for media synchronization control have been proposed. Among them, the VTR (Virtual Time Rendering) algorithm [11], [13], [14] is an effective one that is applicable to various network environments. In this paper, we employ the VTR algorithm for media synchronization.

3.2 Measures for media synchronization quality

In this section, we define application-level QoS parameters for QoS mapping to a user-level QoS parameter. We adopt measures for media synchronization quality used in [11] and [14] as application-level QoS parameters.

First, we consider the following parameters: the coefficient of variation of output interval for audio (\( C_a \)) and for video (\( C_v \)), the average MU rate for audio (\( R_a \)) and for video (\( R_v \)), the mean square error of intra-stream synchronization for audio (\( E_a \)) and for video (\( E_v \)) and the MU loss rate for audio (\( L_{a} \)) and for video (\( L_{v} \)). The parameters introduced so far represent the intra-stream synchronization quality. Second, we calculate the mean square error of inter-stream synchronization (\( E_{int} \)) for inter-stream synchronization quality.

The nine application-level QoS parameters introduced above are candidates for parameters used in QoS mapping discussed later. Since these nine parameters are not independent of each other, not all of the nine is necessary. Conversely, the nine parameters may be insufficient. In these cases, we must try new parameters.
4. ASSESSMENT OF THE USER-LEVEL QoS

4.1 Subjective assessment

In the assessment of the user-level QoS, we also investigate how the application-level QoS affects the user-level QoS. Furthermore, we quantitatively measure subjective effectiveness of the VTR media synchronization algorithm. For this purpose, we compare the user-level QoS with the VTR media synchronization control and that without the control.

We compose audio-video streams as stimuli in our psychological experiment. We prepare an audio-video stream, which is called an original stream. Two types of original streams are used to investigate the effect of the contents of media on QoS mapping. One is a music clip (source A), and the other is a scene in a drama (source B). The media of sources A and B have been encoded so that the specification of them are the same. Table I shows the specification of sources A and B. In Table I, MU means Media Unit, which is the information unit for media synchronization; an audio MU is a packet consisting of 1000 voice samples, and a video MU is a video picture.

Transmitting an original stream over a loaded network, we measure each application-level QoS parameter. We set up the network configuration shown in Fig. 1 and simulate the transmission with NS [15]. The media server transmits an original stream to the media client. The line speed between the two routers is 4.0 Mbps. The load generator generates UDP messages of 1472 bytes each at exponentially distributed inter-arrival times. To realize different QoS at the application-level, we use four kinds of average load: 3.00, 3.15, 3.30, and 3.45 Mbps. While the media server sends an audio-video stream (the original stream), the load generator transmits the interference traffic. The media client receives the original stream with/without the VTR media synchronization control and outputs it. We record the audio-video stream that the media client outputs. In our subjective assessment, we use the stream, which is called the experimental stream.

While we measured application-level QoS parameters for each experimental stream, we carried out subjective assessment. Twenty subjects assessed the experimental streams subjectively by the method of paired comparisons as follows. All pairs of the experimental streams were presented to a subject. The number of pairs is \( C_2 \times C_2 = 28 \). The subject judged whether an offered stream was superior to the other. The first experimental stream of a pair was presented for 15 seconds; then the second one was presented for 15 seconds after five second interval. We repeated this comparison 28 times at random for each subject and measured the proportion of times each experimental stream was judged superior to the other. We call this proportion the experimental proportion. The experimental proportion of the times the \( j \)-th experimental stream was judged superior to the \( i \)-th one is denoted by \( p_{ij}' \).

The subjects are non-experts in the sense that they are not directly concerned with voice and video quality as part of their normal work. They are men and women, and their ages are between 22 and 34.

4.2 Psychological scaling

By Thurstone’s law of comparative judgment, we translate the experimental proportions into psychological scales, which are interval scales. Thurstone’s law comprises five cases (cases I through V). Among these cases, case V is the simplest. When we utilize Thurstone’s law, we first select case V.

Now we present the scaling procedure in case V. From the experimental proportion \( p_{ij}' \), we calculate the standard normal deviate \( Z_{ij}' \) [6]. If the experimental proportion \( p_{ij}' \) takes a value of 0 or 1, \( Z_{ij}' \) becomes negative or positive infinity. In this case, we treat \( Z_{ij}' \) as a missing entry.

