ETSI TS 103 739 V1.2.1 (2017-07)

TECHNICAL SPECIFICATION

Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband wireless terminals (handset and headset) from a QoS perspective as perceived by the user

Reference
RTS/STQ-231-3

Keywords
speech, terminal

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Copyright Notification

Contents

Intellectual Property Rights 5
Foreword 5
Modal verbs terminology 5
Introduction 5
1 Scope 6
2 References 6
2.1 Normative references 6
2.2 Informative references 7
3 Definitions and abbreviations 8
3.1 Definitions 8
3.2 Abbreviations 8
4 Configurations and interfaces 9
4.1 Introduction 9
4.2 Access networks 9
4.3 Additional (radio) links between the terminal and external electroacoustical devices 9
5 Test Configurations 9
5.1 Set-up interface 9
5.2 Set-up for terminals 10
5.3 Acoustical environment 11
5.4 Test signals 11
5.5 Calibration 11
5.6 Environmental conditions for tests 13
5.7 Accuracy of test equipment 13
5.8 Power feeding conditions 14
5.9 Influence of terminal delay on measurements 14
6 Codec independent requirements and associated Measurement Methodologies 14
6.1 Send and receive frequency response 14
6.1.1 Send frequency response 14
6.1.2 Receive frequency response 16
6.2 Send and receive loudness ratings 19
6.2.1 Send Loudness Rating (SLR) 19
6.2.2 Microphone (mic) mute 19
6.2.3 Receive Loudness Rating (RLR) 19
6.2.4 LR stability 20
6.3 Sidetone parameters 20
6.3.1 Introduction 20
6.3.2 SideTone Masking Rating (STMR) 20
6.3.3 Sidetone delay 21
6.4 Send and receive noise 22
6.4.1 Send noise 22
6.4.2 Receive noise 22
6.5 Send and receive distortion 23
6.5.1 Introduction 23
6.5.2 Send Distortion 23
6.5.3 Receive distortion 23
6.6 Stability loss 24
6.7 Terminal Coupling Loss 25
6.8 Double talk performance 26
6.8.1 Introduction 26
6.8.2 Attenuation Range in Send Direction during Double Talk $A_{H,S,dt}$ 27
6.8.3 Attenuation Range in Receive Direction during Double Talk $A_{H,S,dt}$ 28
6.8.4 Detection of echo components during double Talk 28

6.8.5 Minimum activation level and sensitivity of double talk detection 30
6.9 Switching parameters 31
6.9.1 Activation in Send Direction 31
6.9.2 Minimum activation level and sensitivity in Receive direction 31
6.9.3 Automatic level control 31
6.9.4 Silence Suppression and Comfort Noise Generation 31
6.9.5 Non Linear Processing 32
6.10 Background noise performance 32
6.10.1 Performance in send direction in the presence of background noise 32
6.10.2 Speech Quality in the Presence of Background Noise 33
6.10.3 Quality of Background Noise Transmission (with Far End Speech) 33
6.11 Quality of echo cancellation 34
6.11.1 Temporal echo effects 34
6.11.2 Spectral Echo Attenuation 35
6.11.3 Occurrence of Artefacts 35
6.11.4 Variable echo path 35
6.12 Send and receive delay or round trip delay 36
6.13 Objective listening Quality in send and receive direction 38

7 Codec dependent requirements and associated Measurement Methodologies 38
7.1 Speech Coders 38
7.2 Objective listening Quality in send and receive direction 39
7.2.1 Objective listening speech quality MOS-LQO in send direction 39
7.2.2 Objective listening quality MOS-LQO in receive direction 39

8 Requirements and associated Measurement Methodologies (with an additional radio link between the terminal and external electroacoustical devices) 40

Annex A (informative): Bibliography 41

History 42

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

This Technical Specification (TS) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

The present document covers wireless speech terminals. It aims to enhance the interoperability and end-to-end quality with all other types of terminals.

The advanced signal processing of terminals is targeted to speech signals. Therefore, wherever possible speech signals are used for testing in order to achieve realistic test conditions and meaningful results.

1 Scope

The present document provides speech transmission performance requirements for wireless terminals; it addresses all types of wireless terminals, including softphones. The present document addresses handset and headset functions of wideband wireless terminals.

Differently from other standards which define minimum performance requirements, it is the intention of the present document to specify terminal equipment requirements which enable manufacturers and service providers to enable good quality end-to-end speech performance as perceived by the user whatever be the radio link (terminals may implement different radio links with the access network).

When an additional radio link between the terminal and external electroacoustical devices is used (e.g. Bluetooth® link), the present document will address the overall quality.

In the present document objective measurement methodologies and requirements for wireless speech terminals are given.

In addition to basic testing procedures, the present document describes advanced testing procedures taking into account further quality parameters as perceived by the user.

The requirements available in the present document will ensure a high compatibility across access networks with all types of terminals.

It is the aim to optimize the listening and talking quality, conversational performance, as well as the use in noisy environments. Related requirements and test methods will be defined in the present document.

For all the functions, the present document will consider the limitations in audio performance due to different form factors (e.g. size, shape).

Terminals which are not intended to be connected to public networks are outside the scope of the present document.

2 References

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at https://docbox.etsi.org/Reference/.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are necessary for the application of the present document.

[1] Void.

[2] Void.

[3] Recommendation ITU-T P.56: "Objective measurement of active speech level".

[4] Recommendation ITU-T P.57: "Artificial ears".

[5] Recommendation ITU-T P.58: "Head and torso simulator for telephonometry".

[6] Recommendation ITU-T P.64: "Determination of sensitivity/frequency characteristics of local telephone systems".

[7] Recommendation ITU-T P.79: "Calculation of loudness ratings for telephone sets".

[8] Recommendation ITU-T P.340: "Transmission characteristics and speech quality parameters of hands-free terminals".

[9] Recommendation ITU-T P.380: "Electro-acoustic measurements on headsets".

[10] Recommendation ITU-T P.501 Amendment 1 (2012): "Test signals for use in telephonometry".

[11] Recommendation ITU-T P.502: "Objective test methods for speech communication systems using complex test signals".

[12] Recommendation ITU-T P.581: "Use of head and torso simulator for hands-free terminal testing".

[13] IEC 61672: "Electroacoustics - Sound Level Meters".

[14] IEC 61260: "Electroacoustics - Octave-band and fractional-octave-band filters".

[15] ETSI TS 126 171 (V6.0.0): "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR speech codec, wideband; General description (3GPP TS 26.171 version 6.0.0 Release 6)".

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

[17] Recommendation ITU-T G.722: "7 kHz audio-coding within 64 kbit/s".

[18] Recommendation ITU-T G.711.1: "Wideband embedded extension for ITU-T G.711 pulse code modulation".

[19] Recommendation ITU-T G.722.2: "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)".

[20] ETSI TS 103 106: "Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise: Background noise transmission for mobile terminals-objective test methods".

[21] Recommendation ITU-T P.311: "Transmission characteristics for wideband digital handset and headset telephones".

[22] ETSI TS 126 441 (V12.0.0): "Universal Mobile Telecommunications System (UMTS); LTE; EVS Codec General Overview (3GPP TS 26.441 version 12.0.0 Release 12)".

[23] ETSI TS 103 224: "Speech and multimedia Transmission Quality (STQ); A sound field reproduction method for terminal testing including a background noise database".

[24] Recommendation ITU-T P.1010: "Fundamental voice transmission objectives for VoIP terminals and gateways".

[25] Recommendation ITU-T P.863: "Perceptual objective listening quality assessment".

[26] Recommendation ITU-T P.863.1: "Application guide for Recommendation ITU-T P.863".

[27] Recommendation ITU-T G.122: "Influence of national systems on stability and talker echo in international connections".

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI EG 201 377-1: "Speech and multimedia Transmission Quality (STQ); Specification and measurement of speech transmission quality; Part 1: Introduction to objective comparison measurement methods for one-way speech quality across networks".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

artificial ear: device for the calibration of earphones incorporating an acoustic coupler and a calibrated microphone for the measurement of the sound pressure and having an overall acoustic impedance similar to that of the median adult human ear over a given frequency band

codec: combination of an analogue-to-digital encoder and a digital-to-analogue decoder operating in opposite directions of transmission in the same equipment

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 by applying the reverse nominal curve of table 3 of Recommendation ITU-T P.58 [5]

Head And Torso Simulator (HATS) for telephonometry: manikin extending downward 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 human adult and to reproduce the acoustic field generated by the human mouth

Mouth Reference Point (MRP): point located on axis and 25 mm in front of the lip plane of a mouth simulator

nominal setting of the volume control: when a receive volume control is provided, the setting which is closest to the nominal RLR of 2 dB

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

a.c.	alternating current
AM-FM	Amplitude modulation - Frequency modulation
AMR-WB	Adaptive Mode Rate - Wide Band
CDMA	Code Division Multiple Access
CS	Composite Source
CSS	Composite Source Signal
DECT	Digital Enhanced Cordless Telecommunications
DRP	ear Drum Reference Point
ELR	Echo Loudness rating
ERP	Ear Reference Point
EVS	Enhanced Voice Services
FFT	Fast Fourier Transform
G-MOS-LQOw	Overall transmission quality for wideband systems
GSM	Global System for Mobile communication (3GPP)
HATS	Head And Torso Simulator
HRP	HATS Reference Point
LR	Loudness rating
LTE	Long Term Evolution (3GPP)
MOS	Mean Opinion Score
MRP	Mouth Reference Point
NLP	Non-linear processing

N-MOS-LQOw	Transmission quality of the background noise for wideband systems
PN	Pseudo noise sequence
POI	Point Of Interconnect
QoS	Quality of Service
RF	Radio frequency
RLR	Receive Loudness Rating
RMS	Root mean square
SLR	Send Loudness Rating
S-MOS-LQOw	Transmission quality of the speech for wideband systems
STMR	SideTone Masking Rating
TCL	Terminal coupling loss
TOSQA	Telecommunications Objective Speech Quality Assessment
UMTS	Universal Mobil Telecommunications System
VoLTE	Voice over LTE
WB	WideBand
WIFI	Wireless fidelity
WIMAX™	Worldwide Interoperability for Microwave ACCess

4 Configurations and interfaces

4.1 Introduction

The present document is intended to be applicable for different wireless access networks and for additional radio links.

4.2 Access networks

The present document applies to any wireless terminal whatever the network access, e.g. GSM, UMTS, VoLTE, DECT, Bluetooth®, WIFI, WIMAX™ and CDMA.

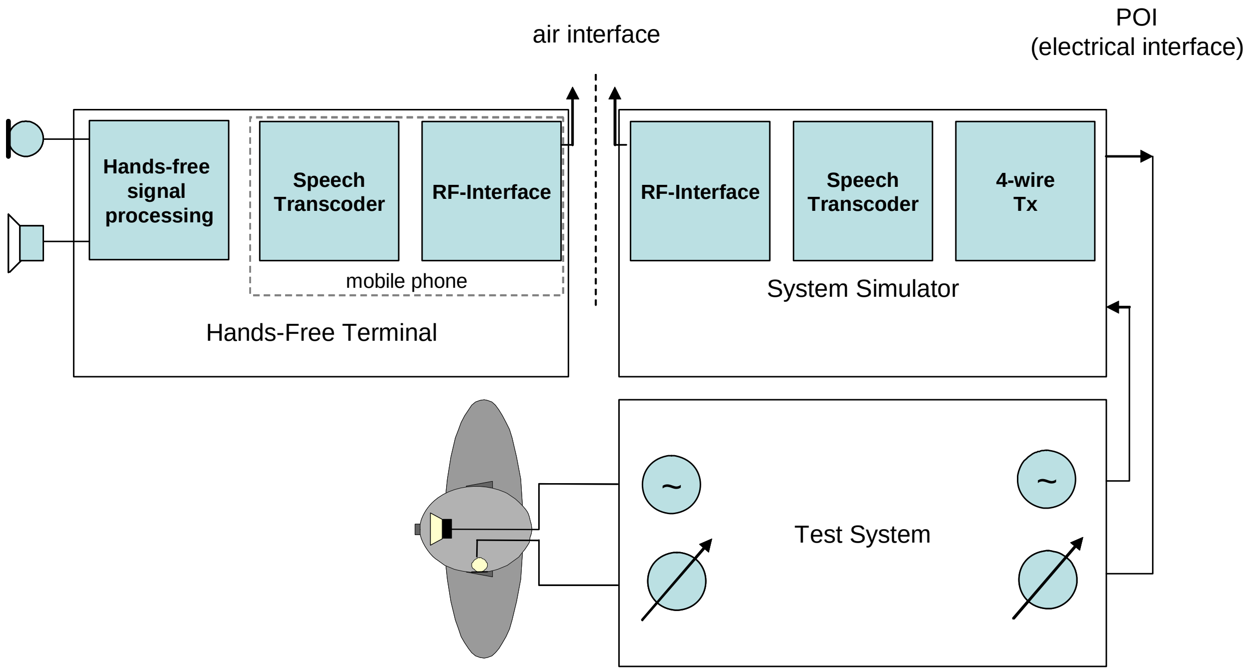
4.3 Additional (radio) links between the terminal and external electroacoustical devices

The whole terminal may include additional (radio) links. The most of the requirements and test methods apply to the whole terminal. When specific requirements or test methods are needed, they can be found in clause 8.

