I n t e r n a t i o n a l   T e l e c o m m u n i c a t i o n   U n i o n

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.311

(03/2011)

SERIES P: TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Voice terminal characteristics

Transmission characteristics for wideband digital handset and headset telephones

Recommendation ITU-T P.311

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.311

Transmission characteristics for wideband digital handset and headset telephones

Summary

Recommendation ITU-T P.311 provides audio performance requirements for wideband (8000 Hz) handset and headset 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, sidetone, stability, echo path and delay. Wideband audio represents a considerable departure from traditional narrow-band telephony, offering significantly improved quality.

Major 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 used for better simulation of real telephone communication.

– Specific headset requirements are incorporated.

– Some technical requirements and related test methods have been modified based on the latest research, e.g., frequency response, distortion, STMR, *D*-factor, and delay.

History

Edition	Recommendation	Approval	Study Group
1.0	ITU-T P.311	1995-04-18	12
2.0	ITU-T P.311	1998-02-27	12
3.0	ITU-T P.311	2005-06-06	12
4.0	ITU-T P.311	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/.

Table of Contents

Recommendation ITU-T P.311

Transmission characteristics for wideband digital handset and headset telephones

1 Scope

This Recommendation provides audio performance requirements and test methods for digital handset and headset telephones capable of transmitting an audio bandwidth extending beyond the conventional telephony bandwidth of 300 Hz 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, video conferencing 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.

Conventional telephone band (300 Hz-3400 Hz) digital handset telephones using encoding according to [ITU-T G.711] and [ITU-T G.726] are covered by [ITU-T P.310].

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

[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 Adaptative 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.380] Recommendation ITU-T P. 380 (2003), *Electro-acoustic measurements on headsets*.

[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*.

[ETSI TS 126 171] ETSI TS 126 171 V9.0.0 (2010), *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 diffuse field equalization: Equalization of the HATS sound pick-up, equalization of the difference, in dB, between the spectrum level of the acoustic pressure at the ear-drum reference point (DRP) and the spectrum level of the acoustic pressure at the HATS reference point (HRP) in a diffuse sound field with the HATS absent using the reverse nominal curve given in Table 3 of [ITU-T P.58].

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

3.1.3 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.

3.1.4 HATS position: The HATS position is the correct handset position in handset measurement, it will give more realistic results compared with LRGP. See Annexes D, E and F of [ITU-T P.64] for detailed information.

3.1.5 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.

3.1.6 wideband telephony: Transmission of speech with a nominal passband wider than 300 Hz-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

ERP Ear Reference Point

LRGP Loudness Rating Guard Ring Position

LSTR Listener Sidetone Rating

MRP Mouth Reference Point

NR Noise Reduction

PN Pseudorandom Noise

RLR Receiving Loudness Rating

SLR Sending Loudness Rating

STMR Sidetone Masking Rating

TCL Terminal Coupling Loss

TCLw Weighted Terminal Coupling Loss

4 Test arrangement

4.1 Test environment

Unless stated otherwise, measurements shall be conducted under anechoic or quiet conditions:

For some headsets or handset terminals with small dimension, an anechoic room will be required. An anechoic room used in wideband test shall be practically free-field down to a lowest frequency of 150 Hz, the handset or headset coupled with the HATS shall lie totally within this free-field volume. Satisfied free-field conditions are met if deviations from the ideal free-field conditions are less than ±1 dB.

Depending on the distance from terminal transducers to mouth and ear, a quiet office room may be sufficient, e.g., for handsets where the artificial mouth and artificial ear are located close to the acoustical transducers. In that case, a test room may be used which meets the following criteria:

1) The difference of sound pressure between the mouth opening and that at 5.0, 7.5 and 10 cm in front of the centre of the lip ring is within ±0.5 dB of that which exists in a known acoustic free-field.
2) The difference of sound pressure between the mouth opening and the ear canal entrance point (EEP) at both the left and right ears of the HATS does not differ by more than ±1 dB from that which exists in a known free-field.

The ambient noise level shall be less than –64 dBPa(A) for all the tests, except for the condition when a simulated or real noise is used as part of the test environment. The test environment shall be free of mechanical disturbances.

