# Quality Assessment of Wideband Speech Communication Services

*Akira Takahashi*†*, Atsuko Kurashima, and Hitoshi Aoki*

**Abstract**

This article describes a method for estimating the subjective quality of wideband (7-kHz) speech solely from physical characteristics of terminals and networks. It enables efficient quality design and management of wideband speech communication services. The validity of this model was shown through experiments.

## 1. Introduction

Wideband speech communication services, which utilize speech with a wider bandwidth (e.g., 7 kHz) than conventional telephone services, are one of the most promising applications over IP (Internet protocol) networks. This is because the quality of IP-telephony services using telephone-band speech has achieved almost the same level as conventional PSTN (public switched telephone network) services and it is desirable to further enhance the quality of service (QoS) so that users can obtain the benefits of broadband IP network services.

Several factors are accelerating the development of wideband speech communication services. First, ITU (International Telecommunication Union) has standardized several wideband speech codecs such as Recommendations G.722, G.722.1, and G.722.2 for use in telecommunications. In addition, broadband IP network services such as “B FLET’S” provided by NTT have popularized IP-telephony services, in which speech codecs are transparent regardless of the speech bandwidth. Moreover, telephone terminals implemented on personal computer platforms, known as “softphones”, remove the limitation caused by telephone-band handset equipment with respect to speech bandwidth. Wider speech bandwidth is also expected from the viewpoint of greater speech intelligibility for hands-free communications.

To provide high-quality wideband speech communication services, it is important to develop quality assessment methodologies that appropriately quantify users’ perceptions of the services and to apply them to quality planning and management of networks and terminals [1]. This article describes a wideband extension of the quality estimation model that we previously developed for telephone-band speech communications [2]. We demonstrate its performance in terms of the consistency between quality estimated by the model and actual subjective quality derived from a subjective quality assessment.

## 2. Objective quality assessment of speech communication services

The quality of speech communications should be discussed in subjective terms. However, since subjective quality assessment is time-consuming and expensive and requires special experimental facilities such as acoustically shielded chambers, we need a method for estimating subjective quality from the physical characteristics of terminals and networks. This is called “objective quality assessment”.

Objective quality assessment methodologies can be categorized into several groups from the viewpoints of aim, measurement procedure, input information, and type of subjective quality, which is called mean opinion score (MOS), to be estimated [1]. For example, objective quality assessment methodologies that exploit network and terminal quality parameters and produce estimates of conversational MOS are called “opinion models”. Opinion models are very useful

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

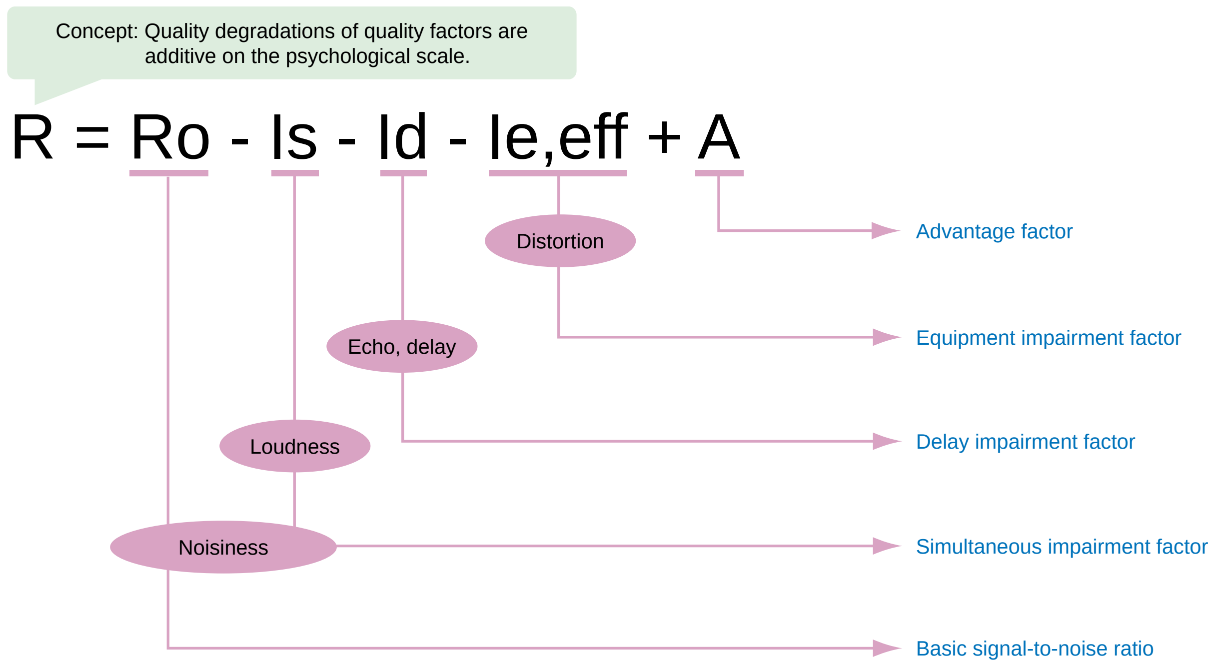



Fig. 1. Concept of the E-model.

