ETSI TS 103 739 V1.1.1 (2009-11)

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
DTS/STQ-00140-3

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

Contents

Intellectual Property Rights

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 (http://webapp.etsi.org/IPR/home.asp).

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.

Foreword

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

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.

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.

In contrast to 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 environment. 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

References are either specific (identified by date of publication and/or edition number or version number) or non-specific.

- For a specific reference, subsequent revisions do not apply.
- Non-specific reference may be made only to a complete document or a part thereof and only in the following cases:
  - if it is accepted that it will be possible to use all future changes of the referenced document for the purposes of the referring document;
  - for informative references.

Referenced documents which are not found to be publicly available in the expected location might be found at http://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.

2.1 Normative references

The following referenced documents are indispensable for the application of the present document. For dated references, only the edition cited applies. For non-specific references, the latest edition of the referenced document (including any amendments) applies.

[1] ETSI EG 202 396-1: "Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database".

[2] ETSI EG 202 396-3: "Speech Processing, Transmission and Quality Aspects (STQ); Speech Quality performance in the presence of background noise Part 3: Background noise transmission - Objective test methods".

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

[4] ITU-T Recommendation P.50: "Artificial voices".

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

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

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

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

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

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

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

[12] ITU-T Recommendation P.501: "Test signals for use in telephonometry".

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

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

[15] IEC 61672: "Electroacoustics - Sound level meters".

[16] ISO 3 (1973): "Preferred numbers - Series of preferred numbers".

[17] ETSI TS 126 171: "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)".

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

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

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

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

2.2 Informative references

The following referenced documents are not essential to the use of the present document but they assist the user with regard to a particular subject area**.** For non-specific references, the latest version of the referenced document (including any amendments) applies.

[i.1] ETSI TS 103 737: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for narrowband wireless terminals (handset and headset) from a QoS perspective as perceived by the user".

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 ITU-T Recommendation P.58 [7]

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): is 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.	alternative current
CSS	Composite Source Signal
D	D-Value of Terminal
DECT	Digital Enhanced Cordless Telecommunications
DRP	ear Drum Reference Point
ERP	Ear Reference Point
FFT	Fast Fourier Transform
HATS	Head And Torso Simulator
HRP	HATS Reference Point
MRP	Mouth Reference Point
POI	Point Of Interconnect
QoS	Quality of Service
RLR	Receive Loudness Rating
SLR	Send Loudness Rating
STMR	SideTone Masking Rating
TCLw	Terminal Coupling Loss (weighted)

4 Configurations and interfaces

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

4.1 Access networks

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

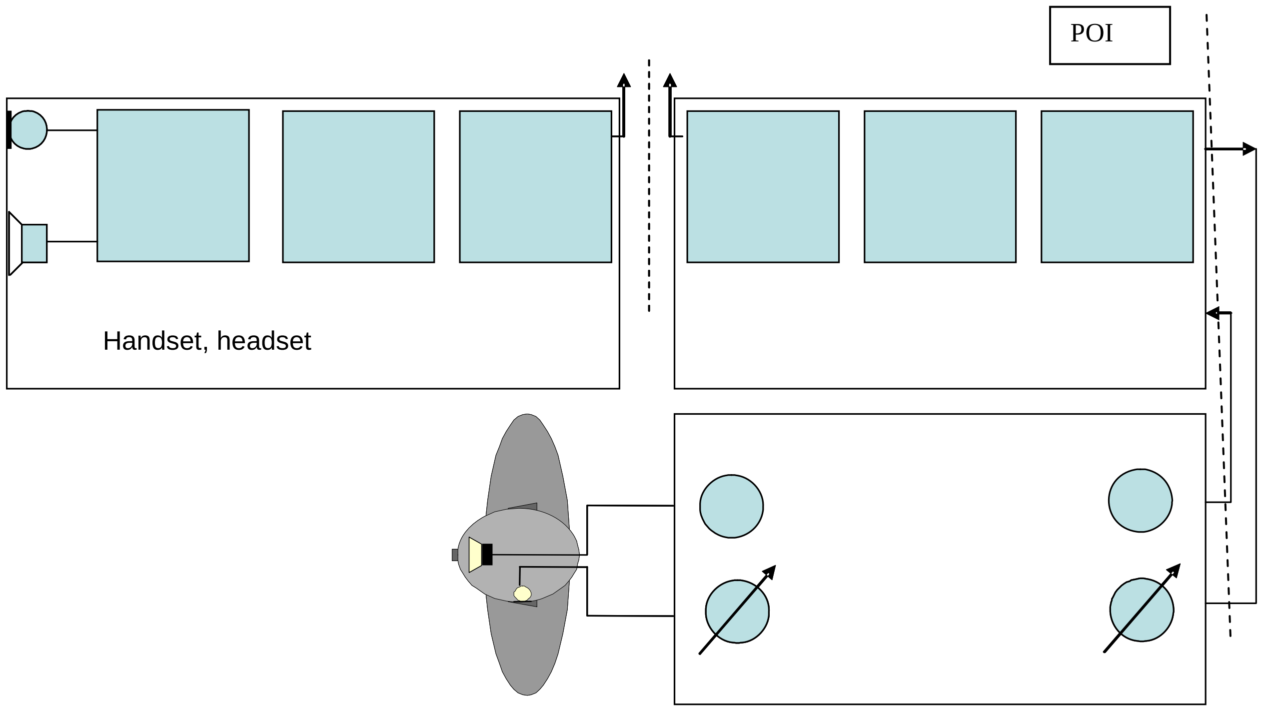
4.2 Additional (radio) links between the terminal and external electroacoustical devices

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 ITU-T Recommendation P.58 [7], appropriate ears are described in ITU-T Recommendation P.57 [6] (type 3.3 and type 3.4 ear), a proper positioning of handsets under realistic conditions is to be found in ITU-T Recommendation P.64 [8].

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: 12,65 kbit/s [18]
- ITU-T Recommendation G.729.1 [18]: 32 kbit/s

Setup for handsets and headsets

When using a handset telephone the handset is placed in the HATS position as described in ITU-T Recommendation P.64 [8]. The artificial mouth shall be conform with ITU-T Recommendation P.58 [7]. The artificial ear shall be conform with ITU-T Recommendation P.57 [6], type 3.3 or type 3.4 ears shall be used.