4.2 Test set-up

4.2.1 General description

To achieve the most realistic simulation of the "average" subscriber, all measurements are performed by using HATS. Although HATS is thoroughly described in [ITU-T P.58], the exact calibration and equalization of HATS can be found in [ITU-T P.581]. Note that the horizontal positioning of the HATS reference plane shall be guaranteed within ±2°.

The appropriate ear simulator is type 3.3 or 3.4 which are detailed in [ITU-T P.57]. The type of ear simulator and application force shall be indicated in the test report.

The interface in Figure 1 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 according to [ITU-T I.412] or IP terminals), to a form compatible with the test equipment. Interfaces can be applied for the sending and receiving signals 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 function 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: the direct approach and the reference codec approach. In the direct approach, the companded digital input or 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 or output bitstream of the telephone set to the equivalent analogue values, so that the 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 bit rate is used, the bit rate recognized as giving the best characteristics should be adopted. 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.

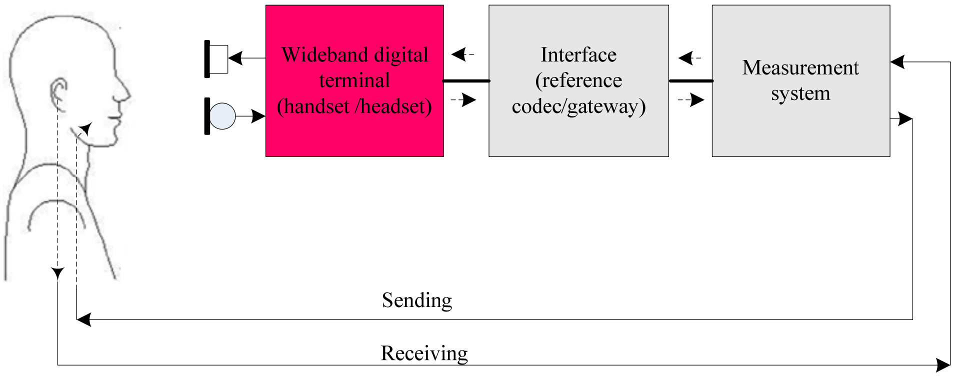

Figure 1 – Test configuration

The set is connected to the interface and is placed in the active call state. Handsets or headsets 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 handsets

Proper positioning of handsets on HATS can be found in Annexes D and E of [ITU T P.64]. Unless stated otherwise, an application force of 8 N is used to handset measurement.

4.2.3 Set-up for headsets

Suggestions for positioning headsets on HATS are given in [ITU-T P.380]. The manufacturer should provide the recommended wearing position (RWP) in the user's guide, which describes a precise way the device should be placed on the user's head. From RWP, the recommended test position (RTP) information can be derived, as close as possible to RWP. Also the RTP description can be provided by the manufacturer which includes information about how the receiving part of the headset should be placed against or inside the ear simulator, and further describes the positioning and orientation of the microphone. If neither RTP nor RWP are provided by the manufacturer, the headset shall be tested in the estimated real use position (ERUP). The test lab shall define an ERUP that closely approximates real use. Natural headband pressure, or other positioning techniques normally used by a real user, shall be used. Because the headset position may affect test results, the test shall be repeated at least 5 times by completely repositioning the headset, following the rules described in [ITU-T P.380].

For binaural headsets, the following applies:

– For headsets that do not have left and right ear wearing indication, they can be tested on HATS that is equipped with one functional ear simulator. The primary receiver (where the microphone system is located) should be positioned on the functional ear simulator for all tests. The secondary receiver shall be moved into position for evaluating frequency response, RLR and distortion measurements.

– For headsets that are designed to be worn in only one manner or fashion, a HATS with two functional ear simulators shall be used. The two receivers shall be tested separately. The secondary receiver need only be tested for frequency response, RLR and distortion.

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 wideband terminals, all test signals which are inserted in the 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 signals (CSS) are 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 in accordance with [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

The signal level refers to the RMS level of the test signal averaged over the complete test sequence length if not described otherwise. The use of test signal levels should be stated in the test report.