because they indicate the overall quality taking into account the conversational quality factors such as delay and echo as well as the listening quality factors such as coding distortion and packet-loss degradation.

ITU standardized an opinion model called the "E-model" in 1998 [3]. Its concept is shown in **Fig. l**. The E-model is a computational model based on the fundamental concept that the degradations in individual quality factors are additive on the psychological scale. Each quality factor is evaluated by applying associated input quality parameters to predetermined functions. Currently, the E-model has 21 input parameters. The overall quality is expressed by "transmission rating factor: R".

In Japan, the Telecommunication Technology Committee (TTC) adopted the E-model as a quality assessment method for telephone-band IP-telephony services and standardized TTC Standard JJ-201.01 in 2003 [4]. JJ-201.01 prescribes how to use the E-model in the evaluation of IP-telephony services as well as other supplementary quality assessment methods such as ITU-T Recommendation P.862 "PESQ" [5]. This standard is referred to when IP-telephony service providers apply for telephone numbers for their services in Japan.

## 3. Quality assessment method for wideband speech

We previously developed an objective quality assessment method specifically for telephone-band speech communication services [2]. This method is applicable not only to quality planning but also to quality benchmarking and management.

From the viewpoint of comparing the quality of wideband services with that of telephone-band ones, it is very important to evaluate both services on the same quality scale. Taking this into account, we extended our previous method to make it applicable to both telephone-band and wideband speech.

### 3.1 Model

The algorithm of our model for estimating subjective conversational quality is shown in **Fig. 2**. It is applicable to conventional telephone-band speech as well as wideband speech. There were four important points in its development:

1) modeling the effects of delay and echo,
2) taking into account the interaction between delay and speech distortion,
3) quantifying the quality enhancement achieved by widening the speech bandwidth, and
4) establishing a subjective quality database for various wideband speech codecs.

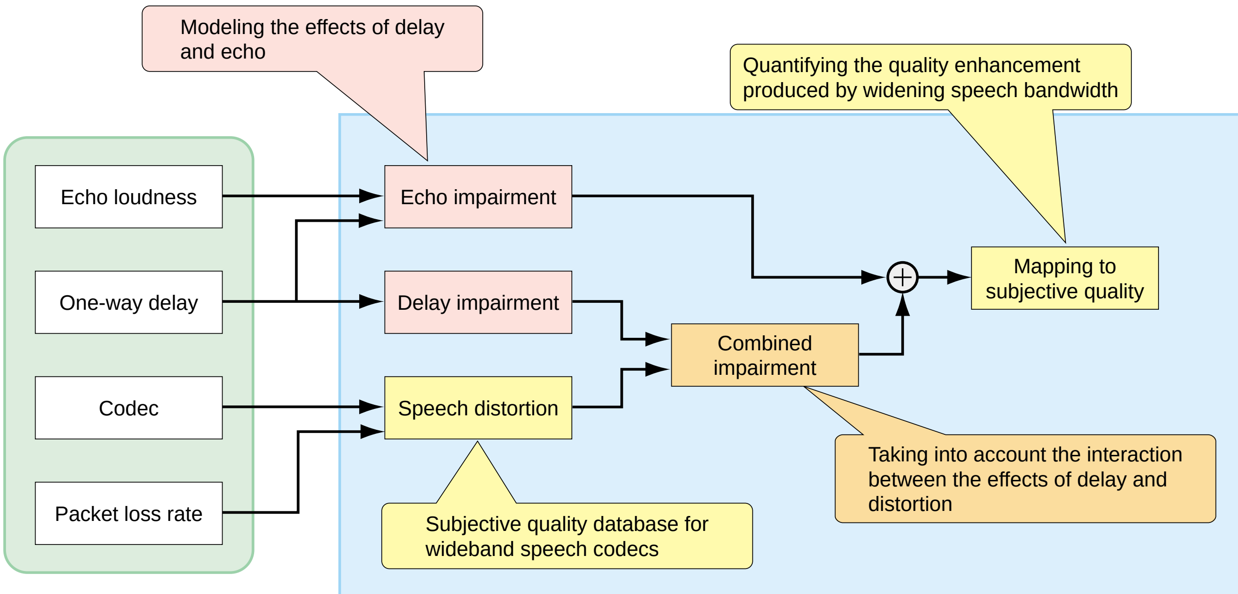



Fig. 2. Algorithm.

