End–to–End Monitoring of Multidimensional User–Level QoS in Audio–Video IP Transmission

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 proposes a method of multidimensional user–level QoS monitoring for audio–video transmission over IP networks in the context of ITU-T Recommendation J.148. In order to assess user–level QoS in a multidimensional way, we utilize the SD (Semantic Differential) method, which is one of the psychometric methods. By comparing results of the multidimensional assessment with overall subjective quality, which is represented by a scalar value, we find major factors which contribute to the user–level QoS. Moreover, in order to clarify implications of the major factors, we perform QoS mapping between user–level and application–level. The mapping functions can estimate the user–level QoS and therefore enables its real–time monitoring. From experimental results, we find that three factors contributes to the overall subjective quality; we then see that the three factors imply audio quality affected by video, video quality affected by audio, and the content type.**

I. Introduction

In audio–video transmission over IP networks, many factors can degrade its quality. For example, a conceivable factor among them is disturbance of temporal structures of audio and video; it is caused by packet loss, packet delay and its jitter. Consequently, it is necessary to investigate *QoS* (*Quality of Service*) of audio–video transmission quantitatively in a multidimensional way. This means that we must use more than one parameter to represent the QoS.

In general, QoS of IP networks has a layered structure. The authors, for example, identifies six levels of QoS: *physical–level*, *node–level*, *network–level*, *end–to–end–level*, *application–level* and *user–level* [1]. The user–level QoS is also referred to as *subjective* or *perceptual* QoS, which is the most important since the users are the ultimate recipients of the services.

Users judge the overall quality of the audio–video transmission in an integrated fashion based on many factors: for example, audio quality, video quality and contents. Therefore, in order to investigate user–level QoS of the audio–video transmission, we must seek some parameters which represent user–level QoS appropriately and clarify the relationship between the overall quality and the parameters. Furthermore, such factors can interact with each other [2]; thus, we must consider the effect of the interaction of the parameters.

Some papers reported user–level QoS assessment of audio–video transmission over IP networks; for instance, see [3]. Many of them used *Mean Opinion Score* (*MOS*) as the user–level QoS parameter. However, MOS is not always adequate to the assessment of audio–video transmission [4], [5]. Then, in [1], [3], [5], [6] and [7], the authors quantitatively assess user–level QoS of audio–video transmission over IP networks with the *psychometric methods* [8], which were proposed in the psychological field to assess human subjectivity. In [5], the user–level QoS is assessed with the method of paired comparison and Thurstone's law of comparative judgment. References [1], [3], [6] and [7] assess the user–level QoS with the *method of successive categories* [8]. The authors confirm the *mutually compensatory property of media* in [1]. References [6] and [7] assess user–level QoS of interactive audio–video applications and effect of monitor sizes on user–level QoS of audio–video transmission, respectively. It should be noted that the real–time estimation method of user–level QoS in [3] can be utilized for end–to–end monitoring of user–level QoS.

The researches mentioned above assess user–level QoS with a scalar QoS parameter; they clarify the relationship between the user–level QoS parameter and many application–level QoS parameters by QoS mapping. However, diversification of networks and that of applications increase the kinds of factors which affect user–level QoS. In addition, some factors can interact with each other. In this situation, only one user–level QoS parameter cannot clarify which factors are influential in user–level QoS. Therefore, we need to investigate user–level QoS in detail by assessing it in a multidimensional way.

We can find few studies on multidimensional user–level QoS assessment of audio–video transmission [9], [10], [11]. For example, Bouch *et al.* proposed a three–dimensional approach to user–level QoS assessment [9]. This method includes three aspects: subjective satisfaction, task performance and user–cost. However, they do not assess subjectivity of audio–video transmission in any multidimensional way.

Regarding the user–level QoS assessment of audio-video transmission, ITU–T has presented an objective multimedia perceptual quality model in Recommendation J.148 [2]. The model treats auditory quality, visual quality and differential delay; it also includes interaction between auditory quality and visual one. However, the recommendation indicates no method to assess multimedia QoS.

This paper proposes a method of multidimensional user–level QoS assessment in audio–video transmission over IP networks in the context of ITU–T Rec. J.148; it utilizes the *SD (Semantic Differential) method* [12] The proposed method can be used for end–to–end monitoring of the user–level QoS. In this sense, the current paper is an extension of [3] to a multidimensional case. Such techniques of user–level QoS monitoring form a basis of future IP networks that guarantee user–level QoS, which is a possible successor of the *Next Generation Network* (*NGN*) [13]. A first–step trial of this kind of study can be found in [14].

The SD method is one of the psychometric methods. This has been widely used for multidimensional assessment of a single medium. For example, in [15] and [16], subjective assessment of sound is performed with the SD method. Also, we can find an application of the SD method with MOS to an audiovisual interactive service over packet networks in [11], which assumes that audio packet loss ratio and video one are independent of each other and that packet delay is constant (i.e., no delay jitter). Therefore, the model does not necessarily reflect actual situations of packet networks.

In addition to multidimensional assessment with the SD method, this paper assesses *overall subjective quality* of the audio–video transmission, which the users finally judge. By QoS mapping, we clarify the relationship between user–level QoS parameters derived by the SD method and the overall subjective quality. In order to find the implication of the user–level QoS parameters, we perform QoS mapping between user–level and application–level. In this paper, we utilize the *multiple regression analysis* for QoS mapping. It should be

noted that the mapping functions thus obtained can be used for real–time monitoring of the user–level QoS, since the application–level QoS parameters adopted in this paper are automatically measurable in real time.

The rest of the paper is organized as follows. Section II presents an overview of ITU–T Rec. J. 148. Section III introduces a method of multidimensional user–level QoS assessment. Section IV explains our experiment. We show experimental results and their considerations in Sec. V. Section VI concludes the paper.

II. Perceptual multimedia QoS assessment model based on ITU–T Recommendation J. 148

Figure 1 depicts the basic form of the multimedia QoS assessment model in ITU–T Rec. J.148.