Unless stated otherwise, the test signal shall be set to –4.7 dBPa at MRP for the sending direction and –16 dBm0 for the receiving direction. The spectrum of acoustic signal produced by the artificial mouth is calibrated under free-field conditions at MRP.

The level of the activation signal shall be –4.7 dBPa at MRP or –16 dBm0 at the interface input (receiving). Alternatively other speech like test signals with the same signal level can be used for activation.

4.5 Accuracy of measurements

Unless specified otherwise, the accuracy of measurements made by test equipment shall not exceed the limits given in Table 1.

Table 1 – Accuracy of measurements

	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%
Application force	±2 N

Unless otherwise specified, the accuracy of the signals generated by the test equipment shall not exceed the limits given in Table 2.

Table 2 – Accuracy of 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 Handset 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 characteristic shall fall between the upper and lower limits given in Table 3 and shown in Figure 2. All sensitivities are in dB on an arbitrary scale.

Table 3 – Sending sensitivity/frequency mask

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	0	
200	5	–5
5000	5	–5
6300	5	–10
8000	5	
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a logarithmic (frequency) – linear (dB) scale.		

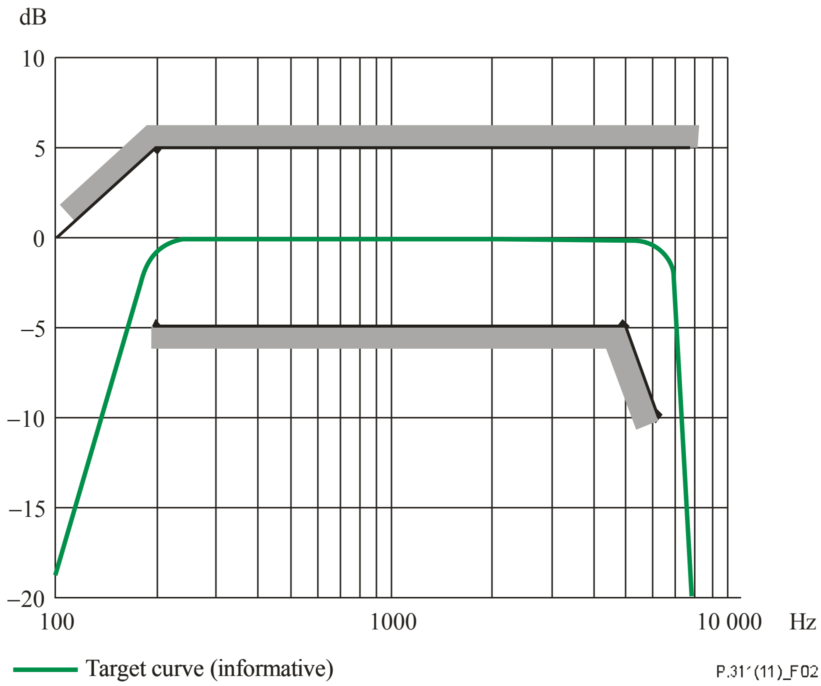

Figure 2 – Handset sending sensitivity/frequency mask

5.1.1.1.2 Measurement method

The measurement set-up is described in clause 4.2.

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 composite source signal (CSS), as defined in [ITU-T P.501], shall be used.

The sending sensitivity/frequency characteristic is measured according to [ITU-T P.64] over a minimum range of 100 Hz to 8000 Hz. The spectrum of acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The testing signal level shall be –4.7 dBPa at MRP. The test signal is averaged over the complete test signal sequence. 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 the MRP, as shown in the equation below:

$$S_{mJ} = 20\log\frac{V_J}{P_m}\ \text{dB rel 1V/Pa} \qquad (1)$$

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 of MRP at *Fi*

NOTE – P_m must be measured in the absence of the "unknown" handset of the test item.

5.1.1.2 Sending loudness rating (SLR)

5.1.1.2.1 Requirement

The nominal value of SLR shall be 8 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 sending sensitivity/frequency in clause 5.1.1.1 with the formula below:

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

where:

m = 0.175

S_{mJ} is the sending sensitivity measured in clause 5.1.1.1

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 –68 dBm0(A).