Recommendations for positioning headsets are given in ITU-T Recommendation P.380 [11]. 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 ITU-T Recommendation P.380 [11].

Unless stated otherwise if a volume control is provided the setting is 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.

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 must 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

Due to the coding of the speech signals, care should be taken when using single frequency for wireless terminals/networks (e.g. GSM/3G) acoustic tests. Appropriate test signals (general description) are defined in ITU-T Recommendations P.50 [4] and P.501 [12]. Normative requirements for the use of test signals from P.501 are for further study.

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 (band limited in receive direction) test signal, 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 P.57 [6] shall be used. The artificial ear shall be positioned on HATS according to ITU-T Recommendation P.58 [7].

The exact calibration and equalization can be found in ITU-T Recommendation P.581 [14].

Setup of background noise simulation

A setup for simulating realistic background noises in a lab-type environment is described in EG 202 396-1 [1].

The EG 202 396-1 [1] 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.

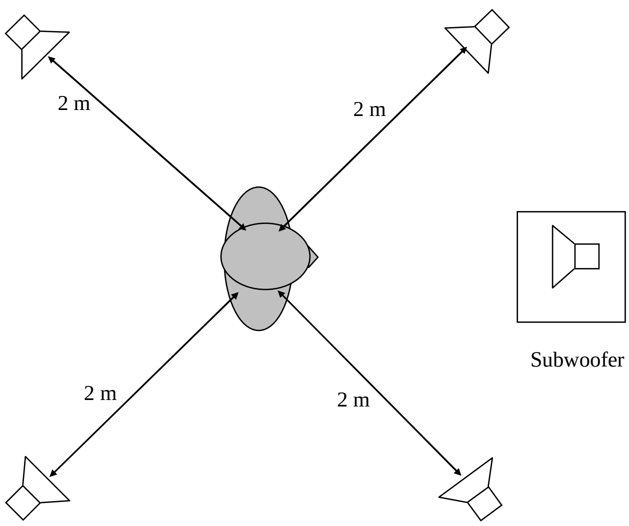


Figure 5.5.1: Loudspeaker arrangement for background noise simulation

The equalization and calibration procedure for the setup is described in detail in EG 202 396-1 [1].

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

The following noises of EG 202 396-1 [1] shall be used.

Table 5.5.1: Noises used for background noise simulation

Recording in pub	Pub_Noise_binaural	30 s	L: 77,8 dB(A) R: 78,9 dB(A)	binaural
Recording at sales counter	Cafeteria_Noise_binaural	30 s	L: 68,4 dB(A) R: 67,3 dB(A)	binaural
Recording in business office	Work_Noise_Office_Callcener_binaural	30 s	L: 56,6 dB(A) R: 57,8 dB(A)	binaural

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 Power	±0,2 dB for levels ≥ -50 dBm
Electrical Signal Power	±0,4 dB for levels < -50 dBm
Sound pressure	±0,7 dB
Time	±0,2 %
Frequency	±0,2 %
Application force	±2 Newton

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 8 kHz
Electrical excitation levels	±0,4 dB **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 [15] 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.

6 Codec independent requirements and associated Measurement Methodologies

6.1 Send and receive frequency response

6.1.1 Send frequency response

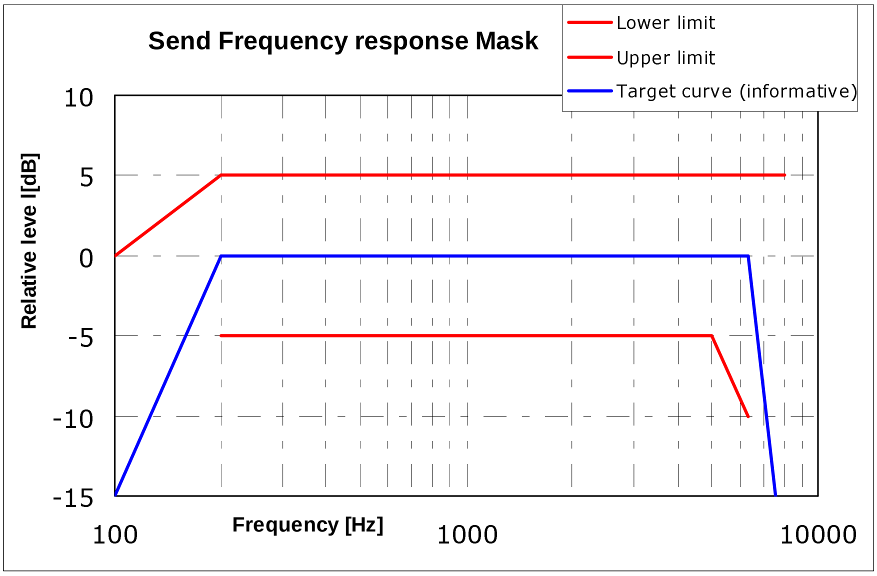
Due to diffuse field equalisation applying in the receive direction a flat curve is preferable in send path.

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

Test method

The test signal to be used for the measurements shall be the artificial voice according to ITU-Recommendation P. 50 [4]. If the signal to noise ratio in high frequency domain is not sufficient Composite Source Signal (CSS) or a speech-like test signal as defined in ITU-T Recommendation P.501 [12] shall be used. The type of test signal used shall be stated in the test report. 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, duration. 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 ITU-T Recommendation P.64 [8]). The application force used to apply the handset against the artificial ear shall be within the range specified in ITU-T Recommendation P.64 [8].

Measurements shall be made at one twelfth-octave intervals as given by the R.40 series of preferred numbers in ISO 3 [16] 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.

