

Perspectives on Multimedia Quality Prediction Methodologies for Advanced Mobile and IP-based Telephony

Nobuhiko Kitawaki

University of Tsukuba

1-1-1, Tennoudai, Tsukuba-shi, 305-8573 Japan.

E-mail: kitawaki@cs.tsukuba.ac.jp

URL <http://www.mmlab.is.tsukuba.ac.jp/~kitawaki/>

ABSTRACT

This paper describes the author's perspective on multimedia quality prediction methodologies for advanced mobile and IP-based telephony, and reports related experiments and trials. First the paper describes the need for perceptual QoS assessment in which various quality factors in advanced mobile and IP-based telephony are analyzed. Then an objective quality prediction scheme is proposed from the viewpoints of quality measurement tools for each quality factor and an opinion model for compound quality factors in mobile and IP-based communications networks. Finally, the author's current trials of measurement tools and opinion models are described.

Keywords: perceptual QoS, quality prediction, opinion model, wideband speech, multimedia, hands-free, speech recognition and synthesis

1. Introduction

The perceptual QoS for advanced mobile and IP-based telephony is very important today from the viewpoints of both customers' satisfaction and providers' management operations. Quality assessment for telephone-band communication has become quite well established, as evidenced by the number of publications and recommendations available [1]. However, advanced mobile and IP-based telephony using PCs is entirely different from conventional telephone-band communication, due to the requirement for hands-free and multi-modal communication, which lead in turn to the need for wideband speech and multimedia.

Mobile communication technology has progressed much, and is expected to be applicable to vehicle communication, which requires hands-free operation. IP-based telephony via the Internet using PCs also requires hands-free communication. In both cases, wideband speech has become increasingly necessary, since hands-free communication using separate microphones and loudspeakers is indispensable, and in a hands-free situation wideband speech is particularly

helpful in enhancing the naturalness of communication.

It is very important to assess perceptual speech, audio, and video quality in multimedia services for optimum communications system design, and effective transmission planning and management to satisfy customers' requirements. This paper describes the author's perspective on, and current work on trials of, multimedia quality prediction methodologies for advanced mobile and IP-based telephony, which are derived from proposals to the ITU-T Study Group 12 for telephone-band digital communications which the author has made since 1981 up to now.

First, the paper describes the need for perceptual QoS assessment, in which various quality factors in advanced mobile and IP-based telephony are analyzed. In the subjective QoS assessment method, the perceptual QoS can usually be assessed by an opinion rating method. However, subjective assessment by opinion rating is time-consuming and expensive. Therefore, the author aims to establish an objective QoS assessment methodology that correlates well with subjective QoS.

An objective quality prediction scheme is then proposed from the viewpoints of quality measurement tools for each quality factor and an opinion model for compound quality factors in mobile and IP-based communications. In particular, this paper describes objective QoS assessment for wideband coded speech distortion including discontinuous impairments such as packet loss, from the viewpoint of diagnostics, and it also describes an opinion model for transmission planning and monitoring from the viewpoint of various compound quality factors affecting perceptual QoS including wideband speech and video.

Finally, the author's current trials of measurement tools and opinion models are described. These include a wideband speech quality measure and artificial voice for 7 kHz-wideband speech, an objective audio quality measure for CD music with varied bandwidth, objective quality assessment methodologies for noise reduction algorithms in speech recognition and synthesis systems, echo canceller algorithms for hands-free communication, and comprehensive opinion models for audio-visual communications.

2. Quality Factors in Advanced Mobile And IP-based Telephony

This section describes the need for perceptual QoS assessment in which various quality factors in advanced mobile and IP-based telephony are analyzed, especially: a) wideband speech, audio, and video for multimedia, b) noise reduction and speech recognition-synthesis for hands-free communication, and c) packet loss, acoustic echo, and pure delay for IP-based communications.

