International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.341

(03/2011)

SERIES P: TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Voice terminal characteristics

Transmission characteristics for wideband digital loudspeaking and hands-free telephony terminals

Recommendation ITU-T P.341

ITU-T

ITU-T P-SERIES RECOMMENDATIONS

TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Voice terminal characteristics	**Series**	**P.30** **P.300**
Reference systems	Series	P.40
Objective measuring apparatus	Series	P.50 P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of speech quality	Series	P.80 P.800
Audiovisual quality in multimedia services	Series	P.900
Transmission performance and QoS aspects of IP end-points	Series	P.1000
Communications involving vehicles	Series	P.1100

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T P.341

Transmission characteristics for wideband digital loudspeaking and hands-free telephony terminals

Summary

Recommendation ITU-T P.341 provides audio performance requirements for wideband audio (8000 Hz) digital loudspeaking and hands-free telephones.

Requirements and test methods are specified for the major audio transmission parameters affecting wideband audio, including sending and receiving levels, frequency response, noise, distortion, stability, echo path and delay. Associated test methods are also given. Wideband audio represents a considerable departure from traditional telephony, offering significantly improved quality.

Main changes over the previous version of this Recommendation (2005) are as follows:

– The structure of the Recommendation is modified; the objective measurement methods originally in Annex A have been transferred into the main body.

– In addition to sinusoidal and noise signals, speech-like stimulus signals, as described in Recommendation ITU-T P.50 and Recommendation ITU-T P.501, are also recommended for the measurements.

– HATS method is introduced for better compliance with practical application.

Some technical requirements and related test methods have been modified according to new development, e.g., frequency response, distortion and delay in sending and receiving direction.

History

Edition	Recommendation	Approval	Study Group
1.0	ITU-T P.341	1995-04-18	12
2.0	ITU-T P.341	1998-02-27	12
2.1	ITU-T P.341 (1998) Cor. 1	1999-09-30	12
3.0	ITU-T P.341	2005-06-06	12
4.0	ITU-T P.341	2011-03-01	12

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

© ITU 2012

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

Recommendation ITU-T P.341

Transmission characteristics for wideband digital loudspeaking and hands-free telephony terminals

1 Scope

This Recommendation provides audio performance requirements and test methods for loudspeaking and hands-free telephones capable of transmitting an audio bandwidth extending beyond the conventional telephony bandwidth of 300 to 3400 Hz, to a bandwidth of approximately 100 Hz to 8000 Hz. Such telephones are known as wideband audio telephones, and will make use of digital encoding schemes such as in [ITU-T G.722]. IP terminals may support other coding algorithms. Wideband audio telephones are expected to be used in new services such as high quality audio conferencing, videoconferencing and multimedia applications.

The requirements listed in this Recommendation are primarily applicable to telephones using ITU-T G.722 encoding at 64 kbit/s, but should also be used as the basis of requirements for other wideband audio encoding schemes. This is still under study in ITU-T.

General information on hands-free terminals, which includes switching characteristics, can be found in [ITU-T P.340] and information on acoustic echo controllers in [ITU-T G.167].

For loudspeaking telephones which do not provide full hands-free operation, the relevant parts of this Recommendation may be used.

Conventional telephone band (300 Hz-3400 Hz) digital hands-free telephones using encoding according to [ITU-T G.711] and [ITU-T G.726] are covered by [ITU-T P.342]. Audio performance requirements for wideband headset terminals are included in [ITU T P.311]. Specifications for car-mounted wideband hands-free terminals are included in [ITU-T P.1110]. Transmission characteristics for narrow-band cordless and mobile digital terminals are included in [ITU-T P.313].

2 Normative references

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T G.114] Recommendation ITU-T G.114 (2003), *One-way transmission time*.

[ITU-T G.122] Recommendation ITU-T G.122 (1993), *Influence of national systems on stability and talker echo in international connections*.

[ITU-T G.167] Recommendation ITU-T G.167 (1993), *Acoustic echo controllers*.

[ITU-T G.711] Recommendation ITU-T G.711 (1988), *Pulse code modulation (PCM) of voice frequencies*.

[ITU-T G.722] Recommendation ITU-T G.722 (1988), *7 kHz audio-coding within 64 kbit/s*.

[ITU-T G.726] Recommendation ITU-T G.726 (1990), *40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)*.

[ITU-T G.729.1] Recommendation ITU-T G.729.1 (2006), *G.729-based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729*.

[ITU-T G.1020] Recommendation ITU-T G.1020 (2006), *Performance parameter definitions for quality of speech and other voiceband applications utilizing IP networks*.

[ITU-T I.412] Recommendation ITU-T I.412 (1988), *ISDN user-network interfaces – Interface structures and access capabilities*.

[ITU-T P.10] Recommendation ITU-T P.10/G.100 (2006), *Vocabulary for performance and quality of service*.

[ITU-T P.50] Recommendation ITU-T P.50 (1999), *Artificial voices*.

[ITU-T P.51] Recommendation ITU-T P.51 (1996), *Artificial mouth*.

[ITU-T P.57] Recommendation ITU-T P.57 (2009), *Artificial ears*.

[ITU-T P.58] Recommendation ITU-T P.58 (1996), *Head and torso simulator for telephonometry*.

[ITU-T P.64] Recommendation ITU-T P.64 (2007), *Determination of sensitivity/frequency characteristics of local telephone systems*.

[ITU-T P.79] Recommendation ITU-T P.79 (2007), *Calculation of loudness ratings for telephone sets*.

[ITU-T P.310] Recommendation ITU-T P.310 (2009), *Transmission characteristics for narrow-band digital handset and headset telephones*.

[ITU-T P.311] Recommendation ITU-T P.311 (2011), *Transmission characteristics for wideband digital handset and headset telephones*.

[ITU-T P.313] Recommendation ITU-T P.313 (2007), *Transmission characteristics for cordless and mobile digital terminals*.

[ITU-T P.340] Recommendation ITU-T P.340 (2000), *Transmission characteristics and speech quality parameters of hands-free terminals*.

[ITU-T P.342] Recommendation ITU-T P.342 (2009), *Transmission characteristics for narrow-band digital loudspeaking and hands-free telephony terminals*.

[ITU-T P.501] Recommendation ITU-T P.501 (2009), *Test signals for use in telephonometry*.

[ITU-T P.581] Recommendation ITU-T P.581 (2009), *Use of head and torso simulator (HATS) for hands-free and handset terminal testing*.

[ITU-T P.1010] Recommendation ITU-T P.1010 (2004), *Fundamental voice transmission objectives for VoIP terminals and gateways*.

[ITU-T P.1110] Recommendation ITU-T P.1110 (2009), *Wideband hands-free communication in motor vehicles*.

[ETSI TS 126 171] ETSI TS 126 171 V9.0.0 (2010), *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; (Speech codec speech processing functions; Adaptive Multi-Rate – Wideband (AMR-WB) speech codec; General description (3GPP TS 26.171 version 9.0.0 Release 9)*.

[IEC 61672-2] IEC 61672-2 (2003), *Electroacoustics – Sound level meters – Part 2: Pattern evaluation tests*.

3 Definitions and abbreviations

3.1 Definitions

This Recommendation defines the following terms:

3.1.1 ear-drum reference point (DRP): Point located at the end of the ear canal, corresponding to the ear-drum position.

3.1.2 ear reference point (ERP): A virtual point for geometric reference located at the entrance to the listener's ear, traditionally used for calculating telephonometric loudness ratings.

