ETSI TS 126 132 V11.1.0 (2013-01)

Technical Specification

Universal Mobile Telecommunications System (UMTS);
LTE;
Speech and video telephony terminal acoustic
test specification
(3GPP TS 26.132 version 11.1.0 Release 11)

Reference

RTS/TSGS-0426132vb10

Keywords

LTE,UMTS

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

Individual copies of the present document can be downloaded from:

http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at

http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:

http://portal.etsi.org/chaircor/ETSI_support.asp

Copyright Notification

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://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.

Foreword

This Technical Specification (TS) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities, UMTS identities or GSM identities. These should be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between GSM, UMTS, 3GPP and ETSI identities can be found under http://webapp.etsi.org/key/queryform.asp.

Contents

Foreword

This Technical Specification has been produced by the 3GPP.

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of this TS, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the specification.

Introduction

The present document specifies test methods to allow the minimum performance requirements for the acoustic characteristics of GSM and 3G terminals when used to provide narrowband or wideband telephony to be assessed.

The objective for narrowband services is to reach a quality as close as possible to ITU-T standards for PSTN circuits. However, due to technical and economic factors, there cannot be full compliance with the general characteristics of international telephone connections and circuits recommended by the ITU-T.

The performance requirements are specified in TS 26.131; the test methods and considerations are specified in the main body of the text.

1 Scope

The present document is applicable to any terminal capable of supporting narrowband or wideband telephony, either as a stand-alone service or as the telephony component of a multimedia service. The present document specifies test methods to allow the minimum performance requirements for the acoustic characteristics of GSM and 3G terminals when used to provide narrowband or wideband telephony to be assessed.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

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

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1]   3GPP TS 26.131: "Terminal Acoustic Characteristics for Telephony; Requirements".

[2]   ITU-T Recommendation B.12 (1988): "Use of the decibel and the neper in telecommunications".

[3]   ITU-T Recommendation G.103 (1998): "Hypothetical reference connections".

[4]   ITU-T Recommendation G.111 (1993): "Loudness ratings (LRs) in an international connection".

[5]   ITU-T Recommendation G.121 (1993): "Loudness ratings (LRs) of national systems".

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

[7]   ITU-T Recommendation G.711 (1988): "Pulse code modulation (PCM) of voice frequencies".

[8]   ITU-T Recommendation P.11 (1993): "Effect of transmission impairments".

[9]   ITU-T Recommendation P.38 (1993): "Transmission characteristics of operator telephone systems (OTS)".

[10]   ITU-T Recommendation P.50 (1993): "Artificial voices".

[11]   3GPP TS 03.58 (Release 1997): "Digital Cellular Telecommunications System (Phase 2+) Characterization test methods and quality assessment for hands-free mobiles".

[12]   IEC Publication 60651: 'Sound Level Meters'.

[13]   ITU-T Recommendation P.51 (1996): "Artificial mouth".

[14]   ITU-T Recommendation P.57 (2005): "Artificial ears".

[15]   ITU-T Recommendation P.58 (1996): "Head and torso simulator for telephonometry."

[16]   ITU-T Recommendation P.79 (2007) with Annex A: "Calculation of loudness ratings for telephone sets."

[17]   3GPP TS 06.77 (R99): 'Minimum Performance Requirements for Noise Suppresser Application to the AMR Speech Encoder'.

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

3GPP TS 26.132 version 11.1.0 Release 11

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

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

[21]   ITU-T Recommendation G.712: "Transmission performance characteristics of pulse code modulation channels".

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

[23]   ITU-T Recommendation O.41: "Psophometer for use on telephone-type circuits".

[24]   ITU-T Recommendation O.131: "Quantizing distortion measuring equipment using a pseudo-random noise test signal".

[25]   ISO 9614: "Acoustics - Determination of sound power levels of noise sources using sound intensity".

[26]   ISO 3745: "Acoustics - Determination of sound power levels of noise sources using sound pressure - Precision methods for anechoic and hemi-anechoic rooms".

[27]   ITU-T Recommendation O.132: "Quantizing distortion measuring equipment using a sinusoidal test signal".

[28]   ETSI TS 103 737: "Transmission requirements for narrowband wireless terminals (handset and headset) from a QoS perspective as perceived by the user".

[29]   ETSI TS 103 738: "Transmission requirements for narrowband wireless terminals (handsfree) from a QoS perspective as perceived by the user".

[30]   ETSI TS 103 739: "Transmission requirements for wideband wireless terminals (handset and headset) from a QoS perspective as perceived by the user".

[31]   ETSI TS 103 740: "Transmission requirements for wideband wireless terminals (handsfree) from a QoS perspective as perceived by the user".

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

[33]   ITU-T Recommendation P.501 Amendment 1 (2012): 'Test signals for use in telephonometry'

[34]   ETSI TS 103 106, "Speech Quality performance in the presence of background noise: Background noise transmission of mobile terminals-Objective test methods'

[35]   ETSI ES 202 396-1, "Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database"

[36]   ETSI EG 202 396-3, "Speech quality performance in the presence of background noise; Part 3: *Background noise transmission – objective test methods*: Background noise simulation technique and background noise database"

[37]   ITU-T Recommendation P.56: 'Objective measurement of active speech level'

[38]   IEC 61672: 'Electroacoustics – sound level meters - part 1: specifications'

3   Definitions, symbols and abbreviations

3.1   Definitions

For the purposes of the present document the term *narrowband* refers to signals sampled at 8 kHz; *wideband* refers to signals sampled at 16 kHz.

For the purposes of the present document, the terms dB, dBr, dBm0, dBm0p and dBA, shall be interpreted as defined in ITU-T Recommendation B.12 [2]; the term dBPa shall be interpreted as the sound pressure level relative to 1 pascal expressed in dB (0 dBPa is equivalent to 94 dB SPL).

A 3GPP softphone is a telephony system running on a general purpose computer or PDA complying with the 3GPP terminal acoustic requirements (TS 26.131 and 26.132).

3.2 Abbreviations

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

ADC	Analogue to Digital Converter
CSS	Composite Source Signal
DAC	Digital to Analogue Converter
DRP	Eardrum Reference Point
DTX	Discontinuous Transmission
EEC	Electrical Echo Control
EEP	Ear Entrance Point
EL	Echo Loss
ERP	Ear Reference Point
FFT	Fast Fourier Transform
HATS	Head and Torso Simulator
LSTR	Listener Sidetone Rating
MRP	Mouth Reference Point
MS	Mobile Station
OLR	Overall Loudness Rating
PCM	Pulse Code Modulation
PDA	Personal Digital Assistant
POI	Point of Interconnection (with PSTN)
PSTN	Public Switched Telephone Network
RLR	Receive Loudness Rating
RMS	Root Mean Squared
SLR	Send Loudness Rating
SS	System Simulator
STMR	Sidetone Masking Rating
SS	System Simulator
TX	Transmission
UE	User Equipment

4 Interfaces

Access to terminals for acoustic testing is always made via the acoustic or air interfaces. The Air Interface is specified by the GSM 05, GSM 45 and 3G 25 series specifications and is required to achieve user equipment (UE) transportability. Measurements can be made at this point using a system simulator (SS) comprising the appropriate radio terminal equipment and speech transcoder. The losses and gains introduced by the test speech transcoder will need to be specified.

The POI with the public switched telephone network (PSTN) is considered to have a relative level of 0 dBr, where signals will be represented by 8-bit A-law, according to ITU-T Recommendation G.711 [7]. Measurements may be made at this point using a standard send and receive side, as defined in ITU-T Recommendations.

Five classes of acoustic interface are considered in this specification:

- Handset UE including softphone UE used as a handset;

- Headset UE including softphone UE used with headset;

- Vehicle Mounted Hands-free UE including softphone UE mounted in a vehicle;

- Desktop-mounted hands-free UE including softphone UE with external loudspeaker(s) used in hands-free mode;

- Hand-held hands-free UE including softphone UE with internal loudspeaker(s) used in hands-free mode.

(See definition of softphone in Clause 3.1)

NOTE: The test setup for a softphone UE shall be derived according to the following rules:

- When using a softphone UE as a handset: the test setup shall correspond to handset mode.

- When using a softphone UE with headset: the test setup shall correspond to headset mode.

- When a softphone UE is mounted in a vehicle: the test setup shall correspond to vehicle-mounted hands-free mode.

- When using a softphone UE in hands-free mode:

   - When using internal loudspeaker(s), the test setup shall correspond to hand-held hands-free.

   - When using external loudspeaker(s), the test setup shall correspond to desktop-mounted hands-free.

5   Test configurations

This section describes the test setups for terminal acoustic testing.

NOTE: If the terminal has several mechanical configurations (e.g., sliding design open or closed), all manufacturer-defined configurations shall be tested.

5.1   Setup for terminals

The general access to terminals is described in figure 1. The preferred acoustic access to GSM and 3G 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 mountings of handset terminals to the HATS in a realistic but reproducible way. Hands-free terminals shall use the HATS or free field microphone techniques in a realistic but reproducible way.

HATS is described in ITU-T Recommendation P.58 [15], appropriate ears are described in ITU-T Recommendation P.57 [14] (Type 3.3), proper positioning of handsets in realistic conditions is found in ITU-T Recommendation P.64, and the test setups for various types of hands-free terminals can be found in ITU-T Recommendation P.581.

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

The preferred way of testing is the connection of a terminal to the system simulator with exact defined settings and access points. The test sequences are fed in either electrically using a reference codec, using the direct signal processing approach, or acoustically using ITU-T specified devices.
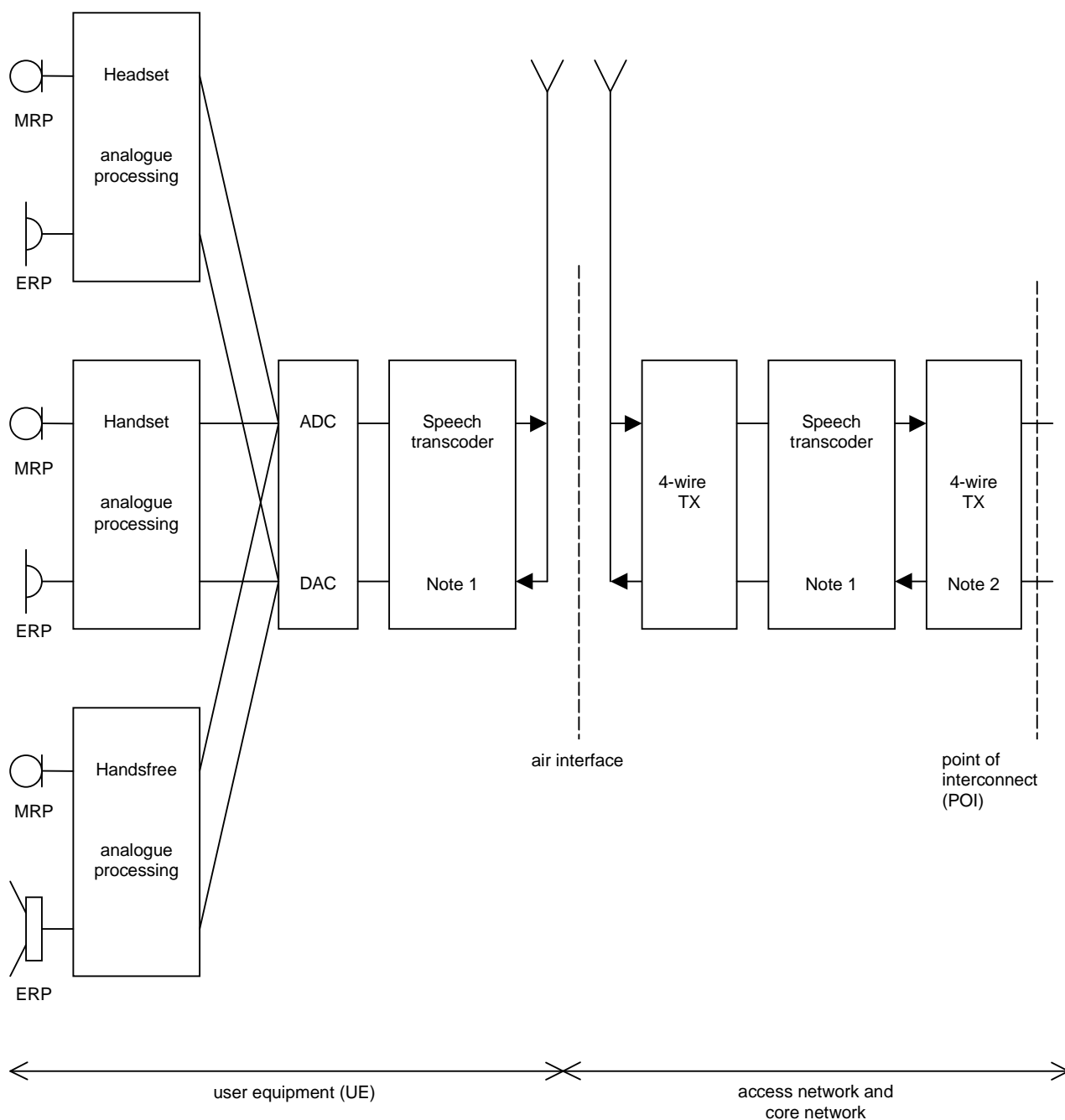
NOTE 1: Includes DTX functionality.
NOTE 2: Connection to PSTN should include electrical echo control (EEC).

