A Method of User–Level QoS Guarantee by Session Control in Audio–Video Transmission over IP Networks

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

***Abstract*— This paper proposes a session control method which provides a mechanism to achieve desirable user–level QoS (i.e., perceptual QoS) of audio–video transmission over IP networks. The proposed method, which is referred to as *GPSQ* (*Guarantee of Psychologically Scaled Quality*), utilizes a user–level QoS parameter of the interval scale in the psychometric analysis. GPSQ involves four kinds of components: media terminals, a SIP server, a QoS manager and bandwidth–controllable routers. From the media terminals, the SIP server obtains the information for this control, which is further delivered to the QoS manager. The QoS manager keeps a database of representative regression lines which express user–level QoS as a function of guaranteed bandwidth of audio and that of video. Using the regression line, the QoS manager calculates necessary guaranteed bandwidth according to user–level QoS specified by the user. Then, the QoS manager sets up routers to guarantee the calculated bandwidth. We implemented GPSQ in a simple experimental network and confirmed its effectiveness.**

I. INTRODUCTION

The widespread use of the IP protocol in broadband networks rapidly increases applications with audio–video transmission over IP networks, which basically provides only best–effort services and therefore cannot guarantee *QoS* (*Quality–of–Service*). In order for the IP network to become a dependable infrastructure of our society, it is strongly required to have some mechanism for QoS guarantee.

Along the protocol stack in IP networks, we can identify six kinds of QoS: *physical–level*, *node–level*, *network–level*, *end–to–end–level*, *application–level*, and *user–level* [1]. Among them, user–level QoS is the quality perceived by the users; consequently, the guarantee of QoS at this level is ideal for the users.

IETF specifies various QoS control schemes for node–level through application–level as RFC: the typical examples are IntServ [2] and DiffServ [3]. However, no scheme can be found at application–level and user–level for audio–video transmission.

ITU–T, on the other hand, copes with the penetration of the IP protocol by leading the standardization activities for the *Next Generation Network* (*NGN*) [4], which is an enhanced IP-based network. ITU–T defines and develops NGN as Recommendations of Y series. Y. 1291 provides an architectural framework for support of QoS from node–level through end–to–end level [5]; examples of QoS control schemes it presents include IntServ and DiffServ. In addition, Y. 1540 describes network–level QoS, which is called *network performance* in the recommendation [6]. The QoS parameters introduced by Y. 1540 includes IPTD(IP packet transfer delay), IPDV(IP packet delay variation), IPLR(IP packet loss ratio), IPER(IP packet error ratio), and IPPT(IP packet throughput). Also, Y. 1541 classifies typical applications into six classes and provisionally specifies the QoS parameter values to be satisfied for five classes out of the six [7]. It should be noted that the QoS parameter values thus specified does not necessarily achieve user–level QoS the users desire in efficient ways; the specified values may require too many resources to satisfy the users' demand or it may not realize the desired user–level QoS.

ITU–T Rec. G.1010 deals with the requirement for user–level QoS in multimedia applications with audio–video transmission over packet networks [8]. G. 1010 identifies eight categories of multimedia applications by considering user expectations and specifies the ranges of packet loss ratio and one–way delay to be satisfied for each category. This can be targets for QoS control in multimedia applications. However, note that user–level QoS depends on other factors than packet loss and delay; for instance, the kind of contents, terminal types, and encoding schemes of audio and video. Consequently, it is not clear how the user–level QoS is achieved with the specified values of packet loss ratio and one–way delay.

It may not be realistic to try to achieve user–level QoS specified by individual users in public IP networks. However, it is desirable to enable a variety of usage of the IP protocol including the one in private networks in order for the IP network to be a true infrastructure. With respect to this issue, Focus Group on NGN (FGNGN) discusses the level of customer manageability, i.e., how resource monitoring and control is performed, ranging from no management to individual management [9].

A prerequisite to enable the guarantee of user–level QoS is the development of methods for assessing user–level QoS quantitatively; the methods should be able to monitor user–level QoS in real time in addition to non–real–time assessment. Since the user–level QoS depends on the user's subjectivity, monitoring the users directly in real time is not realistic. It is practical to estimate user–level QoS from measurable lower–level QoS parameters such as packet loss ratio and delay. Once some method of user–level QoS assessment and that for real–time estimation are developed, they can be foundations for the technology of user–level QoS guarantee. However, no study on this problem had not been found in the literature in the context of audio–video transmission over IP networks.

Then, the authors addressed themselves to the problem; they proposed schemes for user–level QoS assessment by the psychometric method and multivariate analysis [10], [11] and a scheme for the real–time estimation [12]. The estimation scheme utilizes multiple regression lines that predict user–level QoS parameter values from application–level QoS parameter values which can be measured in real time.

