Telephone Speech Quality Standards

for

Wideband IP Phone Terminals (handsets)

CES-Q004-1

1. V.0 March 30, 2007
2. V.0 March 30, 2008
3. V.0 November 10, 2008

March 30, 2009

Communications and Information network Association of Japan
(CIAJ)

Table of contents

Introduction

[Objectives for establishing these Standards]

It is the responsibility of manufacturers or service providers in the information and communication network industry to give users clear information about the speech quality performance of the devices or services they provide. For this reason, when the provision of communication terminals was deregulated more than ten years ago, the Communications and Information network Association of Japan (CIAJ) established the speech quality standards for "*Standard Telephones and Multi-functional Telephones*", and began to provide a compliance test and certification service for speech devices. Manufacturers and vendors have displayed the "C" mark on compliant products as proof of satisfactory speech quality performance for users. CIAJ drafted the speech quality standards for "*Digital Cordless and Mobile Phones*" in 2001 and released it to CIAJ members in the hope of early adoption.

Recently, as the speeds of both Internet access connections and terminals have become faster, IP phones are beginning to be widely used. IP phone service has the advantages of being more economical and more compatible with multimedia than the conventional circuit-switching based telephone services. In particular, IP phones supporting wideband communications have been released in an effort to improve speech quality. In order to accommodate such new services, the CIAJ Telecommunications Quality Committee formed a Working Group to study widedband speech quality problems associated with providing VoIP telephone service and to develop speech quality standards for such services. In order that the standards so developed might be widely accepted by CIAJ member companies, it was necessary to make the standards reasonably achievable with currently available technology. Therefore, the WG surveyed the current situation by conducting speech quality tests, mainly using products from member companies as samples.

In the International Telecommunications Union, the Telecommunication Standardization Sector (ITU-T) had been studying wideband speech quality of VoIP. As a result, ITU-T Recommendation P. 311 was approved in June 2005.

Taking these activities in various standardization organizations into consideration, CIAJ has established the "*Telephone Speech Quality Standards for Wideband IP Phone Terminals*" as follows.

[Notes on the Standards]

These standards were developed by referring to ITU-T Rec. 311 (June 2005). Although the scope of this Recommendation covers headsets and hands-free devices, the scope of application of the CIAJ Standards has placed priority on handsets, on account of the urgency of the need to issue standards and taking into consideration the effort to promote wideband IP telephones by not only terminal manufacturers, but network providers. Among the items specified in the ITU-T Recommendation, receive comfort noise, maximum acoustic pressure, input/output linearity, distortion ratio and stability loss have been classified as items for further study by the CIAJ Telecommunications Quality Committee because the Committee was unable to determine the applicability of these items in the ITU-T Recommendation due to the insufficient time spent on its study within the Committee and on the survey of CIAJ members’ products. Therefore,

P. 311 figures have been listed for reference.

Although these Standards address terminals, some items may be affected by the conditions of the IP network used. To cope with such network factors, end-to-end quality has been specified while setting several sets of typical conditions on the network.

[Use of the Standards]

At present, a compliance mark (the so-called “C Mark”) is defined for “CES-Q001” Standards for analog telephones. The application of a similar mark for the present Standards is being considered, and the application procedure, the evaluation procedure and any other procedures shall be conducted according to operational rules.

1. Scope of application

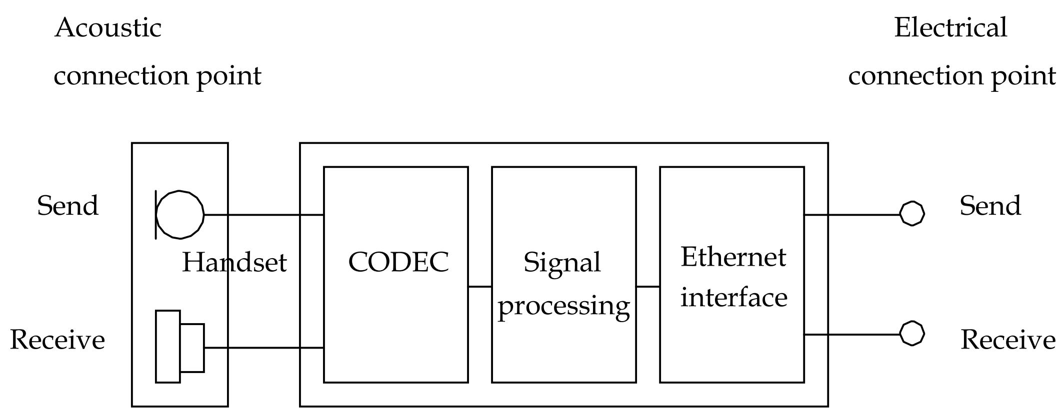
1.1. Devices to which the Standards are to be applied