5 Test Configurations

5.1 Set-up interface

The generic schematic as defined in figure 5.1.1 is applicable to any wireless link.



NOTE: The "whole" terminal includes all the components from "RF interface" to the transducers and may include an additional (radio) link. The air interface considered in the figure is not the additional radio link.

Figure 5.1.1: Set-up interface

5.2 Set-up for terminals

The acoustical access to terminals is the most realistic simulation of the average subscriber. This can be made by using HATS (Head And Torso Simulator) with appropriate ear simulation and appropriate means to fix handset and headset terminals in a realistic and reproducible way to the HATS. HATS is described in Recommendation ITU-T P.58 [5], appropriate ears are described in Recommendation ITU-T P.57 [4] (type 3.3 and type 3.4 ear), a proper positioning of handsets under realistic conditions is to be found in Recommendation ITU-T P.64 [6].

The preferred way of testing a terminal is to connect it to a network simulator with exact defined settings and access points. The test sequences are fed in either electrically, using a reference codec or using the direct signal processing approach and acoustically using ITU-T HATS.

When a coder with variable bit rate is used for testing terminal electroacoustical parameters, the bit rate giving the best characteristics or the most commonly used should be selected, e.g.:

- AMR-WB [15]: 12,65 kbit/s;
- Recommendation ITU-T G.729.1 [16]: 32 kbit/s;
- EVS [22]: 13,2 kbit/s.

Setup for handsets and headsets

When using a handset telephone the handset is placed in the HATS position as described in Recommendation ITU-T P.64 [6]. The artificial mouth shall be conform with Recommendation ITU-T P.58 [5]. The artificial ear shall be conform with Recommendation ITU-T P.57 [4], type 3.3 or type 3.4 ears shall be used.

Recommendations for positioning headsets are given in Recommendation ITU-T P.380 [9]. If not stated otherwise headsets shall be placed in their recommended wearing position. Further information about setup and the use of HATS can be found in Recommendation ITU-T P.380 [9].

Unless stated otherwise if a volume control is provided the setting shall be chosen such that the nominal RLR is met as close as possible.

Unless stated otherwise, the application force of 8 N is used for handset testing. No application force is used for headset.

5.3 Acoustical environment

In general different acoustical environments have to be taken into account: either room noise and background noise are an inherent part of the test environment or room noise and background noise shall be eliminated to such an extent that their influence on the test results can be neglected.

Unless stated otherwise, measurements shall be conducted under quiet and "anechoic" conditions. Considering this, test laboratory, in the case where its test room does not conform to anechoic conditions as given in Recommendation ITU-T P.311 [21], has to present difference in results for measurements due to its test room. In case where an anechoic room is not available the test room has to be an acoustically treated room with few reflections and a low noise level.

Depending on the distance of the transducers from mouth to ear a quiet office room may be sufficient e.g. for handsets where artificial mouth and artificial ear are located close to the acoustical transducers.

However, for some headsets or handset terminals with smaller dimension an anechoic room will be required.

In cases where real or simulated background noise is used as part of the testing environment, the original background noise shall not be noticeably influenced by the acoustical properties of the room.

In all cases where the performance of acoustic echo cancellers shall be tested a realistic room which represents the typical user environment for the terminal shall be used.

5.4 Test signals

Modern wireless terminals often deploy nonlinear and time-varying processing. As such terminals are designed for speech transmission, the most appropriate test signal is real speech. Appropriate test signals (general description) are defined in Recommendation ITU-T P.501 [10].

More information can be found in the test procedures described below.

For testing the wideband telephony service provided by a terminal the test signal used shall be band limited between 100 Hz and 8 kHz with a bandpass filter providing a minimum of 24 dB/Oct. filter roll off, when feeding into the receive direction.

The test signal levels are referred to the average level of the test signal band limited in receive direction, averaged over the complete test sequence unless specified otherwise.

Unless specified otherwise, the test signal level shall be -4,7 dBPa at the MRP.

Unless specified otherwise, the applied test signal level at the digital input shall be -16 dBm0.

5.5 Calibration

Position and calibration of HATS

All the send and receive characteristics shall be tested with the HATS, it shall be indicated what type of ear was used at what application force. For handsets if not stated otherwise 8N application force shall be used.

The horizontal positioning of the HATS reference plane shall be guaranteed within ±2º.

The HATS shall be equipped with a type 3.3 or type 3.4 artificial ear for handsets. For binaural headsets two artificial ears are required. The type 3.3 or type 3.4 artificial ears as specified in Recommendation ITU-T P.57 [4] shall be used. The artificial ear shall be positioned on HATS according to Recommendation ITU-T P.58 [5].

The exact calibration and equalization can be found in Recommendation ITU-T P.581 [12]. If not stated otherwise, the HATS shall be diffuse-field equalized. The inverse nominal diffuse field curve as found in table 3 of Recommendation ITU-T P.58 [5] shall be used.

NOTE: The inverse average diffuse field response characteristics of HATS as found in Recommendation ITU-T P.58 [5] is used and not the specific one that may be provided by the manufacturer of the HATS. Instead of using the individual diffuse field correction, the average correction function is used because, for handset and headset measurements, mostly the artificial ear, ear canal and ear impedance simulations are effective. The individual diffuse-field correction function of HATS includes all diffraction and reflection effects of the complete individual HATS which are not effective in the measurement and potentially would lead to bigger measurement uncertainties than using the average correction.

Setup of background noise simulation

A setup for simulating realistic background noises in a lab-type environment is described in ETSI TS 103 224 [23].

ETSI TS 103 224 [23] contains a description of the recording arrangement for realistic background noises, a description of the setup for a loudspeaker arrangement suitable to simulate a background noise field in a lab-type environment and a database of realistic background noises, which can be used for testing the terminal performance with a variety of different background noises.

The principle loudspeaker setup for the simulation arrangement is shown in figure 5.5.1.

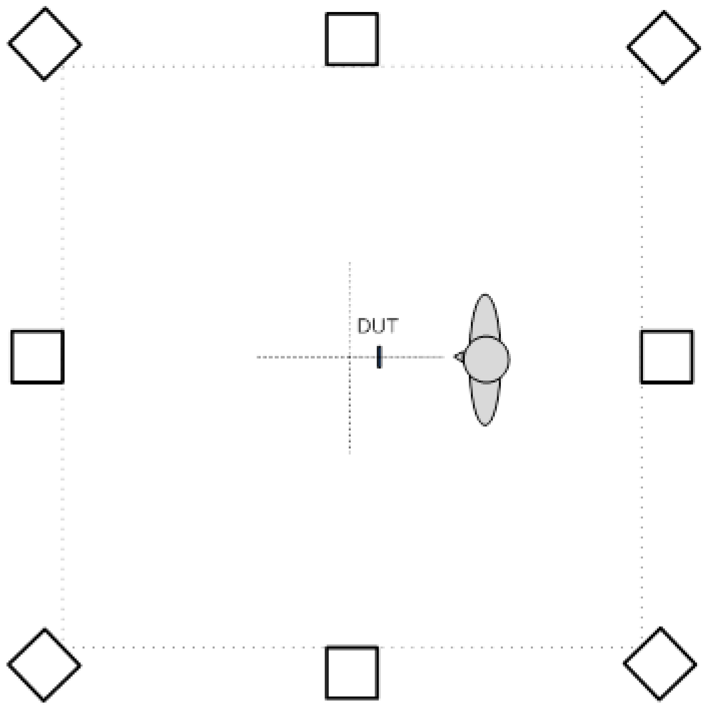


Figure 5.5.1: Loudspeaker arrangement for background noise simulation

The equalization and calibration procedure for the setup is described in detail in ETSI TS 103 224 [23].

If not stated otherwise this setup is used in all measurements where background noise simulation is required.

The following noises of ETSI TS 103 224 [23] shall be used.

Table 5.5.1: Noises used for background noise simulation

Name	Description	Length	Handset Levels
Full-size car 130 km/h (FullSizeCar_130)	HATS and microphone array at co-drivers position	30 s	1: 68,5 dB 2: 68,3 dB 3: 68,8 dB 4: 69,5 dB 5: 69,9 dB 6: 70,5 dB 7: 70,8 dB 8: 71,9 dB
Cafeteria (Cafeteria)	HATS and microphone array inside a cafeteria	30 s	1: 70,0 dB 2: 70,0 dB 3: 70,1 dB 4: 70,7 dB 5: 70,5 dB 6: 70,8 dB 7: 70,6 dB 8: 71,0 dB
Roadnoise (*Roadnoise*)	HATS and microphone array standing outside near a road	30 s	1: 72,8 dB 2: 71,6 dB 3: 72,0 dB 4: 72,9 dB 5: 72,2 dB 6: 73,1 dB 7: 73,0 dB 8: 73,8 dB
Pub Noise (*Pub*)	HATS and microphone array in a pub	30 s	1: 77,2 dB 2: 76,6 dB 3: 75,7 dB 4: 76,0 dB 5: 76,0 dB 6: 76,3 dB 7: 76,0 dB 8: 76,4 dB
Airport departure	HATS and microphone array in an airport gate area	30 s	1: 77,5 dB 2: 78,3 dB 3: 78,7 dB 4: 78,7 dB 5: 78,4 dB 6: 78,8 dB 7: 78,1 dB 8: 78,1 dB

5.6 Environmental conditions for tests

The following conditions shall apply for the testing environment:

a) Ambient temperature: 15 °C to 35 °C (inclusive);

b) Relative humidity: 5 % to 85 %;

c) Air pressure: 86 kPa to 106 kPa (860 mbar to 1 060 mbar).

5.7 Accuracy of test equipment

Unless specified otherwise, the accuracy of measurements made by test equipment shall be better than:

Table 5.7.1: Accuracy of measurements

Item	Accuracy
Electrical Signal Level	±0,2 dB for levels ≥ -50 dBV
Electrical Signal Level	±0,4 dB for levels < -50 dBV
Sound pressure	±0,7 dB
Time	±0,2 %
Frequency	±0,2 %
Application force	±2 Newton
Measured maximum frequency	20 kHz
Clock Accuracy	< 2 ppm

Unless specified otherwise, the accuracy of the signals generated by the test equipment shall be better than:

Table 5.7.2: Accuracy of generated signals

Quantity	Accuracy
Sound pressure level at MRP	±3 dB for 100 Hz to 200 Hz ±1 dB for 200 Hz to 4 kHz ±3 dB for 4 kHz to 14 kHz
Electrical excitation levels	±0,4 dBV across the whole frequency range
Frequency generation	±2 % (see note)
Time	±0,2 %
Specified component values	±1 %
NOTE: This tolerance may be used to avoid measurements at critical frequencies, e.g. those due to sampling and coding operations within the terminal under test.	

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

The sound level measurement equipment shall conform to IEC 61672 [13] Type 1.

5.8 Power feeding conditions

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.

5.9 Influence of terminal delay on measurements

As delay is introduced by the terminal, care shall be taken for all measurements where exact position of the analysis window is required. It shall be checked that the test is performed on the test signal and not any other signal.