This paper proposes a method of user–level QoS guarantee based on the schemes for user–level QoS assessment and real–time estimation. We refer to the method as *GPSQ*(*Guarantee of Psychologically Scaled Quality*); the user–level QoS is expressed in terms of a QoS parameter of the *interval scale*, which is a metric calculated by the psychometric method [13]; we call it the *psychological scale* [10]. The interval scale can represent the human subjectivity more accurately than *MOS* (*Mean Opinion Score*), which is often used as a user–level QoS parameter in technical papers and ITU–T/ITU–R recommendations.

GPSQ provides a framework for user–level QoS guarantee by session control with *SIP* (*Session Initiation Protocol*) [14]. In this paper, we consider how SIP should be utilized for this purpose. We implement GPSQ in a simple experimental system and show the effectiveness in a quantitative way.

The rest of the paper is organized as follows. Section II proposes GPSQ. Section III presents an experimental methodology to examine the effectiveness of GPSQ. Section IV examines experimental results thus obtained and show the effectiveness of GPSQ.

II. GPSQ

A. Principle of GPSQ

The basic principle of GPSQ is reservation of necessary bandwidth to achieve specified user–level QoS by session control. This involves two fundamental issues: (1) how much bandwidth should be allocated, and (2) how bandwidth reservation is made along an end–to–end path. This paper focuses only on the former one as a first step of our approach. We leave the latter issue as future work. The issue has been reported in many publications; those results will be starting points of our research on the latter issue.

Essential ingredients of GPSQ are a SIP server, a QoS manager and bandwidth–controllable routers in addition to media terminals, which send and/or receive audio–video streams. Using SIP, media terminals notify the SIP server of the user's request for audio–video transmission and the terminal's capability concerning media encoding/decoding and display; the request contains information on user–level QoS desired by the user along with other kinds of necessary information.

Resource allocation and management are carried out by the QoS manager in cooperation with the SIP server. The SIP server delivers the information obtained from the media terminals to the QoS manager, which allocates necessary bandwidth to achieve specified user–level QoS. The QoS manager keeps a database of *representative regression lines* each of which expresses user–level QoS for a content type as a function of guaranteed bandwidth of audio and that of video; therefore, it can calculate necessary guaranteed bandwidth for user–level QoS specified by the user. As explained in Subsection II-D, a representative regression line is defined for each *content type*, which contains a group of contents with similar features. Then, the QoS manager sets up routers to guarantee the calculated bandwidth.

B. Session control by SIP and bandwidth allocation by QoS manager

In this paper, we restrict ourselves to one-way audio–video streaming services for simplicity of discussion; that is, we consider one–way audio–video transmission from a media sender to a media recipient. We also suppose that the media sender stores a collection of audio–video streams along with their attributes and statistics; the attribute includes the name, content type and encoding schemes, while the statistics here imply the ones of audio and video bit rates such as the average, variance and maximum.

Figure 1 illustrates a basic call establishment/termination procedure with SIP; it shows only a simple scenario in which the SIP server can accept requests by the users and then the QoS manager can allocate required bandwidth. Referring to Fig. 1, let us explain how the call progresses without going into the syntax and semantics of SIP messages in any depth.

1) The media recipient generates an *INVITE* request to make a call to the media sender. It sends the SIP server the INVITE request with information including IP addresses of the media recipient and sender, the type of the terminal, the name of content it wants the media sender to transmit. The message body of the INVITE request also contains necessary information for the control by the QoS manager, including user–level QoS specified by the user and encoding schemes of audio and video; this is described with *SDP* (*Session Description Protocol*) [15].