Reasonable and appropriate QoS assessment methodologies are required for QoS design and QoS management of advanced mobile and IP-based communications systems from the viewpoints of both the customer's satisfaction and the provider's management operations. From the customer's side, perceptual QoS can be measured by a subjective assessment method using the customer's judgment. From the provider's side, an objective QoS assessment that correlates well with subjective QoS is desired in order to estimate the customer's opinion, as a factor in the design and management of the communications terminals and networks. Therefore objective QoS assessment methodologies may be categorized into: diagnostic assessment tools for terminals and networks; transmission planning tools for communications system design; and in-service and non-intrusive monitoring tools

for communication service management. Of course, it is desired that the estimated QoS is equivalent to the QoS perceived QoS according to the customers' judgment.

Perceptual QoS is affected by many factors in telecommunications systems. Fig.1 shows author's approach to research on objective QoS measurement [2-4]. First, objective QoS measurement methodologies for individual QoS factors are studied, and then an opinion model for compound QoS factors in the communications systems is established.

Fig.2 shows QoS factors affecting perceptual QoS of the advanced mobile and IP-based communication. These factors are featured by multimedia such as wideband speech, audio and video, hands-free quality factors such as noise reduction for speech recognition-synthesis and acoustic echo, and new quality factors due to IP-based network such as packet loss and circuit echo.

Advanced mobile and IP-based telephony is developed by the advent of new coding technologies for wideband speech, audio, and video media, and new transmission technologies based on mobile and IP-based networks. Distinguishable QoS factors affecting perceptual QoS for multimedia services under modern communications networks are coding distortion, delay, echo, and discontinuous distortion such as packet loss, bit error, and frame erasure, etc.

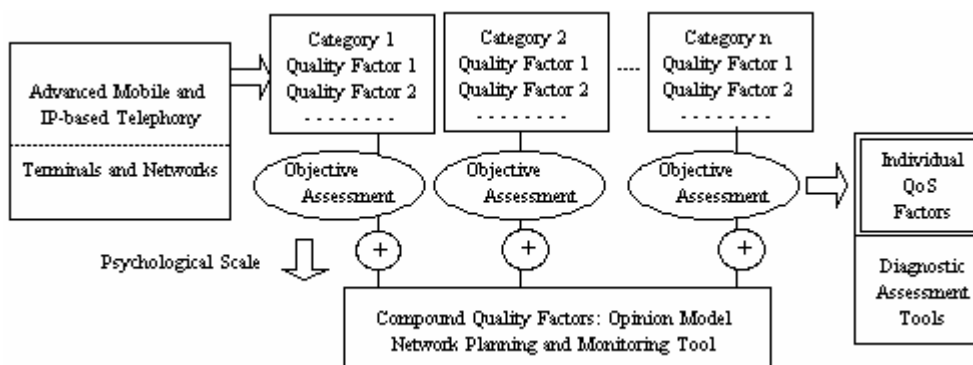


Fig.1 Approach to objective QoS Measurement research in advanced mobile and IP-based telephony.

3. Objective Quality Prediction Scheme

An objective quality prediction scheme is proposed from the viewpoints of quality measurement tools for each quality factor and an opinion model for compound quality factors in mobile and IP-based communications networks. The measurement tools are categorized into a speech-layer objective model and a packet-layer objective model. The opinion model is based on an assumption that all the factors' contributions to quality degradation may be summed on a psychological scale.

3.1 Opinion Model for Compound Quality Factors

Perceptual QoS is affected by various quality factors in telecommunications systems. The author has proposed an opinion model for estimating the MOS from multiple

quality factors in telephone-band communications, which is called OPINE (overall performance index model for network evaluation) [2, 3]. The concept of OPINE is as follows. Relevant quality factors are categorized into several groups according to their quality features, and transformed into internal expressions such as loudness model, Bark spectrum and psychological scale derived from an auditory-psychological process. The MOS can be estimated by summing the psychological performance index (PI) for each group. The OPINE model is described in Supplement 3 of the P-series recommendation in the CCITT Blue Book [4].