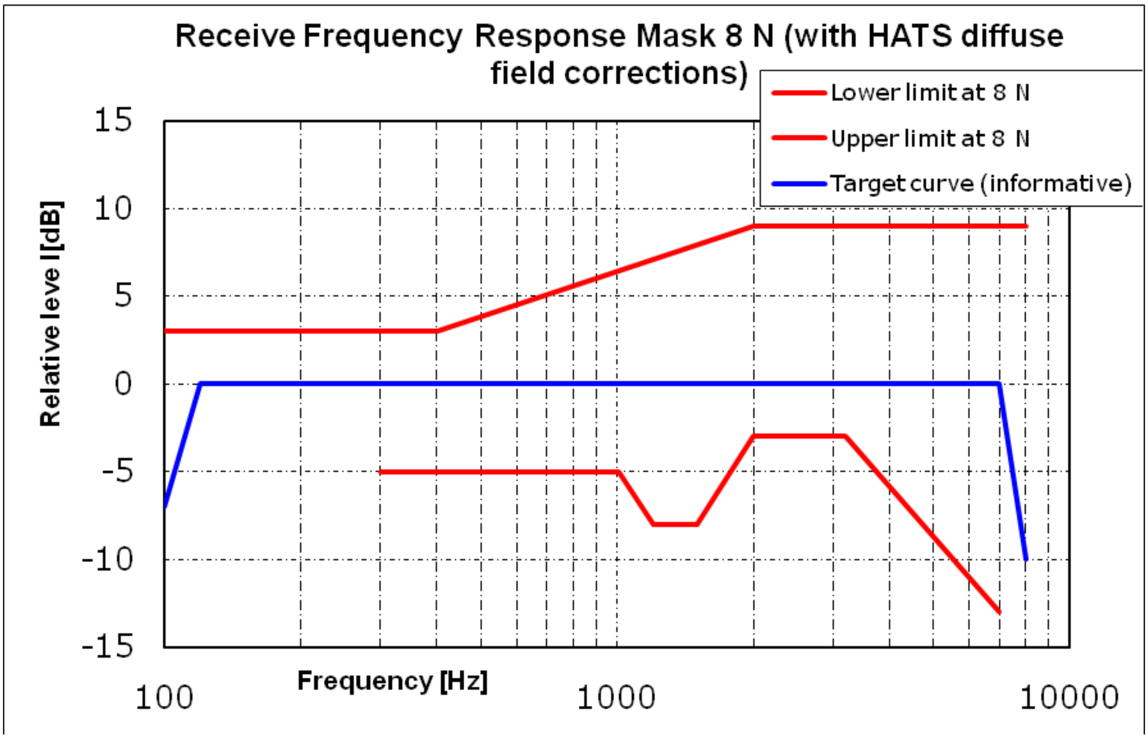


Figure 6.1.2.1: Receive frequency response mask for 8N application force

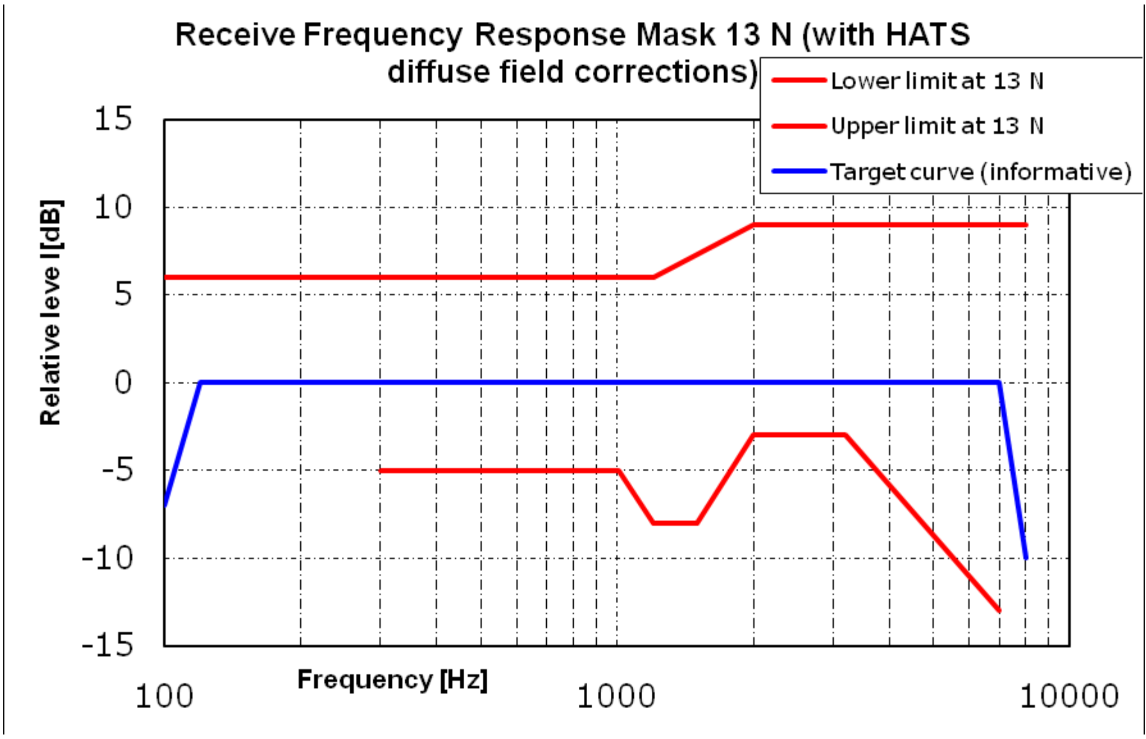


Figure 6.1.2.2: Receive frequency response mask for 13N application force

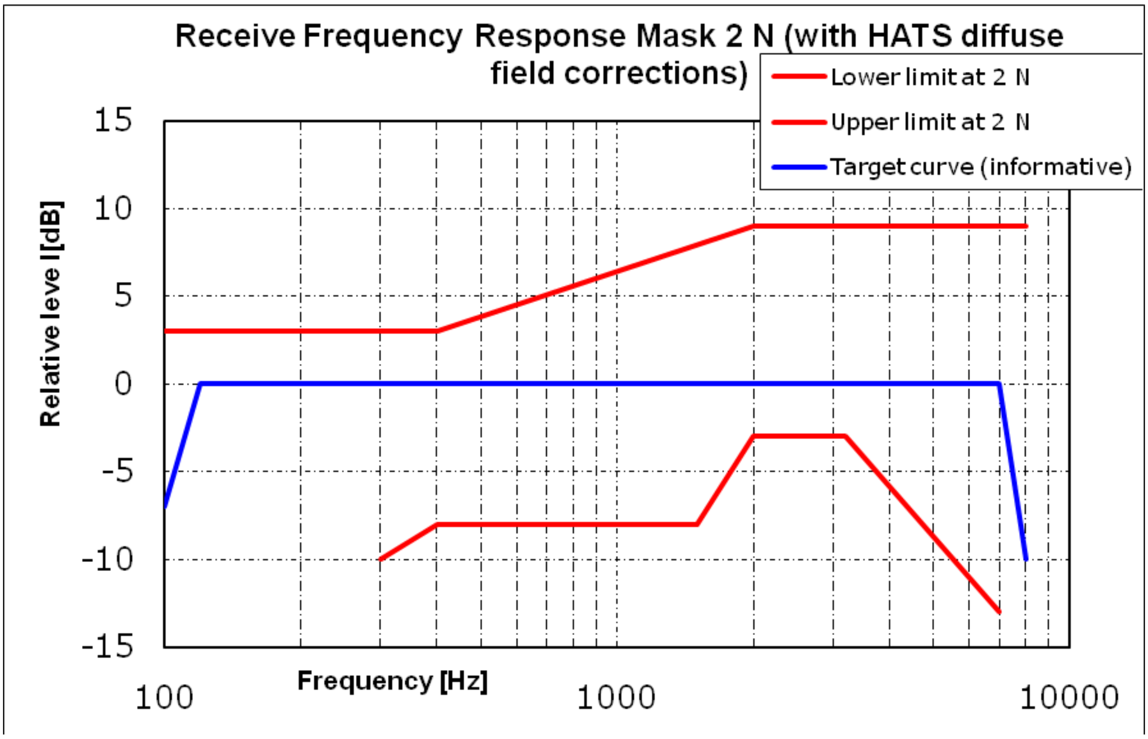


Figure 6.1.2.3: Receive frequency response mask for 2N application force

Test method

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

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

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 artificial voice according to ITU-T Recommendation P.50 [4], duration 20 s (10 s female, 10 s male voice). If the signal to noise ratio in high frequency domain is not sufficient Composite Source Signal (CSS) or a speech-like test signal as defined in ITU-T Recommendation P.501 [12] shall be used. The type of test signal used shall be stated in the test report. The test signal level shall be -16 dBm0, measured according to ITU-T Recommendation P.56 [5] 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 ITU-T Recommendation P.64 [8]). 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 ITU-T Recommendation P.380 [11] 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 the R.40 series of preferred numbers in ISO 3 [16] 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:

- SLR(set) = +8 dB ± 3 dB

Measurement Method

The test signal to be used for the measurements shall be the artificial voice according to ITU-T Recommendation P.50 [4], duration 20 s (10 s female, 10 s male voice). If the signal to noise ratio in high frequency domain is not sufficient Composite Source Signal (CSS) or a speech-like test signal as defined in ITU-T Recommendation P.501 [12] shall be used. The type of test signal used shall be stated in the test report. 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 ITU-T Recommendation P.64 [8]). 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 ITU-T Recommendation P.380 [11] 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 ITU-T Recommendation P.79 [9], 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 ITU-T Recommendation P.79 [9], formula (A - 23b), over bands 1 to 20, using m = 0,175 and the send weighting factors from ITU-T Recommendation P.79 [9], annex A, table A.2.

6.2.2 Receive Loudness Rating (RLR)

Requirement

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

RLR = +2 ± -3 dB.

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.

NOTE: The mechanical design of some UE may make it impossible to seal the ear-piece to the knife edge of the ITU-T artificial ear. Minimal additional methods may be used to provide the seal provided that they do not affect the mounting position of the UE with respect to the Mouth Reference Point and the Ear Reference Point.

For Binaural headset:

RLR (binaural headset) = +8 dB ± 3 dB for each earphone

Measurement Method