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.

5.1.1.3.2 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 interface output is measured in the frequency range from 100 Hz to 8000 Hz with A-weighting according to [IEC 61672-2]. The peak noise level is measured in dBm0(C).

The analysis window is applied directly after stopping the activation signal but taking into account the influence of all acoustical components (reverberations). Care should be taken (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 window has to be adjusted within 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 315 Hz, 510 Hz and 1020 Hz tones separately. The limits shall be as shown in Table 4.

Table 4 – Limit for signal-to-total distortion ratio, sending direction

Frequency (Hz)	Sending level (dBPa at MRP)	Sending ratio (dB)
315	–4.7	28
510	–4.7	32
1020	−20	27
	−15	30
	−10	33
	–4.7	35
	0	35
	+5	30
NOTE – 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.		

5.1.1.4.2 Measurement method

The measurement set-up is described in clause 4.2.

After a correct activation of the system, a sine wave signal at frequencies of 315 Hz, 510 Hz and 1020 Hz is applied at MRP respectively. The signal level shall be calibrated to –4.7 dBPa at MRP for all frequencies, except for the sine wave signal with a frequency of 1020 Hz that shall be applied at MRP at the following levels: –20, −15, −10, –4.7, 0, 5 dBPa. The test signals have to be applied from high levels down to low levels.

The ratio of the signal-to-total distortion power of the digital signal output 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 envelop 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 shall fall between the upper and lower limits given in Table 5, and shown in Figure 3. All sensitivities are in dB on an arbitrary scale.

Table 5 – Receiving sensitivity/frequency mask

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	3	
120	3	–5
200	3	–5
400	3	–5
1010	(Note 1)	–5
1200	(Note 1)	–8
1500	(Note 1)	–8
2000	9	–3
3200	9	–3
7000	9	–13
8000	9	

NOTE 1 – 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 diffuse-field as the reference point instead of ERP.

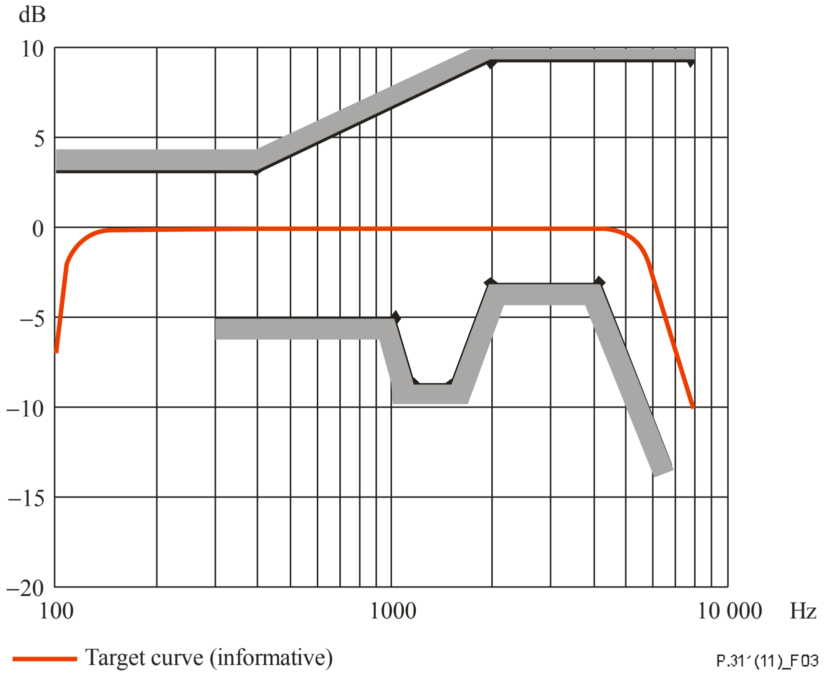

Figure 3 – Handset receiving sensitivity/frequency mask

5.1.2.1.2 Measurement method

The measurement set-up is described in clause 4.2.

The test signal 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, CSS, as 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 digital interface input over the complete test sequence length. The diffuse-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 over 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.