Fig.3 shows the concept of an opinion model for advanced mobile and IP-based telephony derived from the OPINE model for conventional telephone-band communications.

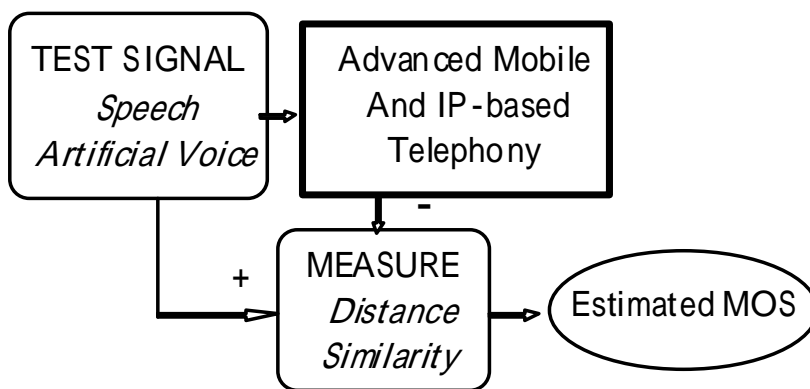


Fig.4 Proposed model for diagnostic objective measurement.

4. Current Trials for QoS Measurement Tools And Opinion Models

4.1 Objective Quality Assessment of Wideband Speech

4.1.1 Wideband Speech Quality Measure

Recommendation P.861 “Objective quality measurement of telephone-band (300-3400 Hz) coded speech” using the PSQM (Perceptual Speech Quality Measure) was standardized in 1996 [6]. PSQM compares the original (input) signal with the degraded output of the device under test using a perceptual model. The key to this process is the transformation of both the original and degraded signals into an internal representation (loudness model and Bark spectrum) that is analogous to the psychological representation of audio signals in the human auditory system, taking account of perceived frequency and loudness. Based on the concept of PSQM,

a new recommendation P.862 “Perceptual Evaluation of Speech Quality (PESQ)” was standardized in 2000, for extending the scope to include IP-based impairment factors.

Wideband-PESQ conforming to proposed draft Recommendation P.862 has been verified from the viewpoint of consistency between subjectively evaluated MOS and objectively estimated MOS, as shown in Fig.5 [7, 8]. The codecs used in this experiment were G.722 “SB-ADPCM,” G.722.1 “MLT,” and G.722.2 “AMR-WB.” Experimental results showed that the correlation between them is strong, and the RMSE is relatively small. It was concluded that Wideband-PESQ is a promising measure for the objective quality assessment of wideband-speech coding.

Objective MOS

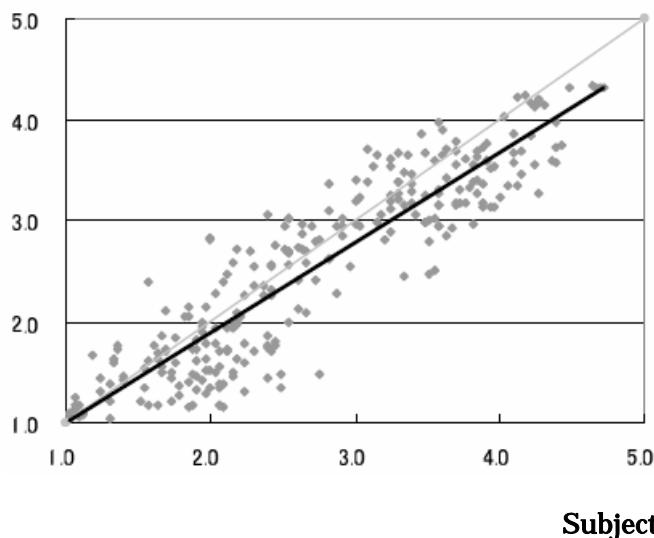


Fig.5 Relationship between objective MOS using wideband PESQ and subjective MOS.