The test signal to be used for the measurements shall be the artificial voice according to ITU-T Recommendation P.50 [4], duration 20 s (10 s female, 10 s male voice). If the signal to noise ratio in high frequency domain is not sufficient Composite Source Signal (CSS) or a speech-like test signal as defined in ITU-T Recommendation P.501 [12] shall be used. The type of test signal used shall be stated in the test report. 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 6.2. The handset is mounted in the HATS position (see ITU-T Recommendation P.64 [8]). 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 ITU-T Recommendation P.57 [6] 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 ITU-T Recommendation P.380 [11] 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 ITU-T Recommendation P.79 [9], 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 ITU-T Recommendation P.79 [9], formula (A - 23c), over bands 1 to 20, using m = 0,175 and the receive weighting factors from table A.2 of ITU-T Recommendation P.79 [9], annex A.

No leakage correction shall be applied for the measurement.

6.2.3 LR stability

For further studies.

6.3 Sidetone parameters

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.

6.3.1 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 must 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 artificial voice according to ITU-T Recommendation P.50 [4]. 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 ITU-T Recommendation P.64 [8]) 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 by the R.40 series of preferred numbers in ISO 3 [16] for frequencies from 100 Hz to 8 kHz inclusive. For the calculation the averaged measured level at each frequency band (ITU-T Recommendation P.79 [9], 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 ITU-T Recommendation P.79 [9], using m = 0,225 and the weighting factors of in table 3 of ITU-T Recommendation P.79 [9].

6.3.2 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 ITU-T Recommendation P.64 [8]).

The test signal is a CS-signal complying with ITU-T Recommendation P.501 [12] 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 ITU-T Recommendation P.501 [12]. 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:

$$\Phi_{xy}(\tau) = \lim_{T\to\infty} \sum_{t=-T/2}^{T/2} S_x(t) S_y(t+\tau) \qquad (2)$$

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:

$$H\{xy(\tau)\} = \sum_{-\infty}^{\infty} \frac{\Phi_{xy}(u)}{\Pi(\tau - u)} \qquad (3)$$

$$E(\tau) = \sqrt{[\phi xy(\tau)]^2 + \{H[\phi xy(\tau)]\}^2} \qquad (4)$$

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

6.3.3 D-factor

This parameter is replaced by the parameter " Background noise performance" defined in clause 6.9.

Information on D-factor is to be found in annex A in TS 103 737 [i.1].

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 a sequence of 4 composite source signals (CSS) as described in ITU-T Recommendation P.501 [12]. 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 ITU-T Recommendation P.64 [8]).

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).

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.

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

An artificial voice according to ITU-Recommendation P.50 [4] or a speech like test signal as described in ITU-T Recommendation P.501 [12] can be used for activation. The activation signal level shall be -16 dBm0.

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

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: Anew 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.

