User–Level QoS Assessment of a Multipoint–to–Multipoint TV Conferencing Application over IP Networks

Yoshihiro Ito and Shuji Tasaka
Department of Computer Science and Engineering, Graduate School of Engineering
Nagoya Institute of Technology, Nagoya 466–8555, Japan
{yoshi, tasaka}@nitech.ac.jp

Abstract—This paper studies a multipoint–to–multipoint TV conferencing application over IP networks and assesses its user–level QoS with two types of QoS mapping. In utilizing the application, the user perceives quality of the communication with every other conferee; we refer to the quality as individual user–level QoS to a conferee. According to the individual user–level QoS, he/she totally judges the quality of the application, which is referred to as overall user–level QoS for the user. The overall user–level QoS of the application can be affected by the individual one to each conferee; therefore, it is difficult to clarify QoS parameters which affect the overall user–level QoS. This paper tackles the problem by utilizing two types of QoS mapping: mapping between the two kinds of user–level QoS and that between user–level QoS and application–level QoS. In this paper, an experiment with a simple task by three conferees was carried out. The user–level QoS is assessed by one of the psychometric methods. As a result of the two types of QoS mapping, we find two interesting results. First, when a user communicates with the other two conferees, the lower individual user–level QoS has more effect on the overall user–level QoS than the higher one. Second, the individual user–level QoS can depend on not only its application–level QoS but also that of the other conferees.

I. INTRODUCTION

As the technology of the Internet advances, many attractive applications have been developed; a TV conferencing application is one of such applications. With it, we can simultaneously send audio and video to many conferees who stay far away from each other via the Internet.

A best effort service is the only service which the current Internet can provide. Consequently, quality of service (QoS) of audio–video transmission cannot be guaranteed over the Internet. Therefore, QoS control is required. In order to control the QoS in audio–video transmission over the Internet appropriately, we need to evaluate the QoS.

Since the function of the Internet has a layered structure, its QoS also has a layered one. For example, reference [1] recognizes six levels of QoS: physical–level, node–level, network–level, end–to–end–level, application–level and user–level.

In an audio–video transmission service, the final addressee is the user. Therefore, the quality perceived by the users ultimately tells degree of goodness in the service. The quality is referred to as user–level QoS or perceptual QoS. Consequently, it is necessary to assess user–level QoS of the TV conferencing applications over the Internet.

The TV conferencing applications have three significant aspects: audio–video transmission, an interactive application and multipoint–to–multipoint communication. These aspects tell us factors which can damage user–level QoS.

First, in audio–video transmission over IP networks, packet loss and its delay are inevitable; they disturb the spatial structure of audio and that of video. The disturbance causes severe degradation of the user–level QoS in the audio–video transmission.

Second, an interactive application requires high interactivity. The interactivity is lowered by long end–to–end delay. Consequently, it degrades user–level QoS.

In general, buffering control is effective in the improvement of the QoS degradation caused by the disturbance of the temporal structure in audio–video transmission. In order to improve worse degradation, we require longer buffering time. However, longer buffering time leads to longer delay. Therefore, the buffering control does not always improve the user–level QoS of interactive applications. In [2], the authors confirm the subjective trade–off caused by buffering control in interactive audio–video transmission.

Finally, in multipoint–to–multipoint communication, many conferees simultaneously communicate with each other. This makes user–level QoS assessment of a TV conferencing application very knotty. The reason for this is as follows. Each conferee perceives quality of the communication with every other conferee; according to the quality, he/she totally judges the perceived quality of the application. The quality which is totally judged is the most important for us. In this paper, we refer to the QoS which a user perceives for the communication with another conferee as individual user–level QoS to the conferee for the user; the perceived QoS which the conferee totally judges is referred to as overall user–level QoS for the user. Consequently, for a user, the individual user–level QoS to each conferee affects the overall one in multipoint–to–multipoint communication. Moreover, the individual user–level QoS can be affected by some kinds of application–level QoS such as media quality and interactivity. Therefore, in order to investigate user–level QoS of a TV conferencing application, we must consider numbers of application–level QoS parameters.

In this paper, we tackle the difficulty in user–level QoS
assessment of a TV conferencing application over IP networks
by utilizing two types of QoS mapping: mapping between
the two kinds of user–level QoS and that between user–level
QoS and application–level QoS. The former clarifies the
relationship between the individual user–level QoS to each
conferee and the overall user–level QoS; the latter extracts
application–level QoS parameters which affect the individual
user–level QoS to each conferee.