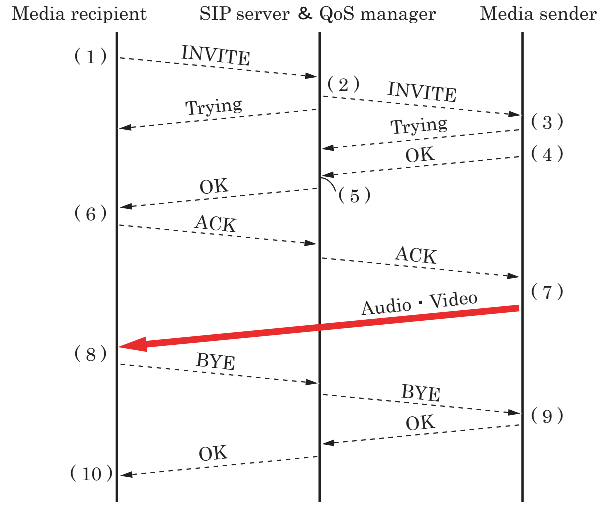

Fig. 1. A basic call establishment/termination procedure with SIP

2) Receiving the INVITE request, the SIP server extracts the control information necessary for the QoS manager from it, modifies and then forwards the INVITE request to the media sender. In the meantime, the SIP server sends a *Trying* response to the media recipient to notify it of receipt of the INVITE request.
3) Once the media sender has received the modified INVITE request, it responds to the SIP server by transmitting Trying. From the INVITE request, the media sender obtains information such as the name of requested content and the IP address of the media recipient.
4) If the media sender can accept the call, it sends the success response *OK* to the SIP server. Encoding schemes and statistics of bit rates of media to be transmitted in addition to the content type are described as the payload of the OK response. If the media sender rejects the INVITE request for any reason, it responds to the SIP server with the *BYE* message.
5) Receipt of the OK response by the SIP server can provide it with all necessary information for the QoS control, which is forwarded to the QoS manager. On the basis of the information, the QoS manager selects an appropriate representative regression line from among the ones in the database. Using the representative regression line and the statistics of the bit rates, the QoS manager calculates necessary bandwidth for audio and that for video so that they can achieve the user–level QoS specified by the media recipient; then, the QoS manager sets up routers to guarantee the pair of bandwidth. Once the bandwidth has been reserved, the QoS manager informs the SIP server of the reservation; accordingly, the SIP server sends an OK response to the media recipient. If the reservation is unsuccessful, a *Not Acceptable* response is sent instead.
6) Once the media recipient has received the OK response, it completes session establishment by sending the SIP server an *ACK* request, which is forwarded to the media sender. If the media recipient receives Not Acceptable, it terminates the session by responding to the SIP server with a BYE message.
7) After the media sender has received the ACK request, it starts to transmit the audio–video streams and continues the transmission until the end of the streams.
8) When the media recipient has completed receiving the audio–video streams, it sends the SIP server a BYE message, which is forwarded to the media sender.
9) The media sender responds to the SIP server by transmitting an OK response in order to terminate the session. Receiving the OK response, the SIP server notifies the QoS manager of the session termination. The QoS manager commands the routers to release the reserved

bandwidth. Then, the SIP server forwards the OK response to the media recipient.
10) Receipt of the OK response by the media recipient completes the session termination process.

C. Bandwidth control at routers

The QoS manager instructs bandwidth–controllable routers on the guarantee of required bandwidth, which is commonly realized with packet schedulers. The packet scheduling algorithm for bandwidth guarantee is typified by *WFQ* (*Weighted Fair Queueing*) and *CBWFQ* (*Class–Based Weighted Fair Queueing*) [16]; they are installed into many of commercially available routers.

D. Database of representative regression lines

This database at the QoS manager is a key to the success of GPSQ; the QoS manager utilizes it to calculate the required bandwidth according to the user–level QoS specified by the user. This subsection explains how we should construct the database.

First of all, we must note that user–level QoS depends on a variety of factors such as reserved bandwidth for audio, that for video, display types, content types, and encoding schemes of audio and video in addition to network conditions. This implies essentially an infinite number of combinations of the factors, which leads to an infinite number of regression lines. Therefore, this straightforward approach is infeasible.

In order for the database to be feasible, we first restrict display types and encoding schemes of audio and video to a certain number of classes, which is practically common. We next classify possible contents treated in the target application into a reasonably finite number of types so that every content in the same type can have a similar form of regression line. For each type, we pick up several contents and calculate their regression lines as in [10] and [12]. Then, we select one from among the regression lines as the representative regression line of the type. Collecting the representative regression lines from all the types for each class of the set of display type and encoding scheme, we can construct the database.

As seen from the above discussion, we are faced with two important technical problems to be solved: (a) how should contents be classified into types? and (b) how should the representative regression line for a type be selected?

Regarding Problem (a), previous studies on perceptual quality assessment of multimedia (typically, audio and video streams) can be informative; for instance, the test plan of the Multimedia (MM) working group in *VQEG* (Video Quality Experts Group) [17]. As the test materials representative of a range of content and applications, the MM group lists the following eight types: (1) video conferencing, (2) movies, (3) sports, (4) music video, (5) advertisement, (6) animation, (7) broadcasting news, and (8) home video. This classification can be a starting point of the solution to Problem (a).

Problem (b) is closely related to Problem (a). The goal of Problem (b) is to obtain a representative regression line that can be as accurate as possible for all possible contents in the type; achievable accuracy depends on what contents are included in the type. Thus, the two problems interact with each other.

Since this paper is a first step of our study on GPSQ, detailed studies on Problems (a) and (b) are left as future work. Adopting a single type of display and a single pair of the encoding schemes, we select several content types, though they are not exhaustive, as stated in the next section. Under this condition, we derive equations (namely, regression lines) which express the user–level QoS parameter as a function of reserved bandwidth of audio and that of video for a few contents, using the methods proposed in [10] and [12].