6.5.1 Send Distortion

Requirement

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

Table 6.5.1.1

Frequency (Hz)	Signal to harmonic distortion ratio limit, send (dB)
315	28
400	32
1 000	35
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 a correct activation of the system, a sinewave signal at frequencies of 315 Hz, 400 Hz, 500 Hz, 630 Hz, 800 Hz and 1 000 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.

An artificial voice according to ITU-Recommendation P.50 [4] or a speech like test signal as described in ITU-T Recommendation P.501 [12] can 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.

6.5.2 Receive distortion

Requirement

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

Table 6.5.2.1

Frequency (Hz)	Signal to distortion ratio limit, receive (dB)
315	20
400	28
500	30
1 000	33

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 and 1 000 Hz is applied to the digital interface respectively. The sinewave signal with a frequency of 1 000 Hz shall be applied to the digital interface at the level of -16 dBm0.

An artificial voice according to ITU-Recommendation P.50 [4] or a speech like test signal as described in ITU-T Recommendation P.501 [12] can be used for activation. Level of this activation signal will be -16 dBm0.

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 10 s artificial voice male and 10 s artificial voice female according to ITU-T Recommendation P.50 [4] is altered. The training sequence level shall be -16 dBm0 in order not to overload the codec.

The test signal is a PN sequence complying with ITU-T Recommendation P.501 [12] 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 3a.

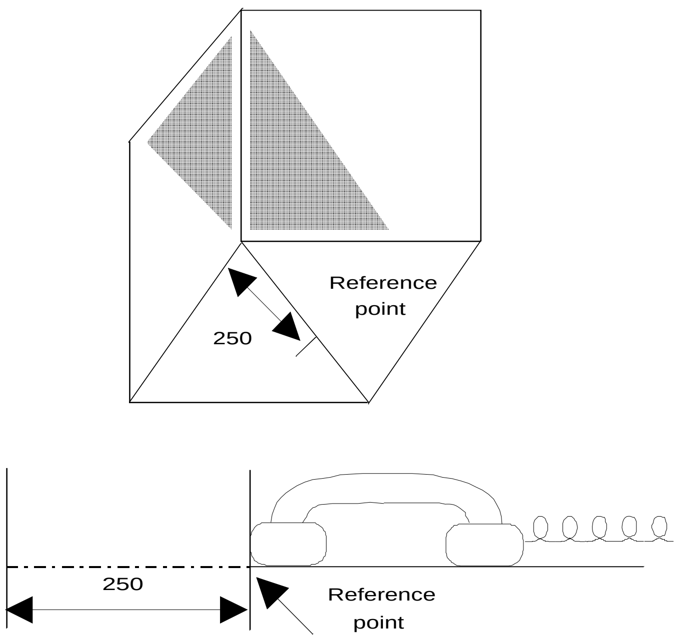
b2) the headset, with the transmission circuit fully active, shall be positioned on the defined surface as follows:

   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 TCLw (or similar parameters)

Requirement

The TCLw shall be ≥ 55 dB.

With the volume control set to maximum TCLw shall be ≥ 46 dB. The volume control shall be set back to nominal after each call unless TCLw ≥ 55 dB can be maintained also with maximum volume setting.

NOTE: A TCLw value of 50 dB is a value currently observed and a value of 55 dB is an achievable value with proper design. It has been noted that residual echo may be perceived even with TCLw higher than 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 ITU-T Recommendation P.64 [8]) and the application force shall be 2N on the artificial ear type 3.3 or type 3.4 as specified in ITU-T Recommendation P.57 [6]. 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 a speech like test signal.

Before the actual test a training sequence consisting of 10 s male artificial voice followed by 10 s female artificial voice according to ITU-T Recommendation P.50 [4] is applied. The training sequence level shall be -16 dBm0 in order not to overload the codec.

The test signal following immediately the training sequence is a PN-sequence complying with ITU-T Recommendation P.501 [12] with a length of 4 096 points (for the 48 kHz sampling rate) and a crest factor of 6 dB. The length of the complete test signal composed of at least four sequences of CSS shall be at least one second (1,0 s). The test signal level is -3 dBm0 (from 50 Hz to 7 kHz). The low crest factor is achieved by random alternation of the phase between -180° and 180°.

The TCLw is calculated according to ITU-T Recommendation G.122 [3], clause B.4 (trapezoidal rule) but using the frequency range of 300 Hz to 6 700 Hz (instead of 300 Hz to 3 400 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. For the measurement a time window has to be applied adapted to the duration of the actual pn-sequence of the test signal (200 ms) choosing the pn-sequence of the third CS-Signal.

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

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

6.8 Double talk performance

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 ITU-T Recommendations P.340 [10] and P.502 [13]):

- 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 6.7.1, 6.7.2 and 6.7.3 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.

6.8.1 Attenuation Range in Send Direction during Double Talk AH,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 6.8.1.1.

The category of the terminal according to table 6.8.1.1 shall be noted in the test report.

Table 6.8.1.1

Category (according to ITU-T Rec. P.340 [10])	1	2a	2b	2c	3
	Full Duplex Capability	Partial Duplex Capability			No Duplex Capability
$A_{H,S,dt}$ [dB]	≤ 3	≤ 6	≤ 9	≤ 12	> 12

In general this table 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 6.8.1.1. A sequence of uncorrelated CS signals is used which is inserted in parallel in send and receive direction.