Figure 1: GSM/3G Interfaces for specification and testing of terminal acoustic characteristics

5.1.1 Setup for handset terminals

When using a handset UE, the handset is placed on HATS as described in ITU-T Recommendation P.64 Annex E [18]. A suitable position shall be defined for each handset UE and documented in the test report. The artificial mouth shall conform to ITU-T Recommendation P.58 [15]. The artificial ear shall conform to ITU-T Recommendation P.57 [14]. Type 3.3 ear shall be used and positioned on HATS according to ITU-T Recommendation P.58 [15].

Position and calibration of HATS
The sending and receiving characteristics shall be tested with the HATS. It shall be indicated what application force was used. If not stated otherwise in TS 26.131, an application force of 8 ± 2 N shall be used.

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

5.1.2 Setup for headset terminals

Recommendations for the setup and positioning of headsets are given in ITU-T Recommendation P.380. If not stated otherwise, headsets shall be placed in their recommended wearing position. Some insert earphones might not fit properly in Type 3.3 ear simulators. For such insert type headsets, an ITU-T Recommendation P.57 [14] Type 2 ear simulator may be used in conjunction with the HATS mouth simulator. The HATS should be equipped with two artificial ears as specified in ITU-T Recommendation P.57 [14]. For binaural headsets two artificial ears are required.

5.1.3 Setup for hands-free terminals

5.1.3.1 Vehicle-mounted hands-free

If not stated otherwise, the artificial head (HATS – head and torso simulator, according to ITU-T Recommendation P.58 [15]) is positioned in the driver's seat for the measurement as shown in figure 3a. The position has to be in line with the average users" position; therefore, all positions and sizes of users have to be taken into account. Typically, all except the tallest 5% and the shortest 5% of the driving population have to be considered. The size of these persons can be derived, e.g., from the 'anthropometric data set' for the corresponding year (e.g., based on data used by car manufacturers). The position of the HATS (mouth/ears) within the positioning arrangement is given individually by each car manufacturer. The position used has to be reported in detail in the test report. If no requirements for positioning are given the distance from the microphone to the MRP is defined by the test lab.

By using suitable measures (e.g., marks in the car, relative position to A-pillar, B-pillar, height from the floor, etc.) an exact reproduction of the artificial head position must be possible at any later time.

NOTE – Different positions of the artificial head may greatly influence the test results. Depending on the application, different positions of the artificial head may be chosen for the tests. It is recommended to check the worst-case position, e.g., those positions where the SNR and/or the speech quality in send may be worst.

Figure 2: void

Figure 3: void
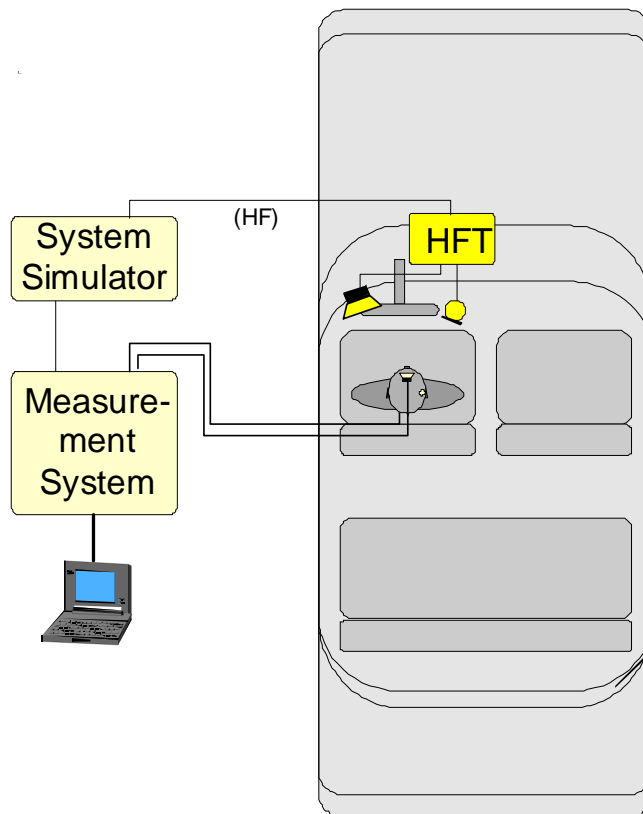
Figure 3a: Test Configuration for vehicle mounted hands-free, using HATS

5.1.3.2   Desktop hands-free

For HATS test equipment, the definition of hands-free terminals and setup for desktop hands-free terminals can be found in ITU-T Recommendation P.581. Measurement setup using a free-field microphone and a discrete P.51 [13] artificial mouth for desktop hands-free terminals can be found in ITU-T Recommendation P.340.The positioning for different types of desktop hands-free terminals is given in ETSI TS 103 738 and ETSI TS 103 740.

5.1.3.3   Hand-held hands-free

Either HATS or a free-field microphone with a discrete P.51 [13] artificial mouth may be used to measure a hand-held hands-free type UE.

If HATS measurement equipment is used, it shall be configured to the hand-held hands-free UE according to figure 4. The HATS should be positioned so that the HATS Reference Point is at a distance d_{HF} from the centre point of the visual display of the Mobile Station. The distance d_{HF} is specified by the manufacturer. A vertical angle θ_{HF} may be specified by the manufacturer. Where it is not specified, the nominal distance d_{HF} shall be 42 cm and θ_{HF} shall be 0º.

> NOTE: The nominal distance of 42 cm corresponds to the distance between the HATS reference point and lip-plane (12 cm) with an additional 30 cm giving a realistic figure as a reference usage of hand-held terminals.
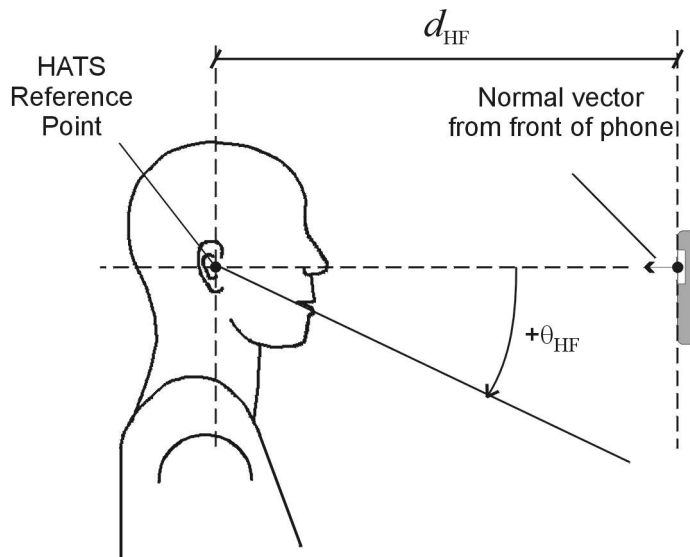
Figure 4: Configuration of hand-held hands-free UE relative to the HATS

If a free-field microphone and a discrete P.51 [13] mouth are used, they shall be configured to the hand-held hands-free UE according to figure 5 for receiving measurements and figure 6 for sending measurements. The microphone should be located at a distance d_{HF} from the centre of the visual display of the UE. The mouth simulator should be located at a distance d_{HF}-12 cm from the centre of the visual display of the UE. The distance d_{HF} is specified by the manufacturer. Where it is not specified the nominal distance d_{HF} shall be 42 cm.
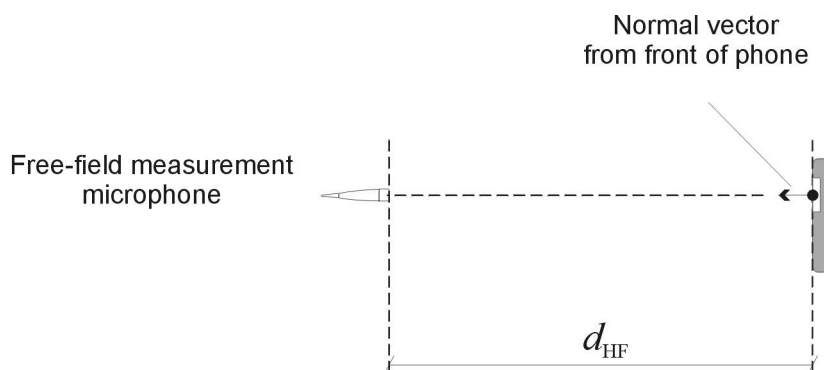
Figure 5: Configuration of hand-held hands-free UE; free-field microphone for receiving measurements
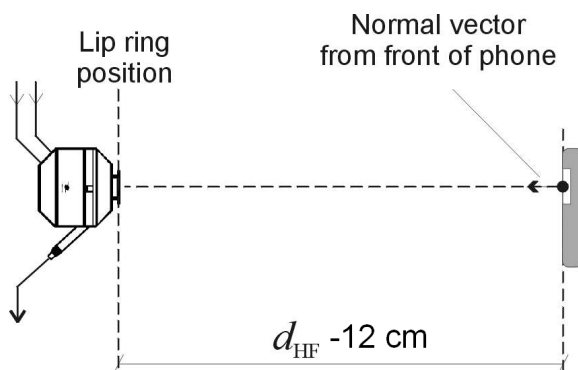
Figure 6: Configuration of hand-held hands-free UE; discrete P.51 artificial mouth for sending measurements

5.1.3.4 Softphone including speakers and microphone

This test setup is applicable to laptop computers or similar devices as seen in figure 7 through figure 11.

Where the manufacturer gives conditions of use, these will apply for testing. If the manufacturer gives no other requirement, the softphone will be positioned according the following conditions:

Measurement with artificial ear and microphone:

Artificial mouth (for sending tests)
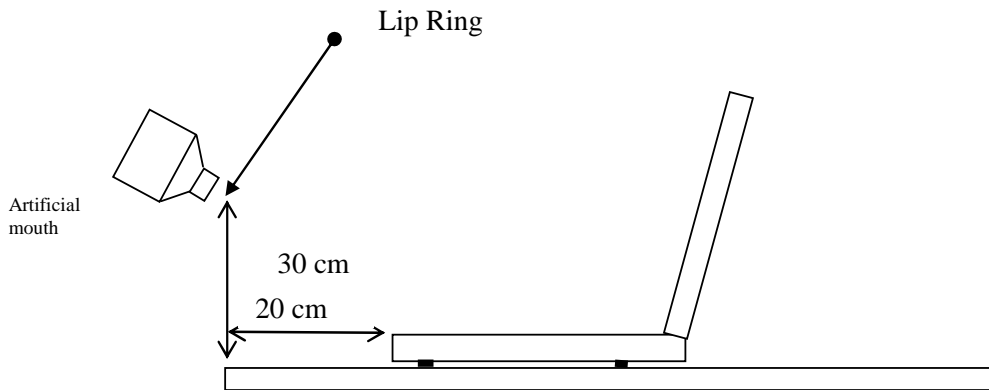
Figure 7: Configuration of a softphone relative to the artificial mouth side view