In order to assess the two kinds of user–level QoS, we utilize
the method of successive categories [3], which is one of the
psychometric methods [4]. On the other hand, we adopt the
multiple regression analysis as the QoS mapping method.

Some papers report user–level QoS assessment of TV conferencing
applications over packet networks; for example, see [5], [6], [7] and [8]. These papers treat user–level QoS of
audio–video transmission between a pair of conferees.

In regard to assessment of user–level QoS of multipoint–
to–multipoint TV conferencing applications, Kurita assesses
overall subjective quality affected by conversation roles [9].
In [9], however, he treats fixed delay and random packet
loss which are added independently; this does not necessarily
reflect actual situations in IP networks. Except for [9], in the
IP network research area, we can find no publication which
treat user–level QoS assessment of multipoint–to–multipoint
IP TV conferencing applications.

This paper is organized as follows. Section II describes
a method of user–level QoS assessment. In Section III,
we present QoS mapping with the multiple regression analysis.
Sections IV and V show our experiment and its results,
respectively. Finally, this paper is concluded in Section VI.

II. USER–LEVEL QoS ASSESSMENT

We consider user–level QoS of a TV conferencing application
for a user from two points of view. One expresses the
individual user–level QoS to each conferee, and the other is
the overall one.

In this paper, we assess the two kinds of user–level QoS
with the method of successive categories; it is one of the
psychometric methods. In the psychological field, the psychometric
methods were proposed to assess human subjectivity. In [10] and [11],
we employ the psychometric methods to assess user–level QoS of audio–video transmission. Reference [10]
utilizes the method of successive categories. In this paper, we
also adopt the method of successive categories as a scheme of
user–level QoS assessment.

The method of successive categories has three phases. First,
subjective scores are measured by the rating scale method [4],
which is one of the psychometric methods. In the rating scale
method, subjects classify objects for evaluation into five or
seven categories. Second, the subjective scores are translated
into an interval scale [4] with the law of categorical judgment
[3]; the law is also one of the psychometric methods. In an
interval scale, the difference between two values is significant;
we can apply almost all the statistical procedures to the interval
scale. Finally, we test the goodness of fit of the obtained
interval scale by Mosteller’s test [12]. The reader is referred
to [10] for further details of the user–level QoS assessment.

In this paper, we term the obtained interval scale the
psychological scale.

III. QoS MAPPING

By utilizing QoS mapping, we can clarify the relationship
between two levels of QoS quantitatively. To carry out QoS
mapping from a lower level to the user–level, the authors
proposed the utilization of the multiple regression analysis
[10] [11]. We consider the user–level QoS parameter as a
criterion variable; QoS parameters at the lower level are
treated as predictor variables. With n lower level QoS parameters A1, A2, ..., An, an estimate $\hat{S}$ of the user–level QoS parameter can be calculated by

$$\hat{S} = \beta_0 + \beta_1 A_1 + \cdots + \beta_n A_n$$

where $\beta_i$ (1 ≤ i ≤ n) is the partial regression coefficient of
the i-th lower level QoS parameter, and $\beta_0$ is the intercept.

Before the multiple regression analysis, we must select some QoS parameters as predictor variables. In [10] and [11],
the authors utilize the principal component analysis to choose
predictor variables.

In this paper, we employ the multiple regression analysis as
the method of QoS mapping.

IV. EXPERIMENT

A. Experimental setup

Figure 1 shows our experimental configuration with three
users. The experimental network consists of three terminals
(terminal 1, terminal 2 and terminal 3), two load generators
(load generator 1 and load generator 2), two load sinks (load
sink 1 and load sink 2) and a multicast router. Terminal 2,
load generator 1 and load sink 1 connect with the multicast
router via a shared Ethernet hub. Similarly, terminal 3, load
generator 2 and load sink 2 connect with the router via another
hub. The transmission rate of each link is set to 10 Mb/s. For
convenience, terminal 1, terminal 2 and terminal 3 are denoted
by T1, T2 and T3, respectively.

Three subjects sit in front of the corresponding terminals
and perform a common task. The task is a simple one. One
subject calls the name of either of the rests. Then, the called
subject becomes a new caller. The subjects repeat the task for
twenty seconds.