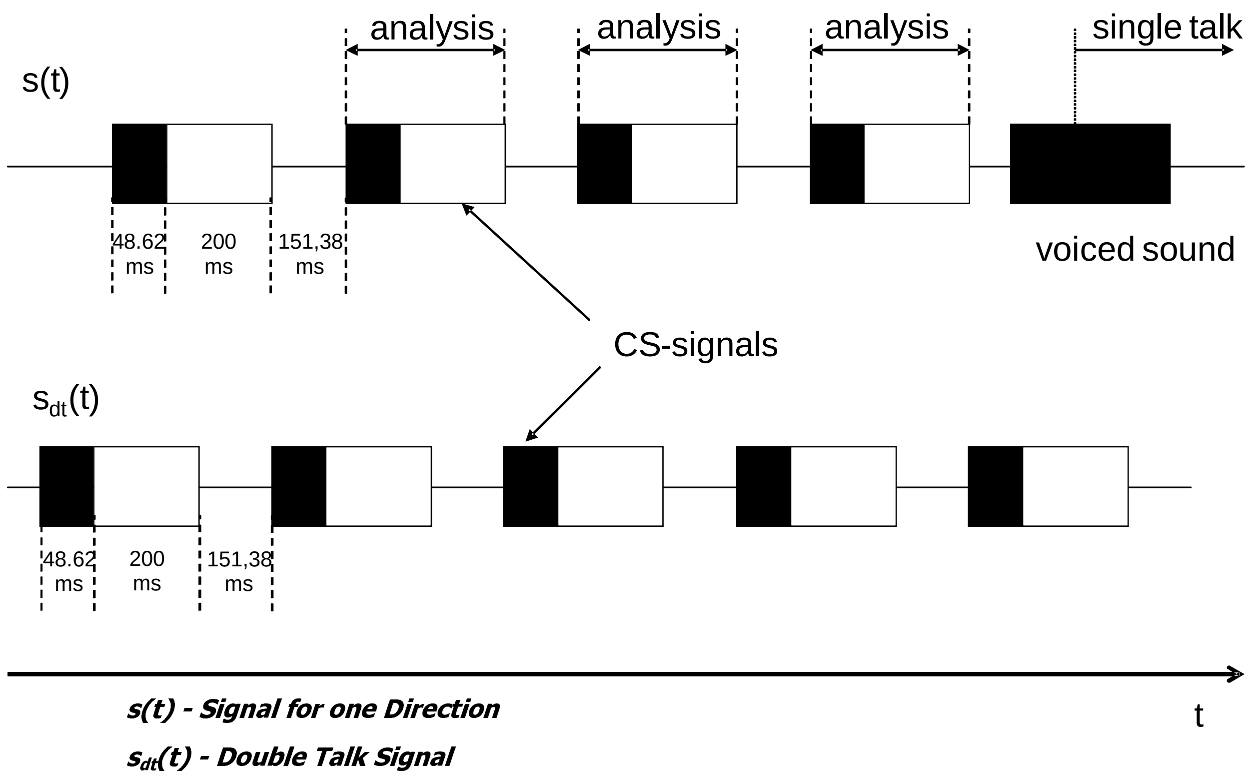


Figure 6.8.1.1: Double Talk Test Sequence with overlapping CS signals in send and receive direction

Figure 6.8.1.1 indicates that the sequences overlap partially. The beginning of the CS sequence (voiced sound, black) is overlapped by the end of the pn-sequence (white) of the opposite direction. During the active signal parts of one signal the analysis can be conducted in send and receive direction. The analysis times are shown in figure 6.8.1.1 as well. The test signals are synchronized in time at the acoustical interface. The delay of the test arrangement should be constant during the measurement.

NOTE: The length of voiced sound of the double talk signal is achieved by repeating one period of the voiced sound for double talk according to ITU-T Recommendation P.501 [12] 10 times and cutting off the initial 3,3 ms of the period of the first voiced sound.

The settings for the test signals are as follows:

Table 6.8.1.2

	Receive Direction (sdt(t))	Send Direction (s(t))
Pause Length between two Signal Bursts	151,38 ms	151,38 ms
Average Signal Level (Assuming an Original Pause length of 101,38 ms)	-16 dBm0	-4,7 dBPa
Active Signal Parts	-14,7 dBm0	-3 dBPa
NOTE: When the test laboratories implement different values (within the accuracy range defined in clause 5.7) it should be indicated in the test report.		

The test arrangement is according to clause 5.

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

The level is determined as level vs. time from the time domain. The integration time of the level analysis is 5 ms. The attenuation is determined from the level difference measured at the beginning of the double talk always with the beginning of the CS-signal in send direction until its complete activation (during the pause in the receive channel). The analysis is performed over the complete signal starting with the second CS-signal. The first CS-signal is not used for the analysis.

6.8.2 Attenuation Range in Receive Direction during Double Talk $A_{H,R,dt}$

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 6.8.2.1 shall be noted in the test report.

Table 6.8.2.1

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

In general this table 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 6.8.1.1. A sequence of uncorrelated CS 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 settings for the test signals are as follows:

Table 6.8.2.2

	Receive Direction (s(t))	Send Direction (sdt(t))
Pause Length between two Signal Bursts	151,38 ms	151,38 ms
Average Signal Level (Assuming an Original pause Length of 101,38 ms)	-16 dBm0	-4,7 dBPa
Active Signal Parts	-14,7 dBm0	-3 dBPa
NOTE: When the test laboratories implement different values (within the accuracy range defined in clause 5.7) it should be indicated in the test report.		

The test arrangement is according to clause 5.

When determining the attenuation range in receive direction the signal measured at the artificial ear referred to the test signal inserted.

The level is determined as level vs. time from the time domain. The integration time of the level analysis is 5 ms. The attenuation is determined from the level difference measured at the beginning of the double talk always with the beginning of the CS-signal in receive direction until its complete activation (during the pause in the send channel). The analysis is performed over the complete signal starting with the second CS-signal. The first CS-signal is not used for the analysis.

6.8.3 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 6.8.3.1 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 the table below are applicable (more information can be found in annex A of the ITU-T Recommendation P.340 [10]).

Table 6.8.3.1

Category (according to ITU-T Rec. P.340)	**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. The measurement signals used are shown in figure 6.8.3.1. A detailed description can be found in ITU-T Recommendation P.501 [12].

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).

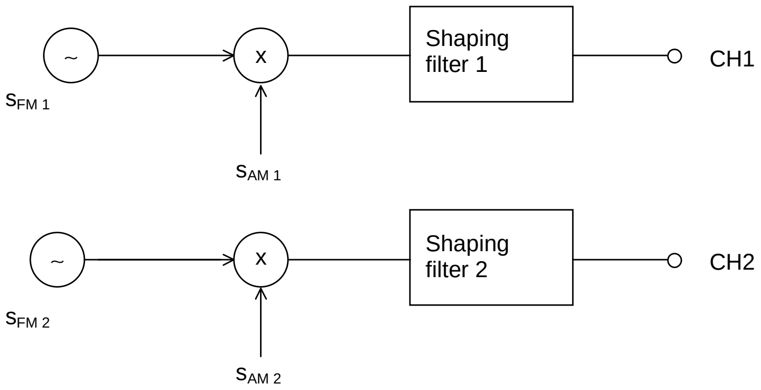


Figure 6.8.3.1: Measurement signals

$$s_{FM1,2}(t) = \sum A_{FM1,2} * \cos(2\pi t n * F_{01,2}) \; ; \; n = 1,2,.... \qquad (5)$$

$$s_{AM1,2}(t) = A_{AM1,2} * \cos(2\pi t F_{AM1,2}); \qquad (6)$$

The settings for the signals are as follows:

Table 6.8.3.2: Parameters of the two Test Signals for Double Talk Measurement based on AM-FM modulated sine waves

Receive Direction				Send Direction		
f_m[Hz]	$f_{mod(fm)}$[Hz]	F_{am}[Hz]		f_m[Hz]	$f_{mod(fm)}$[Hz]	F_{am}[Hz]
125	±2,5	3		150	±2,5	3
250	±5	3		270	±5	3
500	±10	3		540	±10	3
750	±15	3		810	±15	3
1 000	±20	3		1 080	±20	3
1 250	±25	3		1 350	±25	3
1 500	±30	3		1 620	±30	3
1 750	±35	3		1 890	±35	3
2 000	±40	3		2 160	±35	3
2 250	±40	3		2 400	±35	3
2 500	±40	3		2 650	±35	3
2 750	±40	3		2 900	±35	3
3 000	±40	3		3 150	±35	3
3 250	±40	3		3 400	±35	3
3 500	±40	3		3 650	±35	3
3 750	±40	3		3 900	±35	3
4 000	±40	3		4 150	±35	3
4 250	±40	3		4 400	±35	3
4 500	±40	3		4 650	±35	3
4 750	±40	3		4 900	±35	3
5 000	±40	3		5 150	±35	3
5 250	±40	3		5 400	±35	3
5 500	±40	3		5 650	±35	3
5 750	±40	3		5 900	±35	3
6 000	±40	3		6 150	±35	3
6 250	±40	3		6 400	±35	3
6 500	±40	3		6 650	±35	3
6 750	±40	3		6 900	±35	3
7 000	±40	3				
NOTE: Parameters of the Shaping Filter: f ≥ 250 Hz: Low Pass Filter, 5 dB/oct; f < 250 Hz,: High Pass Filter						

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 ITU-T Recommendation P.501 [12]). 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 the table above. The echo attenuation is to be achieved for **each individual frequency band** according to the different categories.

6.8.4 Minimum activation level and sensitivity of double talk detection

For further study.

6.9 Switching parameters

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

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 structure of the test signal is shown in figure 6.9.1.1. The test signal consists of CSS components according to ITU-T Recommendation P.501 [12] with increasing level for each CSS burst.

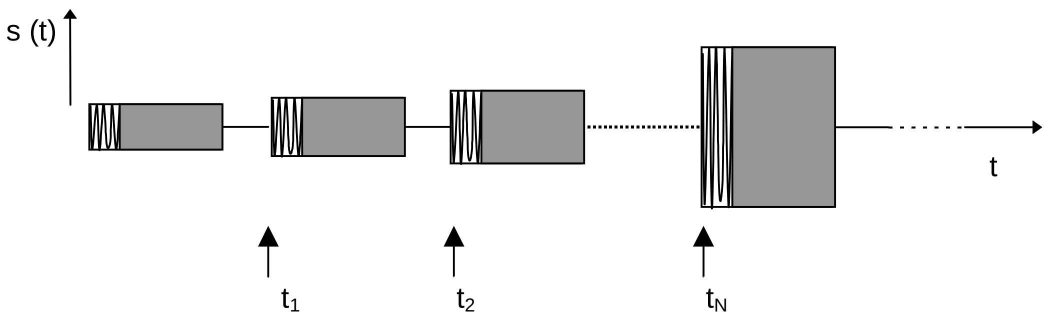


Figure 6.9.1.1: Test Signal to Determine the Minimum Activation Level and the Built-up Time

The settings of the test signal are as follows:

Table 6.9.1.1

	CSS Duration/ Pause Duration	**Level of the first CS Signal (active Signal Part at the MRP)**	**Level Difference between two Periods of the Test Signal**
CSS to Determine Switching Characteristic **in Send Direction**	~250 ms / ~450 ms	-23 dBPa (see note)	1 dB
NOTE: The level of the active signal part corresponds to an average level of -24,7 dBPa at the MRP for the CSS according to ITU-T Recommendation P.501 [12] assuming a pause of about 100 ms.			

It is assumed that the pause length of about 450 ms is longer than the hang-over time so that the test object is back to idle mode after each CSS burst.

The test arrangement is described in clause 5.2.

The level of the transmitted signal is measured at the electrical reference point. The measured signal level is referred to the 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 CSS burst which indicates the first activation of the test object. The time between the beginning of the CSS burst and the complete activation of the test object is measured.

NOTE: If the measurement using the CS-Signal does not allow to clearly identify the minimum activation level, the measurement may be repeated by using a one syllable word instead of the CS-Signal. The word used should be of similar duration, the average level of the word should be adapted to the CS-signal level of the according CS-burst.

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.

6.10 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 -5dB 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 perceptional 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 ITU-T Recommendation P.64 [8]).

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 can be tested based on EG 202 396-3 [2]. The test method described leads to three MOS-LQO quality numbers:

- N-MOS-LQOw: Transmission quality of the background noise
- S-MOS-LQOw: Transmission quality of the speech
- G-MOS-LQOw: Overall transmission quality

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

- N-MOS-LQOw $\geq$ 3,5
- S-MOS-LQOw $\geq$ 3,5
- G-MOS-LQOw $\geq$ 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 ITU-T Recommendation P.64 [8]).

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

The near end speech signal consists of 8 sentences of speech (2 male and 2 female talkers, 2 sentences each). Appropriate speech samples can be found in ITU-T Recommendation P.501 [12]. The preferred language is French since the objective method was validated with French language. The test signal level is -4,7 dBPa at the MRP.

Three signals are required for the tests:

1) The clean speech signal is used as the undisturbed reference (see EG 202 396-3 [2]).

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 EG 202 396-3 [2].

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

Requirement

The test is carried out applying the Composite Source Signal in receive direction. During and after the end of Composite Source Signal bursts (representing the end of far end speech simulation) 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.

Measurement Method

The test arrangement is according to clause 5.2.

The background noises are generated as described in clause 5.5.