Free field microphone (for receiving):
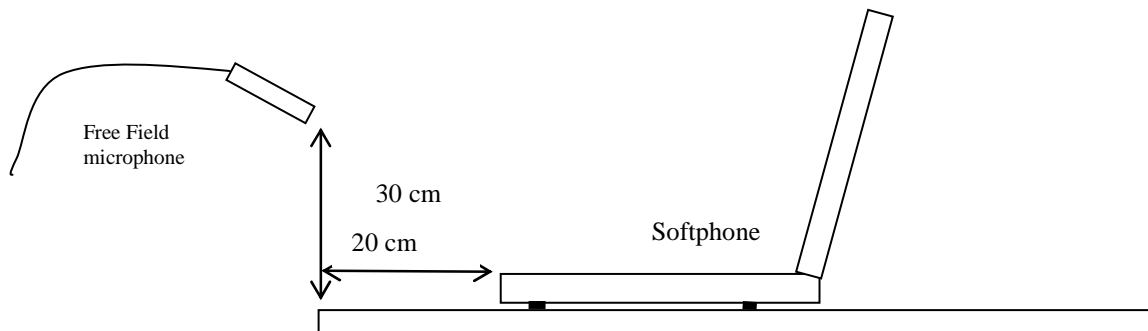
Figure 8: Configuration of a softphone relative to the free field microphone side view

Position of a softphone on the table:

Hands free softphone

Test

20 cm

Microphone (or artificial mouth)
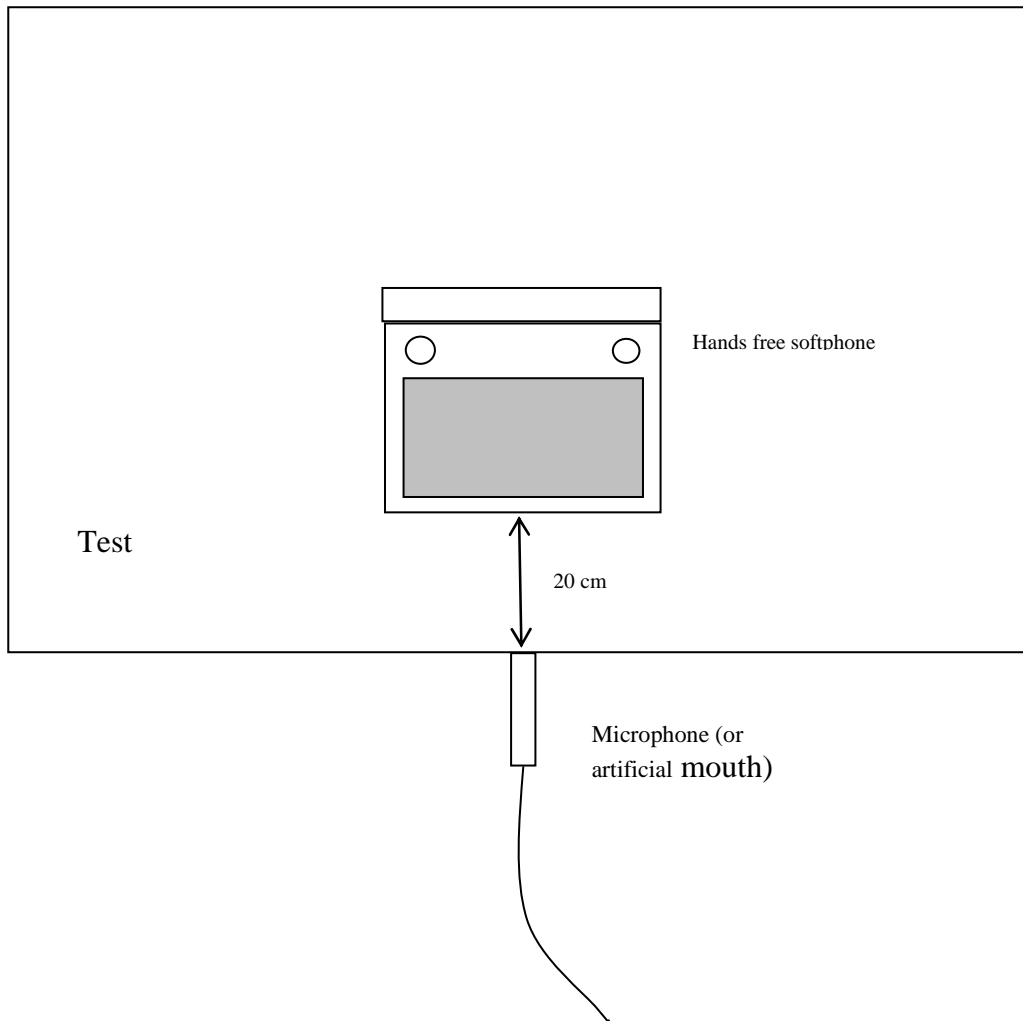
Figure 9: Configuration of a softphone relative to the free-field microphone or artificial mouth viewed from above

Measurement with HATS:

Lip Ring

HATS

30 cm

20 cm
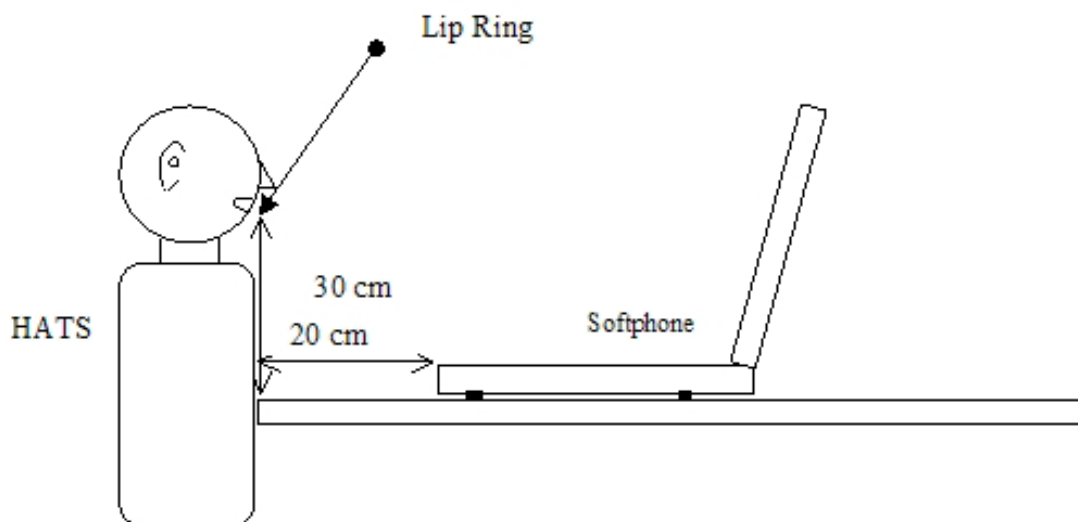
Softphone

Figure 10: Configuration of a softphone relative to the HATS side view
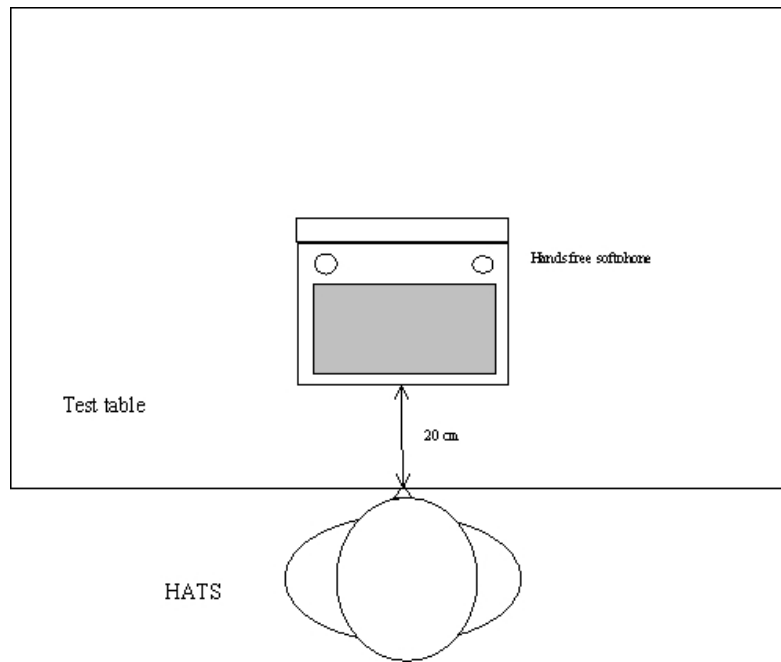
Figure 11: Configuration of a softphone relative to the HATS viewed from above

5.1.3.5 Softphone with separate speakers

This test setup is applicable to laptop computers or similar devices as seen in figure 12 through figure 15.

Where the manufacturer gives conditions of use, these will apply for testing. If the manufacturer gives no other requirement, the softphone will be positioned according to the following conditions:

Where separate loudspeakers are used, the system will be positioned as in figure 12 or figure 13.
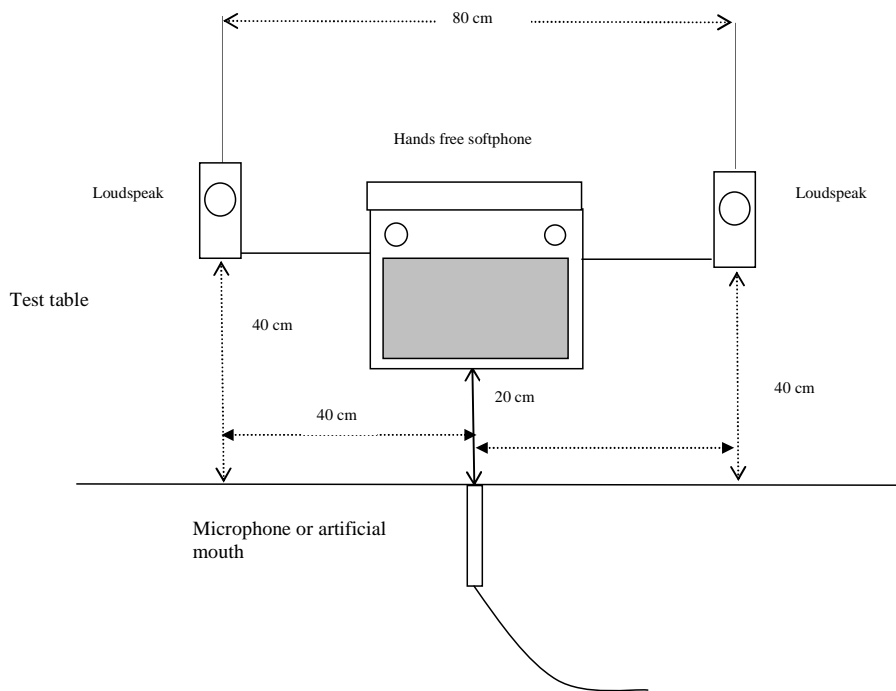
Figure 12: Configuration of a softphone using external speakers relative to microphone or artificial mouth viewed from above

80 cm

Hands free softphone

Loudspeaker

Loudspeaker

Test table

40 cm

40 cm

40 cm

20 cm

40 cm
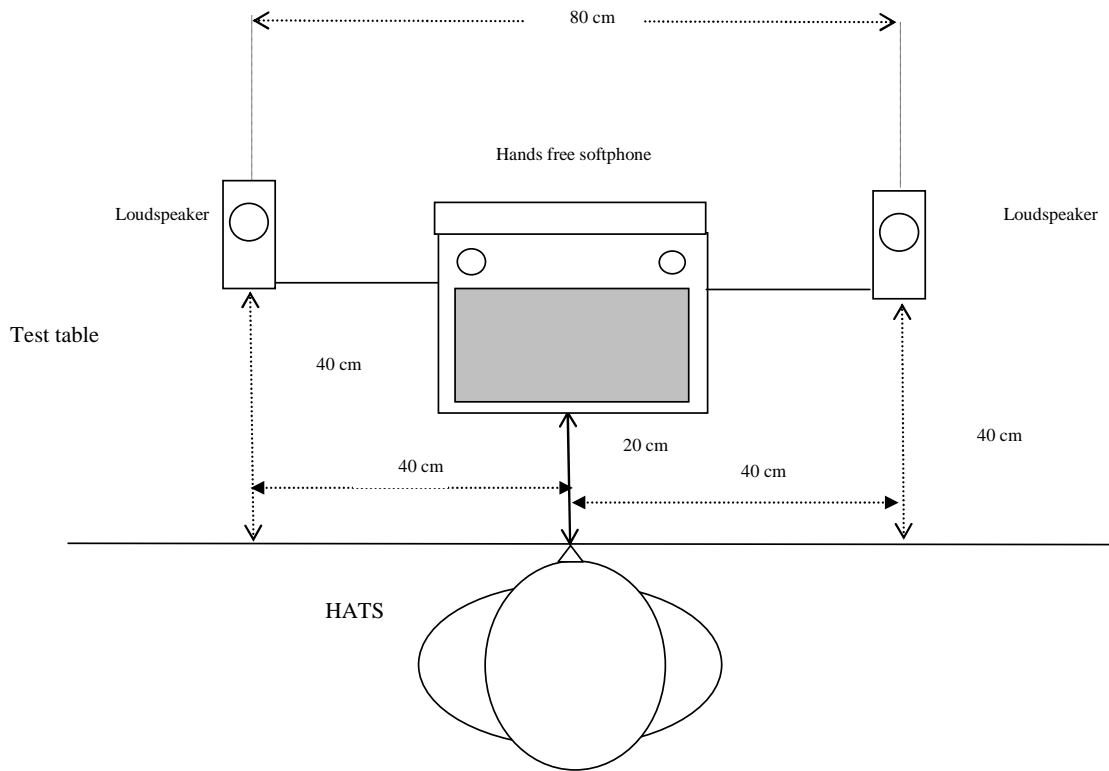
HATS

Figure 13: Configuration of a softphone using external speakers relative to the HATS viewed from above