III. Experimental Methodology

This section explains an experimental methodology to examine the effectiveness of GPSQ. We first show an experimental network, which is a simple network with a single bandwidth–controllable router. We then select content types to be assessed. We further describe experimental methods, which consist of two steps: derivation of the representative regression lines for the database at the QoS manager, and assessment of user–level QoS when the session control is performed with the representative regression lines.

A. Experimental network and contents

1) Network configuration: Figure 2 shows the configuration of the experimental network. It consists of three routers and five PC's, which are used as a media recipient, a media sender, a Web client, a Web server, and a controller that works as both SIP server and QoS manager. We have implemented the SIP server and the QoS manager in a single PC for simplicity. Among the three routers (say Routers 1, 2 and 3), Router 2 is a bandwidth–controllable one (Cisco Systems' 7301), while Routers 1 and 3 are ordinary routers (RiverStone's RS3000) which are used as switching hubs. The links between the routers and ones between a router and a PC are all Ethernet channels.

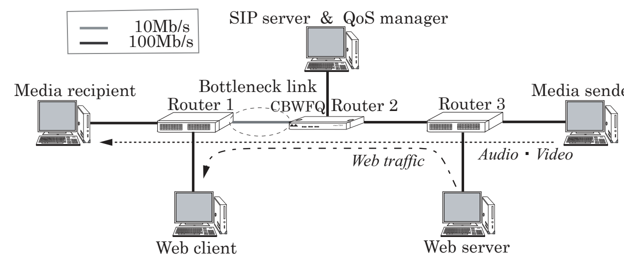

Fig. 2. Configuration of the experimental network

Since we exert the bandwidth control at Router 2, we have set the transmission rates of the links so that the link between Router 1 and Router 2 becomes a bottleneck; the rate is set to 10 Mb/s, while the others are 100 Mb/s.

The media sender transmits a pair of audio and video streams to the media recipient with UDP. We refer to the transmission unit at the application layer as a *media unit* (*MU*); in this paper, we define a video frame as a video MU and a constant number of audio samples as an audio MU. A video MU is usually divided into two or more IP packets, while an audio MU constitutes a single IP packet. For simplicity of discussion as in [12], we do not output a video MU unless all packets of the MU are received correctly.

In the experiment, the media recipient carries out no buffering control on received MU's to absorb their delay jitters. This is because the main purpose of the experiment is to examine the effectiveness of GPSQ; the exercise of buffering control makes it difficult to confirm that the achieved user–level QoS has been supported by GPSQ alone.

As interference traffic for the audio–video streams, Web traffic is transferred from the Web server to the Web client. It is generated according to the configuration of WebStone 2.5 [18]. WebStone is a web server evaluation tool which retrieves files from target web servers continuously.

TABLE I

THE SET OF FILES TO BE RETRIEVED FROM THE WEB SERVER

file name	size [kbyte]	probability
`file500.html`	0.5	0.350
`file5k.html`	5.0	0.500
`file50k.html`	50.0	0.140
`file500k.html`	500.0	0.009
`file5m.html`	5000.0	0.001

Table I. shows the set of files to be retrieved in our experiment. This set contains the same items in `filelist.standard`, which is distributed with WebStone. In this table, `file50k.html`, for instance, means a file of 50 kbytes and retrieved with probability 0.140.

In this experiment, we adopt Apache 2.0 [19] as the Web server. WebStone generates 50 client processes on the PC of Web client. Both Web server and Web client use TCP–NewReno with a window size of 16 kbytes.

In order to reserve the bandwidth, we resort to the CBWFQ packet scheduler at Router 2. CBWFQ classifies traffic flows into a certain number of classes and forms an individual queue for each class. It reserves a certain amount of bandwidth for each class; if the sum of all the reserved bandwidth is smaller than the output link capacity, the remaining capacity is further allocated to each class in proportion to the reserved bandwidth.

A class in CBWFQ usually contains more than one flow. In this experiment, however, we assign the audio flow, video flow and Web traffic flow each to separate classes in order to guarantee the reserved bandwidth for each flow.

2) Content types for assessment: In the experiment, we use three content types: *music video, sports*, and *movies* (*films*). These have been chosen from among the eight types in the VQEG test plan mentioned in Subsection II-D.

B. Derivation of representative regression lines

Let us explain the procedure for the derivation below.

1) Contents: For the three content types selected in the previous subsection, we choose the following contents:

- *music video* (M1): A music video of car chase over a desert with frequent scene change. A group of four males inside a car being chased is singing.
- *sport* (S1): Scenes of a baseball game. A batter is running after he hits a bases–loaded home run. The audio is composed of a commentator's voice and spectators' cheers.
- *movie* (F1): Scenes of a Japanese film. A child and his parents are talking in a room without background music.