3.1.3 free-field equalization: Artificial head is equalized in such a way that for frontal sound incidence in anechoic conditions the frequency response of the artificial head is flat.

3.1.4 group-audio terminal: A speakerphone set primarily designed for use by several users, which will not be equipped with a handset.

3.1.5 hands-free reference point (HFRP): A point located on the axis of the artificial mouth, at 50 cm from the outer plane of the lip ring, where the level calibration is made under free-field conditions. It corresponds to measurement point No. 11 defined in [ITU-T P.51].

3.1.6 hands-free terminal (HFT): A telephone set that does not require the use of hands during the communications session; examples are headset, speakerphone and group-audio terminal.

3.1.7 HATS hands-free reference point (HATS HFRP): Corresponds to a reference point "n" from [ITU-T P.58]: "n" shall be one of the points numbered from 11 to 17 and defined in Table 6A of [ITU-T P.58] (coordinates of far field front point). The HATS HFRP depends on the location(s) of the microphones of the terminal under test: the appropriate axis lip-ring/HATS HFRP shall be as close as possible to the axis lip-ring/speakerphone microphone under test.

3.1.8 head and torso simulator (HATS): Manikin extending from the top of the head to the waist, designed to simulate the sound pick-up characteristics and the acoustic diffraction produced by a median adult, and to reproduce the acoustic field generated by the human mouth.

3.1.9 loudspeaking function: Function of a handset telephone using a loudspeaker associated with an amplifier as a telephone receiver. It shall be used with a handset to transmit sending signals.

3.1.10 speakerphone set: A telephone set using a loudspeaker as a telephone receiver with or without an embedded microphone as a transmitter; it may be used without a handset.

3.1.11 wideband telephony: Transmission of speech with a nominal pass-band wider than 300-3400 Hz, usually understood to be 100 Hz-8000 Hz.

3.2 Abbreviations

This Recommendation also uses the following abbreviations.

Relevant abbreviations in [ITU-T P.10] will also apply:

CSS Composite Source Signal

DRP Ear-Drum Reference Point

HFT Hands-free Terminal

LST LoudSpeaking Terminal

MRP Mouth Reference Point

RLR Receiving Loudness Rating

SLR Sending Loudness Rating

TCL Terminal Coupling Loss

TCLw Weighted Terminal Coupling Loss

4 Test arrangement

4.1 Test environment

4.1.1 Free-field condition

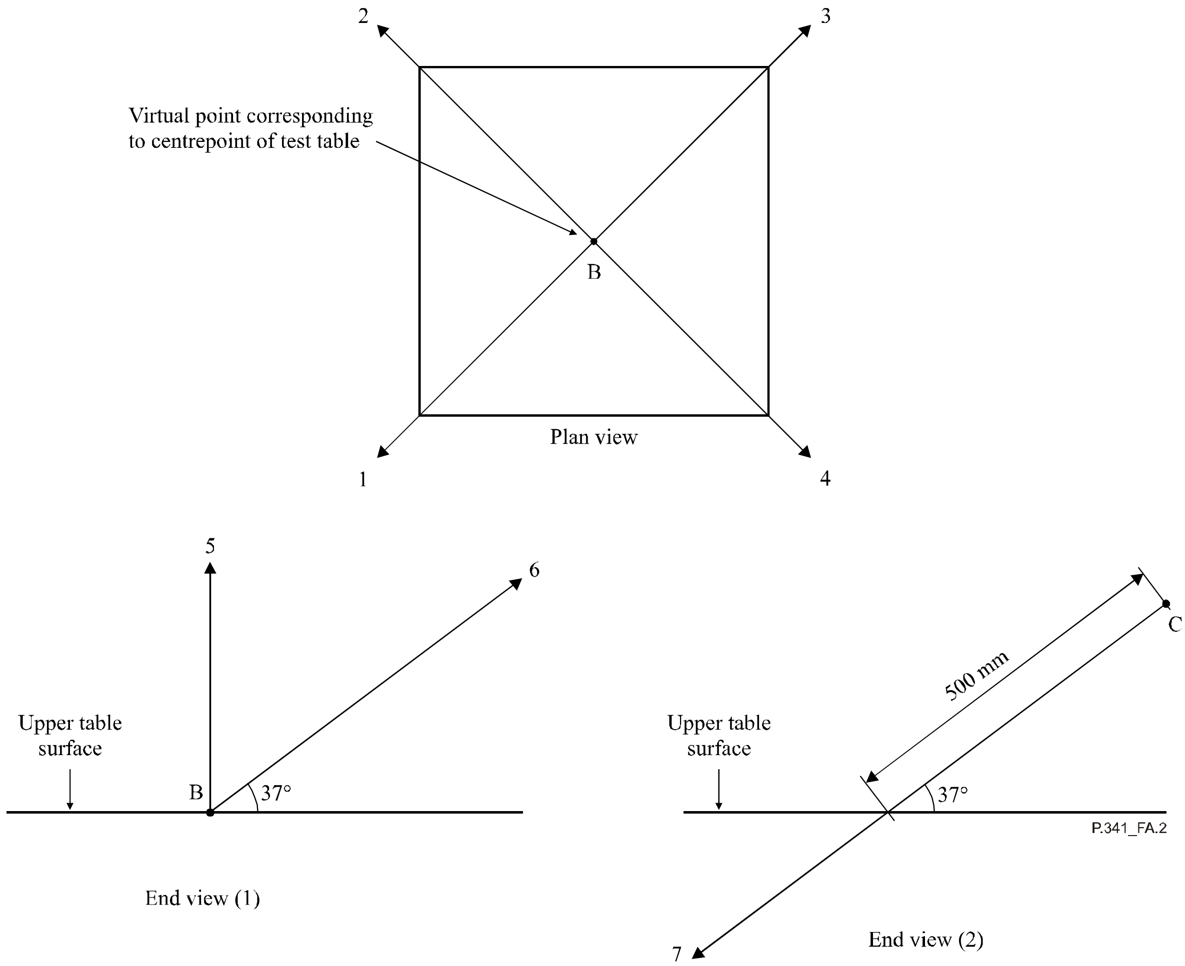
To ensure repeatability of tests, the environment for all measurements shall be free-field (anechoic) down to the lowest frequency of the 1/3-octave band centered at 125 Hz.

Satisfactory free-field conditions are deemed to exist where errors, due to the departure from ideal conditions, do not exceed the limits reported in Table 1, inside a sphere centered at point B in Figure 1, with one-meter radius, in the absence of the test table.

Table 1 – Allowable departure from ideal conditions

1/3 Octave centre frequency (Hz)	Allowable departure (dB)
≤ 630	±1.5
800 to 5000	±1.0
≥ 6300	±1.5

The test signal used for the verification of free-field conditions shall be –20 dBPa at the HFRP. A wideband noise signal shall be used and third octave spectrum measurements shall be carried out at the measurement points. Measurements shall be made along the seven axes numbered 1 to 7 in Figure 1. The sound source (the artificial mouth) shall be placed at positions equivalent to B or C as appropriate. Measurement points along each axis, taken from the lip plane of the artificial mouth, shall be at distances of 315 mm, 400 mm, 500 mm, 630 mm, 800 mm and 1000 mm.