Where an external microphone and speakers are used, the system will be positioned as in figure 14 or figure 15.
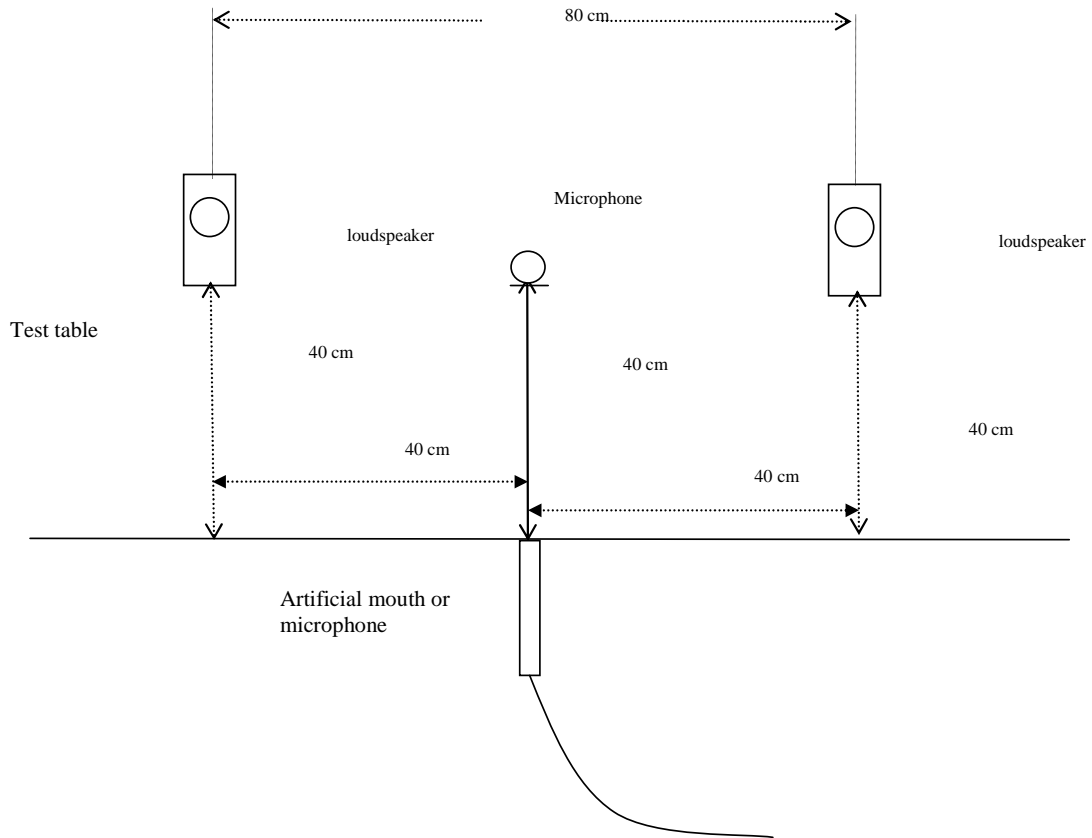
**Figure 14: Configuration of a softphone using
external speakers and a microphone relative to microphone or artificial mouth viewed from above**
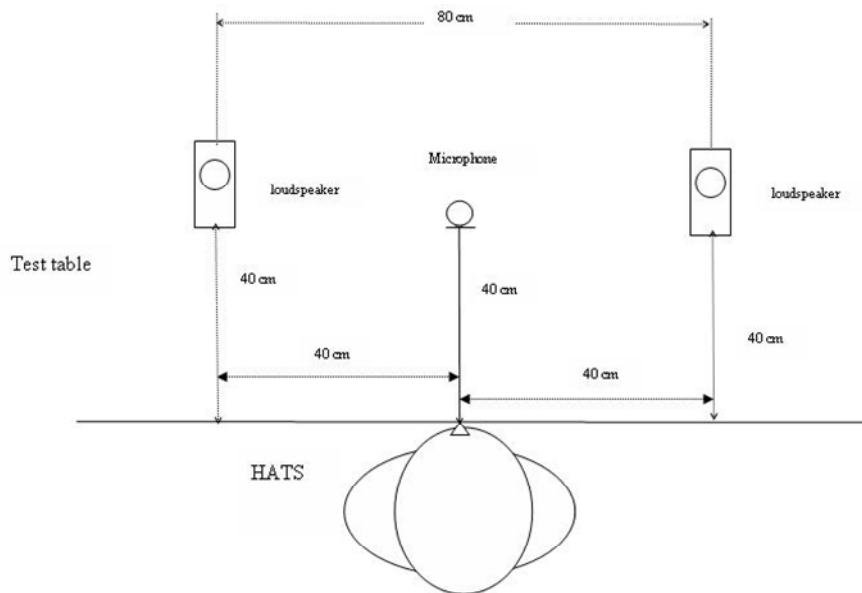
**Figure 15: Configuration of a softphone using
external speakers and a microphone relative to the HATS viewed from above**

5.1.4 Position and calibration of HATS

The horizontal positioning of the HATS reference plane shall be guaranteed within $\pm 2°$ for testing hands-free equipment.

The HATS shall be equipped with a Type 3.3 Artificial Ear. For hands-free measurements the HATS shall be equipped with two artificial ears. The pinnae are specified in Recommendation P.57 [14] for Type 3.3 artificial ears. The pinnae shall be positioned on HATS according to ITU-T Recommendation P.58 [15].

The exact calibration and equalization procedures as well as how to combine the two ear signals for the purpose of measurements can be found in ITU-T Recommendation P.581. If not stated otherwise, the HATS shall be diffuse-field equalized. The reverse nominal diffuse field curve as found in table 3 of ITU-T Recommendation P.58 [15] shall be used. For measurements requiring diffuse-field correction values for closer frequency spacing than that which is specified in ITU-T Recommendation P.58 [15], the interpolation method found in annex A shall be used.

For hand-held hands-free UE, the setup corresponding to 'portable hands-free' in ITU-T Recommendation P.581 should be used.

5.1.5 Test setup for quality in the presence of ambient noise measurements

The setup for simulating realistic ambient noises and the positioning of the HATS in a lab-type environment is described in ETSI EG 202 396-1 [35].

ETSI EG 202 396-1 [35] contains a description of the recording arrangement for realistic ambient noises, a description of the setup for a loudspeaker arrangement suitable to simulate an ambient noise field in a lab-type environment and a database of realistic ambient noises, part of which is used for testing the terminal performance with a variety of conditions.

The equalization and calibration procedure for the test setup are given in detail in ETSI EG 202 396-1 [35].

5.2 Setup of the electrical interfaces

5.2.1 Codec approach and specification

In this approach, a codec is used to convert the digital input/output bit-stream of the system simulator to the equivalent analogue values. With this approach a system simulator simulating the radio link to the terminal under controlled and error-free conditions is required. The system simulator has to be equipped with a high-quality codec with characteristics as close as possible to ideal.

Definition of 0 dBr point:

D/A converter - a Digital Test Sequence (DTS) representing the codec equivalent of an analogue sinusoidal signal with an RMS value of 3,14 dB below the maximum full-load capacity of the codec shall generate 0 dBm across a 600 ohm load;

A/D converter - a 0 dBm signal generated from a 600 ohm source shall give the digital test sequence (DTS) representing the codec equivalent of an analogue sinusoidal signal with an RMS value of 3,14 dB below the maximum full-load capacity of the codec.

Narrowband telephony testing

For testing of a GSM or 3G terminal supporting narrowband telephony, the system simulator shall use the AMR speech codec as defined in the 3GPP TS 26 series of specifications, at the source coding bit-rate of 12,2 kbit/s.

Wideband telephony testing

For testing of a GSM or 3G terminal supporting wideband telephony, the system simulator shall use the AMR-WB speech codec as defined in 3GPP TS 26 series of specifications, at the source coding bit-rate of 12,65 kbit/s.

5.2.2 Direct digital processing approach

In this approach, the digital input/output bit-stream of the terminal connected through the radio link to the system simulator is operated upon directly.

Narrowband telephony testing

For testing of a GSM or 3G terminal supporting narrowband telephony, the system simulator shall use the AMR speech codec as defined in the 3GPP TS 26 series of specifications, at the source coding bit-rate of 12,2 kbit/s.

Wideband telephony testing

For testing of a GSM or 3G terminal supporting wideband telephony, the system simulator shall use the AMR-WB speech codec as defined in the 3GPP TS 26 series of specifications, at the source coding bit rate of 12,65 kbit/s.

5.3 Accuracy of test equipment

Unless specified otherwise, the accuracy of measurements made by test equipment shall exceed the requirements defined in table 1a.

Table 1a: Test equipment measurement accuracy

Item	Accuracy
Electrical Signal Power	± 0,2 dB for levels ≥ -50 dBm
	± 0,4 dB for levels < -50 dBm
Sound pressure	± 0,7 dB
Time	± 5%
Frequency	± 0,2%

Unless specified otherwise, the accuracy of the signals generated by the test equipment shall exceed the requirements defined in table 1b.

Table 1b: Test equipment signal generation accuracy

Quantity	Accuracy
Sound pressure level at MRP	± 1 dB for 200 Hz to 4 kHz
	± 3 dB for 100 Hz to 200 Hz
	± 3 dB for 4 kHz to 8 kHz
Electrical excitation levels	± 0,4 dB (see note 1)
Frequency generation	± 2% (see note 2)
NOTE 1: Across the whole frequency range.	
NOTE 2: When measuring sampled systems, it is advisable to avoid measuring at sub-multiples of the sampling frequency. There is a tolerance of ± 2% on the generated frequencies, which may be used to avoid this problem, except for 4 kHz where only the -2% tolerance may be used.	

The measurements" results shall be corrected for the measured deviations from the nominal level.
The sound level measurement equipment shall conform to IEC 60651 Type 1.

5.4 Test signals

Unless stated otherwise, appropriate test signals for GSM/3G acoustic tests are generally described and defined in ITU-T Recommendation P.501 [22].

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

For testing the narrowband telephony service provided by the UE, the test signal used shall be band limited between 100 Hz and 4 kHz with a bandpass filter providing a minimum of 24 dB/oct. filter roll-off, when feeding into the receiving direction.

For testing the wideband telephony service provided by the UE, 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 receiving direction.

The test signal levels are referred to the average level of the (band limited in receiving direction) test signal, averaged over the complete test sequence, unless specified otherwise. For real speech, the test signal levels are referred to the ITU-T P.56 [37] active speech level of the (band limited in receiving direction) test signal, calculated over the complete test sequence.

5.5 Void

5.5.1 Void

5.5.2 Void

6 Test conditions

6.1 Environmental conditions

6.1.1 Handset and headset terminals

For handset and headset measurements the test room shall be practically free-field down to a lowest frequency of 275 Hz; the handset or headset, including the HATS, shall be totally within this free-field volume. This shall be met if deviations of the ideal free-field conditions are less than ± 1 dB. Qualification of the test room may be performed using the method described in either ISO 3745 Annex A, or ITU-T P. 340 §5.4.

Alternatively, a test room may be used which meets the following two criteria:

1. The relationship between the pressure at the mouth opening and that at 5,0 cm, 7,5 cm and 10 cm in front of the centre of the lip ring is within ± 0.5 dB of that which exists in a known acoustic free-field.

2. The relationship between the pressure at the mouth opening and that at the Ear canal Entrance Point (EEP) at both the left and right ears of the HATS does not differ by more than ± 1 dB from that which exists in a known free-field.