The audio in each content has been encoded with linear PCM using 16 bits per sample for two channels of 24 kHz each, which leads to a bit rate at 1.536 Mb/s. The size of an audio MU is set to 9600 bytes. The video in each content has been encoded so that it can have approximately the same average bit rate; the specifications of the encoded video are shown in Table II.

TABLE II

SPECIFICATION OF VIDEO FOR REPRESENTATIVE REGRESSION LINES

contents	music video(M1)	sport(S1)	movie(F1)
encoding scheme	MPEG–1		
image size [pixel]	320×240		
picture pattern	IBBPBBPBBPBBPBB		
average MU size [byte]	10,425	10,417	10,421
average MU rate [MU/s]	30		
maximum bit rate [Mb/s]	3.040	3.222	3.678
average bit rate [Mb/s]	2.502	2.500	2.501
measurement time [s]	15		

2) Stimuli: Using the above three contents, we first make objects to be presented to assessors (i.e., subjects) for user–level QoS assessment. We refer to the objects as *stimuli.*

Choosing a content, we transfer the corresponding audio–video streams from the media sender to the media recipient over the IP network in Fig. 2. In each transmission, we set the reserved bandwidth for audio and that for video at Router 2 to constant values and repeat the transmission by changing the values of the reserved bandwidth. During each transmission, we record the audio–video streams output at the media recipient. The recorded streams are regarded as the stimuli for the content. The stimuli of the other contents can be obtained in the same way.

For simplicity of discussion on the bandwidth reservation, we allocate enough bandwidth to audio to achieve high quality: a capacity of 1.700 Mb/s. Regarding video, we first made a preliminary experiment on a suitable range of bandwidth; that is, at the smallest bandwidth, 60 % of video MUs are dropped, while at the largest no MU is dropped. As a result, we have obtained the ranges of (2.500 Mb/s, 2.740 Mb/s) for music video M1, (2.350 Mb/s, 3.150 Mb/s) for sport S1, and (2.300 Mb/s, 3.740 Mb/s) for movie F1. For the experiment, we divided each range into 16 uniform intervals; this has resulted in 17 values of the reserved bandwidth for video. The Web traffic uses the remaining bandwidth, which the 50 clients share. Thus, the number of the stimuli to be assessed becomes $17 \times 3 = 51$ because of the 17 bandwidth values for each of the three contents.

3) Method of successive categories: We assess the user–level QoS of the stimuli by means of the *method of successive categories*, which is one of the psychometric methods [13] for obtaining the interval scale. The method of successive categories is composed of two steps: the *rating–scale method* and the *law of categorical judgment.*

In the rating–scale method, an assessor classifies the stimuli into a certain number of categories (e.g., five) each assigned an integer (typically 5 through 1 in order of highly perceived quality) [1]. In this paper, we utilized the following five categories of impairment: "imperceptible" assigned integer 5, "perceptible, but not annoying" 4, "slightly annoying" 3, "annoying" 2, and "very annoying" 1, which are referred to as *Category 5* through *Category 1*, respectively.

We put the stimuli in a random order and presented them to 20 assessors, each using a PC with headphones and a 17 inch–LCD display. The assessors are Japanese males at twenties. They were non–experts in the sense that they were not directly concerned with audio and video quality as a part of their normal work. It took about 20 minutes for an assessor to finish all assessment.

From the measurement results by the rating–scale method, the law of categorical judgment can produce the interval scale, where the intervals between the scale values represent differences between amounts of the sensory attribute measured; furthermore, it can provide the values of the category boundaries [12]. Since the law of categorical judgment is based on assumptions, we have to confirm the goodness of fit for the obtained scale. For a test of goodness of fit, we conduct *Mosteller's test* [13]. Once the goodness of fit has been confirmed, we use the interval scale as the user–level QoS parameter and call it the psychological scale.

4) Regression analysis: Finally, we perform *regression analysis* [20] by defining the reserved bandwidth for video as the *independent variable* and the user–level QoS parameter as the *dependent variable*. Note that the reserved bandwidth for audio is not used as an independent variable since it is kept constant. This produces a regression line which estimates the user–level QoS parameter as a function of reserved bandwidth of video. The set of the regression lines thus obtained is registered into the database at the QoS manager.

See [10] and [12] for further details of the procedure above.