NOTE 1 – Axes 1 to 7 are used in the determination of free-field conditions for 1 m radius sphere.
NOTE 2 – Axes 1 to 4 are in the horizontal plane occupied by the test table surface.
NOTE 3 – Axis 5 is perpendicular to the horizontal plane occupied by the test table surface.
NOTE 4 – Measurements of the free-field sound pressure are made in the absence of the test table.

Figure 1 – Verification of the free-field conditions

4.1.2 Noise level

The broadband noise level shall not exceed –70 dBPa (A). Furthermore, the octave band noise level shall not exceed the limits given in Table 2.

Table 2 – Octave band noise level limits

Octave centre frequency (Hz)	Octave band noise level (dBPa)
63	–45
125	–60
250	–65
500	–65
1000	–65
2000	–65
4000	–65
8000	–65
16000	–65

4.2 Test set-up

4.2.1 General description

The general access to terminals is described in Figure 2. This can be made by using a HATS (head and torso simulator) or a free-field microphone together with an artificial mouth. Therefore, two measurement methods can be applied:

1) the conventional method (using a free-field microphone together with a discrete ITU-T P.51 artificial mouth);
2) the HATS method.

The HATS method aims to achieve the most realistic simulation of the "average" subscriber. All measurement values produced by HATS are intended to be free-field equalized. If not specified, HATS's right ear is used for the receive measurement. More details such as the exact calibration and equalization procedures as well as the combination of the two ear signals for the purpose of measurements can be found in [ITU-T P.581]. Note that the horizontal positioning of the HATS reference plane shall be guaranteed within ±2°.

All effects including test table and measurement equipment are considered a part of the measurement.

NOTE – It is recognized that these two methods may give different results. While the HATS method is intended to simulate the hands-free situation in a more realistic way, the conventional set-up is less sensitive to asymmetric constructions of devices under test for example. Differences in the tests results may result from the different orientation of the artificial mouth, from the diffraction effect of the HATS in sending and receiving and from the different physical position of the artificial head compared to the position of the free-field microphone. These effects and their impact on the measurement are still under study.

The interface in Figure 2 must be capable of converting the digital output stream from the tested set (which may be in various formats, depending on the specific type of telephone set, e.g., ISDN sets or IP terminals), to a form compatible with the test equipment. Interfaces can be applied for the sending and the receiving separately, taking into account telephone sets which are connected to various types of exchanges or gateways. The interface shall be capable of providing the signalling, supervision or routing necessary for the terminal to be working in all test modes.

In general, there are two approaches to evaluate the transmission performance of a wideband digital telephone:

1) the direct approach;
2) the reference codec approach.

In the direct approach, the companded digital input/output bitstream of the telephone set is operated upon directly. In the reference codec approach, a codec is used to convert the companded digital input/output bitstream of the telephone set to the equivalent analogue values, so that existing test procedures and equipment can be used. This codec should be a high-quality codec whose characteristics are as close as possible to ideal. The direct approach is, in principle, the most accurate although the use of the reference codec approach may sometimes be advantageous. For IP terminals, an ideal reference gateway is needed to pack or unpack transmitted packages.

When a coder with variable bite rate is used, we should adopt the bit rate recognized as giving the best characteristics. For example,

- [ITU-T G.722]: 64 kbit/s.
- [ETSI TS 126 171]: 19.85 kbit/s.
- [ITU-T G.729.1]: 32 kbit/s.

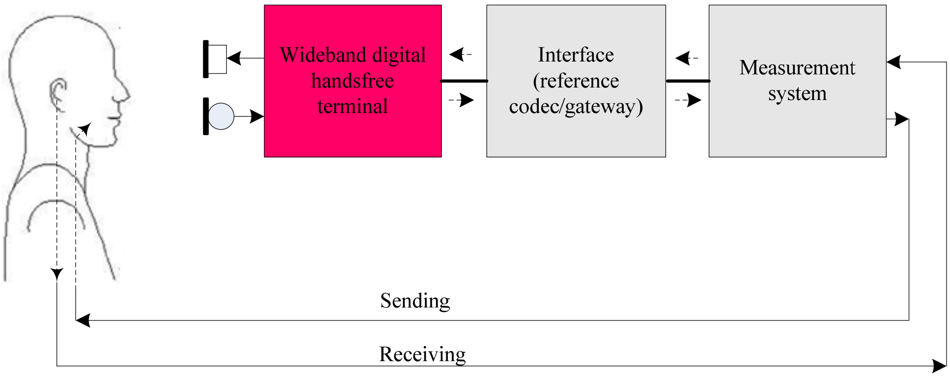


Figure 2 – Test configuration

The terminal is connected to the interface and is placed in the active call state. Terminals fitted with a volume control on receiving shall be set as close as possible to the nominal RLR, and any residential difference from the nominal value will be corrected by the normalization process.

4.2.2 Set-up for desktop hands-free terminal

4.2.2.1 Conventional method

If a free-field microphone together with a discrete ITU-T P.51 artificial mouth is used, the desktop HFT is placed on a test table according to [ITU-T P.340]. As shown in Figure 3, the artificial mouth axis and the microphone axis are coincident with the straight line drawn between point C and point B. For measurements in the sending direction, the artificial mouth is positioned at point C. For receiving direction, the artificial mouth shall be replaced by the free-field microphone. The centre of the microphone grid shall be positioned at point C, with its axis coincident with line CB.

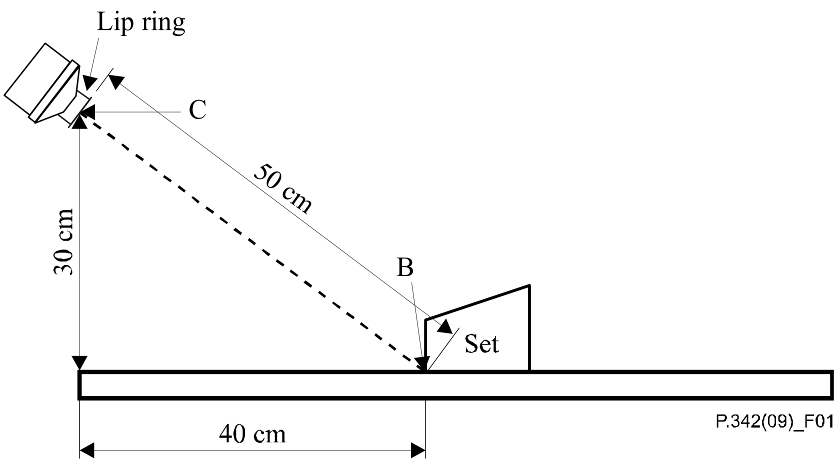


Figure 3 – Measurement configuration for the desktop hands-free terminal using a discrete ITU-T P.51 artificial mouth and a free-field microphone, side view

4.2.2.2 HATS method

When a HATS is used, the centre of the lip-ring of HATS shall also be located at point C as defined in Figure 3, however the reference axis of the mouth ought to be horizontal, as illustrated in Figure 4. For TCLw and stability loss measurement, HATS is positioned but not used.