6 Codec independent requirements and associated Measurement Methodologies

6.1 Send and receive frequency response

6.1.1 Send frequency response

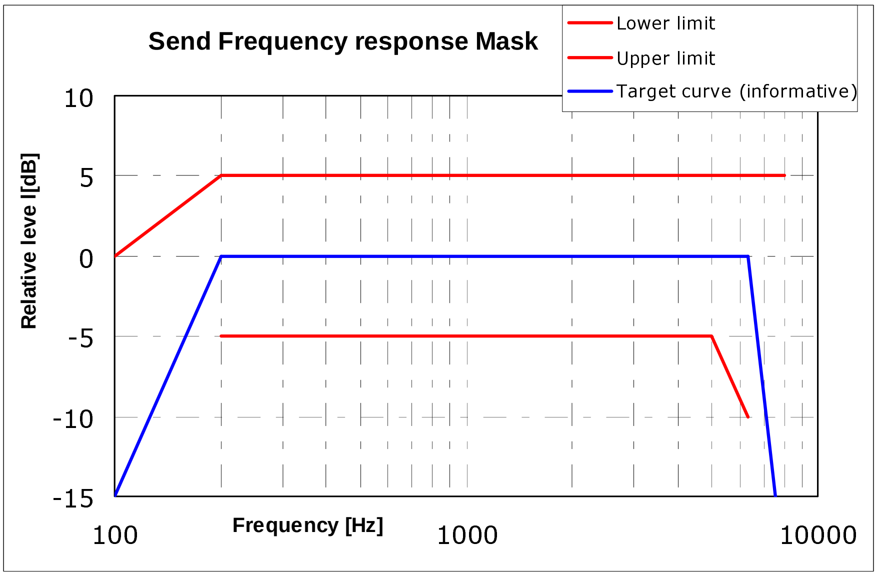
Due to diffuse field equalization applying in the receive direction a flat curve is preferable in send path.

Requirement

The send frequency response of the handset or the headset shall be within a mask as defined in table 6.1.1.1 and shown in figure 6.1.1.1. This mask shall be applicable for all types of handsets and headsets.

Table 6.1.1.1: Send frequency response

Handset-Headset send sensitivity/frequency responseFrequency (Hz)	Upper limit	Lower limit
100	0	
200	5	-5
1 500	5	-5
5 000	5	-5
6 300	5	-10
8 000	5	
NOTE: All sensitivity values are expressed in dB on an arbitrary scale.		



NOTE: The basis for the target frequency responses in send and receive is the orthotelefonic reference response which is measured between 2 subjects in 1 m distance under free field conditions and is assuming an ideal receive characteristic. Under these conditions the overall frequency response shows a rising slope. In opposite to other standards the present document no longer uses the ERP as the reference point for receive but the diffuse-field. With the concept of diffuse-field based receive measurements a rising slope for the overall frequency response is achieved by a flat target frequency response in send and a flat diffuse-field based receive frequency response.

Figure 6.1.1.1: Send frequency response mask

Measurement method

The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [10]. 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.

The handset or headset terminal is setup as described in clause 5.2. The handset is mounted at the HATS position (see Recommendation ITU-T P.64 [6]). The application force used to apply the handset against the artificial ear shall be within the range specified in Recommendation ITU-T P.64 [6].

Measurements shall be made at one twelfth-octave intervals as given by IEC 61260 [14] for frequencies from 100 Hz to 8 kHz inclusive. For the calculation the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.

The sensitivity is expressed in terms of dBV/Pa.

6.1.2 Receive frequency response

Requirement

The receive frequency response of the handset or the headset shall be within a mask as defined in table 6.1.2.1 and shown in figures 6.1.2.1, 6.1.2.2 and 6.1.2.3. The application force for handsets is 2N, 8N and 13N. The mask defined for 8 N application force shall be applicable for all types of headsets.

Table 6.1.2.1: Receive Frequency Response Mask

Frequency	Upper8N	Lower8N	Upper13N	Lower13N	Upper 2N	Lower 2N
100 Hz	3 dB		6 dB		3 dB	
200 Hz	3 dB		6 dB		3 dB	
300 Hz	3 dB	-5 dB	6 dB	-5 dB	3 dB	-10 dB
400 Hz	3 dB	-5 dB			3 dB	-8 dB
1 000 Hz		-5 dB		-5 dB		
1 200H Hz		-8 dB	6 dB	-8 dB		
1 500 Hz		-8 dB		-8 dB		-8 dB
2 000 Hz	9 dB	-3 dB	9 dB	-3 dB	9 dB	-3 dB
3 200 Hz		-3 dB		-3 dB	9 dB	-3 dB
3 400 Hz	9 dB		9 dB		9 dB	
4 000 Hz	9 dB		9 dB		9 dB	
5 000 Hz	9 dB		9 dB		9 dB	
6 300 Hz	9 dB		9 dB		9 dB	
7 000 Hz		-13 dB		-13 dB	9 dB	-13 dB
8 000 Hz	9 dB		9 dB		9 dB	

NOTE 1: The limit curves shall be determined by straight lines joining successive co-ordinates given in the table, where frequency response is plotted on a linear dB scale against frequency on a logarithmic scale is a floating or 'best fit' mask.

NOTE 2: The basis for the target frequency responses in send and receive is the orthotelefonic reference response which is measured between 2 subjects in 1 m distance under free field conditions and is assuming an ideal receive characteristic. This flat response characteristic is shown as the target curve. Under these conditions the overall frequency response shows a rising slope. In opposite to other standards the present document no longer uses the ERP as the reference point for receive but the diffuse-field. With the concept of diffuse-field based receive measurements a rising slope for the overall frequency response is achieved by a flat target frequency response in send and a flat diffuse-field based receive frequency response.

NOTE 3: With current technology it may be difficult or even not possible to achieve the desired frequency response characteristics for handsets with 2N application force.

