Standardization of Speech Quality Assessment of IP Telephony

Akira Takahashi†, Hitoshi Aoki, and Hideaki Yoshino

Abstract

It is essential to develop methods for evaluating, designing, and managing the quality of service (QoS) of IP (Internet protocol) telephony. Here, we introduce recent standardization activities in ITU-T SG12 (International Telecommunication Union, Telecommunication Standardization Sector, Study Group 12), which is studying the end-to-end transmission performance of networks and terminals, from the viewpoints of QoS assessment and IP telephony management, as well as the transmission characteristics of terminals.

1. Achieving better service quality

Since IP (Internet protocol) telephony services exploit IP networks that do not necessarily guarantee the transmission performance, it is extremely important to manage the quality of service (QoS) in addition to designing the networks and terminals. To do this, it is essential to develop methodologies that quantitatively evaluate users' perceptions of services. These are called quality assessment methodologies. **Figure 1** shows their relationships.

† NTT Service Integration Laboratories
Musashino-shi, 180-8585 Japan
E-mail: takahashi.akira@lab.ntt.co.jp

2. Quality assessment methodologies

The QoS of IP telephony should be discussed in terms of subjective quality, which corresponds to users' perceptions of transmitted speech. However, subjective quality assessment is time consuming and expensive, so a method that estimates subjective quality by measuring physical characteristics of transmit-

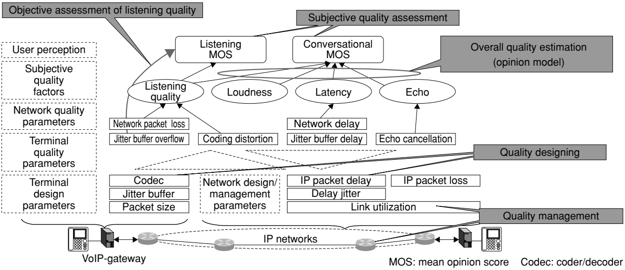

Fig. 1. Study items for QoS of IP telephony services.

ted speech signals is desirable. This is called objective quality assessment. To monitor the QoS of IP telephony in service, a systematic tool for estimating subjective quality is indispensable. The main degradations that affect the QoS of IP telephony are coding distortion, packet loss, non-optimal loudness, delay, and talker echo.

An objective assessment method for loudness has already been established and the methodology, which is called the Loudness Rating, has been standardized as ITU-T Recommendation P.79. The delay time and talker echo can be measured objectively; their impact on subjective quality has been thoroughly studied and the guidelines have been standardized as ITU-T Recommendations G.114 and G.131, respectively. Therefore, it is possible to estimate the effects of these factors on subjective quality by physical measurement.

On the other hand, the effects of coding distortion and packet loss depend on the coding and packetizing schemes and device implementation, so it is difficult to estimate the subjective quality from the codec type and observed packet loss rate. Therefore, an objective speech quality measure that quantifies the amount of distortion caused by coding and packet loss and estimates subjective listening quality has been studied.

SG12 standardized Recommendation P.862 "PESQ (perceptual evaluation of speech quality)" in 2001 and it is currently developing new methodologies based on the following viewpoints.

2.1 Objective quality assessment via acoustic interfaces

Current Recommendation P.862 uses electrical interfaces with devices being tested, assuming the normal electroacoustic characteristics of handsets. However, when evaluating hands-free terminals, it is important to take into account the real electroacoustic characteristics of the terminal in an objective quality assessment because such characteristics vary a lot depending on the terminals. To do this, an objective quality assessment methodology using acoustic interfaces has been studied. Currently, one candidate algorithm has been proposed. At the last SG12 meeting in September 2003, it was pointed out that the scope of the new recommendation should be carefully reconsidered because it has some overlap with Recommendation P.862. Therefore, after revising the scope, SG12 will reissue its invitation for candidate algorithms.

2.2 Single-ended objective quality assessment

According to Recommendation P.862, one must provide original speech as well as degraded speech when determining the amount of speech distortion. However, it is often impossible to obtain original speech for use in objective quality assessment for in-service testing. Therefore, it is desirable to develop a method that uses only degraded speech for estimating speech quality.

After the competition between two candidate algorithms, one of them was selected to proceed for final approval. The draft Recommendation, which is currently called P.SEAM (single-ended assessment method), will be submitted for consent at the next meeting in March 2004.