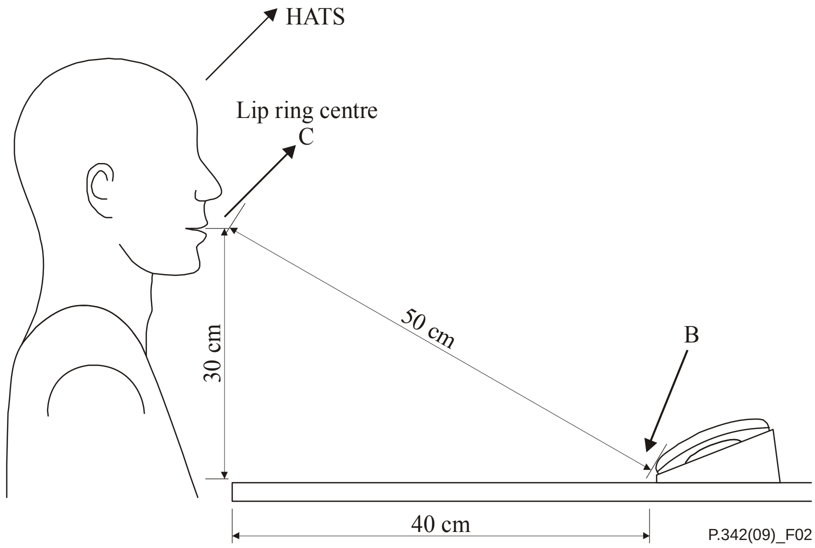


Figure 4 – Measurement configuration for the desktop hands-free terminal using HATS, side view

4.2.3 Set-up for double unit desktop hands-free terminal

For desktop hands-free terminals with a detached microphone and speaker, the standard test position is shown in Figure 5. For desktop hands-free terminals with more pieces, the test arrangement shall be modified to what it is stated in the instruction manual of the terminal under test (TUT).

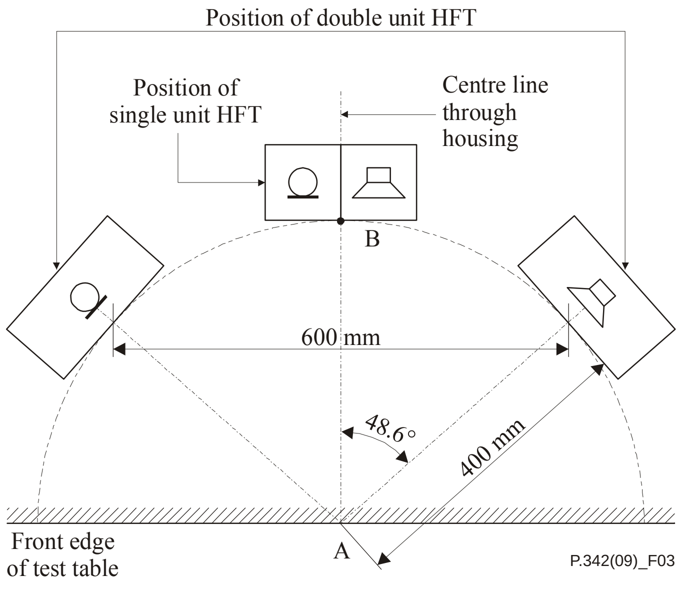


Figure 5 – Measurement configuration for the desktop hands-free terminal with detached microphone and speaker (top view)

4.2.4 Set-up for other types of hands-free terminals

Group-audio terminals, softphones, or speakerphones designed for non-desktop positioning, for example, videophony and multimedia terminals, should be tested with the appropriate position. This position is defined as the recommended test position (RTP). The RTP should be obtained from the manufacturer, and should be based on the product's intended use. Otherwise, (HATS) HFRP may be chosen at point n (far field) from Table 6A of [ITU-T P.58]. Headset terminal positioning is described in [ITU-T P.310]. Car-mounted hands-free terminals test arrangement using HATS method is described in [ITU-T P.581].

Figure 6 gives the test set-up for group-audio terminal using HATS method.

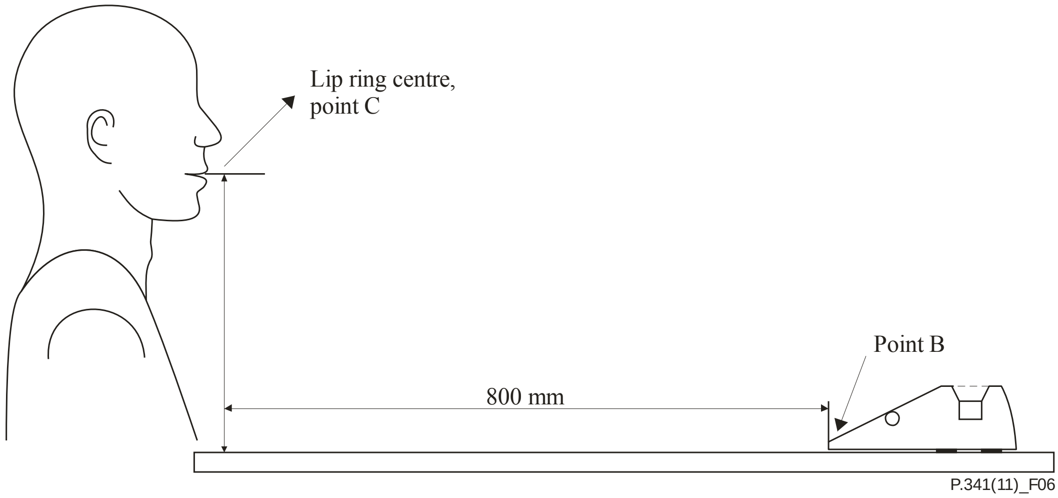


Figure 6 – Measurement configuration for group-audio terminal, sideview

4.2.5 Set-up for handset terminals with loudspeaking function activated

4.2.5.1 Conventional method

Measurements in the sending direction shall be made with the handset placed on HATS as described in [ITU-T P.64].

Receiving measurements of the loudspeaking terminal (LST) are made with the same test position employed in desktop speakerphone terminal measurement, except that the handset is taken off the cradle and placed out of the way during measurement.

TCLw measurements of LST are the same except for the positioning of the handset. Figure 7 shows a recommended test position for making TCL measurements of LST. The handset earphone "centre" shall be placed at point C with the microphone vertical below the earphone. The meaning of "centre" is the centre of the surface of the handset earphone which is placed normally against the ear. This surface is set at 90 degrees relative to the loudspeaker.

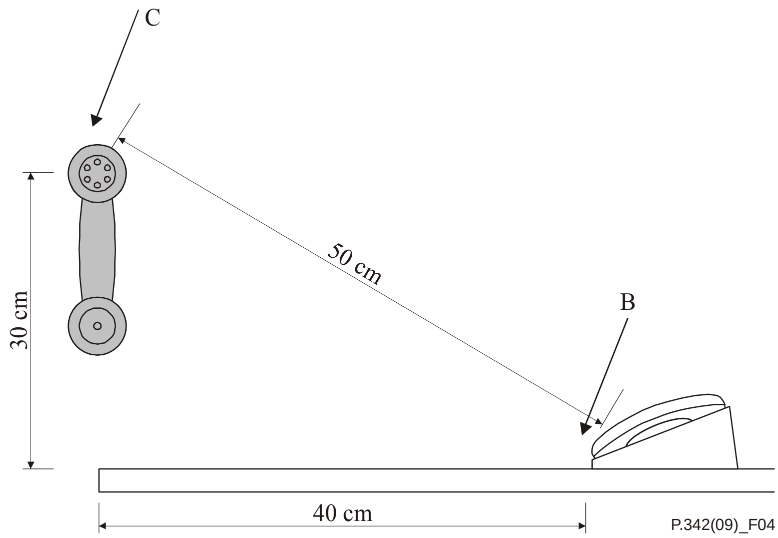


Figure 7 – Standard test position for the LST using conventional method (side view)

4.2.5.2 HATS method

When a HATS is used, the set shall be positioned as shown in Figure 4. The handset is positioned on HATS (the right ear of HATS is used if not specially appointed).