Each terminal transmits a pair of audio and video streams
of the corresponding subject to the others by multicast and...
outputs received audio–video streams with buffering control; it also measures application–level QoS parameters described in Subsection IV–C. In order to measure them, the terminals synchronize their clocks by NTP [13]. The audio and video are transmitted as separate transport streams by using RTP/UDP/IP. Table I shows the specifications of audio and video. In Table I, MU stands for “media unit” and indicates the information unit at the application–level for media synchronization [14].

While each terminal sends a pair of audio and video streams, load generator 1 and load generator 2 transmit load traffic to load sink 1 and load sink 2, respectively. Each load generator generates fixed–size UDP datagrams of 1472 bytes each in its payload at exponentially distributed intervals and sends them to the corresponding load sink.

We change the average amount of load traffic generated by the load generators. For convenience, we denote a combination of the load traffic by \( L_1, L_2 \), where \( L_1 \) and \( L_2 \) indicate the average amount of load traffic generated by load generator 1 and that generated by load generator 2, respectively. To simplify the discussion, we use three combinations of \( L_1 \) and \( L_2 \), which are presented in Table II.

### Table I
**Specifications of audio–video streams.**

<table>
<thead>
<tr>
<th>Coding scheme</th>
<th>Audio</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image size (pixels)</td>
<td>-</td>
<td>552\times288</td>
</tr>
<tr>
<td>Picture format</td>
<td>-</td>
<td>IPPP</td>
</tr>
<tr>
<td>Average MU size (byte)</td>
<td>2.28</td>
<td>2688</td>
</tr>
<tr>
<td>Average bit rate (kb/s)</td>
<td>2.29</td>
<td>645</td>
</tr>
<tr>
<td>Playout time (s)</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

In this experiment, we adopt the simple playout buffering control scheme used in [2]. For simplicity of discussion, we treat nine combinations of the buffering time. Table III shows the buffering time set at each terminal.

### Table II
**Combinations of average load.**

<table>
<thead>
<tr>
<th>Average load [Mb/s]</th>
<th>( L_1 )</th>
<th>( L_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio</td>
<td>6.50</td>
<td>6.50</td>
</tr>
<tr>
<td>Video</td>
<td>8.00</td>
<td>8.00</td>
</tr>
</tbody>
</table>

### Table III
**Combinations of buffering time.**

<table>
<thead>
<tr>
<th>Buffering time [ms]</th>
<th>( T_1 )</th>
<th>( T_2 )</th>
<th>( T_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffering time [ms]</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

B. user–level QoS assessment

By the method of successive categories, we can obtain the two kinds of psychological scales for a user; one indicates the individual user–level QoS to each conferee, and the other expresses the overall user–level QoS. In this paper, we refer to the former as the individual psychological scale to the conferee for the user and the latter as the overall psychological scale for the user.

In the method of successive categories, we utilize five categories of impairment: “imperceptible”, “perceptible, but not annoying”, “slightly annoying”, “annoying”, and “very annoying”.

In the user–level QoS assessment, the subjects are male, and their ages were 20s. The number of subjects is 18.

### Table IV
**Combinations of user–level QoS assessment.**

<table>
<thead>
<tr>
<th>Buffering time [ms]</th>
<th>( T_1 )</th>
<th>( T_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffering time [ms]</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

C. application–level QoS measurement

This paper measures nine application–level QoS parameters which represent media synchronization quality; these parameters are used in [2]. First, we treat the coefficient of variation of output interval; it is defined as the ratio of the standard deviation of the MU output interval of a stream to its average. We denote them by \( E_a \) for audio and by \( E_v \) for video. Second, we employ the average MU delay, which is defined as the average time in seconds from the moment an MU is generated until the instant the MU is output; we denote the average MU delay for audio and that for video by \( D_a \) and \( D_v \), respectively. Third, we utilize the MU loss ratio for audio \( L_a \) and that for video \( L_v \); this is the ratio of the number of lost MUs to the total number of generated MUs. Fourth, we use the mean square error of intra–stream synchronization; it is defined as the average of the squared difference between the output interval of MU at the destination and the generation one at the source; it is denoted by \( E_{\text{int}} \) for audio and by \( E_{\text{int}} \) for video. Finally, we adopt the mean square error of inter–stream synchronization \( E_{\text{int}} \), 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.