IP phone terminals and PC soft phones with handsets (150–7000Hz bandwidth).

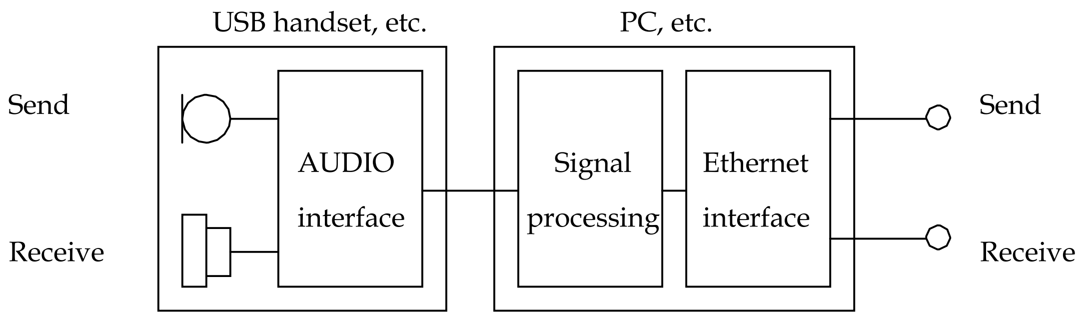
1.2. Interfaces

On the phone terminal side, the acoustic interfaces are the mouthpiece and earpiece of the handset.

On the network side, the interface is Ethernet (IEEE802.3) and is converted into analog by using reference codec ([1]).



IP phone type



PC soft phone type

1.3. Principles for specified values

Standard values are design target values not taking the variance of products into consideration.

[1] : Includes ITU-T Rec. G. 711.1, which is being considered by carriers, with the exception of G. 722 series.

2. Normative references

The standards from which this CES-Q004-1 has been derived are:

ITU-T Rec. P. 311, P. 1010

Regarding loudness rating

ITU-T Recs. P. 48, P. 64 and P. 79

2.1. Overload point for digital signals for IP packet

In the commonly used ITU-T Rec. G. 722 series for wideband voice encoding, the overload point is +9 dBm0, with standard signal level at -15 dBm0.

ITU-T Rec. G. 711 .1, the extension for G. 711 for encoding voice-frequency signals in narrowband, uses +3.17 dBm0 as the overload point and -15 dBm0 as the standard signal level when using the same overload point as G.711 .

3. Send speech characteristics

3.1. Send loudness rating (SLR)

The send loudness rating (SLR) shall be 4±4 dB. The sensitivity shall be based on signal level calculations defined in 2.1 above. LR calculations shall be made according to ITU-T Rec. P. 79 Annex G. If the send sensitivity is variable, the measurement shall be made using the sensitivity recommended by the manufacturer. It is desirable that the physical shape and the dimensions of the handset are close to those defined in ITU-T Rec. P.350.

3.2. Send frequency response

Table 1 and Figure 1 define the mask pattern. The measuring frequency points shall be ISO 1/3 octave band central frequencies. These points shall be within the mask pattern. ([2])

Table 1 ([2])

Frequency(Hz)	Upper Limit(dB)	Lower Limit(dB)
100	6	$-\infty$
125	6	–9
200	6	–6
1000	6	–6
5000	([3])	–9
6300	11	–13
8000	11	$-\infty$

2 Sensitivity is a relative figure, not an absolute figure. Therefore, the points shall be acceptable as long as the shape of the speech frequency characteristics is within the mask pattern when the points are raised or lowered in the direction of sensitivity.