4.1.2 Test signal for Wideband Speech Quality Prediction

We have proposed use of an artificial voice as the test signal for the objective quality assessment of coded speech, because a real speech signal depends on the talker, and a conventional sinusoidal signal is insufficient for evaluating low bit-rate coding [9]. Fig.6 shows the method of generating the artificial voice which has a wide frequency band. The artificial voice reflects the average characteristics of the spoken language such as long-term average spectra, instantaneous amplitude distribution, level distribution of segmental power, spectral distribution of segmental power, voiced/unvoiced structure of the speech waveform, and short-term spectral characteristics. It was standardized as Recommendation P.50 in 1988.

Recent studies showed that the artificial voice can be applied to objective quality measurement for not only for newly developed CELP-type coding and packet loss [10] but also for wideband speech coding using W-PESQ as the objective quality measure [7, 11]. Therefore, the author proposed use of wideband-PESQ and artificial voice for the objective quality measurement of wideband speech coding in ITU-T Study Group 12 [8, 10]

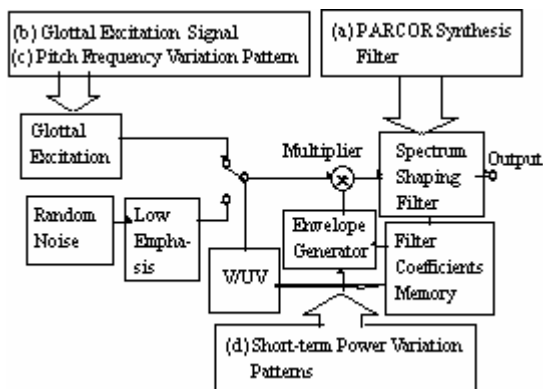


Fig.6 Method of generating the artificial voice

4.2 Objective Measurement of Hands-free Quality Factors

4.2.1 Test Signal used for the Measurement of Acoustic Echo Characteristics

Many test signals are described in Recommendation P.501. However, it is not clear what test signal is appropriate for use in measuring residual acoustic echo characteristics in hands-free telecommunications. The author compared the performance of various test signals for the measurement of residual acoustic echo characteristics expressed as echo return loss enhancement (ERLE) [12]. The signals were real voice, as a reference, white noise, frequency weighted noise, artificial voice, and a composite source signal. Fig.7

shows an example of the ERLE characteristics of artificial female voice (AV-F) and real voice (RV). From the comparative assessment, the ERLE characteristics measured using artificial voice conforming to P.50 were found to be almost equivalent to those of real voice and the most accurate among the test signals evaluated. It was concluded that test signals according to P.50 are satisfactory for the measurement of residual echo characteristics. Therefore, the author proposed the use of P.50 for the measurement of acoustic echo characteristics to ITU-T Study Group 12 [13].

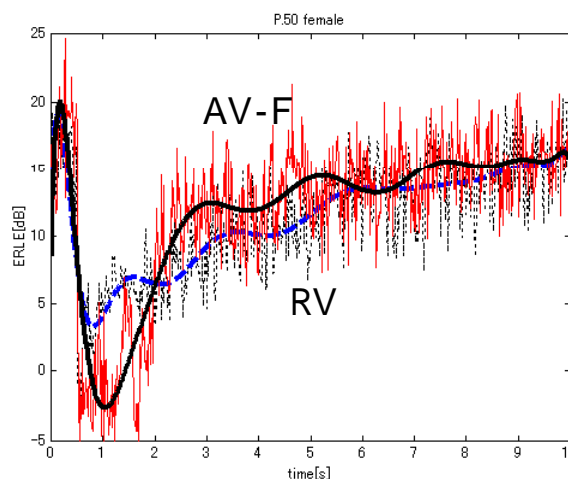


Fig.7 ERLE characteristics measured by AV-F and RV.