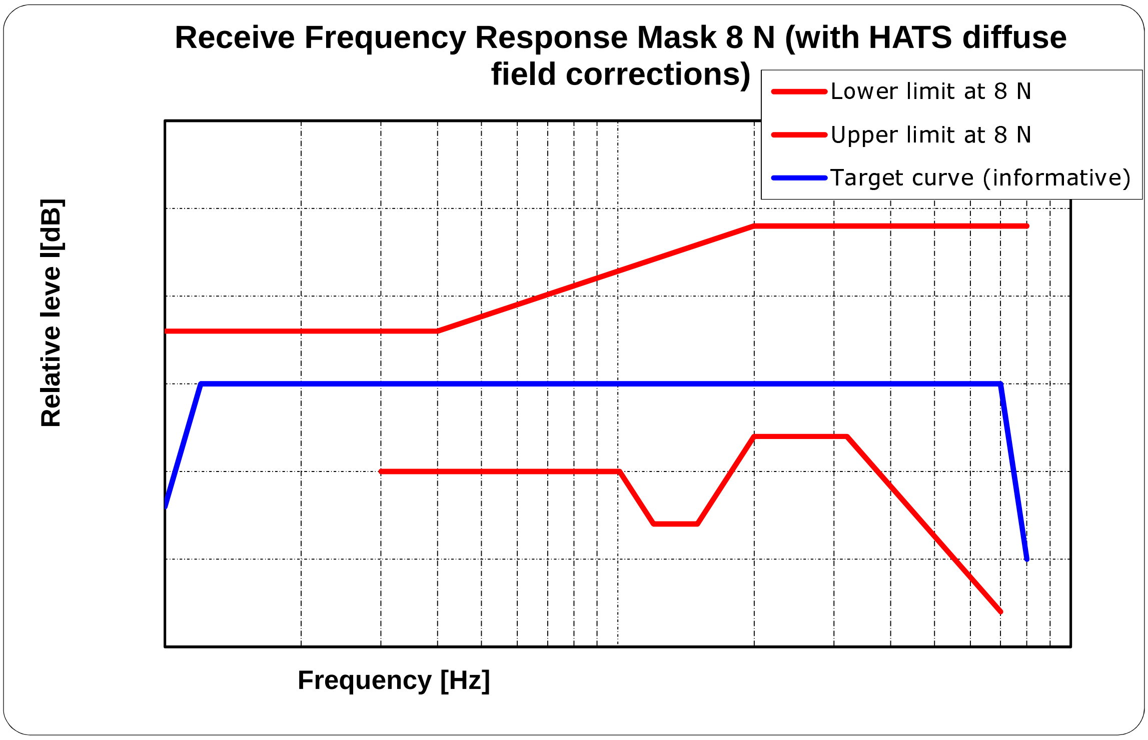


Figure 6.1.2.1: Receive frequency response mask for 8N application force

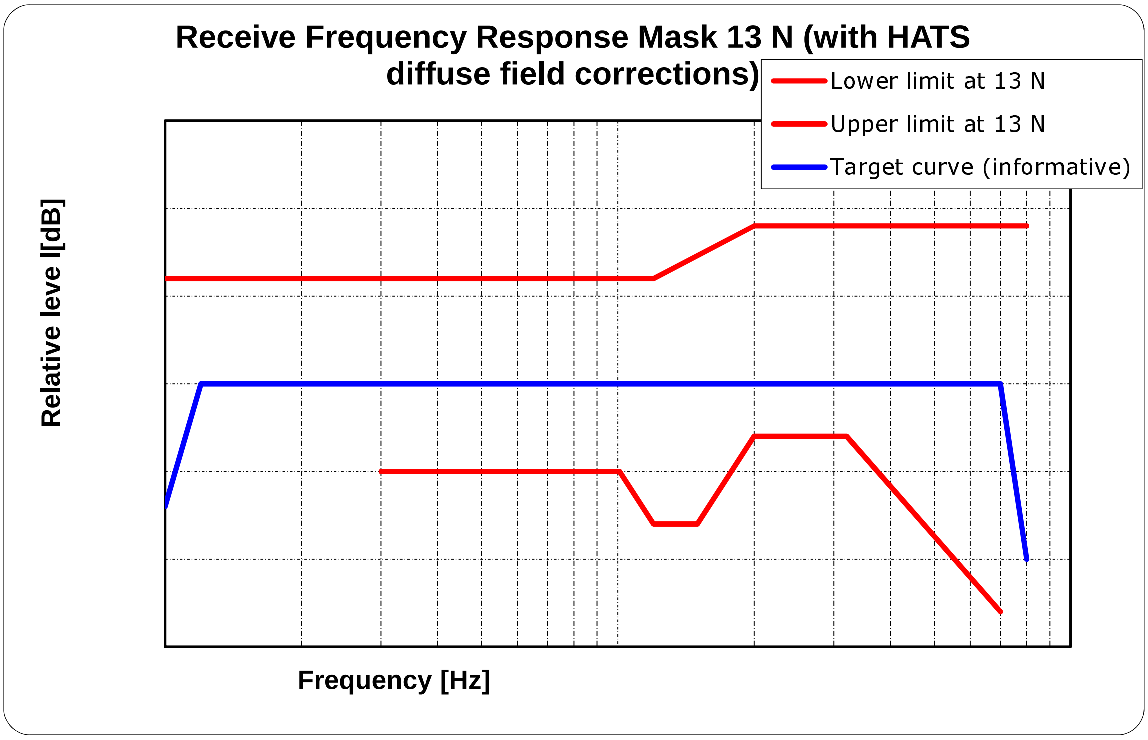


Figure 6.1.2.2: Receive frequency response mask for 13N application force

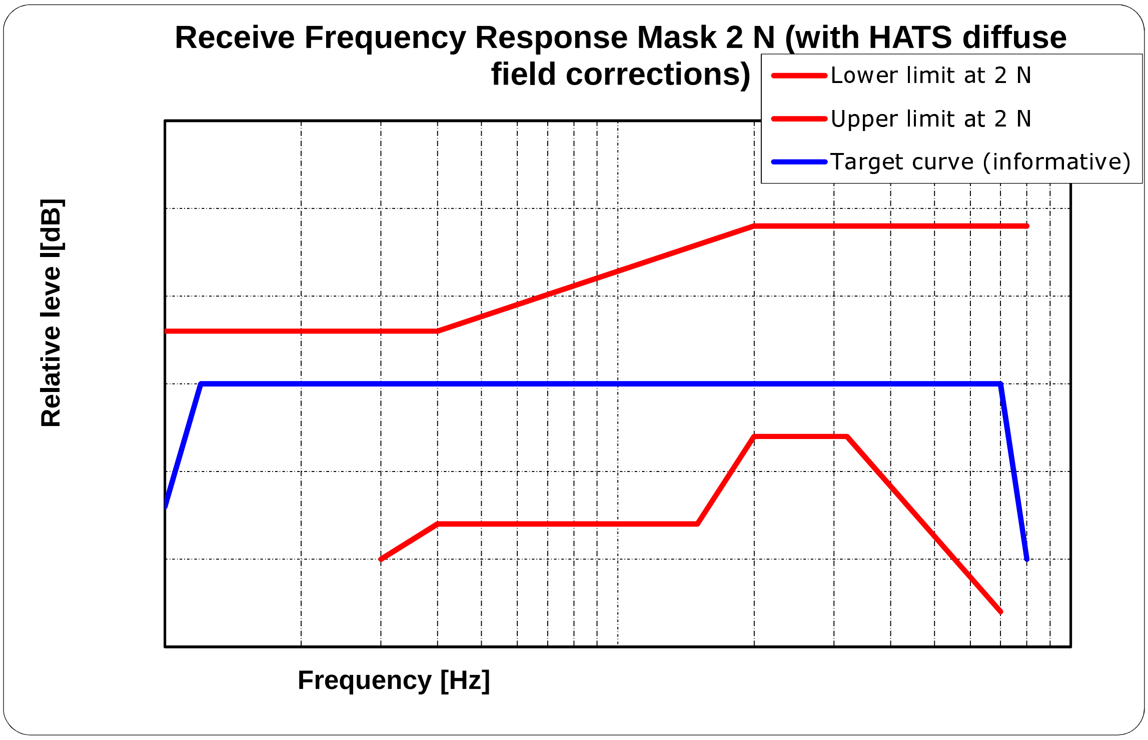


Figure 6.1.2.3: Receive frequency response mask for 2N application force

Measurement method

Receive frequency response is the ratio of the measured sound pressure and the input level.
(dB relative Pa/V)

$$S_{Jeff} = 20 \log (pe_{ff} / v_{RCV}) \text{ dB rel 1 Pa / V} \quad (1)$$

S_{Jeff}	Receive Sensitivity; Junction to HATS Ear with diffuse field correction.
pe_{ff}	DRP Sound pressure measured by ear simulator Measurement data are converted from the Drum Reference Point to diffuse field.
v_{RCV}	Equivalent RMS input voltage.

The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [10]. The test signal level shall be -16 dBm0, measured according to Recommendation ITU-T P.56 [3] at the digital reference point or the equivalent analogue point.

The handset terminal or the headset terminal is setup as described in clause 5.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [6]). The application forces used to apply the handset against the artificial ear is 2N, 8N and 13N.

In case of headset measurements the tests are repeated 5 times, in conformance with Recommendation ITU-T P.380 [9] the results are averaged (averaged value in dB, for each frequency).

The HATS is diffuse-field equalized. The equalized output signal is power-averaged on the total time of analysis. The 1/12 octave band data are considered as the input signal to be used for calculations or measurements.

Measurements shall be made at one twelfth-octave intervals as given by IEC 61260 [14] for frequencies from 100 Hz to 8 kHz inclusive. For the calculation the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

The sensitivity is expressed in terms of dBPa/V.

6.2 Send and receive loudness ratings

6.2.1 Send Loudness Rating (SLR)

Requirement

The nominal value of Send Loudness Rating (SLR) shall be:

$$\mathrm{SLR(set)} = +8 \text{ dB} \pm 3 \text{ dB} \qquad (2)$$

Measurement Method

The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [10]. 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.

The handset or headset terminal is setup as described in clause 5.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [6]). The application force used to apply the handset against the artificial ear is noted in the test report.

In case of headset measurements the tests are repeated 5 times, in conformance with Recommendation ITU-T P.380 [9] the results are averaged (averaged value in dB, for each frequency).

The send sensitivity shall be calculated from each band of the 20 frequencies given in table 1 of Recommendation ITU-T P.79 [7], bands 1 to 20. For the calculation the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.

The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to Recommendation ITU-T P.79 [7], formula (A - 23b), over bands 1 to 20, using m = 0,175 and the send weighting factors from Recommendation ITU-T P.79 [7], annex A, table A.2.

6.2.2 Microphone (mic) mute

Requirement

The SLR (Send Loudness Rating) with mic mute on shall be 50dB higher than with mic mute off.

Measurement method

The terminal will be positioned as described in clause 5.2.

The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [10]. 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.

Calibration is realized as explained in clause 5.5.

The send sensitivity shall be calculated from each band of the 20 frequencies given in table 1 of Recommendation ITU-T P.79 [7], bands 1 to 20. For the calculation the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.

The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to Recommendation ITU-T P.79 [7], annex A.

6.2.3 Receive Loudness Rating (RLR)

Requirement

The nominal value of Receive Loudness Rating (RLR) for handset and monaural headset shall be:

$$\mathrm{RLR} = +2 \text{ dB} \pm 3 \text{ dB} \qquad (3)$$

Where a user controlled receive volume control is provided, the RLR shall meet the selected nominal value for at least one setting of the control. When the control is set to maximum, the RLR shall not be less than (louder than) -13 dB.

With the volume control set to the minimum position the RLR shall not be greater than (quieter than) 18 dB.

For Binaural headset:

$$\text{RLR (binaural headset)} = +8 \text{ dB} \pm 3 \text{ dB for each earphone} \quad (4)$$

Measurement Method

The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [10]. The test signal level shall be -16 dBm0, measured at the digital reference point or the equivalent analogue point. The test signal level is averaged over the complete test signal sequence.

The handset or headset terminal is setup as described in clause 5.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [6]). The application force used to apply the handset against the artificial ear is noted in the test report. The HATS is not diffuse-field equalized. The DRP-ERP correction as defined in Recommendation ITU-T P.57 [4] is applied.

The application force used to apply the handset against the artificial ear is noted in the test report. By default, 8N will be used.

In case of headset measurements the tests are repeated 5 times, in conformance with Recommendation ITU-T P.380 [9] the results are averaged (averaged value in dB, for each frequency).

The receive sensitivity shall be calculated from each band of the 20 frequencies given in table 1 of Recommendation ITU-T P.79 [7], bands 1 to 20. For the calculation the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to Recommendation ITU-T P.79 [7], formula (A - 23c), over bands 1 to 20, using m = 0,175 and the receive weighting factors from table A.2 of Recommendation ITU-T P.79 [7], annex A.

No leakage correction shall be applied for the measurement.

6.2.4 LR stability

For further studies.

6.3 Sidetone parameters

6.3.1 Introduction

The present document covers different types of terminals and different use cases (including noisy environments). STMR requirements are basically defined when using terminals in low noise environments.

6.3.2 SideTone Masking Rating (STMR)

Requirement

The SideTone Masking Rating STMR shall be 16 dB ± 4 dB for nominal setting of the volume control.

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

NOTE: It is preferable to have a constant STMR independent of the volume control setting.

Measurement Method

The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [10]. The spectrum of the 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.

The handset or headset terminal is setup as described in clause 5.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [6]) and the application force shall be 13N on the artificial ear type 3.3 or type 3.4.

Where a user operated volume control is provided, the measurements shall be carried out at the nominal setting of the volume control. In addition the measurement is repeated at the maximum volume control setting.

Measurements shall be made at one twelfth-octave intervals as given in IEC 61260 [14] for frequencies from 100 Hz to 8 kHz inclusive. For the calculation the averaged measured level at each frequency band (Recommendation ITU-T P.79 [7], table 3, bands 1 to 20) is referred to the averaged test signal level measured in each frequency band.

The Sidetone path loss (LmeST), as expressed in dB, and the SideTone Masking Rate (STMR) (in dB) shall be calculated from the formula 5-1 of Recommendation ITU-T P.79 [7], using m = 0,225 and the weighting factors in table 3 of Recommendation ITU-T P.79 [7].

6.3.3 Sidetone delay

Requirement

The maximum sidetone-round-trip delay shall be ≤ 5 ms, measured in an echo-free setup.

Measurement Method

The handset or headset terminal is setup as described in clause 5.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [6]).

The test signal is a CS-signal complying with Recommendation ITU-T P.501 [10] using a pn sequence with a length of 4 096 points (for the 48 kHz sampling rate) which equals to the period T. The duration of the complete test signal is as specified in Recommendation ITU-T P.501 [10]. The level of the signal shall be -4,7 dBPa at the MRP.

The cross-correlation function Φxy(τ) between the input signal $S_x(t)$ generated by the test system in send direction and the output signal $S_y(t)$ measured at the artificial ear is calculated in the time domain:

$$\Phi_{xy}(\tau) = \frac{1}{T} \int_{t=\frac{-T}{2}}^{\frac{T}{2}} S_x(t) \cdot S_y(t+\tau) \qquad (5)$$

The measurement window T shall be exactly identical with the time period T of the test signal, the measurement window is positioned to the pn-sequence of the test signal.

The sidetone delay is calculated from the envelope E(τ) of the cross-correlation function Φxy(τ). The first maximum of the envelope function occurs in correspondence with the direct sound produced by the artificial mouth, the second one occurs with a possible delayed sidetone signal. The difference between the two maxima corresponds to the sidetone delay. The envelope E(τ) is calculated by the Hilbert transformation H {xy(τ)} of the cross-correlation:

$$H\{xy(\tau)\} = \sum_{u=-\infty}^{+\infty} \frac{\Phi_{xy}(u)}{\pi(\tau - u)} \qquad (6)$$

$$E(\tau) = \sqrt{\left[\Phi_{xy}(\tau)\right]^2 + \left[H\{xy(\tau)\}\right]^2} \qquad (7)$$

It is assumed that the measured sidetone delay is less than T/2.

6.4 Send and receive noise

6.4.1 Send noise

Requirement

The maximum noise level produced by the Wireless terminal at the POI under silent conditions in the send direction shall not exceed -68 dBm0(A).

No peaks in the frequency domain higher than 10 dB above the average noise spectrum shall occur.

Measurement Method

For the actual measurement no test signal is used. In order to reliably activate the terminal an activation signal is introduced before the actual measurement. The activation signal shall be the female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [10]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free field conditions at the MRP. The activation signal level shall be -4,7 dBPa, measured at the MRP. The activation signal level is averaged over the complete activation signal sequence. Alternatively other speech like test signals (e.g. artificial voice) with the same signal level can be used for activation.

The handset or headset terminal is set-up as described in clause 5.2. The handset is mounted at the HATS position (see Recommendation ITU-T P.64 [6]).

The send noise is measured at the POI in the frequency range from 100 Hz to 8 kHz. The analysis window is applied directly after stopping the activation signal but taking into account the influence of all acoustical components (reverberations). The averaging time is 1 second. 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 activated condition.

The noise level is measured in dBm0(A).

Spectral peaks are measured in the frequency domain from 100 Hz to 6,3 kHz. The frequency spectrum of the A-weighted idle channel noise is measured by a spectral analysis having a noise bandwidth of 8,79 Hz (determined using FFT 8 k samples/48 kHz sampling rate with Hanning window or equivalent). The idle channel noise spectrum is stated in dB. A smoothed average idle channel noise spectrum is calculated by a moving average (arithmetic mean) 1/3rd octave wide across the idle noise channel spectrum stated in dB (linear average in dB of all FFT bins in the range from 2^-1/6)f to 2^(+1/6)f). Peaks in the idle channel noise spectrum are compared against a smoothed average idle channel noise spectrum.

6.4.2 Receive noise

Requirement

Telephone sets with adjustable receive levels shall be adjusted so that the RLR is as close as possible to the nominal RLR.

The receive noise shall be less than -57 dBPa(A).

Where a volume control is provided, the measured noise shall not be greater than -54 dBPa(A) at the maximum setting of the volume control.