C. Assessment of user–level QoS under the control

For each of the three content types, we prepare three contents which are different from the one used in the derivation of the representative regression line for the content type. The new contents are as follows:

- *music video* (M2): A group of four males different from those in M1 is singing and playing in a TV set. Scene changes are smooth compared to M1.
- *music video* (M3): A man is singing a song in the open. Few scene changes occur.
- *music video* (M4): A girl is singing a song while bathing in a lake.
- *sport* (S2): Scenes of a tennis game by two female players. The audio includes strokes, a commentator's voice and spectators' cheers.
- *sport* (S3): Scenes of a soccer game. At the end of the clip, a player scores a goal. The audio consists of a commentator's voice and spectators' cheers.
- *sport* (S4): Scenes of an 800-meter race in the Olympic Games. Eight athletes are running. A commentator's voice and spectators' cheers are heard.

[1] In the calculation of MOS, we make an implicit assumption that the difference in integer between any two successive categories means the same magnitude of the assessor's sensation (e.g., "5 – 4" has the same meaning as "3 – 2"). We then average the integers for a stimulus over all assessors to get a MOS value. Note that the assumption is not necessarily valid. Thus, MOS is an *ordinal scale* in the strict sense [10].

- *movie* (F2): Scenes of a Japanese movie. A young man and a young woman sitting on a park bench are talking to each other within sound of traffic.
- *movie* (F3): Scenes of a Japanese movie. Three young women who are seated at a round–table are chatting.
- *movie* (F4): Scenes of a Japanese period film. Two men are talking to each other; one is fishing beside the other who stands looking on.

The specifications of video in the above nine contents are shown in Table III. The specification of audio in each content is the same as that of contents M1, S1 and F1.

TABLE III

SPECIFICATION OF ENCODED VIDEO FOR USER–LEVEL QOS ASSESSMENT

contents	music video			sport			movie		
	M2	M3	M4	S2	S3	S4	F2	F3	F4
encoding scheme	MPEG-1								
image size [pixel]	320×240								
picture pattern	IBBPBBPBBPBBPBB								
average MU size [byte]	10,446	10,450	10,458	10,433	10,454	10,425	10,421	10,419	10,446
average MU rate [MU/s]	30								
maximum bit rate [Mb/s]	3.166	2.866	3.845	2.929	3.484	2.819	2.881	2.722	3.613
average bit rate [Mb/s]	2.507	2.508	2.510	2.504	2.509	2.502	2.501	2.503	2.507
measurement time [s]	15								

In this paper, we suppose that the user specifies desired user–level QoS by selecting one category, which we call the *target category*, from among the five (namely, Categories 1 through 5). It is usually the case that a higher target category imposes a higher cost (e.g., charge) on the user. In the experiment in this paper, we selected Categories 5, 4 and 3 as the target category, since Categories 2 and 1 are not desirable from a service–offering point of view.

In the experiment on a specified target category, we set the QoS manager so as to reserve the video bandwidth with which the representative regression line predicts to achieve the average of the upper and lower boundaries of the target category if it is Category 4 or 3. For Category 5, whose upper boundary is positive infinity, we utilize a scale value which makes the 95th percentile of the users statistically assign the corresponding stimulus to Category 5.

On the network under this control, we assess the user–level QoS for the new nine contents in the same way as that in Subsection III-B, though we use different assessors from those in the subsection. Thus, the number of stimuli we want to assess in the experiment on the control scheme becomes $9 \times 3 = 27$ because each of the nine contents is assessed for the three target categories, which imply three different values of the reserved bandwidth for video. However, if we present only these 27 stimuli to the assessors, we cannot calculate the category boundaries, since the stimuli contain nothing corresponding to Category 2 or Category 1.

Then, in the same way as that in Subsection III-B, we prepared 17 stimuli for each content by changing the value of the reserved bandwidth for video; some of the 17 stimuli can be classified into Category 2 or Category 1. This leads to $17 \times 9 = 153$ stimuli. As the stimuli to be presented to assessors, we added those stimuli to the 27 target ones to have $153 + 27 = 180$ stimuli.

The number of assessors in this experiment is 33; they are Japanese males and females at ages between 19 and 25. The necessary time for the assessment for an assessor was about 75 minutes.

IV. EXPERIMENTAL RESULTS

A. Representative regression lines

First, let us evaluate the boundaries of the categories. In order to compare the psychological scales for the three contents on the same basis, we applied the law of categorical judgment to all the measurement results of the three contents together, i.e., the 51 stimuli. Utilizing the method of Mosteller's test [13], we tested the goodness of fit. As a result, we found that the test with a significance level of 0.05 cannot reject the hypothesis that the observed value equals the estimated one. Therefore, we consider the interval scale thus obtained as the user–level QoS parameter, i.e., the psychological scale. Setting the minimum value of the psychological scales to unity (i.e., 1), we also obtained the lower boundary of a category as 4.413 for Category 5, 3.288 for Category 4, 2.612 for Category 3, and 1.894 for Category 2. Consequently, the width of a category is not uniform; this implies that the assumption made in calculating MOS is not correct.