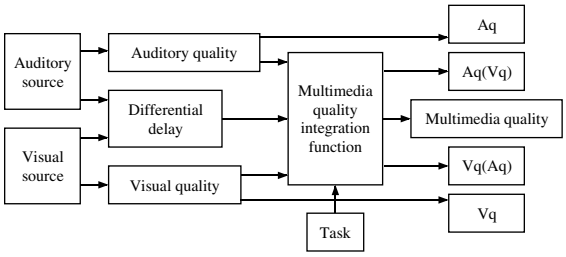

Fig. 1. Basic components of the multimedia model in ITU–T Rec. J. 148.

This model contains one primary output (*multimedia quality*) and four subsidiary ones: Aq, $Aq(Vq)$, Vq and $Vq(Aq)$. The primary output, *i.e.*, the multimedia quality, is a predicted measure of overall auditory–visual quality. Aq and Vq are a prediction of perceptual quality for the audio and that for the video, respectively. $Aq(Vq)$ and $Vq(Aq)$ are perceptual auditory quality taking account of any influence of the video on audio and the visual quality affected by audio, respectively. We usually need only the primary output. However, the primary output can be a function of the subsidiary outputs, especially $Aq(Vq)$ and $Vq(Aq)$. This requires us to obtain a function f which satisfies

$$Q = f(Aq(Vq), Vq(Aq)) \tag{1}$$

where Q is the multimedia quality.

In this paper, we resort to the SD method to extract the subsidiary outputs. We also assess the multimedia quality by the method of successive categories. Then, in order to obtain f, we clarify relationship between the multimedia quality and each subsidiary output by QoS mapping.

III. Multidimensional user–level QoS assessment

In this section, we first introduce the SD method. We then explain the method of successive categories [8] and the principal component analysis we use.

A. SD method

The SD method was proposed by Osgood as a method of measuring meaning. This method can assess an object for evaluation from many points of view with many *pairs of polar terms*. A pair of polar terms consists of one adjective and its opposite one; *e.g.*, quiet and noisy.

In the SD method, how to select pairs of polar terms used for assessment is important. In general, tens of pairs of polar terms are selected by a hearing or a questionnaire. For each selected pair of polar terms, a subjective score of an object for evaluation is measured by the *rating–scale method* [8]. The rating–scale method is also used to measure MOS, which is widely utilized for assessment of a single media. In this paper, we refer to an object for evaluation as a *stimulus*.

In the rating–scale method, subjects classify each stimulus into one of a certain number of categories. 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 appropriate in the strict sense to use the assigned numbers as they are when we calculate the value of the user–level QoS parameter.

B. Method of successive categories

With the psychometric methods, the human subjectivity can be represented by a *measurement scale*. 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 [8]. Since almost all the statistical procedures can be applied to the interval scale and the ratio scale, it is desirable to represent the user–level QoS by an interval scale or a ratio scale. In this paper, we use the interval scale for simplicity of calculation as in [1].

In order to obtain an interval scale as the user–level QoS parameter from the result of the rating–scale method, we first measure the frequency of each category with which the stimulus was placed in the category. With the *law of categorical judgment* [8], we can translate the frequency obtained by the rating–scale method into an interval scale. We refer to the interval scale as the *psychological scale*. See [1] for details of how to apply the law of categorical judgment and a comparison between the psychological scale and MOS.

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 [17] proposed a method of testing the goodness of fit for a scale calculated with Thurstone's law of comparative judgment [8], which is one of psychometric methods. The method can be applied to a scale obtained by the law of categorical judgment. This paper uses Mosteller's method to test the goodness of fit.

C. Principal component analysis

In the SD method, we assess a stimulus with dimensions whose number equals the one of pairs of polar terms. However, many dimensions make analysis of assessment results difficult. The number of dimensions can be reduced by the *principal component analysis* (*PCA*) or the factor analysis. For example, if the number of dimensions is reduced to two or three, we can assess stimuli in a two– or three–dimensional space.

When we use the PCA, we must decide the necessary number of principal components (or dimensions) at first. In this paper, we first calculate cumulative contribution rates; according to the calculated cumulative contribution rates, we decide the necessary number of principal components. We then examine the principal components whose number equals the obtained one. We regard the obtained principal components as user–level QoS parameters.

IV. Experiment

A. Experimental set up

We simulated audio–video transmission with a network simulator and assessed user–level QoS of the transmission. In our experiment, we used ns2 [18] as the network simulator and transmitted six contents:

- two news broadcasts, which are represented by N1 and N2
- two recorded broadcasts of tennis games (T1 and T2)
- two Japanese TV animations (A1 and A2)

We encoded all the contents with a common coding scheme so that application–level QoS of each type of contents becomes almost the same. Table I shows the specification of the transmitted media. In this table, *MU* means the *media unit*, which is the information unit for media synchronization at the application–level. In this paper, we define a video frame as a video MU and a constant number (namely, 1000) of audio samples as an audio MU.

In our experiment, we changed the application–level QoS by transmitting audio–video streams over a loaded network. Figure 2 depicts our network configuration for simulation. In the network, the media server is connected to the media

TABLE I
AUDIO AND VIDEO SPECIFICATIONS.

	audio	video
coding method	G.711 μ–law	MPEG1
image size[pixel]	-	320×240
picture pattern	-	IBBPBBPBBPBBPBB
average MU size [byte]	1000	5000
total MU number	120	300
play time [s]	15	15
average MU rate [MU/s]	8	20
average MU interval [ms]	125	50
coding mode	-	CBR
average bit rate [kb/s]	64	533

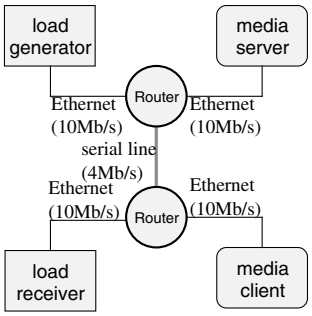

Fig. 2. Network configuration.

client via two routers; it transmitted audio–video streams to the media client. The line speed between the two routers is 4.0 Mb/s. The load generator generated UDP messages of 1472 bytes each at exponentially distributed intervals as load traffic. Each router accommodates the two terminals through individual Ethernets of 10Mb/s. In order to realize different QoS at the application-level, we used three kinds of the average amount of load traffic: 3.00, 3.40 and 3.80 Mb/s. While the media server sent audio–video streams, the load generator transmitted the load traffic. As the amount of load traffic increased, audio broke and video froze frequently in output audio–video streams. We recorded the audio–video streams that the media client had output and treated $3 \times 6 = 18$ recorded audio–video streams as stimuli for user–level QoS assessment.