4.2.2 Performance Prediction of Noise Reduction Algorithm

In vehicle communication and PC-PC communication, noise reduction technology is introduced to suppress the ambient noise. The author proposed a new application of PESQ and artificial voice conforming to Recommendations P.862 and P.50, which is a methodology for the performance estimation of noise reduction algorithms used for noisy speech recognition [14, 15]. For this purpose, recognition experiments using four noise reduction algorithms were performed on the AURORA-2J connected digit speech recognition task. The algorithms were spectral subtraction, temporal domain SVD-based speech enhancement, GMM-based speech estimation, and KLT-based comb-filtering. The training methods used for the algorithms were clean training and multi-condition training.

An example of the relationship between the word accuracy (percentage) and the PESQ score for a number of different noise levels in dB, and the appropriateness of the artificial voice for calculating the PESQ score are shown in Figs.8 and 9 respectively. These results confirmed that there is a strong correlation between the word accuracy and the PESQ score calculated from the real speech and the artificial voice. This method was

proposed to Study Group 12 [15].

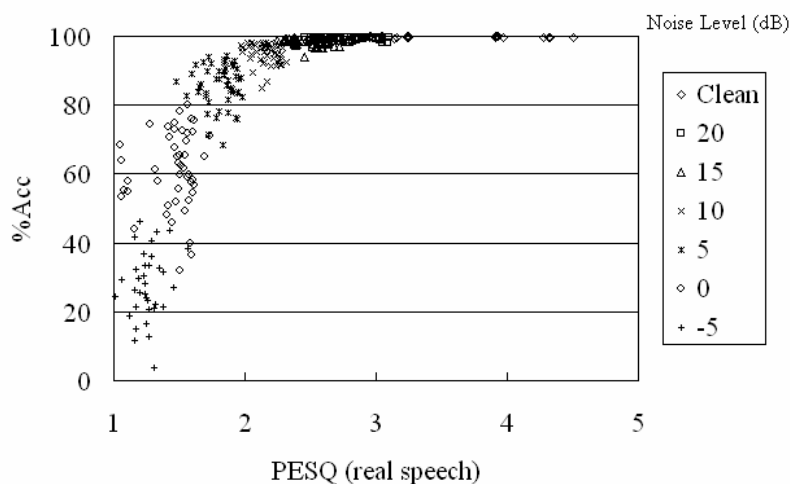


Fig8 Relationship between the word accuracy and the PESQ score in multi-condition training.

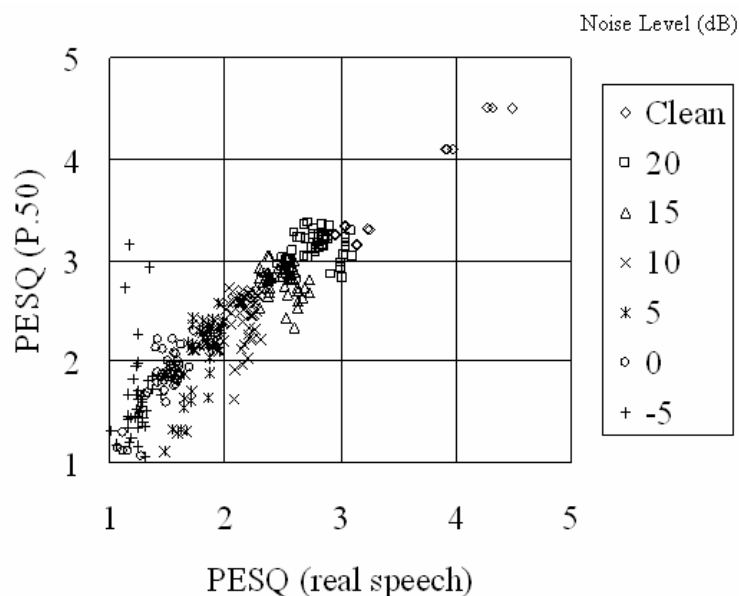


Fig.9 Relationship between the PESQ scores calculated from real speech and the artificial voice.