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

where:

S_{Jedf} is the receiving sensitivity from the interface input or the wideband reference codec input to HATS ear with diffuse field correction at Fi

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

E_J s RMS input voltage at Fi

5.1.2.2 Receiving loudness rating (RLR)

5.1.2.2.1 Requirement

The nominal value of RLR shall be 2 dB, with a tolerance of ±3 dB.

5.1.2.2.2 Calculation

The receiving loudness rating (RLR) shall be calculated according to Annex A of [ITU-T P.79], based on the measurement of receiving sensitivity/frequency in clause 5.1.2.1. Instead of diffuse-field equalization, the DRP-ERP correction as defined in [ITU-T P.57] shall be applied. No leakage correction shall be applied for the measurement.

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

where:

m = 0.175

S_{JE} is the receiving sensitivity from the digital interface input to the output of artificial ear at Fi with DRP/ERP correction, as measured in clause 5.1.2.1

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

5.1.2.3 Receiving noise

5.1.2.3.1 Requirement

When no signal is transmitted, the A-weighted maximum acoustic noise level, measured at DRP with diffuse-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 receiving noise shall not be greater than −57 dBPa(A);
– if a volume control is provided, the measured receiving noise shall not be greater than –54 dBPa (A) at the maximum setting of the volume control;
– 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.

After correct activation, the noise level is measured at DRP of the artificial ear with diffuse-field equalization active in the frequency range from 100 Hz to 8000 Hz. The measured noise signal is A-weighted 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). Care should be taken (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 adjusted within the period where the terminal remains in an activated condition.

5.1.2.4 Receiving distortion

5.1.2.4.1 Requirement

The distortion in the receiving direction shall be measured at DRP with diffuse-field equalization active in terms of the total distortion. The measured distortion power is A-weighted according to [IEC 61672-2]. The limits shall be as shown in Table 6, unless the signal in the artificial ear exceeds +10 dBPa or is less than −50 dBPa.

If a receiving volume control is provided, the requirements apply at a setting as close as possible to the nominal value of RLR, as specified in clause 5.1.2.2.

Table 6 – Limit for signal-to-total distortion ratio, receiving direction

Frequency (Hz)	**Receiving level at the digital interface (dBm0)**	**Receiving ratio (dB)**
315	−16	[20]
510	−16	[28]
1020	–45	[17.5] (Note 2)
	–40	[22.5] (Note 2)
	–30	[30.5]
	–20	[33.0]
	–10	33.5
	–3	31.5
	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 correct activation of the system, a sine wave signal at frequencies of 315 Hz, 510 Hz and 1020 Hz is applied to the digital interface input. The signal level is –16 dBm0 except for the sine wave signal with a frequency of 1020 Hz that shall be applied to the interface at the following levels: –45, –40, –30, –20, –10, –3, 0 dBm0. The test signals have to be applied from high levels to low levels.

The total distortion power shall be measured at DRP with diffuse-field equalization active. The distortion power is A-weighed. The DRP-DF correction, as defined in [ITU-T P.58], is applied. The ratio of the signal-to-total distortion power is calculated.

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.

5.1.3 Sidetone characteristics

5.1.3.1 Sidetone masking rating (STMR)

5.1.3.1.1 Requirement

The nominal value of STMR shall be 18 dB with a tolerance of ±5 dB for nominal setting of the volume control.

For all other positions of the volume control, the STMR must not be below 8 dB.

It is recommended to have the sidetone level independent of the user receive volume control setting.

5.1.3.1.2 Measurement method

The measurement set-up is described in clause 4.2.

The acoustic test signal is applied at MRP. If a receiving volume control is provided, the measurements shall be carried out with the control at a setting as close as possible to the nominal value of RLR, as specified in clause 5.1.2.2.

In addition, the measurement is repeated at the maximum volume control setting to simulate the worst situation.

The sound pressure in HATS ear simulator is measured and corrected to sound pressure levels at ERP using the DRP-ERP correction defined in [ITU-T P.57]. The application force shall be 13 N on the artificial ear.