For TCLw measurement, the handset is positioned on HATS (right ear). For stability loss measurement, the handset is placed at 50 cm beside the terminal, with the transducers facing the table.

4.3 Test signal

In general, a speech-like stimulus signal as described in [ITU-T P.50] and [ITU-T P.501] is preferred for testing. Detailed information about the test signal used can be found in the corresponding clause of this Recommendation. The type of signal used shall be stated in the test report.

For wide-band terminals, all test signals which are inserted in receive direction have to be band-limited. The band limitation is achieved by a bandpass filter in the frequency range between 50 Hz and 8000 Hz providing 24 dB/octave bandpass filtering. According to [ITU-T P.501], when composite source signal (CSS) is used, shaping of the wideband CSS spectrum shall be applied. The shaping response characteristics, described in Figure 7-10 of [ITU-T P.501], are applied. In the send direction, the test signals are used without band limitation.

An ON/OFF modulation (e.g., 250 ms ON and 150 ms OFF) shall be applied if echo control or automatic noise detection mechanisms are involved. If modulated signals are used, excitation levels are referred to the ON component of the signals. CSS, as described in [ITU-T P.501], or switched pink noise are signals which provide the desired ON/OFF modulation. A logarithmically distributed multi-sine wave may be equally well applicable.

An artificial voice according to [ITU-T P.50] or a speech-like test signal as described in [ITU-T P.501] can be used for the activation.

The type of signal used shall be stated in the test report.

NOTE 1 – The use of sine signals may not be appropriate when speech processing and coding systems are implemented in the terminal. For example, in distortion measurement, if a sine wave is not usable, an alternative test signal could be a band-limited noise signal centred on the test frequencies.

NOTE 2 – It should be ensured that the test signal is treated by speech processing algorithms as a speech-like signal, and not a noise-like signal.

NOTE 3 – When measuring digital telephone sets, it is advisable to avoid measuring at sub-multiples of the sampling frequency. There is a tolerance on the frequencies of ±2% which may be used to avoid this problem, except for 8000 Hz where only the −2% tolerance may be used.

4.4 Signal level

4.4.1 General

The use of test signal levels should be stated in the test report. Unless specially defined, the signal level refers to the RMS level of the test signal averaged over the complete test sequence length.

4.4.2 Sending

Unless specified otherwise, the test signal level shall be –4.7 dBPa at MRP defined in [ITU-T P.64].

The signal generated by the artificial mouth is equalized at MRP under free-field conditions at a level of –4.7 dBPa in the frequency range from 100 Hz to 8000 Hz. The spectrum and level recorded at MRP is used as a reference for measurement of sending characteristics.

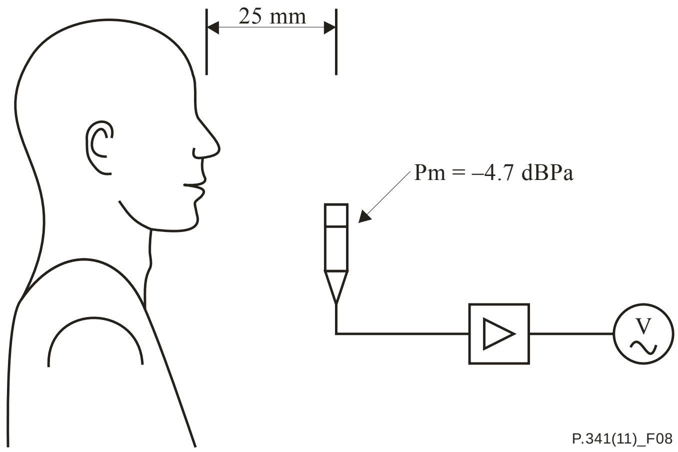


Figure 8 – Calibration and equalization of sound pressure at MRP for HATS

Then the level is adjusted in order to obtain a level of –28.7 dBPa at (HATS) HFRP. Depending on the type of the terminal under test, different levels shall be applied for adjustment as shown in Table 3.

Table 3 – Distance and level used for calibration in the sending direction

	Distance (cm)	**Level (dBPa)**
Desktop terminal	50	–28.7
Group-audio terminal	85	–33.3

4.4.3 Receiving

Unless specially defined, the applied test signal level at the interface input shall be –16 dBm0. All measurement values produced by HATS are intended to be free-field equalized.

4.5 Accuracy of measurements

Unless specified otherwise, the accuracy of measurements made by test equipment shall not exceed the limits given in Table 4.

Table 4 – Accuracy of measurements

Item	**Accuracy**
Electrical signal power	±0.2 dB for levels ≥ –50 dBm
Electrical signal power	±0.4 dB for levels < –50 dBm
Sound pressure	±0.7 dB
Clock accuracy	< 50 ppm
Frequency	±0.2%

Unless specified otherwise, the accuracy of the signals generated by the test equipment shall not exceed the limits given in Table 5.

Table 5 – Accuracy of the signals

Quantity	Accuracy
Sound pressure level at MRP	±3 dB (100 Hz to 200 Hz) ±1 dB (200 Hz to 8000 Hz)
Electrical excitation level	±0.4 dB (Note 1)
Frequency generation	±2% (Note 2)
Clock accuracy	< 50 ppm
Specified component values	±1%
NOTE 1 – Across the whole frequency range. NOTE 2 – When measuring sampled systems, it is advisable to avoid measuring at sub-multiples of the sampling frequency. There is a tolerance of ±2% on the generated frequencies, which may be used to avoid this problem, except for 8000 Hz, where only –2% tolerance may be used.	

The measurement results shall be corrected for the measured deviations from the nominal level.

For terminal equipment which is directly powered from the mains supply, all tests shall be carried out within ±5% of the rated voltage of that supply. If the equipment is powered by other means and those means are not supplied as part of the apparatus, all tests shall be carried out within the power supply limit declared by the supplier. If the power supply is a.c., the test shall be conducted within ±4% of the rated frequency.

Due to the time variability feature of the IP connection, delay variation may impair measurements which are sensitive to the delay variation. In that case, the test lab should make sure that the correct delay value is used.

5 Technical requirements

5.1 Codec independent parameters

5.1.1 Sending characteristics

5.1.1.1 Sending sensitivity/frequency characteristics

5.1.1.1.1 Requirement

The sending sensitivity/frequency characteristics shall fall between the upper and lower limits given in Table 6, and shown in Figure 9. All sensitivities are in dB on an arbitrary scale.

Table 6 – Sending sensitivity/frequency mask

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	4	
125	4	–10
200	4	–4
1000	4	–4
5000	(Note)	–4
6300	9	–7
8000	9	
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a logarithmic (frequency) – linear (dB) scale.		