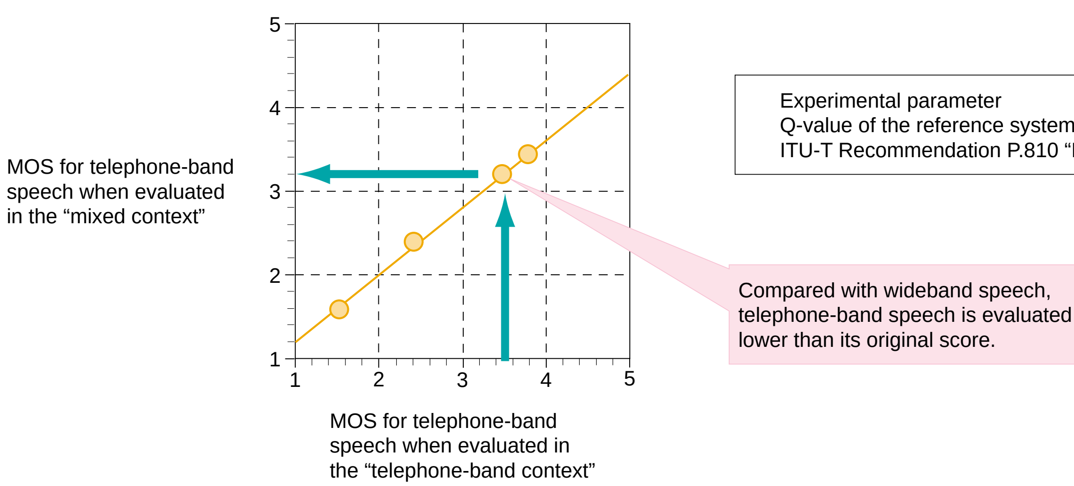



Fig. 3. Change in MOS for telephone-band speech.

Items 1) and 2) were studied in our previous investigation for telephone-band speech [2]. Items 3) and 4) are important, especially from the viewpoint of evaluating wideband speech.

To quantify the quality enhancement achieved by widening the speech bandwidth (Item 3), we carried out two subjective quality experiments. In the first experiment, only telephone-band speech was evaluated, while in the second experiment, both telephone-band and wideband speech were evaluated in the same context. By comparing the MOSs for telephone-band speech obtained in the two experiments, we found that the MOS was lower in the mixed context, in which both wideband and telephone-band speech were evaluated, because users tended to implicitly penalize telephone-band speech due to the band limitation (**Fig. 3**). For instance, a MOS of 3.5 in the telephone-band context corresponds to a MOS of 3.2 in the mixed telephone-band and wideband context.

In addition, we established a subjective quality database for various wideband speech codecs standardized in ITU, which are G.722, G.722.1, and G.722.2 (**Table 1**). The quality of speech codecs is

Table 1. List of wideband speech codecs.

| Codec | Bitrate (kbit/s) | Packet-loss pattern |
|---|---|---|
| G.722 | 64 | Random/bursty |
| G.722.1 | 32 | |
| | 24 | |
| G.722.2 | 6.6, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05, 23.85 | |

characterized by two parameters: one is *Ie*, which represents the distortion caused by speech coding, and the other is *Bpl*, which represents the robustness of a codec against packet loss degradation. We determined the *Ie* and *Bpl* values based on ITU-T Recommendation P.833 [6]. The *Bpl* values were derived for both random and bursty packet-loss patterns.

### 3.2 Validation of model

We investigated the validity of the model by comparing the subjective MOS obtained in actual subjective experiments with the quality estimated by the model. The subjective experiments were carried out in a conversational manner. The experimental parameters were overall loudness rating (OLR), codec, packet-loss rate, one-way end-to-end delay, and talker echo loudness rating (TELR). The results for wideband and telephone-band speech are shown in **Fig. 4**. Here, the estimated MOS for telephone-band speech represents the MOS in the telephone-band context.

From this figure, we can say that the model predicts the subjective quality with high correlation not only for wideband speech but also for telephone-band speech. That is, our model is applicable to the evaluation of loudness, delay, echo, and speech distortion, which are the primary quality factors in IP-telephony services.

Because our method gives a unique quality index, which is the MOS in the mixed context, it is possible to compare the quality of a newly developed wideband speech communication service with that of conventional telephone-band services.