Measurement Method

The handset terminal or the headset terminal is setup as described in clause 5.2.

For the actual measurement no test signal is used. In order to reliably activate the terminal an activation signal is introduced before the actual measurement. The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [10] shall be used for activation. The activation signal level shall be -16 dBm0. The noise level is measured until 10 kHz.

The A-weighted noise level shall be measured at DRP of the artificial ear with the diffuse field equalization active. The noise level is measured until 10 kHz.

Spectral peaks are measured in the frequency domain in the frequency range from 100 Hz to 6,3 kHz. The frequency spectrum of the A-weighted idle channel noise is measured by a spectral analysis having a noise bandwidth of 8,79 Hz (determined using FFT 8 k samples/48 kHz sampling rate with Hanning window or equivalent). The idle channel noise spectrum is stated in dB. A smoothed average idle channel noise spectrum is calculated by a moving average (arithmetic mean) 1/3rd octave wide across the idle noise channel spectrum stated in dB (linear average in dB of all FFT bins in the range from 2^(-1/6)f to 2^(+1/6)f). Peaks in the idle channel noise spectrum are compared against a smoothed average idle channel noise spectrum

NOTE: Care should be taken that only the noise is windowed out by the analysis and the analysis is not impaired by any remaining reverberance or room noise.

6.5 Send and receive distortion

6.5.1 Introduction

The send and receive distortions aim to qualify the harmonic distortion for different signal frequencies.

It is not intended to provide coder-dependant requirements but to assess the electroacoustic performance of the terminal.

NOTE: A new method intended to measure the noise level generated by the equipment in presence of speech signal is currently under study. If it may be implemented in the standard, it could replace at least one of these requirements and test methods.

6.5.2 Send Distortion

Requirement

The ratio of signal to harmonic distortion shall be above the following mask.

Table 6.5.2.1

Frequency (Hz)	Signal to harmonic distortion ratio limit, send (dB)
315	26
400	30
1 000	30
2 000	30
NOTE: The limits for intermediate frequencies lie on straight lines drawn between the given values on a linear (dB) - logarithmic (Hz) scale.	

Measurement method

The terminal will be positioned as described in clause 5.2.

After the correct activation of the system, a sinewave signal at frequencies of 315 Hz, 400 Hz, 500 Hz, 630 Hz, 800 Hz, 1 000 Hz, and 2 000 Hz. The duration of the sine wave shall be less than 1 s. The sinusoidal signal level shall be calibrated to -4,7 dBPa at the MRP.

The signal to harmonic distortion ratio is measured selectively up to 7 kHz.

The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [10] shall be used for activation. Level of this activation signal will be -4,7 dBPa at the MRP.

NOTE: Depending on the type of codec the test signal used may need to be adapted.

6.5.3 Receive distortion

Requirement

The ratio of signal to harmonic distortion shall be above the following mask.

Table 6.5.2.2

Frequency (Hz)	Signal to distortion ratio limit, receive (dB)
315	26
400	30
500	30
630	30
800	30
1 000	30
2 000	30
3 000	30

Measurement method

The terminal will be positioned as described in clause 5.2.

After a correct activation of the system, a digitally simulated sinewave signal at frequencies of 315 Hz, 400 Hz, 500 Hz,1 000 Hz, 2 000 Hz and 3 000 Hz shall be applied to the digital interface at the level of -16 dBm0.

The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [10] shall be used for activation. Level of this activation signal will be -16 dBm0.

The signal to harmonic distortion ratio is measured selectively up to 10 kHz.

6.6 Stability loss

Requirement

With the handset lying on and the transducers facing a hard surface, the attenuation from the digital input to the digital output shall be at least 6 dB at all frequencies in the range of 100 Hz to 8 kHz. In case of headsets the requirement applies for the closest possible position between microphone and headset receiver.

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

Measurement Method

Before the actual test a training sequence consisting of the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [10]. The training sequence level shall be -16 dBm0 in order not to overload the codec.

The test signal is a PN sequence complying with Recommendation ITU-T P.501 [10] with a length of 4 096 points (for the 48 kHz sampling rate) and a crest factor of 6 dB. The duration of the test signal is 250 ms. With an input signal of -3 dBm0, the attenuation from digital input to digital output shall be measured for frequencies from 200 Hz to 4 kHz under the following conditions:

a) the handset or the headset, with the transmission circuit fully active, shall be positioned on one inside surface that is of three perpendicular plane, smooth, hard surfaces forming a corner. Each surface shall extend 0,5 m from the apex of the corner. One surface shall be marked with a diagonal line, extending from the corner formed by the three surfaces, and a reference position 250 mm from the corner, as shown in figure 6.6.1;

b1) the handset, with the transmission circuit fully active, shall be positioned on the defined surface as follows:

1) the mouthpiece and earcap shall face towards the surface;

2) the handset shall be placed centrally, the diagonal line with the earcap nearer to the apex of the corner;

3) the extremity of the handset shall coincide with the normal to the reference point, as shown in figure 6.6.1.

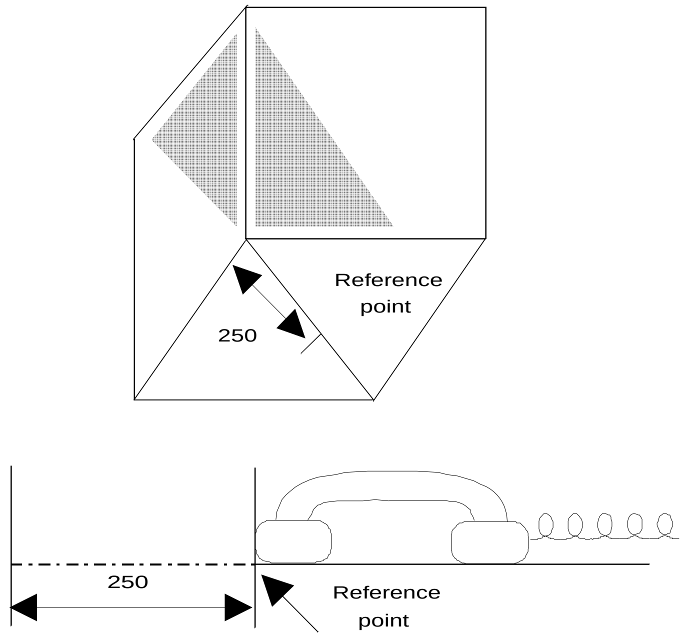
b2) the headset, with the transmission circuit fully active, shall be positioned on the defined surface as follows:

1) the microphone and the receiver shall face towards the surface;

2) for monaural the headset the receiver shall be placed centrally at the reference point as shown in figure 6.6.1;

   for binaural headset, the receivers are placed symmetrically to the diagonal line on both sides of the reference point;

3) the headset microphone is positioned as close as possible to the receiver(s).



NOTE: All dimensions in mm.

Figure 6.6.1

6.7 Terminal Coupling Loss

Requirement

The TCL measured as unweighted Echo Loss shall be ≥ 47 dB for all settings of the volume control (if supplied).

NOTE: A TCL ≥ 50 dB is recommended as a performance objective. Depending on the idle channel noise in the sending direction, it may not always be possible to measure an echo loss ≥ 50 dB.

Measurement Method

The handset or headset terminal is setup as described in clause 5.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [6]) and the application force shall be 2N on the artificial ear type 3.3 or type 3.4 as specified in Recommendation ITU-T P.57 [4]. The ambient noise level shall be less than -64 dBPa(A) for handset and headset terminals. The attenuation from electrical reference point input to electrical reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of Recommendation ITU-T P.501 Amendment 1 [10]. The signal level shall be -10 dBm0.

The TCL is calculated as the difference between the averaged test signal level and the averaged echo level in the frequency range from 100 Hz to 8 000 Hz. For the calculation the averaged measured echo level at each frequency band is referred to the averaged test signal level measured in each frequency band. The first 17,0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences). For the measurement, a time window has to be applied which is adapted to the duration of the actual test signal. For the measurement, a time window has to be applied which is adapted to the duration of the actual test signal. The echo loss is calculated by the equations (7a) and (7b):

$$L_e = C - 10\log_{10}\sum_{i=1}^{N}(A_i + A_{i-1})(\log_{10} f_i - \log_{10} f_{i-1}) \quad (7a)$$

and

$$C = 10\log_{10}(2\,(\log_{10} f_N - \log_{10} f_0) \quad (7b)$$

where

A_0 is the output/input power ratio at frequency $f_0 = 100\ Hz$;

A_1 the ratio at frequency f_i; and

A_N the ratio at frequency $f_N = 8\ 000\ Hz$.

The above equation composed of (7a) and (7b) is a generalized form of the equation defined in clause B.4 of Recommendation ITU-T G.122 [27] for calculating echo loss based on tabulated data, which allows the calculation of echo loss within any frequency range between f_0 and f_N.

The ambient noise level shall be < -64 dBPa(A).

NOTE 1: The extension of the frequency range is for further study.

NOTE 2: Care should be taken when measuring TCL: the echo return not to be masked by the residual noise or the comfort noise when implemented.

6.8 Double talk performance

6.8.1 Introduction

During double talk the speech is mainly determined by 2 parameters: impairment caused by echo during double talk and level variation between single and double talk (attenuation range).

In order to guarantee sufficient quality under double talk conditions the talker Echo Loudness Rating (ELR) should be high and the attenuation inserted should be as low as possible. Terminals which do not allow double talk in any case should provide a good echo attenuation which is realized by a high attenuation range in this case.

The most important parameters determining the speech quality during double talk are (see Recommendations ITU-T P.340 [8] and P.502 [11]):

- Attenuation range in send direction during double talk $A_{H,S,dt}$.
- Attenuation range in receive direction during double talk $A_{H,R,dt}$.
- Echo attenuation during double talk.

The categorization of a terminal is based on the three categories defined in clauses 6.8.2, 6.8.3 and 6.8.4 and this categorization is given by the lowest of the three parameters e.g. if $A_{H,S,dt}$ provides 2a, $A_{H,R,dt}$ 2b and echo loss 1, the categorization of the terminal is 2b.

6.8.2 Attenuation Range in Send Direction during Double Talk $A_{H,S,dt}$

Requirement

Based on the level variation in send direction during double talk $A_{H,S,dt}$ the behaviour of the terminal can be classified according to table 6.8.2.1.

The category of the terminal according to table 6.8.2.1 shall be noted in the test report.

Table 6.8.2.1

Category (according to Recommendation ITU-T P.340 [8])	1	2a	2b	2c	3
	Full Duplex Capability	Partial Duplex Capability			No Duplex Capability
$A_{H,S,dt}$ [dB]	≤ 3	≤ 6	≤ 9	≤ 12	> 12

Table 6.8.2.1 provides a quality classification of terminals regarding double talk performance. However, this does not mean that a terminal which is category 1 based on the double talk performance is of high quality concerning the overall quality as well.

Test signal

The test signal to determine the attenuation range during double talk is the double talk speech sequence as defined in clause 7.3.5 of Recommendation ITU-T P.501 [10] as shown in figure 6.8.2.1. The competing speaker is always inserted as the double talk sequence $s_{dt}(t)$ either in send or receive and is used for analysis.