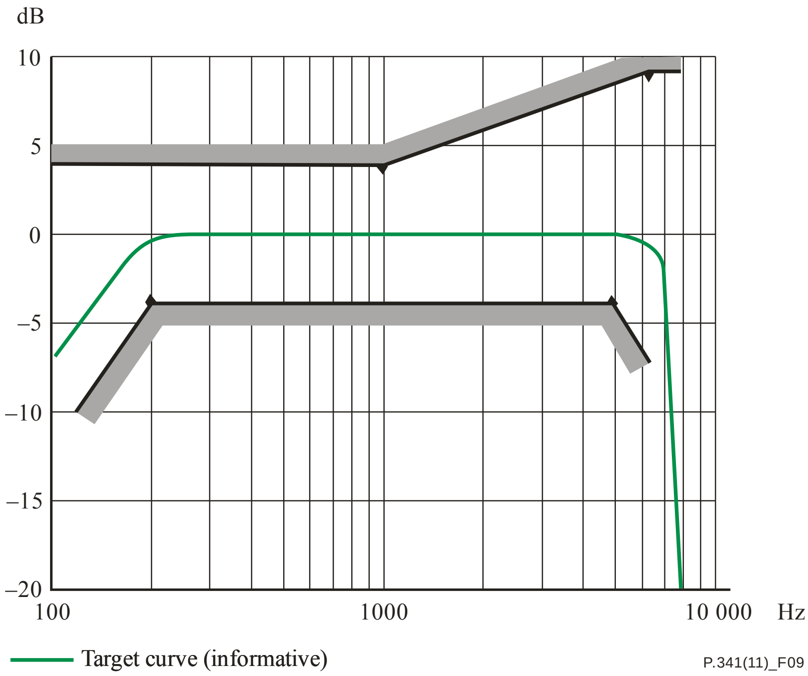


Figure 9 – Hands-free sending sensitivity/frequency mask

5.1.1.1.2 Measurement method

The terminal is set up as specified in clause 4.3.

The test signal to be used for the measurements shall be the artificial voice according to [ITU-T P.50] or a speech-like signal described in [ITU-T P.501].

The testing signal level shall be –4.7 dBPa at MRP. The spectrum of acoustic signal produced by the artificial mouth is calibrated under free-field conditions at MRP. The signal level is adjusted according to clause 4.4.2.

The spectrum at MRP and the actual level at MRP is used as reference to determine the send sensitivity S_{mJ}. The sending sensitivity shall be calculated for each band of the 20 frequencies given in Table A.2 of [ITU-T P.79], bands 1 to 20. The averaged measured level for each frequency band (*Fi*) is referred to the averaged test signal level in each frequency band measured at MRP as shown in the equation below:

$$S_{mJ} = 20\lg(V_J / P_m) + (Corr - 24) \text{ dB rel 1V/Pa} \tag{1}$$

where:

S_{mJ}	is the sending sensitivity from MRP to the interface output at *Fi*
V_j	is RMS voltage of the interface output equivalent at *Fi*
P_m	is the sound pressure at MRP (with a total level of –4.7 dBPa) at *Fi*
(*Corr*–24)	is the correction factor of the artificial mouth as a result of level adjustment described in clause 4.4.2

NOTE – The formula is for the HFRP test position, the correction factor for other types of hands-free terminals can be obtained from Table 3.

5.1.1.2 Sending loudness rating (SLR)

5.1.1.2.1 Requirement

The nominal value of SLR shall be 13 dB, with a tolerance of ±3 dB.

5.1.1.2.2 Calculation

The sending loudness rating (SLR) shall be calculated according to Annex A of [ITU-T P.79], based on the measurement of the sending sensitivity/frequency in clause 5.1.1.1.2:

$$SLR = -\frac{10}{m} \times \lg \sum_{i=1}^{20} 10^{\frac{m}{10}(S_{mJ} - W_{Si})} \quad \text{dB} \tag{2}$$

where:

m	= 0.175
S_{mJ}	is the sending sensitivity measured in clause 5.1.1.1.2
W_{Si}	is the sending weighting factor from Table A.2 of [ITU-T P.79].

5.1.1.3 Sending noise

5.1.1.3.1 Requirement

The noise in the sending direction shall not exceed –64 dBm0 (A).

No peaks in any 1/3-octave band, with a level of 10 dB higher than the average noise spectrum in the frequency domain, shall occur.

5.1.1.3.2 Measurement method

The measurement set-up is described in clause 4.2.

For the actual measurement, no test signal is used. After a correct activation, the noise level at the digital interface output or the wideband reference codec output is measured in the frequency range from 100 Hz to 8000 Hz with A-weighting according to [IEC 61672-2].

The analysis window is applied directly after stopping the activation signal but taking into account the influence of all acoustical components (reverberations). The test house has to ensure (e.g., by monitoring the time signal) that during the test, the terminal remains in activated condition. If the terminal is deactivated during the measurement, the measurement time has to be reduced to the period where the terminal remains in an activated condition.

5.1.1.4 Sending distortion

5.1.1.4.1 Requirement

The distortion in the sending direction shall be measured in terms of the total distortion arising from the application of 1020 Hz tones separately. The limits shall be as shown in Table 7.

Table 7 – Limit for signal-to-total distortion ratio, sending direction

Frequency (Hz)	Sending level (dBPa at MRP)	Sending ratio (dB)
1020	–20	27
	–15	30
	–10	33
	–4.7	35
	0	35
	+5	30
NOTE – The limits for intermediate frequencies lie on straight lines drawn between the given values on a logarithmic (frequency) – linear (dB) scale.		

5.1.1.4.2 Measurement method

The measurement set-up is described in clause 4.2.

After correct activation of the system, a sine wave signal at frequencies of 1020 Hz is applied at MRP. The signal at MRP shall be at the following levels: –20, –15, –10, –4.7, 0, 5 dBPa.

The ratio of the signal-to-total distortion power of the interface output (sending) is measured with A-weighting according to [IEC 61672-2]. The weighting function shall be applied to the total distortion component only (not to the signal component).

NOTE 1 – Depending on the type of codec, the test signal used may need to be adapted. If a sine wave is not usable, an alternative test signal could be a band limited noise signal centred on the above frequencies.

NOTE 2 – It should be ensured that the test signal is treated by speech processing algorithms as a speech-like signal, and not a noise-like signal. Test signal with a time-stationary envelope may be treated by certain algorithms, e.g., noise suppression algorithms, as a noise-like signal.

5.1.2 Receiving characteristics

5.1.2.1 Receiving sensitivity/frequency characteristics

5.1.2.1.1 Requirement

The receiving sensitivity/frequency characteristics for wideband hands-free terminals shall fall between the upper and lower limits given in Table 8, and shown in Figure 10. All sensitivities are in dB on an arbitrary scale.

Table 8 – Hands-free receiving sensitivity/frequency mask

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
125	8	
200	8	–12
250	8	–9
315	7	–6
400	6	–6
5000	6	–6
6300	6	–9
8000	6	
NOTE – The limits for intermediate frequencies lie on straight lines drawn between the given values on a logarithmic (frequency) – linear (dB) scale. NOTE 2 – This Recommendation uses the free-field as the reference point instead of ERP.		