The above-mentioned methodologies estimate the "listening" quality of IP telephony. In addition to speech distortion, which is evaluated as listening quality, there are conversational quality factors such as delay and talker echo. Therefore, to design the overall quality of IP telephony, we need a methodology that estimates overall quality based on various quality parameters, which is called the "Opinion Model". SG12 has already standardized an opinion model called the "E-model" as Recommendation G.107 and is continuing to study ways of improving its estimation accuracy and extending its scope to wideband (7-kHz) speech and/or a hands-free environment.

At the last meeting, some contributions pointed out a problem in the additive property assumed by the E-model. It was agreed to continue the study on modifying the current E-model. This may result in an update to Recommendation G.107. Since the E-model has been widely used as a quality evaluation tool for IP telephony in Japan, such revisions in SG12 should be reflected in Japanese standards, such as TTC (Telecommunication Technical Committee) standard JJ-201.01, in the future.

3. Quality management

To monitor and manage the quality of IP telephony during service, we need a method that estimates subjective quality based on metrics that can be collected from network and terminal devices. SG12 has been studying a method that estimates subjective listening quality based on information obtained from the RTP (realtime protocol) header. It should output intermediate parameters describing the packet-loss pattern and packet-loss rate. These parameters form a subset of those in RTCP-XR (RTP control protocol - extended report) being proposed in IETF (Internet engineering task force). Therefore, if RTCP-XR is standardized and network and/or terminal devices implement it, the algorithm will be able to estimate subjective quality more accurately by directly using it. Currently, two candidate algorithms are being compared. The selection phase is expected to be completed by the end of 2004 and a new draft Recommendation, currently called P.VTQ (voice transmission quality), will be submitted for SG12 consent at the next meeting.

4. Transmission characteristics of IP-telephony terminals

ITU-T SG13, which is studying "multiprotocol and IP-based networks and their internetworking", has standardized Recommendation Y.1541, which defines several classes of IP network performance. However, as shown in Fig. 1, the end-to-end quality of IP telephony depends heavily on terminal characteristics. Therefore, it is important to give guidelines for designing terminal devices.

For this purpose, SG12 has been trying to determine the transmission characteristics of IP telephony terminals. The primary factors in the scope will be delay, talker loudness rating, and echo. Other factors, such as frequency response and side tone, should follow the existing Recommendations for conventional telephones. The draft new Recommendation, which is currently called "P.VoIP (voice over IP)," will be submitted for SG12 consent at the next meeting.

5. Standardization in Japan

Finally, let us introduce standardization activities in Japan. Based on the report of the Study Group on IP network technology, the MPHPT (Ministry of Public Management, Home Affairs, Posts and Telecommunications) amended "Rules for telecommunications numbers", "Regulations for telecommunications facilities for telecommunications business", and other ministerial ordinances in 2002. The amendment determined the requirements for speech quality of IP telephony services based on their R-value, which is the overall transmission rating derived using ITU-T Recommendation G.107 "E-model." In accordance with this, TTC standardized the quality assessment methodologies based on the E-model as TTC standard JJ-201.01. In addition, CIAJ (Communications and Information Network Association of Japan) provided guidelines for IP telephony terminals as CES-Q003-1.

6. Future activities

The activities of SG12 are shifting towards the standardization of quality assessment methodologies of multimedia applications over IP networks. NTT continues to research the quality assessment, design, and management of multimedia services and will contribute to international standardization bodies, including SG12.

Akira Takahashi
Senior Research Engineer, Service Assessment Group, NTT Service Integration Laboratories.
He received the B.S. degree in mathematics from Hokkaido University, Hokkaido in 1988 and the M.S. degree in electrical engineering from the California Institute of Technology in the U.S.A. in 1993. He joined NTT Laboratories in 1988 and has been engaged in the quality assessment of speech and audio telecommunications. Currently, he is primarily working on the quality assessment of speech over IP networks.

Hitoshi Aoki
Research Engineer, Service Assessment Group, NTT Service Integration Laboratories.
He received the B.E. and M.E. degrees in electrical and electronic engineering from Utsunomiya University, Tochigi in 1992 and 1994, respectively. He joined NTT Laboratories in 1994 and has been engaged in the quality assessment of speech and video telecommunications. Currently, he is primarily working on the quality assessment of speech over IP networks.

Hideaki Yoshino
Group Manager, Service Assessment Group, NTT Service Integration Laboratories.
He received the B.S. and M.S. degrees in information science from Tokyo Institute of Technology, Tokyo in 1983 and 1985, respectively. In 1985, he joined NTT Electrical Communications Laboratories and has been researching traffic design and control of communications systems. From 1990 to 1991, he was a visiting scholar in the Faculty of Electrical Engineering, University of Stuttgart, Germany. He is currently researching traffic and service quality for future communications services.