The ambient noise level shall be less than -30 dBPa(A); for idle channel noise measurements the ambient noise level shall be less than –64dBPa(A).

Echo measurements shall be conducted in realistic rooms with an ambient noise level ≤ -64 dBPa(A).

6.1.2 Hands-free terminals

Hands-free terminals should generally be tested in their typical environment of application. Care must be taken that, e.g., noise levels are sufficiently low in order not to interfere with the measurements.

For desktop hands-free terminals the appropriate requirements shall be taken from ITU-T Recommendation P.340.

The broadband noise level shall not exceed –70 dBPa(A). The octave band noise level shall not exceed the values specified in Table 2.

Table 2: P.340 Noise level

Center frequency (Hz)	Octave band pressure level (dBPa)
63	−45
125	−60
250	−65
500	−65
1 000	−65
2 000	−65
4 000	−65
8 000	−65

Echo measurements shall be conducted in realistic rooms with an ambient noise level ≤ -70 dBPa(A).

6.2 System simulator conditions

The system simulator should provide an error-free radio connection to the UE under test. The default speech codec in narrowband, the AMR speech codec, shall be used at its highest bit-rate of 12,2 kbit/s. The default speech codec in wideband, the AMR-WB speech codec, shall be used at 12,65 kbit/s. Discontinuous Transmission (DTX) silence suppression shall be disabled for the purposes of GSM/3G acoustic testing.

7 Narrowband telephony transmission performance test methods

7.1 Applicability

The test methods in this clause shall apply when testing a UE that is used to provide narrowband or wideband telephony, either as a stand-alone service, or as part of a multimedia service.

7.2 Overall loss/loudness ratings

7.2.1 General

The SLR and RLR values for GSM or 3G networks apply up to the POI. However, the main determining factors are the characteristics of the UE, including the analogue to digital conversion (ADC) and digital to analogue conversion (DAC). In practice, it is convenient to specify loudness ratings to the Air Interface. For the normal case, where the GSM or 3G network introduce no additional loss between the Air Interface and the POI, the loudness ratings to the PSTN boundary (POI) will be the same as the loudness ratings measured at the Air Interface.

7.2.2   Connections with handset UE

7.2.2.1   Sending loudness rating (SLR)

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence.

b) The handset terminal is setup as described in clause 5. The sending sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17. 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.

c) The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23b), over bands 4 to 17, using $m = 0,175$ and the sending weighting factors from ITU-T Recommendation P.79 [16], table 1.

7.2.2.2   Receiving loudness rating (RLR)

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.

b) The handset terminal is setup as described in clause 5. The receiving sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

c) The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23c), over bands 4 to 17, using $m = 0,175$ and the receiving weighting factors from table 1 of ITU-T Recommendation P.79 [16].

d) DRP-ERP correction is used. No leakage correction shall be applied.

7.2.3   Connections with desktop and vehicle-mounted hands-free UE

Vehicle-mounted hands-free UE should be tested within the vehicle (for totally integrated vehicle hands-free systems) or in a vehicle simulator, as described in 3GPP TS 03.58 [11].

Free-field measurements for vehicle-mounted hands-free are for further study.

7.2.3.1   Sending loudness rating (SLR)

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence. The broadband signal level is then adjusted to -28,7 dBPa at the HFRP or the HATS HFRP (as defined in ITU-Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as references to determine the sending sensitivity S_{mJ}.

b) The hands-free terminal is setup as described in clause 5. The sending sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17. 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.

c) The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23b), over bands 4 to 17, using $m = 0,175$ and the sending weighting factors from ITU-T Recommendation P.79 [16], table 1.

7.2.3.2 Receiving Loudness Rating (RLR)

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.

b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. The receiving sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17.

For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

c) The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23c), over bands 4 to 17, using $m = 0,175$ and the receiving weighting factors from table 1 of ITU-T Recommendation P.79 [16].

d) No leakage correction shall be applied. The hands-free correction, as described in ITU-T Recommendation P.340 shall be applied. To compute the receiving loudness rating (RLR) for a hands-free terminal (see also ITU-T Recommendation P.340), when using the combination of left and right artificial ear signals from the HATS, the HFL_E has to be 8 dB instead of 14 dB. For further information see ITU-T Recommendation P.581.

7.2.4 Connections with hand-held hands-free UE

7.2.4.1 Sending loudness rating (SLR)

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence. The broadband signal level is then adjusted to -28,7 dBPa at the HFRP or the HATS HFRP (as defined in ITU-T Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as references to determine the sending sensitivity S_{mJ}.

b) The hands-free terminal is setup as described in clause 5. The sending sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17. 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.

c) The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23b), over bands 4 to 17, using $m = 0,175$ and the sending weighting factors from ITU-T Recommendation P.79 [16], table 1.

7.2.4.2 Receiving loudness rating (RLR)

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.

b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. The receiving sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17.

For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

c) The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23c), over bands 4 to 17, using m = 0,175 and the receiving weighting factors from table 1 of ITU-T Recommendation P.79 [16].

d) No leakage correction shall be applied. The hands-free correction as described in ITU-T Recommendation P.340 shall be applied. To compute the receiving loudness rating (RLR) for hands-free terminals (see also ITU-T Recommendation P.340), when using the combination of left and right artificial ear signals from the HATS, the HFL_E has to be 8 dB instead of 14 dB. For further information see ITU-T Recommendation P.581.

7.2.5 Connections with headset UE

Same as for handset.

7.3 Idle channel noise (handset and headset UE)

For idle noise measurements in sending and receiving directions, care should be taken that only the noise is windowed out by the analysis and the result is not impaired by any remaining reverberation or by noise and/or interference from various other sources. Some examples are air-conducted or vibration-conducted noise from sources inside or outside the test chamber, disturbances from lights and regulators, mains supply induced noise including grounding issues, test system and system simulator inherent noise as well as radio interference from the UE to test equipment such as ear simulators, microphone amplifiers, etc.

7.3.1 Sending

The terminal should be configured to the test equipment as described in subclause 5.1.

The environment shall comply with the conditions described in subclause 6.1.

The noise level at the output of the SS is measured with psophometric weighting. The psophometric weighting filter is described in ITU-T Recommendation O.41.

A test signal may have to be intermittently applied to prevent "silent mode" operation of the MS. This is for further study.

The measured part of the noise shall be 170,667 ms (which equals 8192 samples in a 48 kHz sample rate test system). The spectral distribution of the noise is analyzed with an 8k FFT using windowing with ≤ 0,1 dB leakage for non bin-centered signals. This can be achieved with a window function commonly known as a 'flat top window'. Within the specified frequency range, the FFT bin that has the highest level is searched for; the level of this bin is the maximum level of a single frequency disturbance.

To improve repeatability, the test sequence (optional activation followed by the noise level measurement) may be contiguously repeated one or more times.

The total noise powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

The single frequency maximum powers obtained from such repeats shall be averaged. The total result shall be $10*\log_{10}$ of this average in dB.

7.3.2 Receiving

The terminal should be configured to the test equipment as described in subclause 5.1.

The environment shall comply with the conditions described in subclause 6.1.

A test signal may have to be intermittently applied to prevent "silent mode" operation of the MS. This is for further study.

The noise level shall be measured with A-weighting at the DRP with diffuse-field correction. The A-weighting filter is described in IEC 60651.

The measured part of the noise shall be 170,667 ms (which equals 8192 samples in a 48 kHz sample rate test system). The spectral distribution of the noise is analyzed with an 8k FFT using windowing with ≤ 0.1 dB leakage for non bin-centred signals. This can be achieved with a window function commonly known as a 'flat top window'. Within the specified frequency range, the FFT bin that has the highest level is searched for; the level of this bin is the maximum level of a single frequency disturbance.

To improve repeatability, considering the test sequence (optional activation followed by the noise level measurement) may be contiguously repeated one or more times.

The total noise powers obtained from such repeats shall be averaged. The total result shall be $10*\log_{10}$ of this average in dB.

The single frequency maximum powers obtained from such repeats shall be averaged. The total result shall be $10*\log_{10}$ of this average in dB.

7.4   Sensitivity/frequency characteristics

7.4.1   Handset and headset UE sending

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence.

b) The handset terminal is setup as described in clause 5. Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 4 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.

c) The sensitivity is expressed in terms of dBV/Pa.

7.4.2   Handset and headset UE receiving

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.

b) The handset terminal is setup as described in clause 5. Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 4 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.

c) The HATS is diffuse-field equalized. The sensitivity is expressed in terms of dBPa/V. Information about correction factors is available in ITU-T Recommendation P.57 [14].

Optionally, the measurements may be repeated with a 2 N and 13 N application force. For these test cases no normative values apply.

7.4.3   Desktop and vehicle-mounted hands-free UE sending

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence. The broadband signal level is then adjusted to -28,7 dBPa at the HFRP or the HATS HFRP (as defined in ITU-T Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as references to determine the sending sensitivity S_{mJ}.

b) The hands-free terminal is setup as described in clause 5. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 4 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.

c) The sensitivity is expressed in terms of dBV/Pa.

7.4.4 Desktop and vehicle-mounted hands-free UE receiving

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.

b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 4 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.

c) The sensitivity is expressed in terms of dBPa/V.

7.4.5 Hand-held hands-free UE sending

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence. The broadband signal level then is adjusted to -28,7 dBPa at the HFRP or the HATS HFRP (as defined in ITU-T Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as reference to determine the sending sensitivity S_{mJ}.

b) The hands-free terminal is setup as described in clause 5. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 4 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.

c) The sensitivity is expressed in terms of dBV/Pa.

7.4.6 Hand-held hands-free UE receiving

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.

b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave band frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 4 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.

c) The sensitivity is expressed in terms of dBPa/V.

7.5 Sidetone characteristics

7.5.1 Connections with handset UE

The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence.

7.5.1.1 void

7.5.1.2 Connections with handset UE – HATS method

The handset UE is setup as described in clause 5. The application force shall be 13 N on the Type 3.3 artificial ear.

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 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 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 [16], table 4, bands 4 to 17) 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 Rating (STMR), expressed in dB, shall be calculated from formula 5-1 of ITU-T Recommendation P.79 [16], using $m = 0.225$ and the weighting factors in table B.2 (unsealed condition) of ITU-T Recommendation P.79 [16]. No leakage correction (L_E) shall be applied. DRP-ERP correction is used.

7.5.2 Headset UE

The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence.

Measurements shall be made at 1/12-octave intervals as given by the R.10 series of preferred numbers in ISO 3 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 [16], table 4, bands 4 to 17) is referred to the averaged test signal level measured in each frequency band.

The sidetone path loss (L_{meST}), as expressed in dB, shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17. The STMR (in dB) shall be calculated from formula B-4 of ITU-T Recommendation P.79 [16], using $m = 0.225$ and the weighting factors in table B.2 (unsealed condition) of ITU-T Recommendation P.79 [16]. No leakage correction (L_E) shall be applied. DRP-ERP correction is used.

7.5.3 Hands-free UE (all categories)

No requirement other than echo control.

7.5.4 Sidetone delay for handset or headset

The handset or headset terminal is setup as described in clause 5.

The test signal is a CS-signal complying with ITU-T Recommendation P.501 using a PN-sequence with a length, T, of 4 096 points (for a 48 kHz sample rate test system). The duration of the complete test signal is as specified in ITU-T Recommendation P.501. The level of the signal shall be -4,7 dBPa at the MRP.

The cross-correlation function $\Phi xy(\tau)$ 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:

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

The measurement window, T, shall be identical to the test signal period, T, with the measurement window synchronized 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:

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

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

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

7.6 Stability loss

Where a user-controlled volume control is provided it is set to maximum.

Handset UE: The handset is placed on a hard plane surface with the earpiece facing the surface.

Headset UE: The requirement applies for the closest possible position between microphone and headset receiver within the intended wearing position.

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

Hands-free UE (all categories): No requirement other than echo loss.

Before the actual test a training sequence consisting of the British-English single talk sequence described in ITU-T Recommendation P.501 [22] is applied. The training sequence level shall be -16 dBm0 in order to not overload the codec.