When no missing entry exists, a least squares estimate \( Y_{ij}' \) of the psychological scale value can be calculated with \( Z_{ij}' \) by the method in [16]. Gulliksen [9] proposed a least squares solution for the method of paired comparisons with incomplete data. In this paper, we utilize this method. Table II shows obtained psychological scales for sources A and B. In Table II, NC and VTR mean “No control” and “VTR media synchronization control”, respectively.

4.3 Test of goodness of fit for the psychological scale

In this section, we test goodness of fit between the experimental proportion obtained in Subsection 4.1 and that calculated from the psychological scale in the previous subsection by the method in [10]. Let us set the significance level to 0.01. As a result, we find that the null hypothesis that the observed

---

### TABLE I

<table>
<thead>
<tr>
<th>audio</th>
<th>video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coding scheme</td>
<td>ITU-T G.711 or LPC</td>
</tr>
<tr>
<td>Picture pattern</td>
<td>-</td>
</tr>
<tr>
<td>Average bit rate [Mbps]</td>
<td>320 or 240</td>
</tr>
<tr>
<td>Average MU size [bytes]</td>
<td>1000 or 2400</td>
</tr>
<tr>
<td>Total MU number</td>
<td>100 or 300</td>
</tr>
<tr>
<td>Average MU size [bytes]</td>
<td>84 or 300</td>
</tr>
</tbody>
</table>

---

![Network configuration in the simulation.](image.png)

---

### TABLE II

<table>
<thead>
<tr>
<th>control method</th>
<th>average load [Mbps]</th>
<th>psychological scale source A</th>
<th>psychological scale source B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 NC</td>
<td>3.00</td>
<td>1.222</td>
<td>1.485</td>
</tr>
<tr>
<td>2 NC</td>
<td>3.15</td>
<td>1.100</td>
<td>1.210</td>
</tr>
<tr>
<td>3 NC</td>
<td>3.30</td>
<td>0.856</td>
<td>0.921</td>
</tr>
<tr>
<td>4 NC</td>
<td>3.45</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>5 VTR</td>
<td>3.00</td>
<td>2.672</td>
<td>2.243</td>
</tr>
<tr>
<td>6 VTR</td>
<td>3.15</td>
<td>2.381</td>
<td>2.001</td>
</tr>
<tr>
<td>7 VTR</td>
<td>3.30</td>
<td>1.968</td>
<td>1.897</td>
</tr>
<tr>
<td>8 VTR</td>
<td>3.45</td>
<td>0.783</td>
<td>0.667</td>
</tr>
</tbody>
</table>
and the estimated values are the same cannot be rejected for each source. Therefore, we can consider that the proportion \( p''_{ij} \) estimated from \( S''_i \) and \( S''_j \) equals observed one \( p_{ij} \). Thus, we see that it is proper to regard the obtained scale \( S''_i \) as a psychological scale.

5. QoS mapping

In this section, we discuss a translation of the application-level QoS parameters into the user-level QoS parameter by multiple regression analysis. Unless predictor variables are actually independent of each other in multiple regression analysis, multiple colinearity occurs. However, some or all of the nine application-level QoS parameters shown in Subsection 3.2 may be dependent on each other. Therefore, we first examine the correlation coefficient between the application-level QoS parameters and classify them. Then, we select some application-level QoS parameters as predictor variables by the principal component analysis and apply multiple regression analysis.

In this paper, we use the two sources to investigate the effect of the contents of media on the user-level QoS. However, as shown in Table I, the audio and video specifications of the two sources are the same. Therefore, we will study the application-level QoS of sources A and B together in the following subsections.

5.1 Classification of application-level QoS parameters

We measured each application-level QoS parameter while we generated experimental streams of sources A and B. We calculated the correlation coefficient between every pair of the application-level QoS parameters.

From the correlation coefficient, we classified the application-level QoS parameters into three categories a) \( R_a, R_v, L_a, L_v, \) and \( C_v \), b) \( E_c, E_a, \) and \( E_{int} \) and c) \( C_a \) according to the correlation coefficient between them.