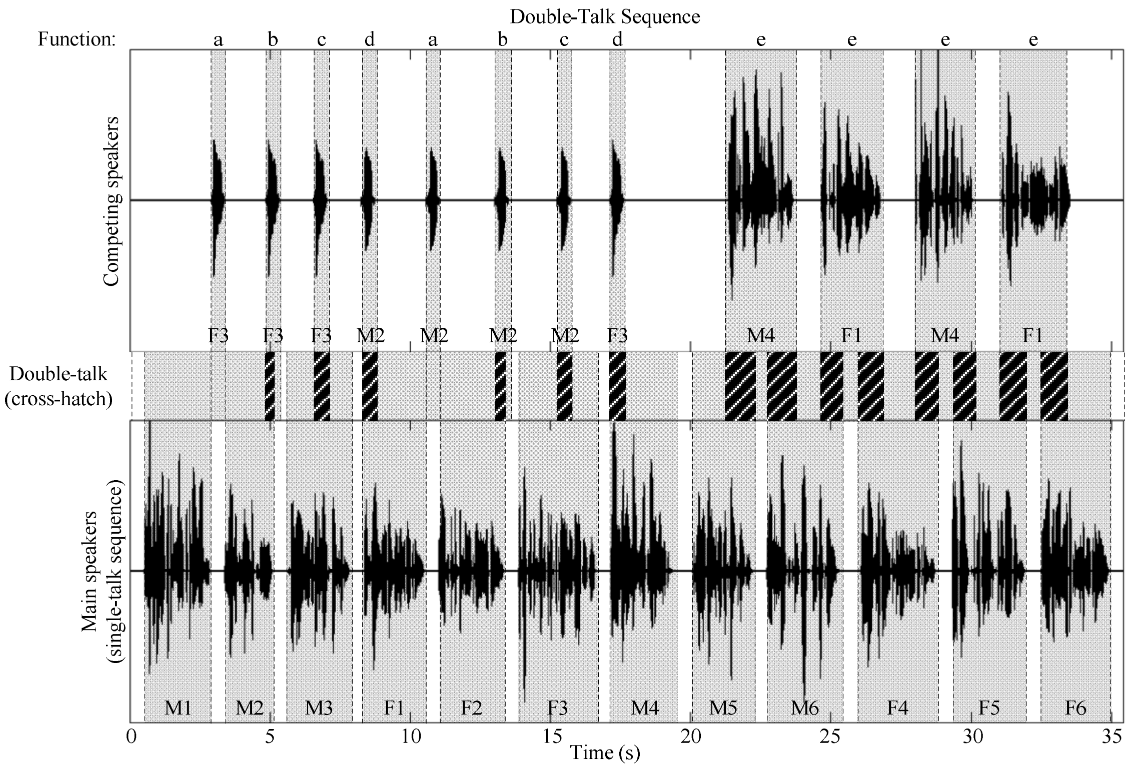


Figure 6.8.2.1: Double Talk Test Sequence with overlapping speech sequences in send and receive direction

Measurement method

The attenuation range during double talk is determined as described in Appendix III of Recommendation ITU-T P.502 [11]. The double talk performance is analysed for each word and sentence produced by the competing speaker. The requirement has to be met for each word and sentence produced by the competing speaker.

The terminal is positioned as described in clause 5.2. Before the actual test, a training sequence for the echo canceller consisting of the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501[10] with a level of -16 dBm0 is applied to the electrical reference point in the receiving direction.

When determining the attenuation range in send direction, the signal measured at the electrical reference point is referred to the test signal inserted.

The attenuation range during double talk is determined as described in appendix III of Recommendation ITU-T P.502 [11]. The double talk performance is analysed for the sequence of words and the sequence of sentences produced by the competing speaker. The requirement has to be met for the sequence of words and the sequence of sentences produced by the competing speaker.

6.8.3 Attenuation Range in Receive Direction during Double Talk $A_{H,S,dt}$

Requirement

Based on the level variation in receive direction during double talk $A_{H,R,dt}$ the behaviour of the terminal can be classified according to table 6.8.2.1.

The category of the terminal according to table 6.8.3.1 shall be noted in the test report.

Table 6.8.3.1

Category (according to Recommendation ITU-T P.340 [8])	**1**	**2a**	**2b**	**2c**	**3**
	Full Duplex Capability	Partial Duplex Capability			No Duplex Capability
$A_{H,R,dt}$ [dB]	≤ 3	≤ 5	≤ 8	≤ 10	> 10

Table 6.8.3.1 provides a quality classification of terminals regarding double talk performance. However, this does not mean that a terminal which is category 1 based on the double talk performance is of high quality concerning the overall quality as well.

Measurement Method

The test signal to determine the attenuation range during double talk is shown in figure 6.8.1. A sequence of speech signals is used which is inserted in parallel in send and receive direction. The test signals are synchronized in time at the acoustical interface. The delay of the test arrangement should be constant during the measurement.

The test arrangement is according to clause 5.

The attenuation range during double talk is determined as described in appendix III of Recommendation ITU-T P.502 [11]. The double talk performance is analysed for each word and sentence produced by the competing speaker. The requirement has to be met for each word and sentence produced by the competing speaker.

6.8.4 Detection of echo components during double Talk

Requirement

Echo Loss during double talk is the echo suppression provided by the terminal during double talk measured at the electrical reference point.

The category of the terminal according to table 6.8.4.1 shall be noted in the test report.

NOTE: The echo attenuation during double talk is based on the parameter Talker Echo Loudness Rating (TELRdt). It is assumed that the terminal at the opposite end of the connection provides nominal Loudness Rating (SLR + RLR = 10 dB).

Under these conditions the requirements given in table 6.8.4.1 are applicable (more information can be found in annex A of the Recommendation ITU-T P.340 [8]).

Table 6.8.4.1

Category (according to Recommendation ITU-T P.340 [8])	1	2a	2b	2c	3
	Full Duplex Capability	Partial Duplex Capability			No Duplex Capability
Echo Loss [dB]	≥ 27	≥ 23	≥ 17	≥ 11	< 11

Measurement Method

The test arrangement is according to clause 5.2.

The double talk signal consists of a sequence of orthogonal signals which are realized by voice-like modulated sine waves spectrally shaped similar to speech. A detailed description can be found in Recommendation ITU-T P.501 [10]. The measurement signals in send direction ($s_1(t)$) and receive direction ($s_2(t)$) are calculated according to figure 6.8.4.1 and equations (8) and (9). The settings for the signals are $F_{1,2}^{FM}(n) = \text{const} = 1$ Hz and $A_{1,2}^{AM}(n) = \text{const} = \frac{2}{3}$, the other settings are given in table 6.8.4.2.

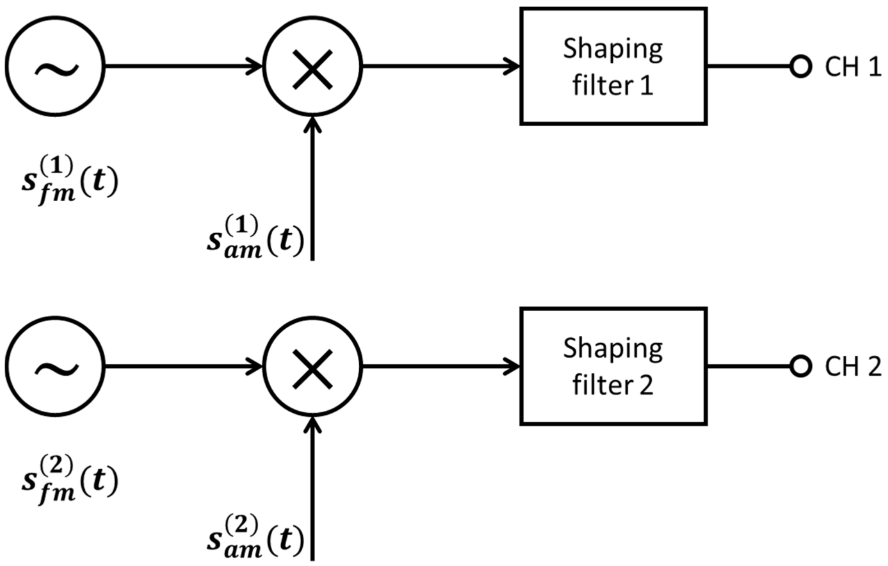


Figure 6.8.4.1: Measurement signals

$$s_{fm}^{(1,2)}(t) = \sum_n A_{1,2}(n) \cdot \cos\left(2\pi f_0^{(1,2)}(n) \cdot t + \mu_{fm}^{(1,2)} \cdot sin\left(2\pi f_{fm}^{(1,2)}(n) \cdot t\right)\right) \quad n = 1, 2, ... \tag{8}$$

$$\text{Where } \mu_{fm}^{(1,2)} = \frac{\Delta f^{(1,2)}(n)}{f_{fm}^{(1,2)}(n)}$$

$$s_{AM}^{(1,2)}(t) = \left(1 + \mu_{AM}^{(1,2)} \cdot \cos\left(2\pi f_{AM}^{(1,2)} \cdot t\right)\right) \tag{9}$$

Three parameters are chosen in a frequency-independent manner: $f_{fm}^{(1,2)}(n) = \text{const} = 1$ Hz, $\mu_{AM}^{(1,2)} = \text{const} = \frac{2}{3}$ and $f_{AM}^{(1,2)}(n) = \text{const} = 3$ Hz. The remaining two parameters are given in table 6.8.4.2. Both shaping filters are identical: a low pass with a cut-off frequency of 250 Hz and a slope of 5 dB/octave.

The amplitudes $A_{1,2}(n)$ that determine the signal levels may be chosen according to the application. Average measurement levels may be chosen, i.e. to -4,7 dB_{Pa} (SND) and -20 dB_V (RCV) for terminal testing. The test signal may be embedded in speech or speech-like sequences.

The settings for the signals are as follows.

Table 6.8.4.2: Parameters of the two Test Signals for Double Talk Measurement based on AM-FM modulated sine waves

Send Direction		ReceiveDirection	
$f_0^{(1)}$ [Hz]	$\pm\Delta f^{(1)}$ [Hz]	$f_0^{(2)}$ [Hz]	$\pm\Delta f^{(2)}$ ([Hz]
125	±2,5	180	±2,5
250	±5	270	±5
500	±10	540	±10
750	±15	810	±15
1 000	±20	1 080	±20
1 250	±25	1 350	±25
1 500	±30	1 620	±30
1 750	±35	1 890	±35
2 000	±40	2 160	±35
2 250	±40	2 400	±35
2 500	±40	2 650	±35
2 750	±40	2 900	±35
3 000	±40	3 150	±35
3 250	±40	3 400	±35
3 500	±40	3 650	±35
3 750	±40	3 900	±35
4 000	±40	4 150	±35
4 250	±40	4 400	±35
4 500	±40	4 650	±35
4 750	±40	4 900	±35
5 000	±40	5 150	±35
5 250	±40	5 400	±35
5 500	±40	5 650	±35
5 750	±40	5 900	±35
6 000	±40	6 150	±35
6 250	±40	6 400	±35
6 500	±40	6 650	±35
6 750	±40	6 900	±35
7 000	±40		
NOTE: Parameters of the Shaping Filter: f ≥ 250 Hz: Low Pass Filter, 5 dB/oct.			

The signals are fed simultaneously in send and receive direction. The level in send direction is -4,7 dBPa at the MRP (nominal level), the level in receive direction is -16 dBm0 at the electrical reference point (nominal level).The test signal is measured at the electrical reference point (send direction). The measured signal consists of the double talk signal which was fed in by the artificial mouth and the echo signal. The echo signal is filtered by comb filter using mid-frequencies and bandwidth according to the signal components of the signal in receive direction (see Recommendation ITU-T P.501 [10]). The filter will suppress frequency components of the double talk signal.

In each frequency band which is used in receive direction the echo attenuation can be measured separately. The requirement for category 1 is fulfilled if in any frequency band the echo signal is either below the signal noise or below the required limit. If echo components are detectable, the classification is based on table 6.8.4.1. The echo attenuation is to be achieved for **each individual frequency band** according to the different categories.

6.8.5 Minimum activation level and sensitivity of double talk detection

For further study.

6.9 Switching parameters

6.9.1 Activation in Send Direction

The activation in send direction is mainly determined by the built-up time $T_{r,S,min}$ and the minimum activation level ($L_{S,min}$). The minimum activation level is the level required to remove the inserted attenuation in send direction during idle mode. The built-up time is determined for the test signal burst which is applied with the minimum activation level.

The activation level described in the following is always referred to the test signal level at the Mouth Reference Point (MRP).

Requirement

The minimum activation level $L_{S,min}$ shall be ≤ -20 dBPa.

The built-up time $T_{r,S,min}$ (measured with minimum activation level) should be ≤ 15 ms.

Measurement Method

The test signal is the "short words for activation" sequence described in clause 7.3.4 of Recommendation ITU-T P.501 [10].

The settings of the test signal are as follows.