The test signal is a PN-sequence complying with ITU-T Recommendation P.501 with a length of 4 096 points (for a 48 kHz sampling rate system) and a crest factor of 6 dB instead of 11 dB. The PN-sequence is generated as described in P.501 with W(k) constant within the frequency range 200-4000 Hz and zero outside this range. The duration of the test signal is 250 ms. With an input signal of -3 dBm0, the attenuation from input to output of the system simulator shall be measured under the following conditions:

   a) The handset or the headset, with the transmission circuit fully active, shall be positioned on a hard plane surface with at least 400 mm free space in all directions; the earpiece shall face towards the surface as shown in figure 15c;

   b) The headset microphone is positioned as close as possible to the receiver(s) within the intended wearing position;

   c) For a binaural headset, the receivers are placed symmetrically around the microphone.
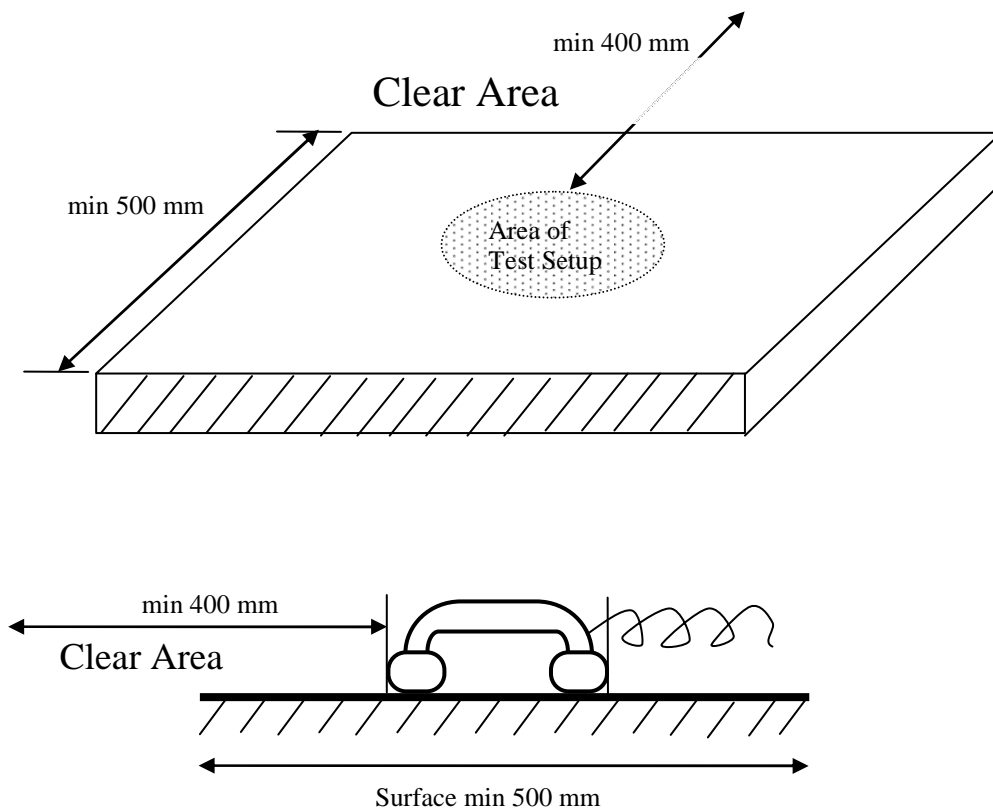
NOTE: All dimensions in mm.

Figure 15c. Test configuration for stability loss measurement on handset or headset UE

The attenuation from input to output of the system simulator shall be measured in the frequency range from 200 Hz to 4 kHz. The spectral distribution of the output signal is analysed with a 4k FFT (for a 48 kHz sample rate test system), thus the measured part of the output signal is 85.333 ms. To avoid leakage effects, the frequency resolution of the FFT must be the same as the frequency spacing of the PN-sequence.

7.7 Acoustic echo control

7.7.1 General

The echo loss (EL) presented by the GSM or 3G networks at the POI should be at least 46 dB during single talk. This value takes into account the fact that UE is likely to be used in a wide range of noise environments.

7.7.2 Acoustic echo control in a hands-free UE

The hands-free UE is setup in a room with acoustic properties similar to a typical 'office-type' room; a vehicle-mounted hands-free UE should be tested in a vehicle or vehicle simulator, as specified by the UE manufacturer (see also 3GPP TS 03.58 [11]). The ambient noise level ≤ 70 dBPa(A). The attenuation from reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, clause B.4 (trapezoidal rule). 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).

The test signal level shall be -10 dBm0.

7.7.3 Acoustic echo control in handset UE

The handset is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A). The attenuation from the reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, clause B.4 (trapezoidal rule). 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).

The test signal level shall be -10 dBm0.

7.7.4 Acoustic echo control in a headset UE

The headset is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A). The attenuation from the reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, clause B.4 (trapezoidal rule). 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).

The test signal level shall be -10 dBm0.

7.8 Distortion

7.8.1 Sending distortion

The handset, headset, or hands-free UE is setup as described in clause 5.

The signal used is a sine-wave signal with a frequency of 1020 Hz. The sine-wave signal level shall be calibrated to the following RMS levels at the MRP: 5, 0, -4,7, -10, -15, -20 dBPa. The test signals have to be applied in this sequence, i.e., from high levels down to low levels.

The duration of the sine-wave signal is recommended to be 360 ms. The manufacturer shall be allowed to request tone lengths up to 1 s. The measured part of the signal shall be 170.667 ms (which equals 2 * 4096 samples in a 48 kHz sample rate test system). The times are selected to be relatively short in order to reduce the risk that the test tone is treated as a stationary signal.

It is recommended that an optional activation signal be presented immediately preceding each test signal to ensure that the UE is in a typical state during measurement. An appropriate speech or speech-like activation signal shall be chosen from ITU-T Recommendations P.501 or P.50 [10]. A recommendation for the use of an activation signal as part of the measurement is defined in figure 16. The RMS level of the active parts of this activation signal is recommended to be equal to the subsequent test tone RMS level. In practice, certain types of processing may be impacted due to the introduction of the activation signal. The manufacturer shall be allowed to specify disabling of the activation signal. It shall be reported whether an activation signal was used or not, along with the characteristics of the activation signal, as specified by the manufacturer.

The ratio of the signal to total distortion power of the signal output of the SS shall be measured with the psophometric noise weighting (see ITU-T Recommendations G.712, O.41 and 0.132). The psophometric filter shall be normalized (0 dB gain) at 800 Hz as specified in ITU-T Recommendation O.41. The weighting function shall be applied to the total distortion component only (not to the signal component).

For measurement of the total distortion component an octave-wide band-stop filter shall be applied to the signal to suppress the sine-wave signal and associated coding artefacts. The filter shall have a lower passband ending at $0.7071 * f_S$, and an upper passband starting at $1,4142 * f_S$, where f_S is the frequency of the sine-wave signal. The

passband ripple of the filter shall be ≤ 0.2 dB. The attenuation of the band-stop filter at the sine-wave frequency shall be ≥ 60 dB. Alternatively, the described characteristics can be implemented by an appropriate weighting on the spectrum obtained from an FFT. The total distortion component is defined as the measured signal within the frequency range 200 Hz to 4 kHz, after applying psophometric and stop filters (hence no correction for the lost power due to the stop filter, known as 'bandwidth correction', shall be applied).

To improve repeatability, considering the variability introduced by speech coding and voice processing, the test sequence (activation signal followed by the test signal) may be contiguously repeated one or more times.. The single signal-to-total-distortion power ratios obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

Figure 16: Recommended activation sequence and test signal.

The activation signal consists of a 'Bandlimited composite source signal with speech-like power density spectrum' signal according to ITU-T Recommendation P.501 with 48,62 ms voiced part (1), 200 ms unvoiced part (2) and 101,38 ms pause (3), followed by the same signal but polarity inverted (4, 5, 6), followed by the voiced part only (7). The pure test tone is applied and after 50 ms settling time (8), the analysis is made over the following 170,667 ms (9).

NOTE 1: Void.

NOTE 2: In order to ensure that the correct part of the signal is analyzed, the total delay of the terminal and SS may have to be determined prior to the measurement.

NOTE 3: For hands-free terminals tested in environments defined in subclause 6.1.2, care should be taken that the reverberation in the test room, caused by the activation signal, does not affect the test results to an unacceptable degree, referring to subclause 5.3.

7.8.2 Receiving

The handset, headset, or hands-free UE is setup as described in clause 5.

The signal used is a sine-wave signal with frequency of 1020 Hz. The signal shall be applied at the signal input of the SS at the following levels: 0, -3, -10, -16, -20, -30, -40, -45 dBm0. The test signals have to be applied in this sequence, i.e., from high levels down to low levels.

The duration of the sine-wave signal is recommended to be 360 ms. The manufacturer shall be allowed to request tone lengths up to 1 s. The measured part of the signal shall be 170.667 ms (which equals 2 * 4096 samples in a 48 kHz sample rate test system). The times are selected to be relatively short in order to reduce the risk that the test tone is treated as a stationary signal.

It is recommended that an optional activation signal be presented immediately preceding each test signal to ensure that the UE is in a typical state during measurement. An appropriate speech or speech-like activation signal shall be chosen from ITU-T Recommendations P.501 or P.50 [10]. A recommendation for the use of an activation signal as part of the measurement is defined in figure 17. The RMS level of the active parts of this activation signal is recommended to be equal to the subsequent test tone RMS level for low and medium test levels. To avoid saturation of the SS speech encoder, it is recommended for high test levels that the activation signal level be adjusted such that its peak level equals the peak level of the test tone. In practice, certain types of processing may be impacted due to the introduction of the activation signal. The manufacturer shall be allowed to specify disabling of the activation signal. It shall be reported

whether an activation signal was used or not, along with the characteristics of the activation signal, as specified by the manufacturer.

The ratio of the signal to total distortion power shall be measured at the applicable acoustic measurement point (DRP with diffuse-field correction for handset and headset modes; free field for hands-free modes) with psophometric noise weighting (see ITU-T Recommendations G.712, O.41 and 0.132). The psophometric filter shall be normalized to have 0 dB gain at 800 Hz as specified in ITU-T Recommendation O.41. The weighting function shall be applied to the total distortion component only (not to the signal component).

For measurement of the total distortion component an octave-wide band-stop filter shall be applied to the signal to suppress the sine-wave signal and associated coding artefacts. The filter shall have a lower passband ending at $0,7071 * f_S$, and an upper passband starting at $1,4142 * f_S$, where f_S is the frequency of the sine-wave signal. The passband ripple of the filter shall be ≤ 0.2 dB. The attenuation of the band-stop filter at the sine-wave frequency shall be ≥ 60 dB. Alternatively, the described characteristics can be implemented by an appropriate weighting on the spectrum obtained from an FFT. The total distortion component is defined as the measured signal within the frequency range 200 Hz to 4 kHz, after applying psophometric and stop filters (hence no correction for the lost power due to the stop filter, known as 'bandwidth correction', shall be applied).

To improve repeatability, considering the variability introduced by speech coding and voice processing, the test sequence (activation signal followed by the test signal) may be contiguously repeated one or more times. The single signal-to-total-distortion power ratios obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

Figure 17: Recommended activation sequence and test signal.

The activation signal consists of a 'Bandlimited composite source signal with speech-like power density spectrum' signal according to ITU-T Recommendation P.501 with 48,62 ms voiced part (1), 200 ms unvoiced part (2) and 101,38 ms pause (3), followed by the same signal but polarity inverted (4, 5, 6), followed by the voiced part only (7). The pure test tone is applied and after 50 ms settling time (8), the analysis is made over the following 170,667 ms (9).

NOTE 1: Void.

NOTE 2: In order to ensure that the correct part of the signal is analyzed, the total delay of the terminal and SS may have to be determined prior to the measurement.

NOTE 3: For hands-free terminals tested in environments defined in subclause 6.1.2, care should be taken that the reverberation in the test room, caused by the activation signal, does not affect the test results to an unacceptable degree, referring to subclause 5.3.

7.9 Ambient noise rejection

Handset and headset UE:

NOTE: The ambient noise rejection method has been updated to support basic testing of UEs that include noise suppression. Other methods are being investigated that may be more suitable for proper characterization of ambient noise rejection by UEs.

a) A 1/2-inch pressure microphone is calibrated using a known sound source and mounted at the MRP, without the HATS present. The measurement system is calibrated to enable the sound pressure levels at the microphone to be determined in 1/3-octave bands.

b) For calibration of the background noise generation system, flood the room in which the measurement is to be made with a band-limited (100 Hz to 8 kHz) pink-noise, accurate within ± 3 dB. The level at the MRP shall be adjusted to 70 dBSPL(A) (-24 dBPa(A)). The tolerance on this level is ± 1 dB. The resulting sound spectrum is P_{rn} (dBPa), measured in 1/3-octave bands.