3 This upper limit will not be designated, since it is the point of change between 1000 Hz and 6300 Hz.

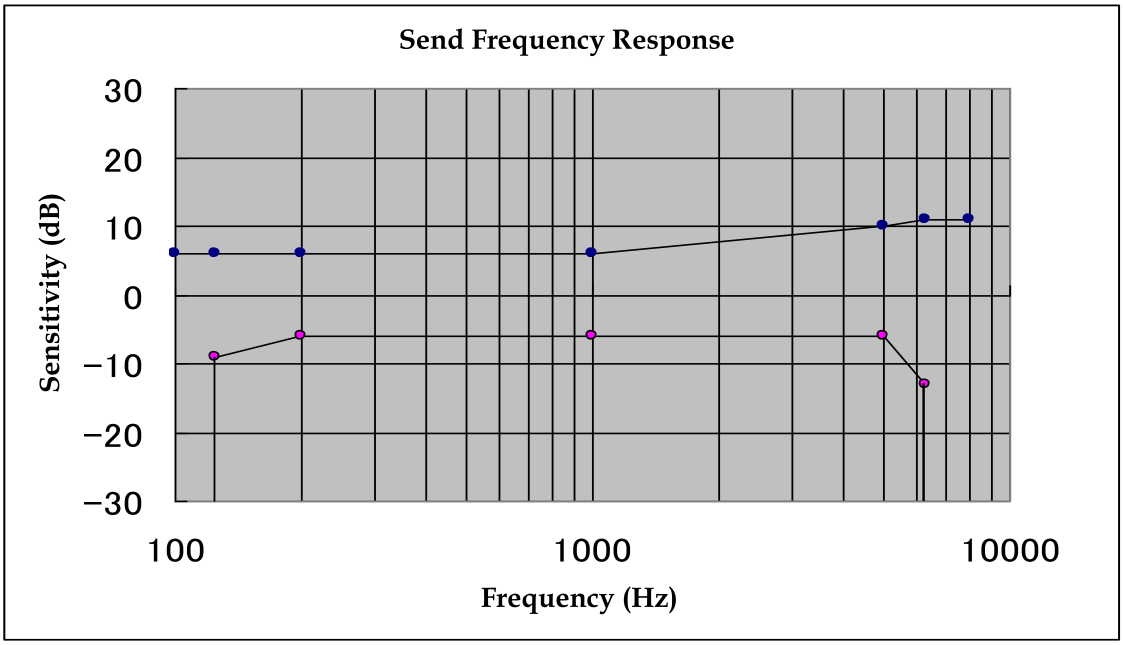


Figure 1 (2)

3.3. Send noise

The send noise shall be -68 dBm0A or lower. The measurement shall be made in accordance with CES-Q004M-1 "*Measuring method*." Any noise reduction capabilities, such as a noise canceller, an echo canceller or AGC, shall be enabled. If the send sensitivity is variable, the measurement shall be made using the sensitivity recommended by the manufacturer.

3.4. Distortion (Reference values)

For sine wave acoustical input signals of 200 Hz, 1 kHz and 6 kHz, the ratio of distortion of the digital signal output for IP packet payload must be equivalent or above the numbers in the table below.

Sending level（relative）(Reference is -4.7dBPa)	**Lower limit of distribution（d B）**		
	200 Hz	**1 kHz**	**6 kHz**
+18（+12(4)） to −20	29.0	35.0	29.0
−30	25.0	26.5	25.0
−46	11.0	12.5	11.0

2 Sensitivity is a relative figure, not an absolute figure. Therefore, the points shall be acceptable as long as the shape of the speech frequency characteristics is within the mask pattern when the points are raised or lowered in the direction of sensitivity.

4 The overload point will be encoding format +3.17 dBmO, the same as G. 711

4. Receive speech characteristics

4.1. Receive loudness rating (RLR)

The receive loudness rating (RLR) shall be 2±4 dB. If a receive sound level setting function is provided, the nominal setting designated by the manufacturer shall be selected.