Table 6.9.1.1

	Single word Duration/ Pause Duration	Level of the first single word (active Signal Part at the MRP)	Level Difference between two Periods of the Test Signal
Single word to Determine Switching Characteristic in Send Direction	~600 ms / ~400 ms	-24 dBPa (see note)	1 dB
NOTE 1: The level of the active signal part corresponds to an average level of -24,7 dBPa at the MRP for the single word according to Recommendation ITU-T P.501 [10]. NOTE 2: The signal level is determined for each utterance individually according to Recommendation ITU-T P.56 [3].			

It is assumed that the pause length of about 400 ms is longer than the hang-over time so that the test object is back to idle mode after each single word

The test arrangement is described in clause 5.2.

The level of the transmitted signal is measured at the electrical reference point. The test signal is filtered by the transfer function of the test object. The measured signal level is referred to the filtered test signal level and displayed vs. time. The levels are calculated from the time domain using an integration time of 5 ms.

The minimum activation level is determined from the single word which indicates the first activation of the test object. The time between the beginning of the single word and the complete activation of the test object is measured.

6.9.2 Minimum activation level and sensitivity in Receive direction

For further study.

6.9.3 Automatic level control

For further study.

6.9.4 Silence Suppression and Comfort Noise Generation

For further study.

6.9.5 Non Linear Processing

Additional requirements may be needed in order to further investigate the effect of NLP implementations on the users' perception of speech quality.

6.10 Background noise performance

6.10.1 Performance in send direction in the presence of background noise

Requirement

The level of comfort noise shall be within in a range of +2 and -5 dB compared to the original (transmitted) background noise. The noise level is calculated with A-weighting.

NOTE 1: It is advisable that the comfort noise matches the original signal as good as possible (from a perceptual point of view).

NOTE 2: Input for further specification necessary (e.g. on temporal matching).

The spectral difference between comfort noise and original (transmitted) background noise shall be within the mask given through straight lines between the breaking points on a logarithmic (frequency) - linear (dB sensitivity) scale as given in table 6.10.1.1.

Table 6.10.1.1: Requirements for Spectral Adjustment of Comfort Noise (Mask)

Frequency	Upper Limit	Lower Limit
200 Hz	12 dB	-12 dB
800 Hz	12 dB	-12 dB
800 Hz	10 dB	-10 dB
2 000 Hz	10 dB	-10 dB
2 000 Hz	6 dB	-6 dB
4 000 Hz	6 dB	-6 dB
8 000 Hz	6 dB	-6 dB
NOTE: All sensitivity values are expressed in dB on an arbitrary scale.		

Measurement Method

The background noise simulation as described in clause 5.5 is used.

The handset terminal is set-up as described in clause 5.2. The handset is mounted at the HATS position (see Recommendation ITU-T P.64 [6]).

First the background noise transmitted in send is recorded at the POI for a period of at least 20 s.

In a second step a test signal is applied in receive direction consisting of an initial pause of 10 s and a periodical repetition of the Composite Source Signal in receive direction (duration 10 s) with nominal level to enable comfort noise injection simultaneously with the background noise. For the measurement the background noise sequence has to be started at the same point as it was started in the previous measurement. Alternatively other speech like test signals (e.g. artificial voice) with the same signal level can be used.

The transmitted signal is recorded in send direction at the POI.

The power density spectra measured in send direction without far end speech simulation averaged between 10 s and 20 s is referred to the power density spectrum measured in send direction determined during the period with far end speech simulation in receive direction averaged between 10 s and 20 s. Level and spectral differences between both power density spectra are analysed and compared to the requirements.

6.10.2 Speech Quality in the Presence of Background Noise

Requirement

Speech Quality for wideband systems shall be tested based on ETSI TS 103 106 [20].

For the background noises defined in clause 5.5 the following requirements apply:

- N-MOS-LQOw ≥ 3,5.
- S-MOS-LQOw ≥ 3,5.
- G-MOS-LQOw ≥ 3,5.

NOTE: It is recommended to test the terminal performance with other types of background noises if the terminal is likely to be exposed to other noises than specified in clause 5.5.

Measurement Method

The background noise simulation as described in clause 5.5 is used. The handset terminal is set-up as described in clause 5.2. The handset is mounted at the HATS position (see Recommendation ITU-T P.64 [6]).

The background noise should be applied for at least 5 s in order to adapt noise reduction algorithms in advance of the test.

The near end speech signal consists of 16 sentences of speech (2 male and 2 female talkers, 4 sentences each). An appropriate measurement sequence in American English is provided in annex C of ETSI TS 103 106 [20]. The test signal level is -1,7 dBPa at the MRP.

Three signals are required for the tests:

1) The clean speech signal is used as the undisturbed reference (see ETSI TS 103 106 [20]).
2) The speech plus undisturbed background noise signal is recorded at the terminal's microphone position using an omni directional measurement microphone with a linear frequency response between 50 Hz and 12 kHz.
3) The send signal is recorded at the electrical reference point.

N-MOS-LQOw, S-MOS LQOw and G-MOS LQOw are calculated as described in ETSI TS 103 106 [20].

6.10.3 Quality of Background Noise Transmission (with Far End Speech)

Requirement

The test is carried out applying a speech signal in receive direction. The test is carried out by comparing the noise level transmitted in the sending direction under reference conditions with no far end speech, to the noise level transmitted in the sending direction under test conditions including far end speech. During and after the end of the speech signal the signal level in send direction should not vary more than 10 dB (during transition to transmission of background noise without far end speech). The measurement is conducted for all types of background noise as defined in clause 5.5.

NOTE: The intention of this measurement is to detect impairments (modulations, switching and others) influencing the background noise transmitted from the terminal under test when a signal from the distant end (receiving side of the terminal under test) is present. Under these test conditions no modulation of the transmitted signal should occur. Modulation, switching or other type of impairments might be caused by an improper behaviour of a nonlinear processor working in conjunction with the echo canceller and erroneously switching or modulating the transmitted background noise.

Measurement Method

The test arrangement is according to clause 5.2.

The background noises are generated as described in clause 5.5.

First, the reference measurement is conducted without inserting the signal at the far end. At least 10 s of noise is analysed. The transmitted background signal level versus time is calculated using a time constant of 35 ms. This is the reference signal.

In a second step the test measurement is conducted using exactly the identical background noise, but with inserting the CS-signal at the far end. The background noise signal shall start at the same point in time as was used for the reference measurement without the far end signal. The background noise should be applied for at least 10 s in order to allow adaptation of the noise reduction algorithms and should be mixed with speech like signal e.g. CSS. After at least 10 s from the start of the background noise, to allow convergence of noise reduction algorithms, a CSS, or other appropriate speech test signal according to Recommendation ITU-T P.501 [10] is applied in receive direction with a duration of ≥ 2 CSS periods. The test signal level in the receiving direction is -16 dBm0 at the electrical reference point.

For both reference and test conditions, the send signal is recorded at the electrical reference point and the test signal level versus time is calculated using a time constant of 35 ms.

The difference in level in the send direction is determined during the time interval when the CS-signal is applied and after it stops. The level difference is defined as the difference of the recorded signal levels vs. time between the reference condition measured without far end signal and the test condition signal measured with far end signal. The difference should not exceed the requirement at any time within the analysis interval, defined as the time between the start of the arrival of the far end speech at the device under test and 500 msec after the end of the far end speech.

6.11 Quality of echo cancellation

6.11.1 Temporal echo effects

Requirement

This test is intended to verify that the system will maintain sufficient echo attenuation during single talk. The measured echo attenuation during single talk should not decrease by more than 6 dB from the maximum measured during the TCL test.

Measurement Method

The test arrangement is according to clause 5.2.

The background noises are generated as described in clause 5.5.

First, the reference measurement is conducted without inserting the signal at the far end. At least 10 seconds of noise is analysed. The background level versus time is calculated using a time constant of 35 ms. This is the reference signal.

In a second step the same measurement is conducted, but with inserting the speech signal at the far end. The exactly identical background noise signal is applied. The background noise signal should start at the same point in time as was used for the reference measurement without the far end signal. The background noise should be applied for at least 5 seconds in order to allow adaptation of the noise reduction algorithms. After at least 5 seconds a series of the female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [10] is applied in receive direction with duration of at least 10 s. The test signal level in the receiving direction is -16 dBm0 at the electrical reference point.

For both reference and test conditions, the send signal is recorded at the electrical reference point and the level versus time is calculated using a time constant of 35 ms.

The level variation in send direction is determined during the time interval when the speech signal is applied and after it stops. The level difference is determined from the difference of the recorded signal levels versus time between reference signal and the signal measured with far end signal.

NOTE : Care should be taken not to confuse noise or comfort noise with residual echo. In cases of doubt the measured echo signal should be compared to the residual noise signal measured under the same conditions without inserting the receive signal. If the level vs. time analysis leads to the identical result it can be assumed that no echo but just comfort noise is present.

6.11.2 Spectral Echo Attenuation

Requirement

The echo attenuation vs. frequency shall be below the tolerance mask given in table 6.11.2.1.

Table 6.11.2.1: Echo attenuation limits

Frequency	Limit
100 Hz	-41 dB
1 300 Hz	-41 dB
3 450 Hz	-46 dB
5 200 Hz	-46 dB
7 500 Hz	-37 dB
8 000 Hz	-37 dB
NOTE: The limit at intermediate frequencies lies on a straight line drawn between the given values on a log (frequency) - linear (dB) scale.	

During the measurement it should be ensured that the measured signal is really the echo signal and not the Comfort Noise which may possibly be inserted in send direction in order to mask the echo signal.

Measurement Method

The test arrangement is according to clause 5.2.

Before the actual measurement a training sequence is fed in consisting of 10 s CS signal according to Recommendation ITU-T P.501 [10]. The level of the training sequence is -16 dBm0.

The test signal consists of a periodically repeated Composite Source Signal. The measurement is carried out under steady-state conditions. The average test signal level is -16 dBm0, averaged over the complete test signal. 4 CS signals including the pauses are used for the measurement which results in a test sequence length of 1,4 s. The power density spectrum of the measured echo signal is referred to the power density spectrum of the original test signal. The analysis is conducted using FFT analysis with 8 k points (48 kHz sampling rate, Hanning window).

The spectral echo attenuation is analysed in the frequency domain in dB.

6.11.3 Occurrence of Artefacts

For further study.

6.11.4 Variable echo path

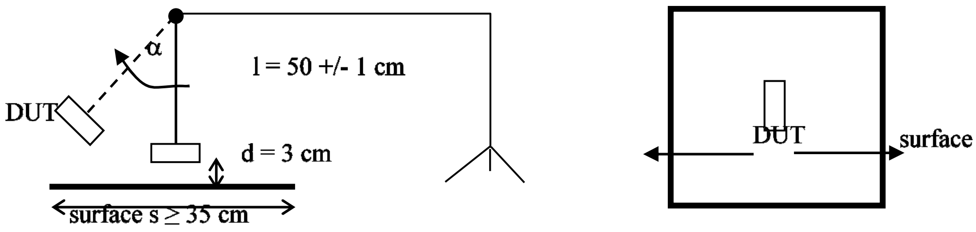
Requirement

This test is intended to verify that the system will maintain sufficient echo attenuation during single talk with dynamic changing echo paths. The measured echo level over time during single talk shall not be more than 10 dB above the minimum noise level during the measurement.

Measurement method

The handset is positioned d = 3 cm above a horizontal hard surface, facing the surface with speaker and microphone. The surface shall be at least 35 x 35 cm The handset is fixed like a pendulum with a non-elastic cord 3 cm above the centre of the horizontal surface, see figure 6.11.4.1. The pivot is 55 ± 1 cm above the hard plate.

Test setup for headsets: tbd.



side view **top view**

Figure 6.11.4.1: Positioning of handset under test

The "handset-pendulum" is displaced at least to the edge of the hard surface. The test signal playback shall start with the release of the displaced handset under test.

As test signal the compressed real speech signal described in clause 7.3.3 of Recommendation ITU-T P.501 [10] is used. The signal level shall be -10 dBm0. The terminal volume control is set to nominal RLR. The first 4 sentences of the test signal are used to allow full convergence of the echo canceller. The next 4 sentences (from 10,75 s to 22,5 s) are used for the analysis. The echo signal level is analysed over time The echo signal level is analysed for 11,75 s, using a time constant of 35 ms.