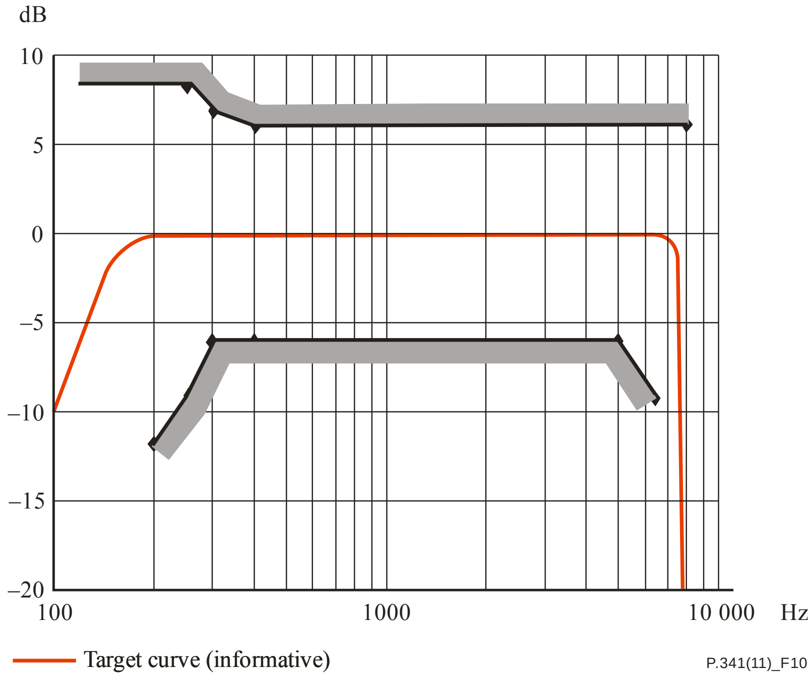


Figure 10 – Hands-free receiving sensitivity/frequency mask

5.1.2.1.2 Measurement method

The HFT is placed on the test table as specified in clause 4.3.

The test signal to be used for the measurements shall be the artificial voice according to [ITU-T P.50]. If the signal-to-noise ratio in the high frequency domain is not sufficient, the CSS defined in [ITU-T P.501] shall be used.

The receiving sensitivity/frequency characteristic is measured according to [ITU-T P.64] over a minimum range of 100 Hz to 8000 Hz. The test signal level shall be –16 dBm0 at the interface input (receiving) over the complete test sequence length. The free-field correction of HATS, as defined in [ITU-T P.581], is applied using the reverse nominal curve given in Table 3 of [ITU-T P.58]. The equalized output signal is power-averaged on the total time of analysis.

The receiving sensitivity shall be calculated for each band of the 20 frequencies given in Table A.2 of [ITU-T P.79], bands 1 to 20. For calculation, the averaged measured level for each frequency band is referred to the averaged test signal level in each frequency band.

$$S_{Jeff} = 20\lg\left(P_e^{'} / E_J\right) \quad \text{dB rel 1 Pa/V} \tag{3}$$

where:

S_{Jeff} is the receiving sensitivity from the interface input (receiving) to HATS ear with free-field correction at *Fi*

$P_e^{'}$ is free-field equalized sound pressure at *Fi*, converted from the measurement data P_e at DRP

E_J is RMS input voltage at *Fi*

5.1.2.2 Receiving loudness rating (RLR)

5.1.2.2.1 Requirement for desktop HFT

The nominal value of RLR shall be 5 dB, with a tolerance of ±3 dB.

In case the HFT has a volume control, it shall be satisfied that:

1) RLR ≤ –2 dB at the upper part of the volume range.
2) The range of volume control ≥ 15 dB.

5.1.2.2.2 Requirement for group-audio HFT

The nominal value of RLR shall be 5 dB, with a tolerance of ±3 dB.

In case the HFT has a volume control, it shall be satisfied that:

1) RLR ≤ –6 dB at the upper part of the volume range.
2) The range of volume control ≥ 19 dB.

5.1.2.2.3 Measurement method

The receiving loudness rating (RLR) shall be calculated according to Annex A of [ITU-T P.79], based on the measurement of the receiving sensitivity/frequency in clause 5.1.2.1.2.

$$RLR = -\frac{10}{m} \times \lg \sum_{i=1}^{20} 10^{\frac{m}{10}(S_{JE} - W_{Ri})} \tag{4}$$

where:

m = 0.175

S_{JE} is the receiving sensitivity measured in clause 5.1.2.1.2

W_{Ri} is the receiving weighting factor from Table A.2 of [ITU-T P.79]

According to [ITU-T P.340], the calculated RLR shall be corrected by subtracting 14 dB and without the L_e factor for hands-free terminals measurement when only using one artificial ear of HATS.

According to [ITU-T P.581], if both HATS artificial ears are used, the equalized output signal of each ear is power-averaged on the total time of analysis. The "right" and "left" signals are voltage-summed for each 1/3 octave frequency band. The correction factor has to be 8 dB, instead of 14 dB.

NOTE – The 8 dB correction results from the 14 dB correction, as specified in [ITU-T P.340], subtracted by 6 dB due to the voltage summation of the signals measured at the two artificial ears.

5.1.2.3 Receiving noise

5.1.2.3.1 Requirement

When no signal is transmitted or driven by a signal corresponding to a "quietcode" value for the decoder, the A-weighted maximum acoustic noise level with free-field equalization active shall be as follows:

– if no user-controlled volume control is provided or when the volume control is set to nominal RLR value, the measured receiving noise shall not be greater than –54 dBPa(A);

– the level in any 1/3-octave band, between 100 Hz and 10 kHz shall not exceed a value of –64 dBPa;

– no peaks in any 1/3-octave band with the level of 10 dB higher than the average noise spectrum in the frequency domain shall occur.

5.1.2.3.2 Measurement method

The measurement set-up is described in clause 4.2 and any comfort noise shall be disabled.

A signal corresponding to a "quietcode" value for the decoder is applied at the interface. After a correct activation, the equalized output of the right ear is measured in the frequency range from 100 Hz to 10 kHz with A-weighting according to [IEC 61672-2].

The analysis window is applied directly after stopping the activation signal but taking into account the influence of all acoustical components (reverberations). The test house has to ensure (e.g., by monitoring the time signal) that during the test the terminal remains in an activated condition. If the terminal is deactivated during the measurement, the measurement time has to be reduced to the period where the terminal remains in activated condition.

5.1.2.4 Receiving distortion

5.1.2.4.1 Requirement

The distortion in the receiving direction shall be measured in terms of the total distortion (harmonic, quantizing and noise) arising from the application of 1020 Hz tones. The limits shall be as shown in Table 9. If a receiving volume control is provided, the requirement applies at a setting as close as possible to the nominal value of RLR.

Table 9 – Limit for signal-to-total distortion ratio, receiving direction

Frequency (Hz)	Receiving level at the digital interface (dBm0)	Receiving ratio (dB)
1020	–45	[17.5] (Note 2)
	–40	[22.5] (Note 2)
	–30	[30.5]
	–20	[33.0]
	–10	33.5
	–3	31.2
	0	25.5
NOTE 1 – The limits for signal-to-total distortion ratio for intermediate receiving levels lie on straight lines drawn between the given values on a linear (dB receiving level) – linear (dB ratio) scale. NOTE 2 – For levels –40 and –45 dBm0, the stated limits are recommendations; hence a lower signal-to-distortion ratio shall not be regarded as a failing result. However, the obtained results shall be reported. NOTE 3 – By convention within this table, values enclosed in square brackets remain preliminary definitions and therefore are not in force.		

5.1.2.4.2 Measurement method

The measurement set-up is described in clause 4.2.

After a correct activation of the system, a sine wave signal at frequency of 1020 Hz is applied to the interface input. The signal shall be at the following levels: –45, –40, –30, –20, –10, –3, 0 dBm0. The test signals have to be applied from high levels down to low levels.

The total distortion power shall be measured with A-weighting according to [IEC 61672-2]. The ratio of the signal-to-total distortion power is calculated.

NOTE – Depending on the type of codec, the test signal used may need to be adapted. If a sine wave is not usable, an alternative test signal could be a band limited noise signal centred on the above frequencies.