5.2 QoS mapping from application-level QoS

Using the correlation coefficient, we carried out the principal component analysis. As a result, we see that the cumulative contribution rate for the first two principal components is 98.8\%.

Therefore, we first select two application-level QoS parameters out of the nine as predictor variables. Accordingly to the classification of the nine application-level QoS parameters shown in the previous subsection, we select two QoS parameters as predictor variables of multiple regression analysis. Since \( C_a \), slightly correlates with the parameters in categories a) and b), it is not appropriate as a predictor variable owing to multi-colinearity. Therefore, we select a parameter from each of categories a) and b) as a predictor variable.

Let us choose \( E_a \) and \( C_v \) as predictor variables according to the contribution rates to each principal component. Then, we obtain

\[
\hat{S}_A = 2.415 - 0.00365E_a - 0.729C_v \quad (1)
\]

\[
\hat{S}_B = 2.204 - 0.00223E_a - 0.633C_v \quad (2)
\]

where \( \hat{S}_A \) and \( \hat{S}_B \) denote estimated values of the psychological scales for sources A and B, respectively.

The contribution rates adjusted for degrees of freedom for sources A and B are 0.941 and 0.974, respectively. Therefore, we can estimate a psychological scale with accuracy, using these two application-level QoS parameters. It should be noted that Eqs. (1) and (2) clearly show that the regression line depends on the contents.

5.3 Control gain of media synchronization

In this subsection, we perform multiple regression analysis with other parameters. We consider the average amount of interference traffic \( L \) [Mbps]. Moreover, using a dummy variable \( C \), we take the existence of the media synchronization control into account as a predictor variable as follows:

\[
C = \begin{cases} 
0 & \text{(No Control)} \\
1 & \text{(VTR)} 
\end{cases} \quad (3)
\]

\[
\hat{S}_A = 11.534 - 3.330L + 1.156C \quad (4)
\]

\[
\hat{S}_B = 11.198 - 3.192L + 0.798C \quad (5)
\]

The contribution rates adjusted for degrees of freedom for sources A and B become 0.880 and 0.844, respectively. These values show that the goodness of fit is a little worse than that of the previous results.

We make the following observations from Eqs. (4) and (5). First, the psychological scale decreases as the average amount of interference traffic increases. Next, for source A, for instance, the media synchronization control can increase the value of the scale by 1.156 from no control. In an interval scale, we can select an arbitrary value as the unit of the scale.

Therefore, we convert the value of 1.156 into an amount of interference traffic, which becomes \( 1.156/3.330 = 0.347 \) Mbps. That is, for source A, the VTR media synchronization control subjectively lightens the average amount of interference traffic by 0.347 Mbps. We refer to this effect as control gain by media synchronization. Similarly, the control gain by media synchronization for source B is 0.250 Mbps. Consequently, the VTR media synchronization algorithm is more effective for source A than source B.

6. Conclusions

In this paper, we proposed a user-level QoS assessment scheme for audio-video transmission with the psychometric methods. As a preliminary study, we utilized the method of paired comparisons and the law of comparative judgment. By experiment, we constructed the psychological scale that represents the user-level QoS for audio-video transmission. The psychological scale is an interval scale and fit the experimental result. In addition, we discussed a method of mapping from QoS at lower levels to user-level QoS. As a preliminary study of the QoS mapping, we proposed a mapping method to translate the application-level QoS parameters into the user-level QoS parameter by multiple regression analysis. As a consequence, we estimated the psychological scale with two
application-level QoS parameters selected by the principal component analysis.

The results shown in this paper are applicable various fields. However, when we study the application of our results, we must consider that the psychological scale is an interval scale; the absolute values of a psychological scale are meaningless, and only a difference between two values is meaningful. Therefore, we must study a method to determine required values of the psychological scale. An example of such methods is as follows. We first prepare a stimulus as the reference; we then compare the psychological scale value for this reference with the one for an object. We will study such applications in future work.

ACKNOWLEDGMENT

This work was supported by the Grant-In-Aid for Scientific Research of Japan Society for the Promotion of Science under Grant 14350200.

REFERENCES