To ensure that the sound field is sufficiently diffuse, the following applies:

The diffuse sound-field is calibrated in the absence of any local obstacles. The averaged field shall be uniform to within ± 3 dB within a radius of 0,15 m of the MRP, when measured in 1/3-octave bands from 100 Hz to 3,15 kHz.

NOTE 1: The pressure intensity index, as defined in ISO 9614, may prove to be a suitable method for assessing the diffuse field.

NOTE 2: Where more than one loudspeaker is used to produce the desired sound field, the loudspeakers must be fed with non-coherent electrical signals to eliminate standing waves and other interference effects.

c) The speech test signal to be used for measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated at the MRP. The test signal level shall be +1,3 dBPa, measured at the MRP. The resulting sound spectrum is Po dBPa, measured in 1/3-octave bands.

d) Mount the MS under test, according to subclause 5.1.1.

e) Set up a send speech path between the MS and the SS and reproduce the artificial speech signal and pink noise signal simultaneously according to the sequence in figure 17a:

Figure 17a: Conditioning and test sequence for ambient noise rejection.

NOTE: Windows I, II, III are used only for conditioning of the MS. Window IV is used for the actual measurement of the MS speech send sensitivity and window V for the measurement of the MS noise send sensitivity.

f) To determine the send noise sensitivity, measure as a function of frequency in 1/3-octave bands (index j), the electrical output V_{jrn}, (expressed as dB rel. 1 V) at the audio output of the SS for the applied noise acoustic pressure P_{jrn} (expressed as dB rel. 1 Pa) at the MRP. This measurement is performed during the period determined by window V shown in figure 17a. Since the MS sending sensitivity is not defined above 3,4 kHz,

the measurement shall be cut off at 3,4 kHz. For the bands below 315 Hz, the noise level shall be referenced to the speech level at 315 Hz to yield the DELSM.

The room send noise sensitivity is expressed as: $Sm_{jrn} = V_{jrn}$ (dBV) - P_{jrn} (dBPa) dB rel. 1V/Pa

g) To determine the send speech sensitivity, measure as a function of frequency in 1/3-octave bands (index j) the electrical output, V_j, (expressed as dB rel. 1 V) at the audio output of the SS for the applied speech acoustic pressure, P_{j0}, (expressed as dB rel. 1 Pa), at the MRP. This measurement is performed during the period determined by window IV shown in figure 17a.

The speech sending sensitivity is expressed as;

Sm_{js} (dB) = V_j (dBV) - P_{jo} (dBPa) dB rel. 1 V/Pa.

h) The difference between the room noise sensitivity and the speech sending sensitivity, DELSM (Δ_{jSM}), in each 1/3-octave band for the MS is determined as:

DELSM (Δ_{jSM}) = Sm_{jrn} - Sm_{js} (dB)   (for j = 1 to 2, $Sm_{js} = Sm_{3s}$).

i) The Ambient noise rejection, ANR, is calculated as a single figure value according to the following formula:

$$ ANR = -\frac{4}{5} \sum_{i=1}^{13} \Delta_{jSM} \cdot 10^{-0.0175 W_{jsi}} $$

j   = The index of 1/3-octave bands centered at frequencies from 200 Hz to 3150 Hz inclusive.

W_{jsi} = the sending weighting factors from ITU-T Recommendation P.79 [16], table 1 for the jth 1/3-octave band centre frequency.

Hands-free UE (all categories):

No test method for hands-free operations.

7.10   Delay

7.10.1   Delay in sending direction (Handset UE)

The handset terminal is setup as described in clause 5.1.1.

The delay shall include all entities in sending direction from MRP to the POI, but shall exclude the delays introduced by the test equipment.
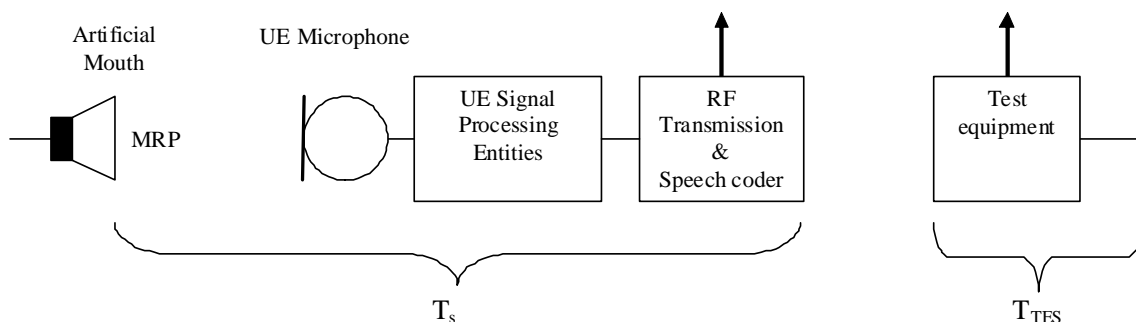
Figure 17b1: Different entities contributing to the delay in sending direction

The delay in sending direction, measured from MRP to POI, is $T_s + T_{TES}$.

All test equipment delays, for the network type, codec type and bitrate used according to clause 5, (including radio access, speech codec, A/D and D/A conversions etc.) are included in T_{TES}. The values used for testing (typical value considering variations due to interleaving etc.) as declared by the test equipment manufacturers shall be reported along with the measurement results.

1. For the measurements, a Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] 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 32 k samples (with 48 kHz sampling rate). The test signal level is -4,7 dBPa at the MRP.

2. The reference signal is the original signal (test signal). The setup of the handset/headset terminal is made corresponding to clause 5.1.

3. 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 subtracting the test equipment delay T_{TES}.

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

7.10.1a   Delay in sending direction (headset UE)

The delay shall include all entities in sending direction from MRP to the POI, but shall exclude the delays introduced by the test equipment.
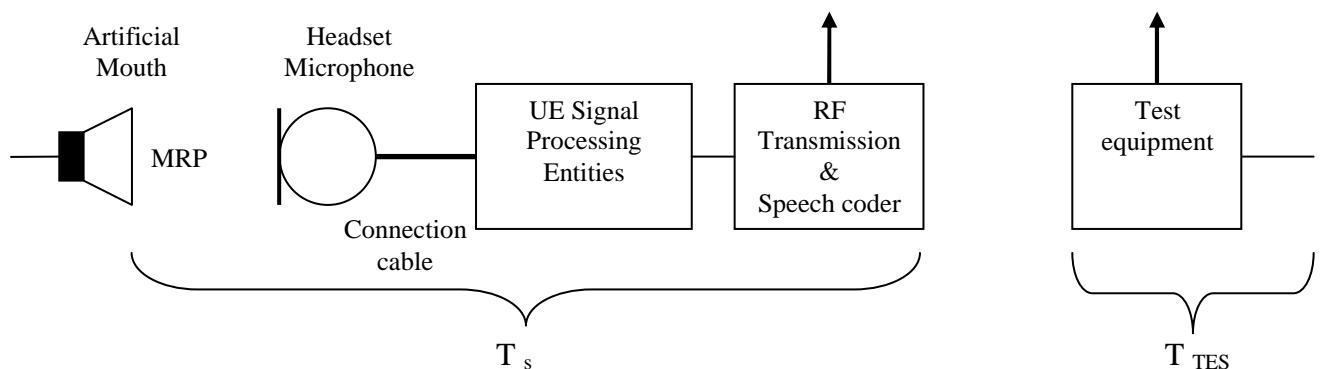
Figure 17b2: Different entities contributing to the delay in sending direction with a headset connected via cable

Note: The test setup only applies to headsets connected by wire. Wireless headsets (e.g. connected by Bluetooth) are currently out of scope.

The test method is the same as for handset UE (clause 7.10.1).

7.10.2   Delay in receiving direction (handset UE)

The handset terminal is setup as described in clause 5.

The delay shall include all entities in receiving direction from the POI to the DRP, but shall exclude the delays introduced by the test equipment.
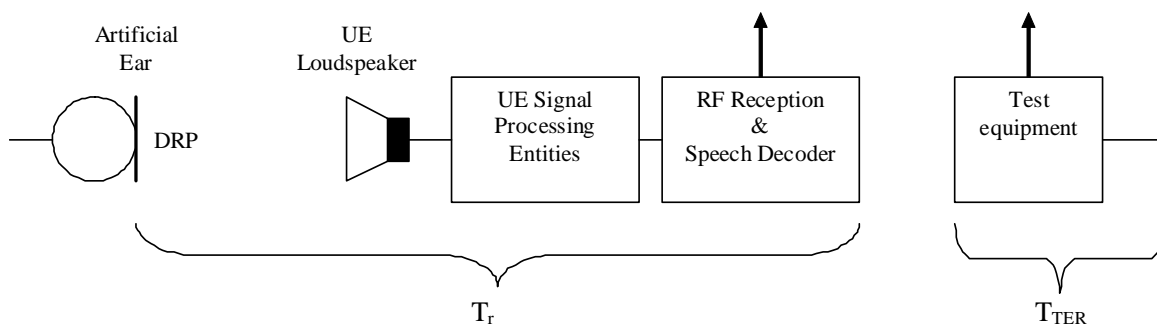
Figure 17b3: Different entities contributing to the delay in receiving direction

The delay in receiving direction, measured from POI to DRP, is $T_r + T_{TER}$.

All test equipment delays, for the network type, codec type and bitrate used according to clause 5, (including radio access, speech codec, A/D and D/A conversions etc.) are included in T_{TER}. The values used for testing (typical value considering variations due to interleaving etc.) as declared by the test equipment manufacturers shall be reported along with the measurement results.

1. For the measurements a Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] 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 32 k samples (with 48 kHz sampling rate).The test signal level is -16 dBm0 measured at the digital reference point or the equivalent analogue point.

2 The reference signal is the original signal (test signal). The setup of the handset/headset terminal is in correspondence to clause 5.1.

3. 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 subtracting the test equipment delay T_{TER}.

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

7.10.2a   Delay in receiving direction (headset UE)

The delay shall include all entities in receiving direction from the POI to the DRP, but shall exclude the delays introduced by the test equipment.
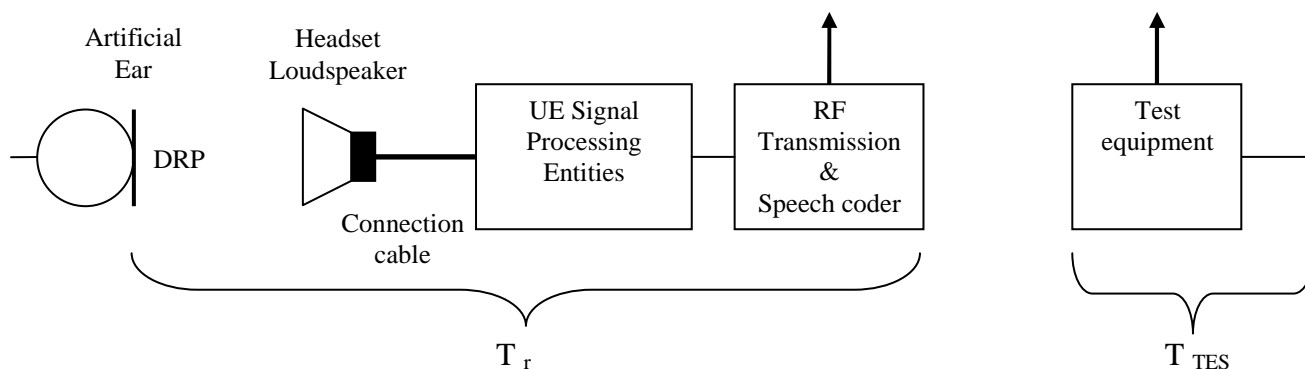
Figure 17b4: Different entities contributing to the delay in receiving direction with a headset connected via cable

Note: The test setup only applies to headsets connected by wire. Wireless headsets (e.g. connected by Bluetooth) are currently out of scope.

The test method is the same as for handset UE (clause 7.10.2).

7.11 Echo control characteristics

7.11.1 Test set-up and test signals

The device is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A).

The test shall be performed with the British-English 'long' double-talk and conditioning speech sequences from ITU-T Recommendation P.501 [22], with the signals in the receiving direction band limited according to clause 5.4.

A description of the test stimuli is presented in Table 2a and Table 2b. The test sequence is composed of an initial conditioning sequence of 23,5 s and a double talk sequence of 35 s. For the analysis, the double talk sequence is divided into two segments, a first double-talk sequence with single short near-end words (0 – 20 s), and a second double-talk sequence with continuous double talk (20 – 35 s).