B. Selection of pairs of polar terms

We selected pairs of polar terms of Japanese adjectives for the SD method. In order to investigate how subjects recognize audio quality and video one separately, we explicitly specified the type of medium (audio or video) for stimulus by adding the type to each adjective. When we could not find any appropriate adjective, we adopted a verb instead. First, we collected 90 pairs of polar terms. Second, we picked up 30 pairs which represent audio or video quality from among them. Third, we selected 20 pairs of polar terms from among the above 30 pairs by a preliminary experiment.

Note that this experiment was performed in Japanese. Then, this paper has translated the used Japanese pairs of polar terms into English. Therefore, the meanings of adjectives or verbs written in English here slightly differ from those of Japanese ones.

Table II shows the 20 selected pairs of polar terms in a form of *English adjective or verb–opposite one*: For convenience, we have assigned a unique identification number j to each pair of polar terms where $j = 1, \cdots, 20$.

C. Multidimensional assessment with SD method

Subjects assessed 18 stimuli with the rating–scale method for each pair of polar terms. In order to express the degree of implication of the adjectives or verbs, we used two kinds of adverb and one adjective: *very*, *slightly* and *neutral*. As a result, we have five terms for each adjective or verb: "one adjective or verb with *very*", "the one with *slightly*", "*neutral*", "the opposite one with *slightly*" and "the one with *very*". For convenience, we refer to the terms as categories 5, 4, 3, 2 and 1. We assume that the term means higher quality as the category number becomes larger.

TABLE II
PAIRS OF POLAR TERMS.

j	Pair of polar terms
1	Audio is quiet–Audio is noisy
2	Video is bright–Video is dark
3	Audio is powerful–Audio is weak
4	Video flows–Video jolts
5	Audio is clear–Audio is distorted
6	Video is clean–Video is dirty
7	Audio is calm–Audio is riotous
8	Audio and video are in sync –Audio and Video are out of sync
9	Video is sharp–Video is blurred
10	Audio is rich–Audio is poor
11	Video does not break down–Video breaks down
12	Audio is uninterrupted–Audio is interrupted
13	Video is merry–Video is gloomy
14	Audio is comfortable–Audio is uncomfortable
15	Video is stable–Video is unstable
16	Audio is not hoarse–Audio is hoarse
17	Video has an impact–Video does not have an impact
18	Audio is beautiful–Audio is not beautiful
19	Video is natural–Video is artificial
20	Video does not freeze–Video freezes

The subjects are male and female, and their ages were twenties. The number of subjects are 58. Since subjects must assess stimuli for every pair of polar terms in the SD method, it can take long time to finish the assessment. The long assessment time may cause subjects to forget the impression of the stimuli. Thus, we allowed the subjects to confirm the same stimulus for many times.

We apply the method of successive categories to all the results of the SD method. We then obtain interval scale values for every pair of polar terms. The obtained interval scales become user–level QoS parameters. If we can apply the method of successive categories to all the results obtained in the SD method, the number of the user–level QoS parameters equals the one of the pairs of polar terms. In order to decrease the number, we perform the PCA of the obtained user–level QoS parameters and get the principal components. We regard the obtained principal components as user–level QoS parameters again.

D. Overall quality assessment

The subjects also assessed overall subjective quality of audio–video transmission by the method of successive categories. Twenty–four subjects from among the above mentioned ones assessed the overall subjective quality by the rating–scale method. The subjects were provided the same stimuli as ones which were used in the SD method. In the rating–scale method, we defined five categories of *impairment*:"*imperceptible*" assigned integer 5, "*perceptible, but not annoying*" 4, "*slightly annoying*" 3, "*annoying*" 2, and "*very annoying*" 1.

We apply the law of categorical judgment to the obtained result. We refer to the obtained interval scale as the *overall psychological scale* and regard it as the user–level QoS parameter which indicates overall subjective quality.

E. QoS mapping

We investigate the relationship between the principal components obtained by the SD method and the overall quality by QoS mapping from user–level to user–level. In this paper, we adopt multiple regression analysis as the QoS mapping method. That is, we regard the principal components and the overall quality as predictor variables and the criterion variable, respectively. Moreover, we clarify the meanings of the principal components by QoS mapping from application–level to user–level. Similarly, we treat the application–level

QoS parameters and the principal components as predictor variables and criterion variables, respectively.

V. Experimental results

A. Psychological scale for each pair of polar terms

We applied the rating–scale method to each pair of polar terms. As a result, for pair 20, namely, "Video does not freeze–Video freezes", all the subjects assigned two stimuli into category 1. Therefore, we cannot calculate interval scale values of the two stimuli for pair 20 by the law of categorical judgment. Thus, we have decided to remove pair 20. As a result of elimination of pair 20, pairs 1 through 19 in Table II have remained.

In the following, we denote the psychological scale corresponding to pair j of polar terms as *psychological scale* j.

We performed Mosteller's test of goodness of fit [17] for the obtained interval scales. As a result of the test, the null hypothesis that the obtained interval scales fit 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.