Second, we show the psychological scale. For convenience of derivation of representative regression lines at the next step, we introduce the *normalized reserved variable* X as follows:

$$X \triangleq (B_v - m_v)/\sigma_v \tag{1}$$

where B_v is the reserved bandwidth for video, and m_v and σ_v are the average of the video bit rate and its standard deviation, respectively.

We show the psychological scale versus normalized reserved variable for the contents M1, S1 and F1 in Fig. 3, where the lower boundaries of the categories are also plotted as straight lines parallel to the abscissa. In this figure, we observe that as the value of the normalized reserved variable becomes larger, the psychological scale tends to increase for every content.

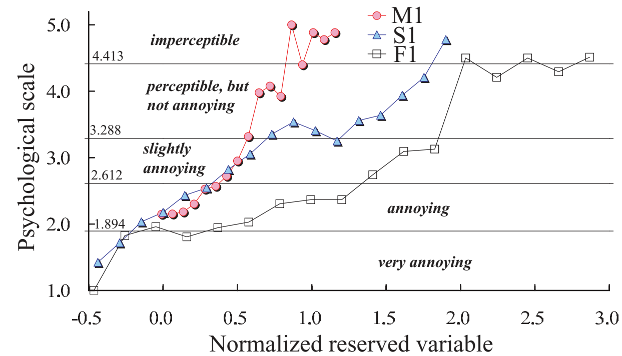

Fig. 3. Psychological scale versus normalized reserved variable

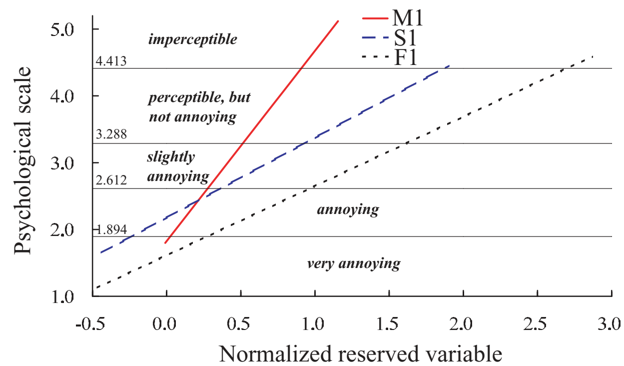

Fig. 4. Estimate of psychological scale versus normalized reserved variable

Next, defining X as the independent variable and the psychological scale for a content as the dependent variable, we perform regression analysis and obtain

$$\hat{I}_M = 1.826 + 2.849\,X \tag{2}$$
$$\hat{I}_S = 2.178 + 1.195\,X \tag{3}$$
$$\hat{I}_F = 1.613 + 1.036\,X \tag{4}$$

where $\hat{I}_M$, $\hat{I}_S$, and $\hat{I}_F$ denote an estimate of the psychological scale for the music video, that for the sport, and that for the movie (film), respectively. The contribution rates adjusted for degrees of freedom of Eqs. (2), (3) and (4) become 0.966, 0.974 and 0.954, respectively. These values mean that the above equations can estimate the psychological scale values with high accuracy.

Figure 4 plots Eqs. (2) through (4). In this figure, we notice that the increasing rate of the psychological scale (i.e., the slope of the line) and the location of the line on the plane depend on the content; the increasing rate of the psychological scale for M1 is the highest, while those for S1 and F1 are comparable. Utilizing this figure, we can calculate the necessary reserved bandwidth for video to achieve a specified value of the psychological scale for a given content type.

B. User–level QoS under the control

Using Eqs. (1) through (4), we can calculate the reserved bandwidth of video for each target category of each content. Table IV shows the calculated reserved bandwidth. With the values in Table IV, we carried out the experiment on assessment of the user–level QoS under the control.

TABLE IV

RESERVED BANDWIDTH FOR TARGET CATEGORY [MB/S]

target category	music video			sport			movie		
	M2	M3	M4	S2	S3	S4	F2	F3	F4
Category 3	2.614	2.632	2.681	2.624	2.842	2.594	2.710	2.611	2.693
Category 4	2.700	2.731	2.818	2.764	3.230	2.702	2.851	2.684	2.820
Category 5	2.864	2.920	3.078	3.031	3.967	2.907	3.118	2.823	3.061

In the same way as that in the previous subsection, we obtained the psychological scales for the target categories of the nine contents. As a result of Mosteller's test, by removing some values, the hypothesis that the observed value equals the calculated value cannot be rejected with a significance level of 0.01; we removed the stimuli of M4 for Categories 4 and 5. The lower boundary of Category 5, that of Category 4, that of Category 3 and that of Category 2 become 4.980, 3.937, 3.058, and 1.876, respectively.