The sidetone sensitivity shall be calculated for each band of the 20 frequencies given in Table 3 of [ITU-T P.79], bands 1 to 20. The averaged measured level at ERP for each frequency band is referred to the averaged test signal level at MRP in the corresponding frequency band.

STMR is then calculated according to the formula below:

$$STMR = -\frac{10}{m} \times \log_{10} \sum_{M=1}^{20} 10^{(m/10)(S_{meST} - L_E - W_{ML})} \qquad (5)$$

where:

m = 0.225

S_{meST} is the sidetone sensitivity from MRP to ERP via the electrical sidetone path at Fi

W_{ML} is the sidetone weighting factor (unsealed) from Table B.2 of [ITU-T P.79]

L_E = 0

5.1.3.2 *D*-factor

5.1.3.2.1 Requirement

The recommended value for the stationary pink noise is: D-factor(pink) ≥ 0 dB. As a long-term objective, the value of +3 dB is recommended.

NOTE 1 – The parameter of D-factor is meaningful on the presupposition that the noise suppression in the terminal is not affected by any transmission of speech signal. In fact, if the simultaneously transmitted speech will influence the noise reduction (NR) processing, especially in the sending direction, an advanced measurement method should be used to determine NR effect.

NOTE 2 – This clause applies to terminals providing narrow-band and wideband telephony. However, the use of D-factor is defined only over narrow-band frequency range. Thus, the test method for D-factor is the same for either narrow-band or wideband telephony.

NOTE 3 – This Recommendation requires proper D-factor value only for stationary pink noise. For other realistic noises, the D-factor value may be different and the requirement depends on the customer's requirement.

5.1.3.2.2 Measurement method

The measurement method is defined according to Annex E of [ITU-T P.79].

The D-factor is computed directly from measurements of the difference Δ_{Sm} between the sending sensitivities for diffuse and direct sound, S_{si} (*diff*) and S_{si} (*direct*), respectively.

$$\Delta_{Sm} = S_{si}(diff) - S_{si}(direct) \qquad (6)$$

Both S_{si} (*diff*) and S_{si} (*direct*) can be measured according to clause 5.1.1.1, except that diffuse sound is used as the input signal when measuring S_{si} (*diff*). The sound pressure level of the diffuse sound field shall be adjusted to –24 dBPa(A) at MRP in the absence of HATS. The tolerance on the level shall be ±1 dB. Measurements are made on one-third octave bands for the 14 bands centred at 200 Hz to 4000 Hz (bands 4 to 17). During the test, sufficient adaptation time is needed to guarantee a stable condition.

The D-factor is computed as a weighted average of Δ_{Sm}:

$$D = -\sum_{i=1}^{N} K_i . \Delta_{Sm} \qquad (7)$$

The coefficients K_i are given in Table E.1 of [ITU-T P.79].

NOTE – For sets with linear microphones and circuitry, D-factor is unrelated to the ambient noise level. For sets with non-linear microphones and/or non-linear circuitry, D-factor depends on the room noise level.

5.1.4 Echo path loss characteristics

5.1.4.1 Weighted terminal coupling loss (TCLw)

5.1.4.2 Requirement

The TCLw shall be ≥ 55 dB at the nominal setting of the user selectable volume control.

With the volume control set to maximum position, TCLw shall be ≥ 46 dB. The volume control shall be automatically set back to nominal after each call unless TCLw ≥ 55 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.

5.1.4.2.1 Measurement method

The measurement set-up is described in clause 4.2.

The test is performed with the handset mounted at HATS position with the application force of 2 N on the artificial ear. The test shall be done in quiet conditions with the ambient noise level less than –64 dBPa(A).

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 is a pseudorandom noise (PN) sequence as defined in [ITU-T P.501] with a length of 4096 points (for the 4.8 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.

5.1.4.2 Stability loss

5.1.4.2.1 Requirement

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

With the handset lying on a hard surface with the transducers facing that surface, the 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.

5.1.4.2.2 Measurement method

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 under the following conditions.