Table III shows the *psychological intervals* between categories for each pair of polar terms; the psychological interval is defined as the difference between the upper boundary of one category and the one of the next higher category. For example, 1–2 indicates the difference between the upper boundary of category 1 and the one of category 2. Since the upper boundary of category 5 is infinity, we do not include the interval between category 4 and category 5. From Table III, we find that the

TABLE III
Psychological interval between category boundaries for each pair of polar terms.

j	Psychological interval		
	1-2	2-3	3-4
1	1.606	0.563	1.157
2	0.556	1.209	1.374
3	0.723	1.242	1.577
4	0.797	0.495	1.099
5	1.654	0.724	1.188
6	0.963	0.853	1.272
7	1.290	1.382	0.948
8	0.789	0.421	0.731
9	1.126	0.921	1.270
10	0.980	1.298	1.521
11	0.739	0.685	1.100
12	1.193	0.443	0.902
13	0.886	1.531	1.109
14	1.092	1.155	1.140
15	0.910	0.555	1.174
16	1.370	0.780	1.064
17	0.877	1.228	1.028
18	1.389	0.902	1.222
19	0.908	0.613	1.113

psychological intervals are not uniform for each pair of polar terms. This means that the translation by the law of categorical judgment is mandatory.

We now show the psychological scale value of every stimulus for each pair of polar terms. First, we plot the psychological scale value of every stimulus for pairs 1 through 10 of polar terms in Figs. 3 through 5. Figures 3, 4 and 5 plot the value when the amount of load traffic are 3.0, 3.4 and 3.8 Mb/s, respectively. We then display the psychological scale value for pairs 11 through 19 of polar terms in Figs. 6, 7 and 8, which correspond to 3.0, 3.4 and 3.8 Mb/s load traffic, respectively. In these figures, we set the minimum value of the psychological scale to the origin, for convenience.

From Figs. 3 through 8, we see that the psychological scale values for almost all the pairs of polar terms decrease as the amount of load traffic increases. However, the degree of the decrease depends on the pair of polar terms and contents. For example, for pair 4 of polar terms "Video flows–Video jolts", the psychological scale value extremely decreases as the amount of load traffic increases. On the other hand, for pair 2 of polar terms "Video is bright–Video is dark", the psychological scale value decreases only slightly.

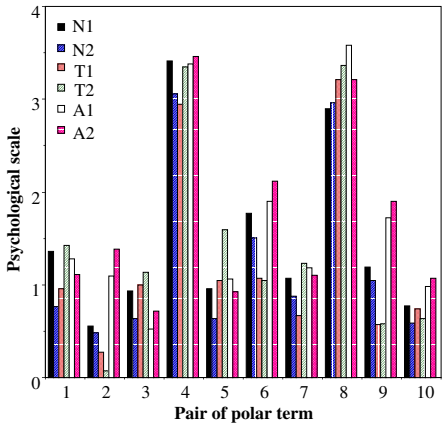

Fig. 3. Psychological scale j (3.0Mb/s load).

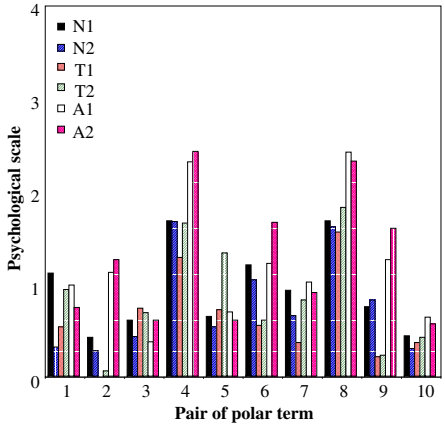

Fig. 4. Psychological scale j (3.4Mb/s load).

B. Result of PCA

We performed PCA of the psychological scales obtained in the previous subsection and calculated the cumulative contribution rates. Figure 9 displays the cumulative contribution rate as a function of the number of dimensions. From Fig. 9, we see that the cumulative contribution rate is 87.8 % when the first and second principal components are adopted. Similarly, the cumulative contribution rate of the first three principal components becomes 93.6 %. By adding the third principal component, the cumulative contribution rate significantly increases. However, the addition of the fourth principal component does not contribute to increase of the cumulative

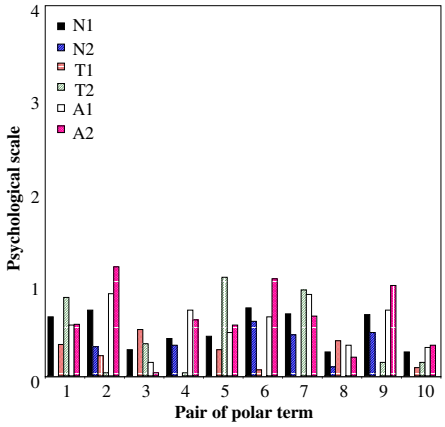

Fig. 5. Psychological scale j (3.8Mb/s load).

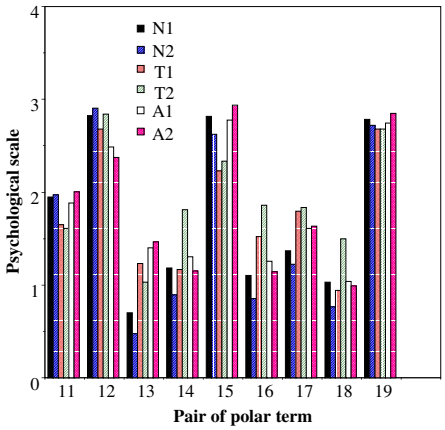

Fig. 6. Psychological scale j (3.0Mb/s load).

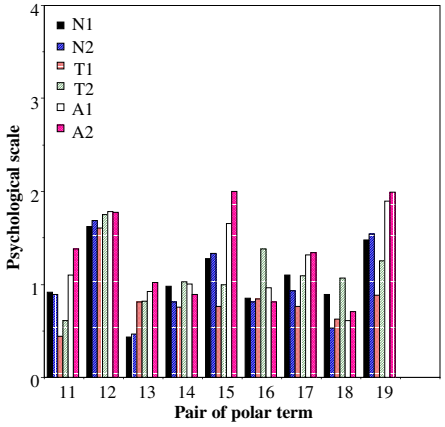

Fig. 7. Psychological scale j (3.4Mb/s load).

contribution rate very much. Therefore, we treat the first three principal components as user–level QoS parameters. These three user–level QoS parameters can express 93.6 % of the amount of information which all the psychological scales have. That is, the 19 psychological scales can be summarized into the three principal components with high accuracy.