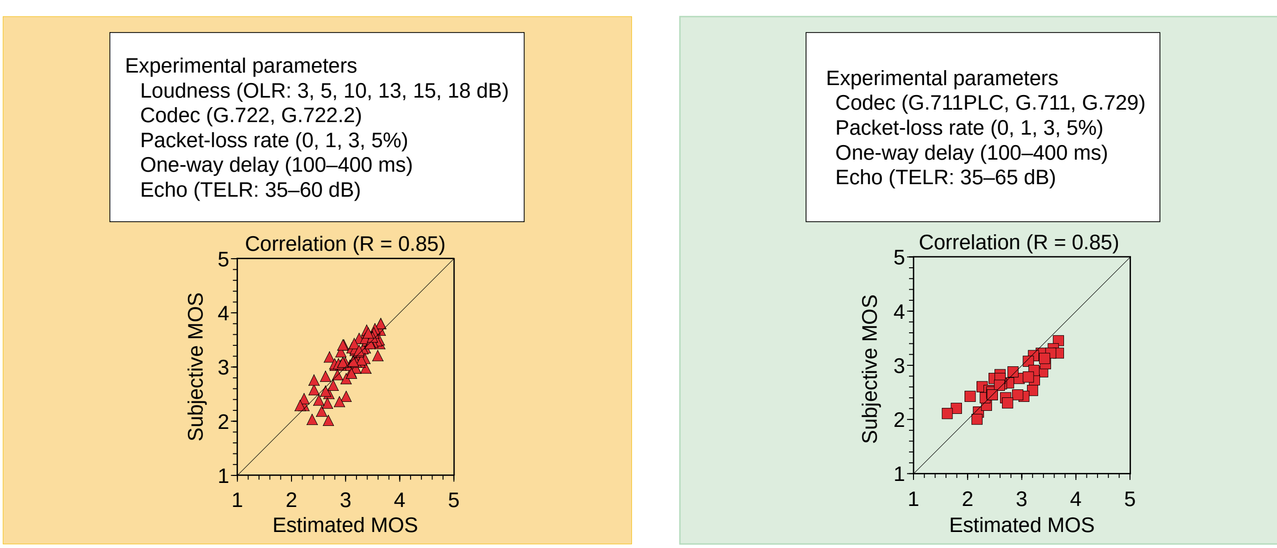
## 4. Conclusion

In this article, we described a quality estimation model for wideband speech that can be applied to conventional telephone-band speech. Experimental results showed that its estimates correlate very well with the subjective quality, which represents actual users' perceptions of services. This model is a powerful tool for designing and managing the quality of wideband speech communication services taking into account users' opinions. We intend to further enhance this model from the viewpoint of evaluating hands-free communications.

## References


[1] A. Takahashi, H. Yoshino, and N. Kitawaki, "Perceptual QoS assessment technologies for VoIP," IEEE Commun. Mag., July 2004.

[2] A. Takahashi, "Opinion Model for Estimating Conversational Quality of VoIP," IEEE ICASSP'04, May 2004.


(A) Wideband (B) Telephone-band

Fig. 4. Relationship between estimated and subjective MOS values.

[3] ITU-T Recommendation G.107, “The E-model, a computational model for use in transmission planning,” Mar. 2005.

[4] TTC Standard JJ-201.01, “A method for speech quality assessment of IP telephony,” June 2005.

[5] ITU-T Recommendation P.862, “Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” Feb. 2001.

[6] ITU-T Recommendation P.833, “Methodology for derivation of equipment impairment factors from subjective listening only tests,” Feb. 2001.



**Akira Takahashi**

Senior Research Engineer, Supervisor, 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, U.S. in 1993. He joined NTT Laboratories in 1988 and has been engaged in research into the quality assessment of speech and audio telecommunications. Currently, he is primarily working on the quality assessment of speech over IP networks. He has been contributing to ITU-T SG12 since 1994. He received the Telecommunication Technology Committee Award in Japan in 2004 and the ITU-AJ Award in Japan in 2005. He is a member of IEEE and the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan.


**Atsuko Kurashima**

Research Engineer, Service Assessment Group, NTT Service Integration Laboratories.


She joined Musashino Electrical Communication Laboratories, Nippon Telegraph and Telephone Public Corporation (now NTT) in 1980. She is currently researching objective quality assessment of telephony speech. She received the OHM Technology Award in Japan in 2005. She is a member of IEICE.


**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. He has been contributing to ITU-T SG12 since 2003. He is a member of IEICE and the Acoustical Society of Japan.