Method 1:

a) The handset, with the speech transmission circuit fully active shall be positioned on one inside surface that is part of three perpendicular plane, smooth and hard surfaces. Each surface shall extend 500 mm from the apex of the corner. One surface shall be marked with a diagonal line extending from the corner and with a reference position 250 mm from the corner formed by the three surfaces, as shown in Figure 4.

b) The handset shall be positioned on the defined surface as follows:

   1) the mouthpiece and ear-cap shall face towards the surface;

   2) the handset shall be placed centrally above the diagonal line, with the ear-cap nearer to the apex of the corner;

   3) the extremity of the handset shall coincide with the perpendicular to the reference point, as shown in Figure 4.

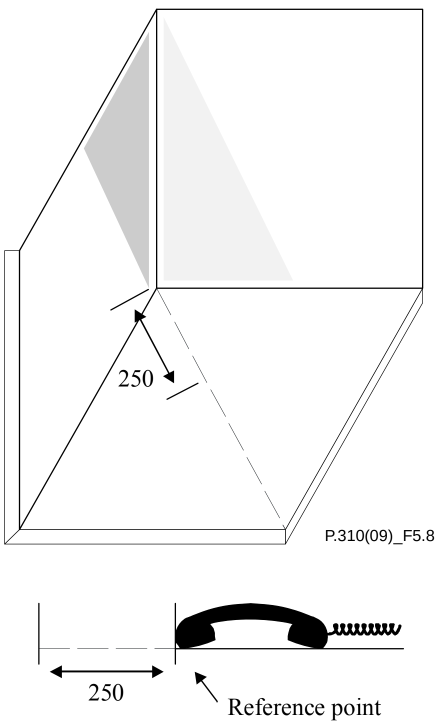

Figure 4 – Reference corner and reference position

NOTE – All dimensions are in mm.

Method 2:

The handset, with the transmission circuit fully active, is placed with the ear-cap and mouthpiece facing a hard, smooth surface free of any other object for 0.5 m.

Stability is the least value of the attenuation in the band 100 Hz to 8000 Hz.

5.2 Codec relevant parameters

5.2.1 Delay

5.2.1.1 Description

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

Certain aspects of this delay can be optimized in VoIP telephones, such as the internal hardware or firmware delay 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 and/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 a 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: $T_s \leq 25$ ms
- Category B: $T_s \leq 35$ ms
- Category C: $T_s \leq 50$ ms
- Category D: $T_s > 50$ ms

For ITU-T G.722 coding algorithm (20 ms payload's length in one packet), at least Category B 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 the 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 B 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.

6 Headset technical requirements

6.1 General description

The requirements in this clause apply to the headset and terminal together. It is not intended to be a specification for a headset separate from the terminal.

Except for parameters in this clause, other requirements which have been covered in handset requirements are also applicable for headset.

Different from the handset test, the test report shall individually provide the test results for each of the five repetitions, plus any additional statistic analysis as required.

6.2 Codec independent parameters

6.2.1 Receiving characteristics

6.2.1.1 Receiving sensitivity/frequency characteristics

The requirement for the handset is applicable for the headset. The difference is that with current technology, it may be difficult or even not possible to achieve the desired frequency response characteristics for headsets below 250 Hz.

6.2.1.2 Receiving loudness rating (RLR)

6.2.1.2.1 Requirements

– For monaural earphones, the nominal value of RLR shall be 2 dB, with a tolerance of ±3 dB.

– For binaural earphones, the nominal value of RLR shall be 8 dB for each earphone, with a tolerance of ±3 dB.

6.2.2 Echo path loss characteristics

6.2.2.1 Stability loss

6.2.2.1.1 Requirement

In the case of headsets, the requirement in clause 5.1.4.2.1 applies for the closest possible position between a microphone and a headset receiver.

NOTE – Depending on the type of headset, it may be necessary to repeat the measurement in different positions.

6.2.2.1.2 Measurement method

The method described in clause 5.1.4.2.2 is used. The headset, with the transmission circuit fully active, shall be positioned on the defined surface as follows:

– the microphone and the receiver shall face towards the surface.

– the headset receiver shall be placed centrally at the reference point as shown in Figure 4.

– the headset microphone is positioned as close as possible to the receiver.

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