The principal component loadings of each psychological scale are shown in Table IV. Figures 10 and 11 plot the principal component loading values. In Fig. 10, the abscissa indicates the first principal loading, and the ordinate represents the second one. The abscissa and the ordinate of Fig. 11 are the first principal loading and the third one, respectively.

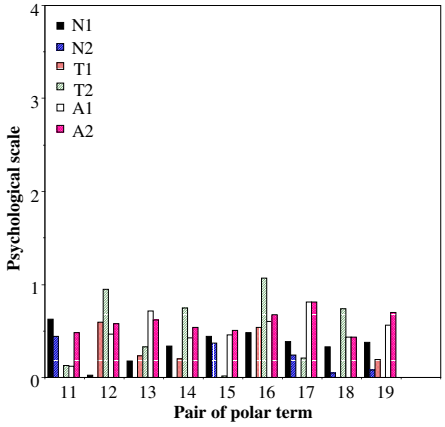

Fig. 8. Psychological scale j (3.8Mb/s load).

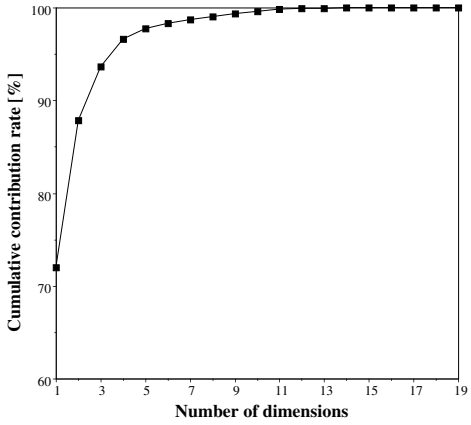

Fig. 9. Cumulative contribution rate versus number of dimensions.

TABLE IV
PRINCIPAL COMPONENT LOADING OF PSYCHOLOGICAL SCALE j.

psychological scale j	loading		
	First	Second	Third
1	0.860	-0.211	0.299
2	0.275	0.869	0.331
3	0.749	-0.501	-0.316
4	0.971	0.097	-0.194
5	0.731	-0.579	0.316
6	0.801	0.560	-0.084
7	0.771	0.093	0.488
8	0.958	0.003	-0.240
9	0.643	0.744	0.118
10	0.942	0.234	0.030
11	0.911	0.229	-0.254
12	0.925	-0.176	-0.266
13	0.828	0.096	0.177
14	0.927	-0.308	0.120
15	0.945	0.208	-0.224
16	0.802	-0.532	0.173
17	0.954	0.042	-0.005
18	0.881	-0.380	0.204
19	0.962	0.112	-0.226

From Fig. 10, we find that the first principal component highly correlates with psychological scales corresponding to many pairs of polar terms; in particular, “Video flows–Video jolts”, “Video is natural–Video is artificial”, “Audio and video are in sync–Audio and video are out of sync”, “Video has an impact–Video does not have an impact”, “Video is stable–Video is unstable” and “Audio is rich–Audio is poor”. Therefore, it is not easy to judge what the first principal component means. However, many of the psychological scales which highly correlate with the first principal component seem to concern *dynamics of the media*. Consequently, the first principal component is considered to indicate *quality concerned with dynamics of audio* or *that of video*.

The second principal component correlates with the psychological scales corresponding to the pair of polar terms “Video is bright–Video is dark” and “Video is sharp–Video is blurred”. That is, the second principal component seems to mean *brightness or sharpness of video*.

Finally, Fig. 11 reveals that the third principal component correlates with the psychological scale for pairs of polar terms “Audio is calm–Audio is riotous”. Consequently, this component shows *calm of audio*.

Note that we have tried to clarify the meaning of each principal component by examining Figs. 10 and 11. However, we do not intend to clarify the meaning only by the principal component analysis. We will quantitatively investigate the meaning by QoS mapping later.

Table V and Figs. 12 and 13 show the principal component scores for each stimulus. In Fig. 12, the abscissa is the first principal component score, and the ordinate is the second one. Similarly, in Fig. 13, the abscissa and ordinate indicate the first principal component score and the third one, respectively. In these figures, each stimulus is denoted by “the content type–average amount of load traffic”. For example, A1–3.80 means

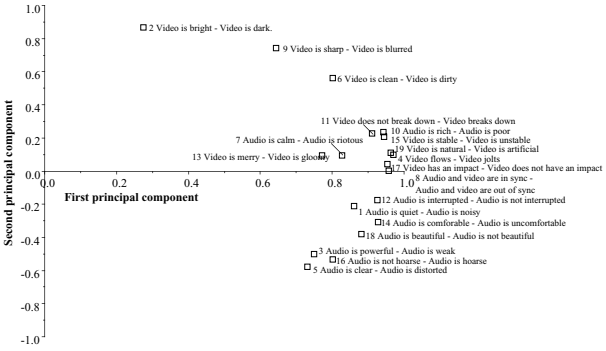

Fig. 10. First and second principal component loadings.

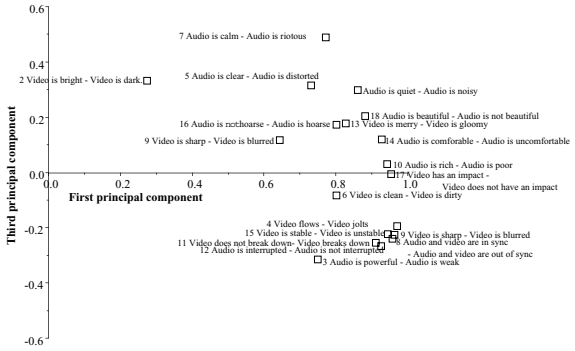

Fig. 11. First and third principal component loadings.

the stimulus of A1 with the average amount of load traffic of 3.80Mb/s.