First the measurement is conducted without inserting the signal at the far end. At least 10 s of noise is analysed. The background signal 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 CS-signal at the far end. The exactly identical background noise signal is applied. The background noise signal must start at the same point in time which was used for the measurement without 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 a Composite Source Signal according to ITU-T Recommendation P.501 [12] is applied in receive direction with a duration of ≥ 2 CSS periods. The test signal level is -16 dBm0 at the electrical reference point.

The send signal is recorded at the electrical reference point. The test signal 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 CS-signal is applied and after it stops. The level difference is determined from the difference of the recorded signal levels vs. time between reference signal and the signal measured with far end signal.

6.10.4 Quality of Background Noise Transmission (with Near End Speech)

Requirement

The test is carried out applying a simulated speech signal in send direction. During and after the end of the simulated speech signal (Composite Source Signal bursts) the signal level in send direction should not vary more than 10 dB.

Measurement Method

The test arrangement is according to clause 5.2.

The background noises are generated as described in clause 5.5. The background noise should be applied for at least 5 s in order to allow adaptation of the noise reduction algorithms.

The near end speech is simulated using the Composite Source Signal according to ITU-T Recommendation P.501 [12] with a duration of ≥ 2 CSS periods. The test signal level is -4,7 dBPa at the MRP.

The send signal is recorded at the electrical reference point. The test signal level versus time is calculated using a time constant of 35 ms.

First the measurement is conducted without inserting the signal at the near end. The signal level is analysed vs. time. In a second step the same measurement is conducted but with inserting the CS-signal at the near end. The level variation is determined by the difference between the background noise signal level without inserting the CS-signal and the maximum level of the noise signal during and after the CS-bursts in send direction.

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 TCLw test.

Measurement Method

The test arrangement is according to clause 5.2.

The test signal consists of periodically repeated Composite Source Signal according to ITU-T Recommendation P.501 [12] with an average level of -5 dBm0 as well as an average level of -25 dBm0. The echo signal is analysed during a period of at least 2,8 s which represents 8 periods of the CS signal. The integration time for the level analysis shall be 35 ms, the analysis is referred to the level analysis of the reference signal.

The measurement result is displayed as attenuation vs. time. The exact synchronization between input and output signal has to be guaranteed.

NOTE 1: In addition tests with more speech like signals should be made, e.g. ITU-T Recommendation P.50 [4] to see time variant behaviour of EC. However, for such tests the simple broadband attenuation based test principle as described above cannot be applied due to the time varying spectral content of the speech like signals.

NOTE 2: The analysis is conducted only during the active signal part, the pauses between the Composite Source Signals are not analysed. The analysis time is reduced by the integration time of the level analysis (35 ms).

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 possibly may 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 ITU-T Recommendation P.501 [12]. 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.

6.12 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 recognised that the end to end delay should be as small as possible in order to ensure high quality of the communication.

The delay T_{rtd} in send direction T_s plus the delay in receive direction T_r shall be less than 70 ms if the handset or headset terminal is implemented in conjunction with the speech coder and the RF-transmission. If the handset or headset terminal is connected via additional radio link the delay in send direction T_s plus the delay in receive direction T_r shall be less than 70 ms plus the delay of the radio link and in case of Bluetooth link 120 ms.

NOTE 1: Those limits are based on the assumption that the mobile phone signal processing is deactivated and does not introduce any additional processing delay.

NOTE 2: Half of the round trip delay corresponds to the mean one-way delay.

As the actual delay depends on the codec implementations, complementary requirements and test methods are defined in clause 7.

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+ $t_{Systeem}$.

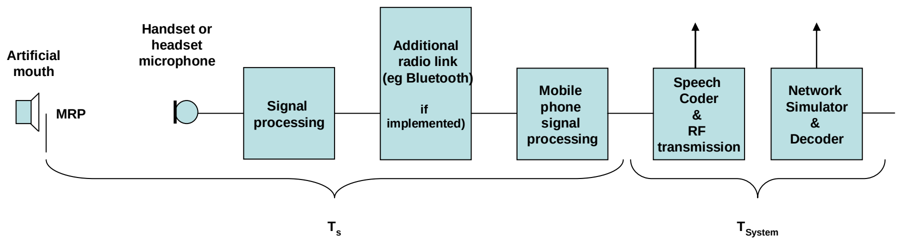


Figure 6.12.1: Different blocks contributing to the delay in send direction

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 ITU-T Recommendation P.501 [12] 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 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}.

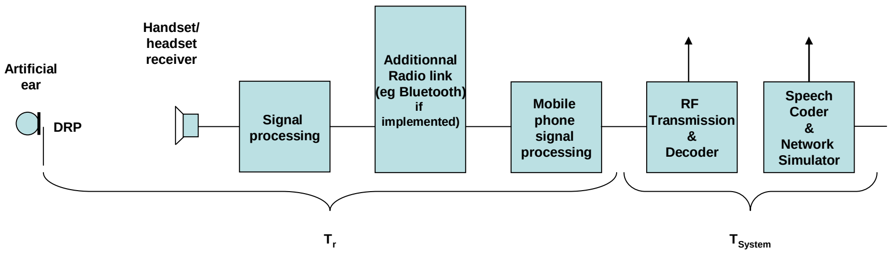


Figure 6.12.2: Different blocks contributing to the delay in receive direction

The system delay t_{System} is depending 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 ITU-T Recommendation P.501 [12] 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 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-LQO_M > 4,0.

- Optimum target value: MOS-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 defines a list of speech coders implemented in the terminals (non-exhaustive).

Table 7.1: Speech coders

System	Codec
GSM 850, 900, 1 800, 1 900	AMR-WB (ITU-T G.722.2)@12,65 kbit/s [21]
UMTS (WCDMA)	AMR-WB (ITU-T G.722.2) @12,65 kbit/s [21]
Voice over Data Network (VoDN)	ITU-T G.722 [19] ITU-T G.729.1 [18] ITU-T G.711.1 [20]

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 Send and receive delay or round trip delay

To be completed in the next version of the present document.

7.3 Objective listening Quality in send and receive direction

For further study (this clause will be updated when the relevant quality model will be available).

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 8.1.0 Release 8)".
- ETSI TS 126 132: "Universal Mobile Telecommunications System (UMTS); LTE; Speech and video telephony terminal acoustic test specification (3GPP TS 26.132 version 8.1.0 Release 8)".
- 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".

History

Document history		
V1.1.1	November 2009	Publication