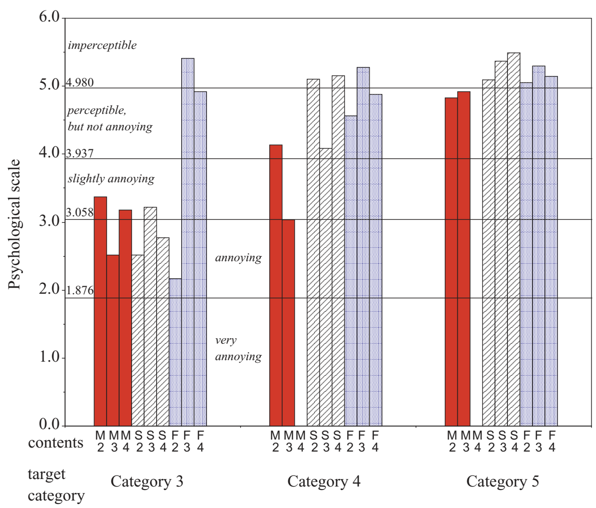

Fig. 5. Experimental results of the control.

We display the obtained psychological scale values for each target category in Figure 5, where we make the following observations. First, for two content types, sport and movie (film), we approximately achieved the desired user–level QoS when the target category is either Category 5 or Category 4. For music video, on the other hand, the psychological scales are less than the desired ones. Second, when the target category is Category 3, we obtained the desired user–level QoS for only M2, M4 and S3. Consequently, for the content types of sport and movie, GPSQ is effective when the users select Category 4 or Category 5 as the target category, For the content type of music video, the control is insufficient.

The reason why the control is ineffective for music video may be the sensitivity of the estimated psychological scale to the reserved bandwidth for video. Note that the slope of the line for music video in Fig. 4 is steep; this implies that even a small estimation error can cause a large discrepancy.

The discrepancy for the target category of Category 3 may be due to the linear approximation for the estimation of the psychological scale.

The accuracy of the estimation is affected by a variety of factors including the classification method of content types, the estimation method of the psychological scale, and the selection of the representative regression lines. These should be for further study.

V. CONCLUSIONS

This paper proposed GPSQ, which is a method of guaranteeing user–level QoS in audio–video transmission over IP networks. We have conducted a simple experiment to examine the effectiveness of GPSQ and have seen that GPSQ can approximately achieve user–level QoS specified by the user for some content types.

Since this paper is a first step of our study on GPSQ, the results of achievable user–level QoS were obtained in a limited situation. For example, the user–level QoS under the control of GPSQ was measured only for three contents in each type. Also, the reserved bandwidth for audio was kept constant during the experiment. We should examine various values of reserved bandwidth for audio so that we can utilize the mutually compensatory property between audio and video [10] for enhancement of user–level QoS.

In order to validate the effectiveness of GPSQ in more common situations, we need to address ourselves to many issues as future work. From among the issues, we first point out the construction of the database of representative regression line, which should accommodate not only stored media as studied in the current paper but also live media in order to support interactive applications. We also have to study methods of bandwidth reservation along an end–to–end path.

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. Ishibashi, "Mutually compensatory property of multimedia QoS," in *Conf. Rec. IEEE ICC2002*, pp. 1105–1111, Apr./May 2002.

[2] IETF RFC 1633, "Integrated Services in the Internet Architecture: An Overview," June 1994.

[3] IETF RFC 2475, "Architecture for Differentiated Services," Dec. 1998.

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

[5] ITU-T Rec. Y.1291, "An architectural framework for support of Quality of Service in packet networks," May 2004.

[6] ITU-T Rec. Y.1540, "Internet protocol data communication service – IP packet transfer and availability performance parameters," Dec. 2002.

[7] ITU-T Rec. Y.1541, "Network performance objectives for IP–based services," May 2002.

[8] ITU-T Rec. G.1010, "End–user multimedia QoS categories," Nov. 2001.

[9] C.-S. Lee, "NGN Focus Group Activities," *NGN Technical Workshop*, Jeju, Korea, Mar. 2005.

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

[11] 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.

[12] 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.

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

[14] IETF RFC 3261, "SIP: Session Initiation Protocol," June 2002.

[15] IETF RFC 2327, "SDP: Session Description Protocol," Apr. 1998.

[16] S. Vegesna, "IP Quality of Service," Cisco Press, 2001.

[17] The Video Quality Experts Group, "The Video Quality Experts Group Web Site," http://www.its.bldrdoc.gov/vqeg/.

[18] Mindcraft Inc., "WebStone Benchmark Information," http://www.mindcraft.com/webstone/.

[19] The Apache Software Foundation, "The Apache HTTP Server Project," http://www.apache.org/.

[20] W. J. Dixon and F. J. Massey Jr, *Introduction to statistical analysis*, McGraw-Hill, N. Y., 1984.