In Fig. 12, we find that the first principal component score decreases as the average amount of load traffic increases regardless of contents. On the other hand, the second principal component score scarcely depends on the average amount of load traffic. It depends on the content type. Also, it is difficult to find main factors of the third principal component from Fig. 13.

C. Overall quality assessment

We assessed the overall perceptual quality for 24 subjects by the rating–scale method. Table VI shows the result. By applying the law of categorical judgment to the result in Table VI, we calculated the interval scale which indicates the overall perceptual quality. We then performed Mosteller's test. As a result of the test, we found that the null hypothesis that the obtained interval scale fits the observed data cannot be rejected at significance level 0.05. Therefore, we regard the obtained

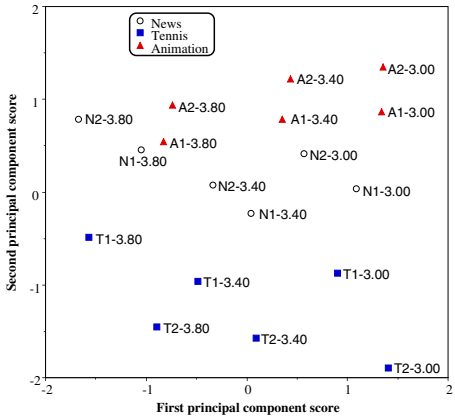

Fig. 12. First and second principal component scores.

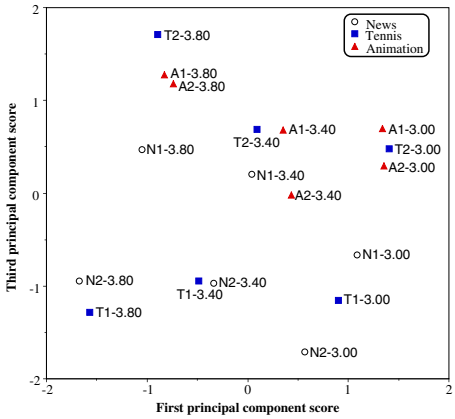

Fig. 13. First and third principal component scores.

interval scale as the overall psychological scale. Table VII and Fig. 14 show the obtained overall psychological scale.

From Fig. 14, we see that the overall psychological scale decreases as the average amount of load traffic increases for all the contents. Furthermore, the decreasing rate depends on the content type.

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

In order to clarify the meaning of each principal component obtained in Subsection V-B quantitatively, we perform QoS mapping from application–level to user–level. Before multiple regression analysis, we must select some application–level QoS parameters as predictor variables. In this paper, as in [1],

TABLE V
PRINCIPAL COMPONENT SCORE OF EACH STIMULUS.

stimulus	content	load [Mb/s]	Principal component score		
			First	Second	Third
1	N1	3.00	1.086	0.033	-0.662
2	N1	3.40	0.040	-0.228	0.203
3	N1	3.80	-1.047	0.458	0.467
4	N2	3.00	0.569	0.412	-1.711
5	N2	3.40	-0.338	0.075	-0.969
6	N2	3.80	-1.675	0.785	-0.944
7	T1	3.00	0.905	-0.871	-1.153
8	T1	3.40	-0.488	-0.958	-0.942
9	T1	3.80	-1.569	-0.487	-1.280
10	T2	3.00	1.409	-1.897	0.478
11	T2	3.40	0.095	-1.570	0.691
12	T2	3.80	-0.893	-1.454	1.708
13	A1	3.00	1.341	0.865	0.696
14	A1	3.40	0.351	0.786	0.684
15	A1	3.80	-0.828	0.544	1.279
16	A2	3.00	1.352	1.351	0.291
17	A2	3.40	0.432	1.217	-0.017
18	A2	3.80	-0.742	0.941	1.181

TABLE VI
RESULT OF OVERALL PERCEPTUAL QUALITY MEASUREMENT BY THE RATING–SCALE METHOD.

stimulus	content	load [Mb/s]	category				
			1	2	3	4	5
1	N1	3.00	0	0	0	6	18
2	N1	3.40	0	5	10	9	0
3	N1	3.80	7	13	4	0	0
4	N2	3.00	0	0	2	6	16
5	N2	3.40	0	8	10	5	1
6	N2	3.80	13	8	3	0	0
7	T1	3.00	0	0	0	2	22
8	T1	3.40	3	12	7	0	2
9	T1	3.80	21	3	0	0	0
10	T2	3.00	0	0	1	7	16
11	T2	3.40	1	14	6	2	1
12	T2	3.80	23	1	0	0	0
13	A1	3.00	0	0	0	4	20
14	A1	3.40	1	2	9	8	4
15	A1	3.80	10	12	1	0	1
16	A2	3.00	0	0	0	4	20
17	A2	3.40	0	3	8	9	4
18	A2	3.80	11	11	2	0	0

TABLE VII
OVERALL PSYCHOLOGICAL SCALE.

stimulus	content	load [Mb/s]	overall psychological scale
1	N1	3.00	5.493
2	N1	3.40	3.847
3	N1	3.80	2.218
4	N2	3.00	5.355
5	N2	3.40	3.348
6	N2	3.80	1.800
7	T1	3.00	6.202
8	T1	3.40	2.954
9	T1	3.80	0.581
10	T2	3.00	5.529
11	T2	3.40	3.071
12	T2	3.80	0.000
13	A1	3.00	5.786
14	A1	3.40	3.917
15	A1	3.80	2.279
16	A2	3.00	5.786
17	A2	3.40	4.102
18	A2	3.80	1.788

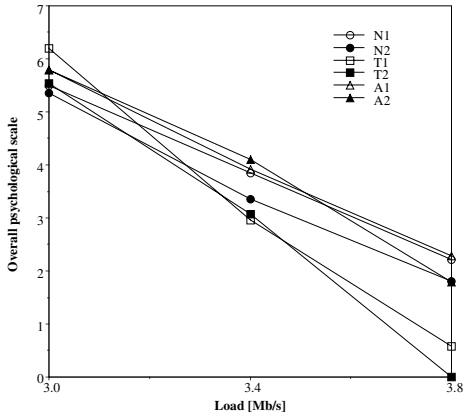

Fig. 14. Overall psychological scale versus average load.