4.3 Multimedia in Multi-modal Systems

4.3.1 Optimum Quality Design of Audio Coding at Fixed Low Bit-Rate

Recently, high quality and low bit-rate audio coding have been developed, and used in network services such as the online distribution of music, and the playing of music on PCs. The author proposed an optimum method of design of audio coding at a fixed low bit-rate, based on choosing

the most appropriate bandwidth for encoding the audio samples and taking into account the effects of coding distortion and the limited frequency bandwidth [16]. Also, an objective quality measurement algorithm was proposed, as shown in Fig.10, which features arbitrary selection of the bandwidth. It was concluded that it is possible to select the most suitable bandwidth without relying on the designer's subjective intuition, which is obviously not always reliable.

4.3.2 Wideband, Multi-modal, and Multi-party

Fig.11 shows the framework of quality assessment research for next-generation communications services. The key words for such services will be “wideband,” “multi-modal,” and “multi-party.” As the bandwidths of core and access networks rapidly become broader, telecommunications applications have more bandwidth available for speech, audio, and video data, and this leads to higher-quality services. Assessing the quality of such services based on the simple opinion ratings that have been used for the PSTN and for VoIP might be insufficient. In consequence, we need to compare services not only on a one-dimensional scale, like the MOS, but also on a multi-dimensional scale that characterizes the QoS in a way that takes the richness of the services into account. There have been some studies of psychological factors that affect high-quality audio and video [17]. In evaluating the quality levels of multi-modal services, we need to take into account the interaction between media, as well as the quality of individual media, into account. The conventional targets of quality assessment have been one-to-one communications and one-way content-delivery services. IP telephony and video-streaming services are typical examples. With the advent of high-speed wired and wireless networks, multi-party communications services, such as instant messaging, teleconferencing, and distributed collaboration services are being deployed. Such services are multi-point (users are geographically dispersed), real-time, and interactive. The important points in evaluating the quality of multi-party services are the heterogeneous communications environments of the users and the synchronization of user streams.

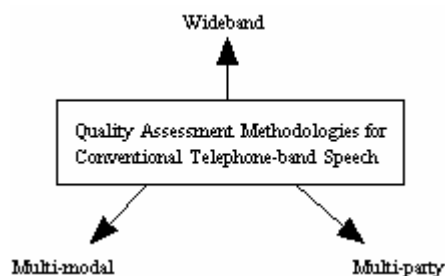


Fig.11 Framework for the development of quality assessment research.

5. Conclusions

This paper has described the author’s perspective on multimedia quality prediction methodologies for advanced mobile and IP-based telephony, which primarily focused upon objective quality assessment, and described associated practical experiments and trials. First, the paper described the need for perceptual QoS assessment. Then, an objective quality prediction scheme was proposed from the viewpoints of quality measurement tools for each quality factor and an opinion model for compound quality factors. Finally, the author’s current trials of measurement tools and opinion models were described.

References

- [1] Akira Takahashi, Hideaki Yoshino, and Nobuhiko Kitawaki: “Perceptual QoS Assessment Technologies for VoIP,” IEEE Communications Magazine, (in press).
- [2] N. Osaka, K. Kakehi, S. Iai, and N. Kitawaki: “A Model for Evaluating Talker Echo and Sidetone in a Telephone Transmission Network,” IEEE Trans. Commun., vol. 40, no. 11, pp. 1684-1692, 1992.
- [3] N. Osaka, N. Kitawaki, and K. Kakehi: Calculation method of OPINE, CCITT COM XII-No.235, 1981-1984, August 1984.
- [4] CCITT P-Series Recommendations, Blue Book Vol.V, Supplement No.3, 1988.
- [5] N. Kitawaki, K. Itoh, M. Honda, and K. Kakehi: “COMPARISON OF OBJECTIVE SPEECH QUALITY MEASURES FOR VOICEBAND CODECS”, IEEE Int. Conf. on Acoust., Speech, Signal Processing, ICASSP’82, pp. S9.5.1-9.5.4, May 1982.
- [6] A. Takahashi, N. Kitawaki, P. Usai, and D. Atkinson: PERFORMANCE EVALUATION OF OBJECTIVE QUALITY MEASURE FOR CODED SPEECH,