5.1.3 Echo path loss characteristics

5.1.3.1 Weighted terminal coupling loss (TCLw)

5.1.3.1.1 Requirement

The TCLw shall be ≥ 46 dB at the nominal setting of the user selectable volume control.

With the volume control set to maximum, TCLw shall be ≥ 40 dB. The volume control shall be set back to nominal after each call automatically unless TCLw ≥ 46 dB can be maintained also with maximum volume setting.

NOTE – The echo impairment perceived by the person at the opposite end of the connection from a telephone set is a function of the magnitude of the talker echo signal as well as the talker echo path delay. The echo signal becomes more disturbing as the talker echo path delay increases. In consideration of the increasing delays introduced by modern networks, a higher TCLw value than that specified here may be necessary for proper operation with these networks.

5.1.3.1.2 Measurement method

The measurement set-up is described in clause 4.2. For hands-free measurement, HATS is positioned but not used. For loudspeaking measurement, the handset is positioned on HATS (right ear).

Before the actual test, an activation signal described in clause 4.3 is applied to the interface input as a training sequence.

The test signal may be a pseudorandom noise (PN) sequence as defined in [ITU-T P.501] with a length of 4096 points (for the 48 kHz sampling rate) and a crest factor of 6 dB. The low crest factor is achieved by random alternation of the phase between –180° and 180°. The length of the complete test signal composed of at least four sequences of CSS shall be at least one second. The test signal level is –3 dBm0. The calibration shall be determined during the ON portions of the signal. The test signal shall be band-limited to 50 Hz-7000 Hz.

The attenuation from the interface input (receiving) to the interface output (sending) is measured. The TCLw is calculated according to [ITU-T G.122] but using the frequency range of 300 Hz to 6700 Hz (instead of 300 Hz to 3400 Hz). For the calculation, the averaged measured echo level at each frequency band is referred to the averaged test signal level in each frequency band. For the measurement, a time window has to be applied adapted to the duration of the actual PN sequence of the test signal (200 ms) choosing the PN sequence of the third CSS signal.

5.1.3.2 Stability loss

5.1.3.2.1 Requirement

Stability loss is the minimum loss from the interface input (receiving) to the interface output (sending), at any test frequency.

Stability loss shall be at least 6 dB at all frequencies in the range 100 Hz to 8000 Hz and at all settings of the receiving volume control, if provided.

5.1.3.2.2 Measurement method

For desktop hands-free terminals, the test set-up described in TCLw measurement is used.

For loudspeaking mode terminals, the handset is placed at 50 cm beside the terminal with the transducer facing the table.

For HFT with double or more units, the different pieces of the HFT shall be placed as close as possible to each other, but without modifying the normal use configuration of the HFT.

The same test sequence employed in TCLw measurement is applied in stability loss measurement.

The attenuation from the interface input (receiving) to the interface output (sending) is measured for frequencies from 100 Hz to 8000 Hz. Stability is the least value of the attenuation in the band 100 Hz to 8000 Hz.

5.2 Codec dependent parameters

5.2.1 Delay

5.2.1.1 Description

Latency of packet-based network is much longer than traditional digital network, especially when there is congestion in the transmission path. Under this circumstance, this Recommendation addresses delay contributed by VoIP terminal to help reduce the total end-to-end delay.

Certain aspects of delay can be optimized in VoIP telephones, such as the internal delay of the hardware or firmware and the optimization of the jitter buffer operation, which must trade off the impairment of packet loss against the expected delay variation of the far-end telephone or the network. While other aspects, such as packetization and depacketization, are also important sources of delay, they are a function of the selected codec and the number of speech frames per packet, so they cannot be optimized in VoIP telephones. Therefore, this Recommendation now specifies delay in terms of categories for network planning purposes, similar to [ITU-T P.1010].

Delay may be time variant. Therefore constant monitoring of the actual delay may be required when evaluating the range of delay which can be observed in a given connection.

5.2.1.2 Sending delay

5.2.1.2.1 Requirement

For a VoIP terminal, send delay is defined as the one-way delay from the acoustical input (mouthpiece) of this VoIP terminal to its interface to the packet-based network. The total send delay is the upper bound on the mean delay and takes into account the delay contributions of all of the elements shown in Figure 2 of [ITU-T G.1020].

Four categories are given in this Recommendation:

- Category A: Ts ≤ 25 ms
- Category B: Ts ≤ 35 ms
- Category C: Ts ≤ 50 ms
- Category D: Ts > 50 ms

For ITU-T G.722 coding algorithm (20 ms payload's length in one packet), at least Category C is required. For terminals which include a wireless transmission link at least, Category C is recommended.

NOTE – The functions that contribute the sending delay are the encoding, packetization (payload's length in each IP packet) and signal processing.

5.2.1.2.2 Measurement

The measurement set-up is described in clause 4.2.

The test signal to be used for the measurements shall be a composite source signal (CSS) as described in [ITU-T P.501]. The test signal consists of the voiced part followed by a pseudorandom noise sequence with a minimum periodicity of 500 ms. The spectrum of acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be –4.7 dBPa, measured at the MRP. The test signal level is averaged over the complete test signal sequence.

NOTE – If the expected delay is higher than 500 ms, a pseudorandom noise sequence with a higher periodicity should be used.

The delay is calculated using the cross-correlation function between the signal at the digital interface input and the signal at MRP. The measurement is corrected by the delay introduced by the test equipment. The delay is expressed in milliseconds, determined from the maximum of the cross-correlation function.

5.2.1.3 Receiving delay

5.2.1.3.1 Requirement

For a VoIP terminal, receive delay is defined as the one-way delay from the interface to the packet-based network of this VoIP terminal to its acoustical output (earpiece). The total receiving delay is the upper bound on the mean delay and takes into account the delay contributions of all of the elements shown in Figure 2 of [ITU-T G.1020].

Four categories are given in this Recommendation:

- Category A: Tr ≤ 30 ms
- Category B: Tr ≤ 65 ms
- Category C: Tr ≤ 100 ms
- Category D: Tr > 100 ms

For ITU-T G.722 coding algorithm (20 ms payload's length in one packet), at least Category C is required. For terminals which include a wireless transmission link, e.g., WiFi or Bluetooth link, at least Category C is recommended.

NOTE – The functions that contribute the receiving delay are the encoding, depacketization (payload's length in each IP packet), jitter buffering and signal processing.

5.2.1.3.2 Measurement

The measurement set-up is described in clause 4.2.

The test signal to be used for the measurements shall be a composite source signal (CSS) as described in [ITU-T P.501]. The test signal consists of the voiced part followed by a pseudorandom noise sequence with a minimum periodicity of 500 ms. The test signal level shall be –16 dBm0, measured at the electrical test point. The test signal level is averaged over the complete test signal sequence. The delay is calculated using the cross-correlation function between the signal at the digital interface output and the signal at DRP acquired by artificial ear. The measurement is corrected by the delay introduced by the test equipment. The delay is expressed in milliseconds, determined from the maximum of the cross-correlation function.

[ITU-T G.114] recommends that the maximum end-to-end delay should be less than 150 ms, to minimize the effect on the dynamics of conversations. However, it is desirable to keep the delays as low as possible. Considering the various coding schemes and different network applications, it is recommended that any combination of different vendor sending and receiving delay shall be under 100 ms.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Terminals and subjective and objective assessment methods**
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2012