[3], [6] and [7], we regard measures of *media synchronization* quality as candidates of the predictor variables.

In general, media synchronization is classified into *intra–stream synchronization* and *inter–stream synchronization*. The former keeps the continuity of a single stream (audio or video), while the latter is synchronization between an audio stream and the corresponding video stream.

In order to represent media synchronization quality, references [7] and [19] use nine application–level QoS parameters, all of which are automatically measurable. We also use them in this paper. 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. 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; this is defined as the average number of (either audio or video) MUs output in a second at the destination. Third, we treat 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. Finally, we adopt 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 eight 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.

In order to perform multiple regression analysis, we must pick up some application–level QoS parameters as predictor variables from among the nine parameters. We then classify the application–level QoS parameters by the principal component analysis.

As a result of PCA, we see that the cumulative contribution rate for the first two principal components is 97.9 %. This means that the first two principal components can present 97.9 % of information involved by the nine application–level QoS parameters. Therefore, we adopt the first and second principal components. Table VIII displays the principal component loading of each principal component.

TABLE VIII
PRINCIPAL COMPONENT LOADINGS FOR EACH APPLICATION–LEVEL QOS PARAMETER.

	First	Second
R_v	-0.999	-0.036
L_v	0.999	0.034
E_v	-0.352	0.911
E_{int}	0.888	0.453
C_v	0.964	0.186
R_a	-1.000	0.003
L_a	0.999	-0.021
E_a	-0.752	0.584
C_a	0.989	0.134

From Table VIII, we find that the nine parameters can be classified into two groups:
group a) R_a, L_a, E_a, C_a, R_v, L_v, E_{int} and C_v
group b) E_v
The parameters in group a) highly correlate with the first principal component. On the other hand, the parameter in group b) highly correlates with the second principal component.

In order to avoid the effect of multi–collinearity, we select one application–level QoS parameter from each group. Consequently, the number of the combination of the application–level QoS parameters becomes $8 \times 1 = 8$.

In this paper, we first perform multiple regression analysis with all combinations of the application–level QoS parameters as predictor variables for each principal component of user–level QoS. Then, we select a combination which indicates the highest contribution rate adjusted for degrees of freedom. Tables IX through XI present contribution rates adjusted for degrees of freedom for the first through third principal components, respectively.

TABLE IX
CONTRIBUTION RATE ADJUSTED FOR DEGREE OF FREEDOM FOR EACH COMBINATION OF INDEPENDENT VARIABLES (FIRST PRINCIPAL COMPONENT OF PSYCHOLOGICAL SCALES).

	R_v	L_v	E_{int}	C_v	R_a	L_a	E_a	C_a
E_v	0.89	0.88	0.88	0.93	0.86	0.85	0.42	0.90

TABLE X
CONTRIBUTION RATE ADJUSTED FOR DEGREE OF FREEDOM FOR EACH COMBINATION OF INDEPENDENT VARIABLES (SECOND PRINCIPAL COMPONENT OF PSYCHOLOGICAL SCALES).

	R_v	L_v	E_{int}	C_v	R_a	L_a	E_a	C_a
E_v	0.62	0.62	0.62	0.58	0.63	0.64	0.75	0.60

TABLE XI
CONTRIBUTION RATE ADJUSTED FOR DEGREE OF FREEDOM FOR EACH COMBINATION OF INDEPENDENT VARIABLES (THIRD PRINCIPAL COMPONENT OF PSYCHOLOGICAL SCALES).

	R_v	L_v	E_{int}	C_v	R_a	L_a	E_a	C_a
E_v	0.14	0.14	0.10	0.15	0.13	0.14	0.27	0.14

From Table IX, we see that two combinations, (C_v, E_v) and (C_a, E_v), show high contribution rates adjusted for degrees of freedom for the first principal component. In this paper, we select a parameter regarding audio and one concerning video. Since E_v is common to the two combinations and concerns video, we choose C_a as the other predictor variable. The regression line for the first principal components is

$$\hat{U}_1 = 3.163 - 1.595 \times 10 C_a + 2.499 \times 10^{-4} E_v \qquad (2)$$

where $\hat{U}_1$ is an estimate of the first principal component. By the statistical test [20], we have confirmed that both C_a and E_v are statistically significant at significance level 0.01.

By investigating the standardized partial regression coefficient of each predictor variables, we found that C_a contributes to the first principal component more than E_v. Therefore, we can regard the first principal component as the factor which indicates audio perceptual quality affected by video quality. This means that the first principal component corresponds to $Aq(Vq)$ in Eq. (1).

Similarly, from Table X, we select (E_a, E_v) as predictor variables for the second principal component. The regression line becomes

$$\hat{U}_2 = -8.958 + 4.861 \times 10^{-2} E_v - 1.002 \times 10^{-2} E_a \quad (3)$$

where $\hat{U}_2$ is an estimate of the second principal component. The result of the statistical test shows that E_v and E_a are statistically significant at significance level 0.01. The standardized partial regression coefficient of E_v and that of E_a show that E_v contributes to the second principal component more than E_a. Consequently, the second principal component can be regarded as the video perceptual quality affected by audio quality, that is, $Vq(Aq)$.