V. EXPERIMENTAL RESULTS

A. User–level QoS

Tables IV, V and VI show the calculated psychological scales assessed at T1, T2 and T3, respectively. In these tables, the first column from the right side indicates the overall psychological scale values; the second column and third one show the individual psychological scales.
By way of example, we plot the psychological scale for the load traffic of (8.0, 8.0) at T1 in Fig. 2, that at T2 in Fig. 3 and that at T3 in Fig. 4.

From Figs. 2, 3 and 4, we find that the overall psychological scale is not always close to the average of the two individual psychological scales; it is closer to the lower value than the higher one. This means that the lower individual user-level QoS has more effect on the overall user-level QoS than the higher one. In the following subsection, we examine this observation quantitatively by QoS mapping.

B. Mapping between the two kinds of user-level QoS

With the multiple regression analysis, we perform the mapping between the two kinds of user-level QoS, that is, the mapping from the individual psychological scales to the overall one. In this analysis, we consider the overall psychological scale and the individual ones as the criterion variable and predictor variables, respectively.

In QoS mapping, we must discriminate between the two individual psychological scales in order to use them as predictor variables. In the previous subsection, we found that the lower individual psychological scale has more effect on the overall one than the higher individual one. Then, we define two variables: $U_{MAX}$ and $U_{MIN}$. The variable $U_{MAX}$ corresponds to the higher individual psychological scale, while $U_{MIN}$ the lower one. We treat $U_{MAX}$ and $U_{MIN}$ as predictor variables.
As a result of the multiple regression analysis, we obtained the following multiple regression line:

\[ U_O = -0.115 + 0.297 U_{MAX} + 0.742 U_{MIN} \]  

(2)

where \( U_O \) indicates the estimate of the overall psychological scale value. The multiple correlation coefficient of the line becomes 0.985; this value tells us that the overall psychological scale is precisely determined by the individual ones.

In Eq. (2), the coefficient of \( U_{MIN} \) is larger than that of \( U_{MAX} \). That is, the lower individual psychological scale has more effect on the overall psychological scale than the higher one; this result has already been shown in the previous subsection.

C. QoS mapping between application–level and user–level

Before QoS mapping between user–level and application–level, we must select some application–level QoS parameters among those described in Subsection IV-C as predictor variables so that predictor variables hardly correlate with each other.

Then, we carried out the principal component analysis in the same manner as that in [10]. As a result, we see that the cumulative contribution rate of the first two principal components and that of the first three ones become 86.8% and 97.0%, respectively. Therefore, we employ the first three principal components. Table VII represents the obtained principal component loadings.

<table>
<thead>
<tr>
<th>principal component loading</th>
<th>first</th>
<th>second</th>
<th>third</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L_a )</td>
<td>0.888</td>
<td>0.389</td>
<td>-0.173</td>
</tr>
<tr>
<td>( E_a )</td>
<td>0.397</td>
<td>0.197</td>
<td>0.138</td>
</tr>
<tr>
<td>( C_a )</td>
<td>0.275</td>
<td>0.049</td>
<td>-0.046</td>
</tr>
<tr>
<td>( D_a )</td>
<td>-0.341</td>
<td>0.179</td>
<td>0.310</td>
</tr>
<tr>
<td>( L_v )</td>
<td>0.139</td>
<td>0.583</td>
<td>-0.708</td>
</tr>
<tr>
<td>( E_v )</td>
<td>0.263</td>
<td>0.377</td>
<td>0.227</td>
</tr>
<tr>
<td>( C_v )</td>
<td>0.941</td>
<td>0.245</td>
<td>-0.002</td>
</tr>
<tr>
<td>( D_v )</td>
<td>-0.480</td>
<td>0.806</td>
<td>0.344</td>
</tr>
<tr>
<td>( E_{int} )</td>
<td>0.902</td>
<td>0.107</td>
<td>0.307</td>
</tr>
</tbody>
</table>

According to Table VII, we categorized the nine application–level QoS parameters into three groups:

- **group a)** \( L_a, E_a, C_a, E_v, C_v \) and \( E_{int} \)
- **group b)** \( D_a \) and \( D_v \)
- **group c)** \( L_v \)

The parameters in group a) and that in group c) are correlated with the first principal component and the third one, respectively; they relate to media quality. The parameters which belong to group b) highly correlate with the second principal component; they are concerned with interactivity.