The measurement result is displayed as echo level versus time

No level peak should be more than 10 dB above the minimum noise level during the measurement.

6.12 Send and receive delay or round trip delay

Requirement

Send and receive delays are tested separately but the requirement is defined for the combination of send and receive delays (round-trip delay).

It is recognized that the end to end delay should be as small as possible in order to ensure high quality of the communication.

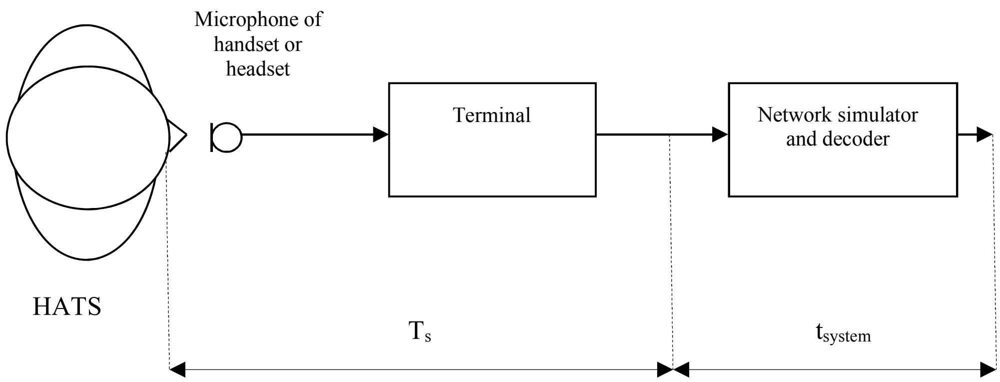
The round-trip delay T_{rtd} (sum of send and receive delay) shall be less than 100 ms (category B in Recommendation ITU-T P.1010 [24].

From the users perspective, a value less than 50 ms (category A in Recommendation ITU-T P-1010 [21]) is preferred.

Measurement method

- **Send direction**

The delay in send direction is measured from the MRP to POI. The delay measured in send direction is T_s+ t_{System}.



NOTE : The delay should be minimized! This can, e.g. be accomplished by designing the speech decoder output, the additional radio link, and the signal processing in a way, that sample-based processing and frame-based processing interoperate by using common buffers at their interfaces.

Figure 6.12.1: Different blocks contributing to the delay in send direction

The system delay t_{System} is depending on the transmission method used and the network simulator. The delay t_{System} shall be known:

1) For the measurements a Composite Source Signal (CSS) according to Recommendation ITU-T P.501 [10] is used. The pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal level is -4,7 dBPa at the MRP.

   The reference signal is the original signal (test signal).

   The setup of the handset/headset terminal is in correspondence to clause 5.2.

2) The delay is determined by cross-correlation analysis between the measured signal at the electrical access point and the original signal. The measurement is corrected by delays which are caused by the test equipment.

3) The delay is measured in ms and the maximum of the envelope of the cross-correlation function is used for the determination.

- **Receive direction**

The delay in receive direction is measured from POI to the Drum Reference Point (DRP). The delay measured in receive direction is $T_r + t_{System}$.

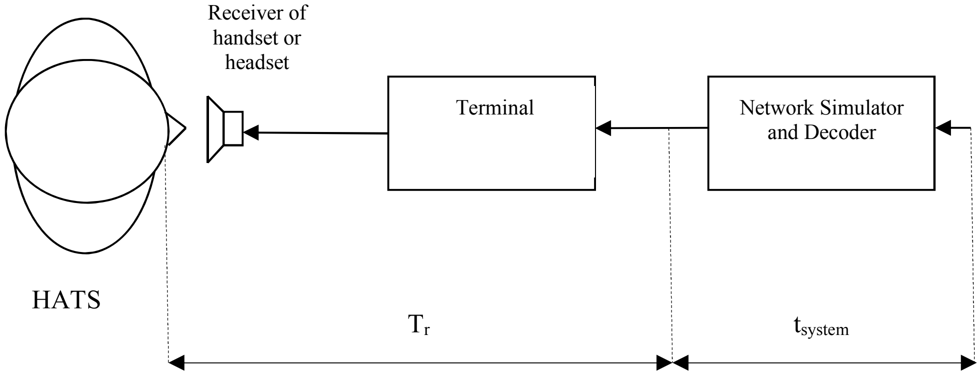


Figure 6.12.2: Different blocks contributing to the delay in receive direction

The system delay t_{System} depends on the transmission system and on the network simulator used. The delay t_{System} shall be known:

1) For the measurements a Composite Source Signal (CSS) according to Recommendation ITU-T P.501 [10] is used. The pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal level is -16 dBm0 at the electrical interface (POI).

2) The reference signal is the original signal (test signal).

3) The test arrangement is according to clause 5.2.

4) The delay is determined by cross-correlation analysis between the measured signal at the DRP and the original signal. The measurement is corrected by delays which are caused by the test equipment.

5) The delay is measured in ms and the maximum of the envelope of the cross-correlation function is used for the determination.

6.13 Objective listening Quality in send and receive direction

The aim is to provide the best listening quality whatever the implementation is:

- Provisional target value: $MOS\text{-}LQO_M > 4{,}0$.
- Optimum target value: $MOS\text{-}LQO_M > 4{,}3$.

As the actual listening quality depends on the codec implementations, specific requirements and test methods are defined in clause 7.

This clause will be updated when the relevant quality model will be available.

7 Codec dependent requirements and associated Measurement Methodologies

7.1 Speech Coders

The present document is intended to be applicable for different speech coders implemented in access networks and additional links.

Table 7.1.1 defines a list of speech coders implemented in the terminals (non-exhaustive).

Table 7.1.1: Speech coders

System	Codec
GSM 850, 900, 1 800, 1 900	AMR-WB (ITU-T G.722.2)@12,65 kbit/s [19] EVS WB @ 13,2 kbit/s [22]
UMTS (WCDMA), VoLTE	AMR-WB (ITU-T G.722.2) @12,65 kbit/s [19] EVS WB @ 13,2 kbit/s [22]
Voice over Data Network (VoDN)	Recommendation ITU-T G.722 [17] Recommendation ITU-T G.722.2 [19] Recommendation ITU-T G.729.1 [16] Recommendation ITU-T G.711.1 [18]

The objective is to minimize the impact of transcodings on the quality. Care should also be taken to avoid as far as possible to cascade different speech processing.

7.2 Objective listening Quality in send and receive direction

7.2.1 Objective listening speech quality MOS-LQO in send direction

The listening speech quality tests are conducted under clean network conditions.

Requirements

The requirements for the listening speech quality are as follows.

Table 7.2.1.1

Speech coder	MOS-LQOS (P.863 [25])	MOS-LQOM (TOSQA 2001 [i.1])
AMR-WB (ITU-T G.722.2)@12,65 kbit/s [19]	4,0	(ffs)
EVS WB @ 13,2 kbit/s [22]	(ffs)	(ffs)
Recommendation ITU-T G.722 [17]	3,9	(ffs)
Recommendation ITU-T G.729.1 [16]	4,1	(ffs)
Recommendation ITU-T G.711.1 [18]	(ffs)	(ffs)

NOTE 1: Recommendation ITU-T P.863 [25] is using a superwideband scale. Insufficient experience is available so far with this method. Therefore the numbers for MOS-LQS are provisional and may be updated with a later revision of the present document.

Measurement method

Objective listening speech quality is measured using Recommendation ITU-T P.863 [25] in superwideband mode.

The test signal to be used for the measurements shall be 4 sentence pairs (male/female) fulfilling the requirements of Recommendation ITU-T P.863.1 [26]. The 4 sentence pairs are taken from Recommendation ITU-T P.501 [10], annex C. It shall be stated, which sentence pairs were used. The test signal level is averaged over all sentence pairs (4 sentence pairs). The measurement is done 4 times, every time using another pair of the speech sentences. The result of the measurement is the averaged value of all 4 measurements.

NOTE 2: For the use of Recommendation ITU-T P.863 [25] the following applies (see Recommendation ITU-T P.863.1 [26]):

Superwideband Context (MOS-LQOS):

- Reference Signal Superwideband flat filtered 50 Hz to 14 kHz.
- Test Signal Superwideband flat filtered 50 Hz to 14 kHz.

NOTE 3: An alternative test method is TOSQA 2001 (ETSI EG 201 377-1 [i.1]). With TOSQA, terminals used in narrowband and wideband mode should be measured based on MOS-LQOM.

7.2.2 Objective listening quality MOS-LQO in receive direction

The listening speech quality tests are conducted under clean network conditions as well as with network impairments simulated. In addition to the listening speech quality tests the delay is measured.

Requirements

The requirement for the listening speech quality and the delay under clean network conditions are as follows.

Table 7.2.2.1

Speech coder	MOS-LQOS (P.863)	MOS-LQOM (TOSQA 2001)
AMR-WB (ITU-T G.722.2)@12,65 kbit/s [19]	(ffs)	(ffs)
EVS WB @ 13,2 kbit/s [22]	(ffs)	(ffs)
Recommendation ITU-T G.722 [17]	(ffs)	(ffs)
Recommendation ITU-T G.722.2 [19]	(ffs)	(ffs)
Recommendation ITU-T G.729.1 [16]	(ffs)	(ffs)
Recommendation ITU-T G.711.1 [18]	(ffs)	(ffs)

NOTE 1: Not sufficient experience is available so far with Recommendation ITU-T P.863 [25] and TOSQA 2001 (ETSI EG 201 377-1 [i.1]) for measuring handsfree terminals. Therefore the numbers for MOS-LQOS and MOS-LQON are for further study.

Measurement method

Objective listening speech quality is measured using Recommendation ITU-T P.863 [25] in superwideband mode.

The test signal to be used for the measurements shall be 4 sentence pairs (male/female) fulfilling the requirements of Recommendation ITU-T P.863.1 [26]. The 4 sentence pairs are taken from Recommendation ITU-T P.501 [10], annex C. It shall be stated, which sentence pairs were used. The test signal level is averaged over all sentence pairs (4 sentence pairs). The measurement is done 4 times, every time using another pair of the speech sentences. The result of the measurement is the averaged value of all 4 measurements.

NOTE 2: For the use of Recommendation ITU-T P.863 [25] the following applies (see Recommendation ITU-T P.863.1 [26]):

Superwideband Context (MOS-LQOS):

- Reference Signal Superwideband flat filtered 50 Hz to 14 kHz.
- Test Signal Wideband flat low pass filtered 7,8 kHz.

NOTE 3: An alternative test method is TOSQA 2001 (ETSI EG 201 377-1 [i.1]). With TOSQA, terminals used in narrowband and wideband mode should be measured based on MOS-LQOM.

For the performance tests with network impairments the following settings are used.

Table 7.2.2.2: Network conditions for electrical-acoustical measurements (speech samples)

Table 7.2.2.2 is for further study.

8 Requirements and associated Measurement Methodologies (with an additional radio link between the terminal and external electroacoustical devices)

The intention is to provide requirements and test methods for the complete chain.

Annex A (informative): Bibliography

ETSI TS 126 131: "Universal Mobile Telecommunications System (UMTS); LTE; Terminal acoustic characteristics for telephony; Requirements (3GPP TS 26.131 version 13.3.0 Release 13)".

ETSI TS 126 132: "Universal Mobile Telecommunications System (UMTS); LTE; Speech and video telephony terminal acoustic test specification (3GPP TS 26.132 version 13.3.0 Release 13)".

ETSI ES 202 737: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for narrowband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user".

ETSI ES 202 738: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for narrowband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user".

ETSI ES 202 739: "Speech and multimediaTransmission Quality (STQ); Transmission requirements for wideband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user".

ETSI ES 202 740: "Speech and multimediaTransmission Quality (STQ); Transmission requirements for wideband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user".

ETSI EN 300 176-2: "Digital Enhanced Cordless Telecommunications (DECT); Test specification; Part 2: Speech".

Recommendation ITU-T P.50: "Artificial voices".

History

Document history		
V1.1.1	November 2009	Publication
V1.1.2	September 2010	Publication
V1.2.1	July 2017	Publication