Finally, Table XI shows very low contribution rates. This means that the third principal component is not related to the application–level QoS parameters. From Fig. 13, whose ordinate indicates the third principal component score, we see that the third principal component slightly concerns the content type. Then, referring to Fig. 13, we define a dummy variable C as follows:

$$C = \begin{cases} 0 & \text{(Content is N2)} \\ 1 & \text{(Content is T1)} \\ 2 & \text{(Content is N1)} \\ 3 & \text{(Content is A2)} \\ 4 & \text{(Content is T2)} \\ 5 & \text{(Content is A1)} \end{cases} \quad (4)$$

Instead of the nine application–level QoS parameters, we regard C as a predictor variable. As a result, we obtain

$$\hat{U}_3 = -1.229 + 0.492C \quad (5)$$

where $\hat{U}_3$ is an estimate of the third principal component. The contribution rate adjusted for degrees of freedom of the above regression line becomes 0.855. As a result, we see that the third principal component depends on the content type, which cannot be expressed by the application–level QoS.

E. QoS mapping between user–level QoS

In order to clarify how the overall perceptual quality is expressed as a function of the principal components, we perform multiple regression analysis. As a result of multiple regression analysis, we obtain

$$\hat{I} = 3.559 + 1.738U_1 + 2.457 \times 10^{-1} U_2 - 4.995 \times 10^{-1} U_3 \quad (6)$$

where $\hat{I}$, U_1, U_2 and U_3 are the overall psychological scale, the first principal component, the second principal component and the third principal component, respectively. The contribution rate adjusted for degrees of freedom becomes 0.964. We statistically test whether U_1, U_2 and U_3 make a significant contribution to the multiple regression line. The result of the statistical test shows that all partial regression coefficients are statistically significant at significance level 0.01.

As discussed in the previous subsection, U_1 and U_2 correspond to $Aq(Vq)$ and $Vq(Aq)$, respectively. Moreover, I represents the multimedia quality. Consequently, Eq. (6) is a possible form of Eq. (1), which we want to obtain. Note that the function f in this case includes U_3; this means that the overall perceptual quality can be affected by the content type. Furthermore, by applying Eqs. (2), (3) and (5) to Eq. (6), we can estimate the overall psychological scale from the application–level QoS parameters, which are automatically measurable. Thus, the method proposed in this paper enables each receiving terminal to monitor the multidimensional user–level QoS, in particular, $Aq(Vq)$, $Vq(Aq)$ and I.

VI. Conclusions

This paper proposed a method of monitoring multidimensional user–level QoS of audio–video IP transmission in the framework of the perceptual multimedia quality model recommended in ITU–T J.148. As a result, we identified the subsidiary outputs $Aq(Vq)$ and $Vq(Aq)$. These outputs can be estimated from application–level QoS parameters. We then assessed overall perceptual quality and clarified how the overall perceptual quality is expressed as a function of the subsidiary outputs by QoS mapping. The obtained results showed that we can estimate the overall perceptually quality from application–level QoS parameters. Moreover, we quantitatively showed the effect of the contents on the overall perceptual quality.

Future work includes detailed study on the effect of content types on user–level QoS and applications of multidimensional user–level QoS monitoring to future IP networks that guarantee user–level QoS.

Acknowledgment

This work was supported by the Grant–In–Aid for Scientific Research of Japan Society for the Promotion of Science under Grant 17360179.

References

[1] S. Tasaka and Y. Ito, "Psychometric analysis of the mutually compensatory property of multimedia QoS," *Conf. Rec. IEEE ICC2003*, pp. 1880–1886, May 2003.
[2] ITU-T Rec. J.148, "Requirements for an objective perceptual multimedia quality model," May 2003.
[3] S. Tasaka and Y. Ito, "Real–time estimation of user–level QoS of audio–video transmission over IP networks," in *Conf. Rec. IEEE ICC2006*, June 2006.
[4] A. Watson and M. A. Sasse, "Measuring perceived quality of speech and video in multimedia conferencing applications," *Proceedings of ACM Multimedia'98*, pp. 55–60, Sept. 1998.
[5] Y. Ito and S. Tasaka, "Quantitative assessment of user–level QoS and its mapping," *IEEE Trans. Multimedia*, vol.7, no.3, pp. 572–584, June 2005.
[6] 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. ICC 2004*, pp. 2214-2220, June 2004.
[7] Y. Ito and S. Tasaka, "Effect of monitor size on user–level QoS of audio–video transmission over IP networks in ubiquitous environments," *Conf. Rec. PIMRC2005*, May 2005.
[8] J. P. Guilford, *Psychometric methods*, McGraw-Hill, New York, 1954.
[9] A. Bouch and M. A. Sasse, "The case for predictable media quality in networked multimedia applications," in *Proc. ACM/SPIE MMCN'00*, pp. 188–195, 2000.
[10] F. Wilson, I. Wakeman and W. Smith, "Quality of service parameters for commecial application of videotelephony," *Proc. Human Factors in Telecommunications'93*, pp. 139–148, 1993.
[11] K. Yamagishi and T. Hayashi, "Opinion model using psychological factors for interactive multimodal services," *IEICE Trans. Commun.*, vol.E89–B, no.2, pp. 281–288, Feb. 2006.
[12] C. E. Osgood, "The nature and measurement of meaning." *Psychological Bulletin*, vol. 49, no. 3, pp. 197–237 May 1952.
[13] C.–S. Lee and D. Knight, "Realization of the next–generation network," *IEEE Commun. Mag.*, vol.43 No.10 pp.34–41 ! $ Oct. 2005.
[14] 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.
[15] S. Namba, S. Kuwano, K. Kinoshita and K. Kurakata, "Loudness and timbre of broad–band noise mixed with frequency modulated sounds," *J. Acoust. Soc. Jpn. (E)*, vol. 13, no. 1, pp. 49–58, 1992.
[16] S. Namba, S. Kuwano, M. Koyasu, "The measuremet of temporal stream of hearing by continuous judgements–In the case of the evaluation of helicopter noise," *J. Acoust. Soc. Jpn. (E)*, vol. 14, no. 5, pp. 341–352, 1993.
[17] 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.
[18] URL http://www.isi.edu/nsnam/ns/
[19] 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.
[20] M. G. Bulmer, *Principles of statistics*, Dover publications, Inc., N. Y., 1979.