We pick up one parameter from each group. Consequently, the number of combinations of the predictor variables becomes \( 6 \times 2 \times 1 = 12 \). For the 12 combinations of predictor variable, we carried out the multiple regression analysis and obtained the multiple regression lines. Table VIII shows the multiple correlation coefficient adjusted for degrees of freedom of the regression line for each combination.

<table>
<thead>
<tr>
<th>( D_a, L_v )</th>
<th>( E_a )</th>
<th>( C_a )</th>
<th>( E_v )</th>
<th>( C_v )</th>
<th>( E_{int} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D_a, L_v )</td>
<td>0.95</td>
<td>0.92</td>
<td>0.91</td>
<td>0.92</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Table VIII indicates that we can obtain the highest multiple correlation coefficient adjusted for degrees of freedom by utilizing two combinations: \( (L_a, D_a, L_v) \) and \( (L_a, D_v, L_v) \). The result of the statistical test tells us that \( L_v \) makes no significant contribution to the multiple regression lines with a significance level of 0.01. Consequently, either \( (L_a, D_a) \) or \( (L_a, D_v) \) is the most appropriate for the combination of predictor variables. In this paper, we select a parameter regarding audio and one concerning video. Therefore, we choose \( L_a \) and \( D_v \) as predictor variables.

Then, we carried out the multiple regression analysis again. Equation (3) shows the obtained multiple regression line.

\[ U_I = 2.684 - 7.936 \times 10^{-2} L_a - 1.088 \times 10^{-3} D_v \]  

(3)

where \( U_I \) indicates the estimate of the individual psychological scale value. The multiple correlation coefficient adjusted for degrees of freedom is 0.93. From Eq. (3), we see that the individual perceived quality mainly depends on the two application–level QoS parameters; one \( (L_a) \) is concerned with the media quality, and the other \( (D_v) \) relates to interactivity.

Although we have found the application–level QoS parameters which affect the individual psychological scale, there may be room for consideration of other parameters since the multiple correlation coefficient of Eq. (3) is not extremely high. However, the previous result of the principal component analysis shown in this subsection has revealed that it is unnecessary to consider any application–level QoS parameters except for \( L_a \) and \( D_v \) as predictor variables. This suggests that the application–level QoS parameters measured at the two terminals other than the subject’s one may affect the individual psychological scale.

Since the interactivity concerns response time for a call, the delay at the other terminals can affect the interactivity. In this paper, we then consider \( D_v \) which is measured at the other terminals. For convenience, we denote the subject’s terminal by \( S \); \( A \) and \( R \) represent the terminal for the assessment of the individual psychological scale and the rest, respectively. We also denote the \( D_v \) measured at \( X(= S, A, R) \) for communication with \( Y(= S, A, R) \) as \( D_{vxY} \). Consequently, there are six kinds of \( D_{vxY} \). Note that \( D_{vSA} \) means \( D_v \); we then treat five kinds of \( D_{vXY} \) except for \( D_{vSA} \).

In addition to \( L_a \) and \( D_v \), we use the five parameters \( (D_{vAS}, D_{vSR}, D_{vRS}, D_{vRA} \) and \( D_{vAR} \)) as candidates of the predictor variables. Before the multiple regression analysis, we carried out the principal component analysis for the seven parameters \( (L_a, D_v, D_{vAS}, D_{vSR}, D_{vRS}, D_{vRA} \) and \( D_{vAR} \)). As a result, we saw that the cumulative contribution rate for the first three components and that for the first four components become 84.6% and 94.7%, respectively. Then, we adopt the first
four principal components. Table IX represents the principal component loadings for each component.