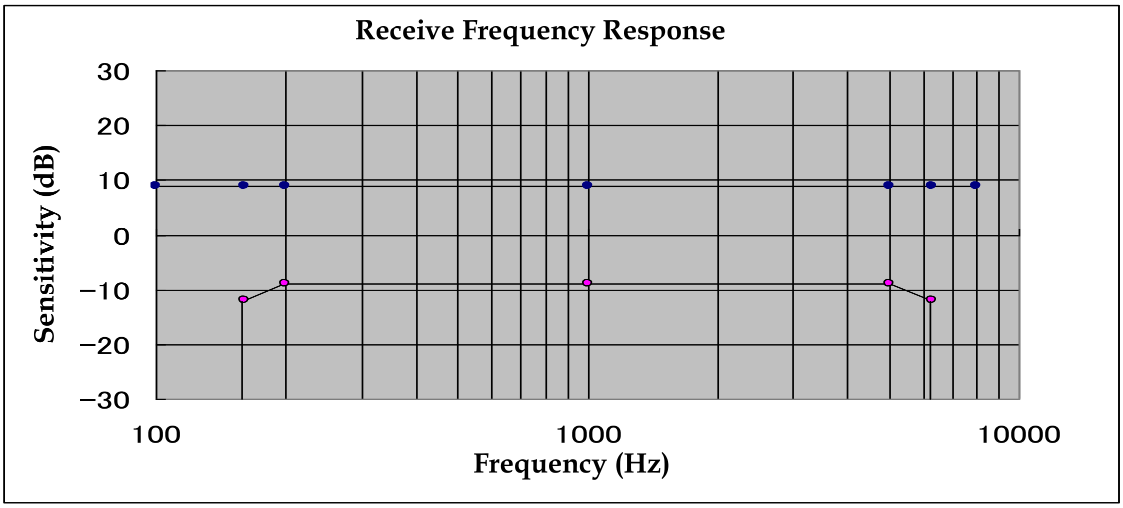
The sensitivity shall be based on signal level calculations defined in section 3 above. LR calculations shall be made according to ITU-T Rec. P. 79 Annex G. It is desirable that the physical shape and the dimensions of the handset are close to those defined in ITU-T Rec. P. 350.

4.2. Receive frequency response

Table 2 and Figure 2 define the mask pattern. The measuring frequency points shall be ISO 1/3 octave band central frequencies. These points shall be within the mask pattern. ([2])

Table 2 ([2])

Frequency(Hz)	Upper Limit(dB)	Lower Limit(dB)
100	9	–∞
160	9	–12
200	9	–9
1000	9	–9
5000	9	–9
6300	9	–12
8000	9	––∞



[2] Sensitivity is a relative figure, not an absolute figure. There fore, the points shall be acceptable as long as the shape of the speech frequency characteristics is within the mask pattern when the points are raised or lowered in the direction of sensitivity.

Figure 2

ITU-T Rec. P. 311 sets the upper and lower limits at within 5 dB, but the rule shall be relaxed for the time being, taking into consideration the technology currently available. ([2A])

4.3. Receive noise

The receive noise shall be -59 dBPa(A) or lower. If a receive sound level adjustment function is provided, the nominal position shall be selected. The measurement shall be made in accordance with CES-Q004M-1 "*Measuring method*." Any noise reduction capabilities, such as a noise canceller, an echo canceller or AGC, shall be enabled.

4.4. Distortion (Reference values)

For sine wave acoustical input signals of 200 Hz, 1 kHz and 6 kHz, the ratio of distortion of the output sound signal must be equivalent or above the numbers in the table below.

Digital data on IP packet payload (dBm0) Reference is -15 d Bm0	**Lower limit of distribution(dB)**		
	200 Hz	**1 kHz**	**6 kHz**
+8(-3([4])) to –30	29.0	35.0	29.0
–40	25.0	26.5	25.0
–56	11.0	12.5	11.0

5. Sidetone characteristics

5.1. Sidetone masking rating (STMR)

The sidetone masking rating (STMR) shall be 13 - 22 dB. If a receiving sound level adjustment function is provided, the nominal position shall be selected. The measurement shall be made in accordance with CES-Q004M-1 "*Measuring method*." If a sending sensitivity setting function is provided, the standard setting shall be selected for the measurement.

6. Echo

The return that causes talker echo is specified as the talker echo loudness rating (TELR).

[2A] ITU-T Rec. 311 also states: "*It is recognized that it may be difficult for telephone sets equipped with receivers designed according to the current technology to meet the recommended requirement, and still comply with the recommended TCLw limit*."