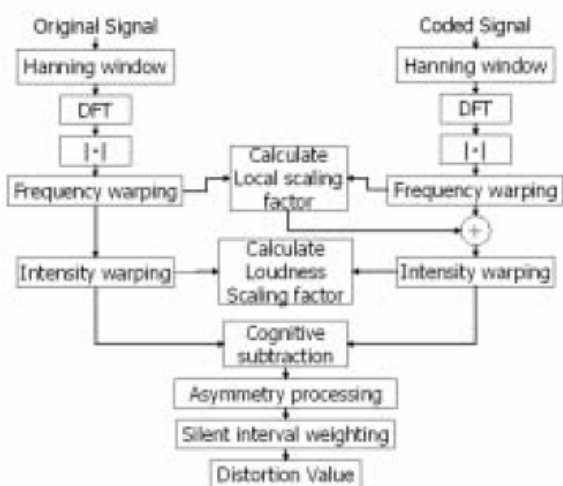


Fig.10 Objective audio quality measurement algorithm.

European Conf. on Speech Communication and Technology, EuroSpeech '97, pp. 437-440, September 1997.

[7] Nobuhiko Kitawaki, Kou Nagai, and Takeshi Yamada: "Objective Quality Assessment of Wideband Speech Coding using W-PESQ Measure and Artificial Voice," 2004 IEEE Workshop on Measurement of Speech and Audio Quality in Networks, Proc. MESAQIN 2004, pp.1-6, May 2004.

[8] Nobuhiko Kitawaki: "Verification of Wideband-PESQ," ITU-T COM12-D188, 2001-2004, March 2004.

[9] K. Itoh, N. Kitawaki, H. Nagabuchi, and H. Irii: "A New Artificial Speech Signal for Objective Quality Evaluation of Speech Coding Systems," IEEE Trans. Commun., vol.42, no.2/3/4, pp.664-672, February/March/April 1994

[10] Nobuhiko Kitawaki: "Artificial voice for objective quality measurement of CELP coding and packet loss using PESQ measure," ITU-T COM12-D191, 2001-2004, March 2004.

[11] Nobuhiko Kitawaki: "Artificial voice applied for objective quality measurement of wideband-speech coding using PESQ measure," ITU-T COM12-D190, 2001-2004, March 2004.

[12] Nobuhiko Kitawaki, Takeshi Yamada, and Futoshi Asano: "Comparative Assessment of Test Signals Used for Measuring Residual Echo Characteristics," IEICE, Trans. Communication, vol. E-86-B, no. 3, pp. 1102-1108, March 2003.

[13] Nobuhiko Kitawaki: "Test signal used for measuring residual acoustic echo characteristics," ITU-T COM12-C65, 2001-2004, March 2004.

[14] Takeshi Yamada, and Nobuhiko Kitawaki: "A PESQ-based Performance Prediction Method for Noisy Speech Recognition," International Congress on Acoustics, ICA2004, Proc. pp.Tu.P2.9, April 2004.

[15] Nobuhiko Kitawaki, and Takeshi Yamada: "Performance estimation methodology of noise reduction algorithm used for noisy speech recognition by PESQ and artificial voice," ITU-T COM 12-C66, 2001-2004, March 2004.

[16] Kazuhisa Kotegawa, Nobuhiko Kitawaki, Takeshi Yamada, and Futoshi Asano: "OPTIMUM BANDWIDTH CHOOSING METHOD FIXED AT LOW BIT-RATE," 2002 IEEE Speech Coding Workshop, Proc. pp. 44-46, October 2002.

[17] T. Tachi, S. Iai, and N. Kitawaki: "Proposal of Selection Method of Test Pictures in HDTV Subjective Quality Assessments," IEEE MULTIMEDIA'92, pp.67-68, April 1992.