**TABLE IX PRINCIPAL COMPONENT LOADING.**

<table>
<thead>
<tr>
<th>principal component loading</th>
<th>first</th>
<th>second</th>
<th>third</th>
<th>fourth</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_a$</td>
<td>-0.17</td>
<td>0.71</td>
<td>0.09</td>
<td>0.68</td>
</tr>
<tr>
<td>$D_v$</td>
<td>0.76</td>
<td>-0.25</td>
<td>-0.48</td>
<td>0.26</td>
</tr>
<tr>
<td>$D_{vAS}$</td>
<td>0.73</td>
<td>0.63</td>
<td>-0.09</td>
<td>-0.24</td>
</tr>
<tr>
<td>$D_{vSR}$</td>
<td>0.76</td>
<td>-0.34</td>
<td>-0.42</td>
<td>0.17</td>
</tr>
<tr>
<td>$D_{vRS}$</td>
<td>0.76</td>
<td>-0.24</td>
<td>0.49</td>
<td>0.11</td>
</tr>
<tr>
<td>$D_{vRA}$</td>
<td>0.75</td>
<td>0.63</td>
<td>-0.04</td>
<td>-0.26</td>
</tr>
<tr>
<td>$D_{vAR}$</td>
<td>0.75</td>
<td>-0.21</td>
<td>0.55</td>
<td>0.09</td>
</tr>
</tbody>
</table>

From Table IX, we classify the seven parameters into the following four groups:

- **group A** $L_a$
- **group B** $D_v$ and $D_{vSR}$
- **group C** $D_{vAS}$ and $D_{vRA}$
- **group D** $D_{vRS}$ and $D_{vAR}$

The parameter in group A) correlates with the second principal component and the fourth one; the parameters which belong to groups B), C) and D) correlate with the first principal component. Moreover, there is a correlation between the parameters in group C) and the second principal component; the parameters in group B) and those in group D) have negative correlation with the third principal component and positive correlation with it, respectively.

We selected one parameter from each group; the number of combinations of predictor variables becomes $1 \times 2 \times 2 \times 2 = 8$. Here, we give $L_a$ and $D_v$ priority. Consequently, we do not treat $D_{vSR}$, which highly correlates with $D_v$. Therefore, the number of combinations of predictor variables is 4.

As a result of the multiple regression analysis, we obtained four multiple regression lines. Table X shows the multiple correlation coefficient adjusted for degrees of freedom each regression line.

**TABLE X MULTIPLE CORRELATION COEFFICIENTS ADJUSTED FOR DEGREES OF FREEDOM.**

<table>
<thead>
<tr>
<th>$L_a$ $D_v$ $D_{vRS}$ $D_{vAR}$</th>
<th>0.93</th>
<th>0.94</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_a$ $D_v$ $D_{vAS}$ $D_{vRA}$</td>
<td>0.93</td>
<td>0.94</td>
</tr>
</tbody>
</table>

From Table X, we see that combination ($L_a$, $D_v$, $D_{vRA}$, $D_{vAR}$) slightly improves the multiple correlation coefficient adjusted for degrees of freedom over the others. As a result of statistical test, we saw that $D_{vAR}$ does not make any significant contribution with a significance level of 0.01; therefore, we removed $D_{vAR}$. We carried out the multiple regression analysis again. Equation (4) shows the obtained regression line.

$$U_1 = 2.730 - 7.762 \times 10^{-2} L_a - 8.837 \times 10^{-4} D_v - 5.429 \times 10^{-4} D_{vRA}$$

The multiple correlation coefficient adjusted for degrees of freedom is 0.94; thus, it was slightly improved by considering $D_{vRA}$. However, the statistical test tells us that $D_{vRA}$ makes significant contribution to the regression line. That is, $D_{vRA}$ affects the individual psychological scale at $S$. This means that the individual psychological scale can be affected by not only the application–level QoS at the subject’s terminal but also those at the other terminals.

**VI. CONCLUSIONS**

This paper studied user–level QoS assessment of a TV conferencing application over IP networks. We utilized two types of QoS mapping: mapping between the two kinds of user–level QoS and one between user–level QoS and application–level QoS. By experiment, we made the following observations. First, in the TV conferencing application, the lower individual user–level QoS has more effect on the overall user–level QoS than the higher one. Second, the individual user–level QoS depends on not only application–level QoS parameters measured at the user’s terminal but also those measured at the other terminals.

We have some issues to be studied as our future work. First, in our experiment, we used one specific task. However, the user–level QoS of a TV conferencing application can be affected by the kind of task. Therefore, we will treat other tasks. Second, according to the obtained result, we will investigate a method for calculation of appropriate buffering time.

**ACKNOWLEDGMENT**

The authors thank Koji Sonohara for his assistance in the experiment. This work was supported by the Grant–In–Aid for Scientific Research of Japan Society for the Promotion of Science under Grant 17360179.

**REFERENCES**