[4] The overload point will be encoding format +3.17 dBmO, the same as G. 711

6.1. Weighted terminal coupling loss

Since the IP phone type and the PC soft phone type terminals use four wires, the only coupling that arises is acoustic coupling from the earpiece to the microphone. The acoustic coupling level is expressed as the weighted terminal coupling loss (TCLw).

TELR is the sum of the TCLw, and SLR and RLR.

The TELR of the IP phones shall be 58 dB or higher at the normative recommended value of receiving sound level adjustment.

6.2. Stability loss (Reference values)

The stability loss from the digital input to the digital output shall be determined for all frequencies in the range of 100 Hz to 8000 Hz.

The stability loss shall be measured with the handset lying on a hard surface.

Rules concerning these measuring methodology shall be discussed in the future.

Stability loss for IP phones shall be no less than 6 dB at all receive loudness settings.

7. Network performance loss due to traffic congestion

Types of distortion that arise in a VoIP-based telephone service include those generated within the terminal, such as speech coding distortion, and those that vary depending on the network performance objectives, such as distortion due to packet loss. The delay time within a terminal is dependent on the relationship between the delay variation within the network and the terminal's jitter buffer performance. Therefore, a requirement for these quality factors cannot be defined in a terminal in isolation of a connected network.

7.1. Network performance objectives

Network load conditions are defined in Table 3.

Table 3

Element	IP network performance objectives
Average delay time, T (ms)	70
Maximum delay time, Ta (ms)	67 . 10
Maximum delay variation, ΔT: Δt_{max} (ms)	20
Average delay variation: Δt_{ave} (ms)	2 . 90
Packet loss ratio (%): Ppl	0 . 1

The delay variation is a change in the instantaneous delay time of the network from Ta to Ta+ΔT. The probability of ΔT is assumed to follow an exponential distribution. The event probability of to the maximum Δt max for delay variation ΔT shall be 99.9%.

The probability of Ppl is assumed to follow a uniform distribution (random loss). Busty loss is left as an item for further study.

7.2. Terminal voice latency

The terminal voice latency is the average end-to-end voice latency, under the performance on the IP network shown in Table 3, minus the average delay time of the IP network. If the terminal voice latency fluctuates, the average value is used. The average delay time of the IP network is the sum of the absolute delay time and the average delay variation of the IP network.

Send delay time should be 35 mS or less when there is no network load. (target: 30 mS or less)

Receive delay time, when there is network loads described in Table 3, shall be 65 mS or less (target: 50 mS or less)

Although it is possible to realize these targets, the requirement is provisionally relaxed in consideration of the actual performance of the products available on the market.

7.3. Transmission rating factor (Reference values)

In existing telephone bandwidths, the "*R Value*" is adopted as the index to quantify the transmission rating factor, but for wideband voice communications, there is no ITU-T recommendation. Thus, the ITU-T Rec. P. 862.2 Wide Band PESQ (WB-PESQ) shall be used. The WB-PESQ for end-to-end acoustic input/output measurement conditions for both sending and receiving shall be 3.5 or above.

Telecommunications Quality Committee

Chair	Hiroshi Asada	Business Communication Systems Department1, PC Development center Toshiba Corporation PC&Network Company

Working Group on Operation Requirements

Chair	Shinji Naito	Enterprise System Department, Development Promotion headquarters, Nakayo Telecommunications, Inc.
Members	Yoshimasa Kamogashira	Engineering Division, Iwatsu Electric Co., Ltd.
	Hiroshi Asada	Business Communication Systems Department1, PC Development center Toshiba Corporation PC&Network Company
	Hiroshi Kuboki	eSound Business Depertment Telecom Systems Division Oki Electric Industry Co., Ltd.
	Kunihiko Kimura	Research & Development Div Hardware Development Dept SAXA,Inc.
	Ogiyama Tetsuya	Network Development Div. NEC Infrontia Corp.
	Katsuichi Osakabe	Sound Network Division YAMAHA CORPORATION
Ad-hoc member	Manabu Okamoto	NTT Cyber Space Laboratories NIPPON TELEGRAPH AND TELEPHONE CORPORATION
Secretary	Tadahiro Higuchi	Communications and Information Network Association of Japan