The sending speech during double-talk and the 'near-end speech only' are recorded individually, with the 'near-end speech only' sequence recorded with silence in the receiving direction. The time-alignment of the two recorded sequences is performed off-line during the analysis.

Table 2a: Test stimuli for recording of Echo Canceller operation

|	Conditioning	Single words (segment 1) and full sentence (segment 2) double talk
Far-end signal	`FB_female_conditioning_seq_long.wav`	`FB_male_female_single-talk_seq.wav`
Artificial mouth signal	`FB_male_conditioning_seq_long.wav`	`FB_male_female_double-talk_seq.wav`

Table 2b: Test stimuli for reference "near-end speech only" recording.

	Conditioning	Single words (segment 1) and full sentence (segment 2) double talk
Far-end signal	`FB_female_conditioning_seq_long.wav`	`silence`
Artificial mouth signal	`FB_male_conditioning_seq_long.wav`	`FB_male_female_double-talk_seq.wav`

The level of the signal of the artificial mouth shall be -4,7 dBPa measured at the MRP. In order to obtain a reproducible time alignment as seen by the UE, the artificial mouth signal shall be delayed by the amount of the receiving direction delay. For the purpose of this alignment, the receiving direction delay for handset and headset modes is defined from the system simulator input to the artificial ear. For hands-free modes, the downlink delay is defined from the system simulator input to the acoustic output from the UE loudspeaker.

The level of the downlink signal shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point.

7.11.2 Test method

The test method measures the duration of any level difference between the sending signal of a double-talk sequence (where the echo canceller has been exposed to simultaneous echo and near-end speech) and the sending signal of the same near-end speech only. The level difference is classified into eight categories according to Figure 17b5 and Table 2c, representing various degrees of 'Full duplex operation', 'Near-end clipping', and 'Residual echo'.

NOTE 1: The limits for specifying the categories in Figure 17b5 and Table 2c are provisional pending further analysis and validation.

NOTE 2: The categories in Figure 17b5 and Table 2c are labelled in a functional order and the subjective impression of the respective categories is for further study.

NOTE 3: To reduce potential issues associated with low-frequency test room noise, a [4th]-order high-pass filter with a cut-off frequency of [100] Hz can be applied before the level computation.
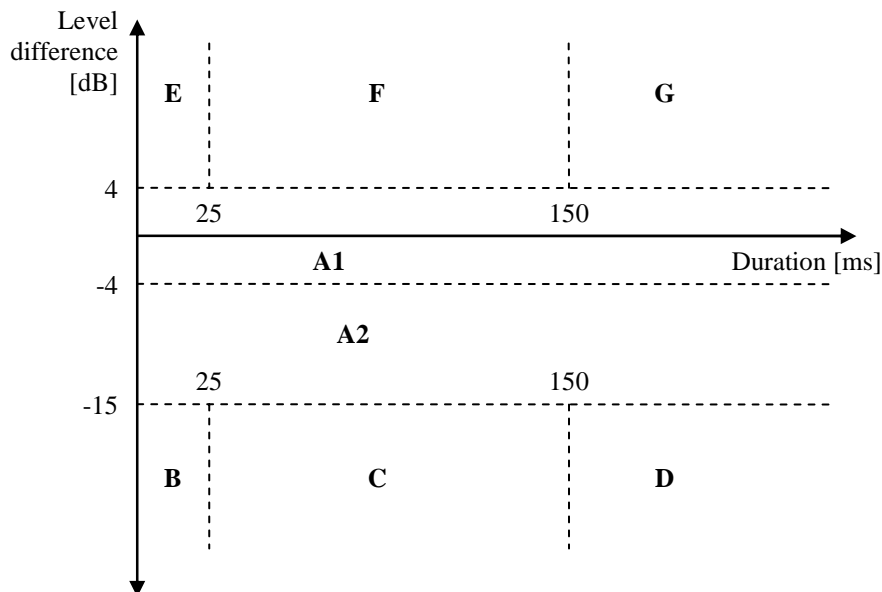
Figure 17b5: Classification of echo canceller performance

Table 2c: Categories for echo canceller performance classification

Category	Level difference (ΔL)	Duration (D)	Description
A1	-4 dB ≤ ΔL < 4 dB		Full-duplex and full transparency
A2	-15 dB ≤ ΔL < -4 dB		Full-duplex with level loss in Tx
B	ΔL < -15 dB	D < 25 ms	Very short clipping
C	ΔL < -15 dB	25 ms ≤ D < 150 ms	Short clipping resulting in loss of syllables
D	ΔL < -15 dB	D ≥ 150 ms	Clipping resulting in loss of words
E	ΔL ≥ 4 dB	D < 25 ms	Very short residual echo
F	ΔL ≥ 4 dB	25 ms ≤ D < 150 ms	Echo bursts
G	ΔL ≥ 4 dB	D ≥ 150 ms	Continuous echo

A pseudo-code reference of the test method including test scripts and test-vectors is presented in clause C.3 and outlined in the following sub clauses.

7.11.2.1 Signal alignment

For the analysis of the signal level difference, the send signal during double-talk and the near-end only signal are aligned using a correlation analysis as described in clause C.3.2.

7.11.2.2 Signal level computation and frame classification

The analysis is based on the digital level measured with a meter according to IEC 61672 [38] with a time constant of 12,5 ms, sampled at 5 ms intervals corresponding to the evaluated frames.

The 'double-talk' frames are defined as the frames where both the far-end (receiving direction) signal includes active speech (extended with a hang-over period of 200 ms) and the near-end signal is composed of active speech. Active speech is defined to be detected using a speech level meter according to ITU-T P.56, and frames within -15.9 dB from the active speech level are classified as active speech frames.

The 'far-end single-talk adjacent to double-talk' frames are similarly defined using a speech level meter according to ITU-T P.56 as the frames with active far-end speech (extended with a hang-over period of 200 ms) and no active near-end speech (extended with a hang-over period of 200 ms).

A reference implementation of the signal level computation and frame classification is presented in clause C.3.3.

7.11.2.3 Classification into categories

The analysis and classification into the categories according to Figure 17b5 and Table 2c is performed according to the reference implementation described in clause C.3.4 and C.3.4.

The frames are first categorized according to the level categories defined in Table 2c. To determine the durations, the amount of adjacent frames falling into the same level category is determined.

The classification is then performed individually for the following situations:

- frames classified as 'double-talk' from segment 1 of the double-talk sequence (see clause 7.11.1)

- frames classified as 'far-end single-talk adjacent to double-talk' from segment 1 of the double-talk sequence

- frames classified as 'double-talk' from segment 2 of the double-talk sequence

- frames classified as 'far-end single-talk adjacent to double-talk' from segment 2 of the double-talk sequence

To determine the percentage values for each category (A1, A2, B, C, D, E, F, and G) within each situation, the number of frames falling into the respective category is divided by the total number of frames within the situation in question.

To determine the averaged level difference of the frames for each category (A1, A2, B, C, D, E, F, and G) within each situation, the sum of the level difference (in dB) of the frames falling into the respective category is divided by the total number of frames within the situation in question.

7.12 Quality (speech quality, noise intrusiveness) in the presence of ambient noise

The speech quality in sending for narrowband systems is tested based on ETSI TS 103 106 [34]. This test method leads to three MOS-LQOn quality numbers:

N-MOS-LQOn: Transmission quality of the background noise

S-MOS-LQOn: Transmission quality of the speech

G-MOS-LQOn: Overall transmission quality

The test arrangement is given in clause 5.1.5. The measurement is conducted for 8 noise conditions as described in Table 2d. The measurements should be made in the same unique and dedicated call. The noise types shall be presented according to the order specified in Table 2d.

Table 2d: Noise conditions used for ambient noise simulation

Description	File name	Duration	Level	Type
Recording in pub	Pub_Noise_binaural_V2	30 s	L: 75,0 dB(A) R: 73,0 dB(A)	Binaural
Recording at pavement	Outside_Traffic_Road_binaural	30 s	L: 74,9 dB(A) R: 73,9 dB(A)	Binaural
Recording at pavement	Outside_Traffic_Crossroads_binaural	20 s	L: 69,1 dB(A) R: 69,6 dB(A)	Binaural
Recording at departure platform	Train_Station_binaural	30 s	L: 68,2 dB(A) R: 69,8 dB(A)	Binaural
Recording at the drivers position	Fullsize_Car1_130Kmh_binaural	30 s	L: 69,1 dB(A) R: 68,1 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 a cafeteria	Mensa_binaural	22 s	L: 63,4 dB(A) R: 61,9 dB(A)	Binaural
Recording in business office	Work_Noise_Office_Callcenter_binaural	30 s	L: 56,6 dB(A) R: 57,8 dB(A)	Binaural

1) Before starting the measurements a proper conditioning sequence shall be used. The conditioning sequence shall be comprised of the four additional sentences 1- 4 described in ETSI TS 103 106 [34], applied to the beginning of the 16-sentence test sequence. The conditioning signal level is -1,7 dBPa at the MRP, measured as the active speech level according to ITU-T P.56 [37].

NOTE: The sequence of speech samples concatenated for the test signal, consisting of alternating talkers in the sending direction, reduces the overall test time but may represent an unrealistic behaviour for certain voice enhancement technologies. Alternative concatenations are for further study.

2) The send speech signal consists of the 16 sentences of speech as described in ETSI TS 103 106 [34]. The test signal level is -1,7 dBPa at the MRP, measured as the active speech level according to ITU-T P.56 [37]. Three signals are required for the tests:

   – The clean speech signal is used as the undisturbed reference (see ETSI TS 103 106 [34], ETSI EG 202 396-3 [36]).

   – The speech plus undisturbed background noise signal is recorded at the terminal's microphone position using an omnidirectional measurement microphone with a linear frequency response between 50 Hz and 12 kHz.

   – The send signal is recorded at the POI.

3) N-MOS-LQOn, S-MOS-LQOn and G-MOS-LQOn are calculated as described in ETSI TS 103 106 [34] on a per sentence basis and averaged over all 16 sentences. The results shall be reported as average and standard deviation.

4) The measurement is repeated for each ambient noise condition described in Table 2d.

5) The average of the results derived from all ambient noise types is calculated.

8 Wideband telephony transmission performance test methods

8.1 Applicability

The test methods in this clause shall apply when testing a UE that is used to provide narrowband or wideband telephony, either as a stand-alone service, or as part of a multimedia service.

The application force used to apply the handset against the artificial ear shall be 8 ± 2 N. For the headset case, the application of the headset shall comply with ITU-T Recommendation P.57 [14].

8.2 Overall loss/loudness ratings

8.2.1 General

The SLR and RLR values for GSM or 3G networks apply up to the POI. However, the main determining factors are the characteristics of the UE, including the analogue to digital conversion (ADC) and digital to analogue conversion (DAC). In practice, it is convenient to specify loudness ratings to the Air Interface. For the normal case, where the GSM or 3G network introduce no additional loss between the Air Interface and the POI, the loudness ratings to the PSTN boundary (POI) will be the same as the loudness ratings measured at the Air Interface.

8.2.2 Connections with handset UE

8.2.2.1 Sending loudness rating (SLR)

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence.

b) The handset terminal is setup as described in clause 5. The sending sensitivity shall be calculated from each band of the 20 frequencies given in table G.1 of ITU-T Recommendation P.79 Annex A [16], 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.

c) The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23b), over bands 1 to 20, using m = 0,175 and the sending weighting factors from ITU-T Recommendation P.79 Annex A [16], table A2.

8.2.2.2 Receiving loudness rating (RLR)

a) The test signal to be used for the measurements shall be the British-English single talk sequence described ITU-T Recommendation P.501 [22]. The test signal level shall be 16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.

b) The handset terminal is setup as described in clause 5. The receiving sensitivity shall be calculated from each band of the 20 frequencies given in table A.2 of ITU-T Recommendation P.79 Annex A [16], 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.

c) The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23c), over bands 1 to 20, using m = 0,175 and the receiving weighting factors from table A.2 of ITU-T Recommendation P.79 Annex A [16].

d) DRP-ERP correction is applied. No leakage correction shall be applied.

8.2.3 Connections with desktop and vehicle-mounted hands-free UE

Vehicle-mounted hands-free UE should be tested within the vehicle (for the totally integrated vehicle hands-free systems) or in a vehicle simulator, as described in 3GPP TS 03.58 [11].

Free-field measurements for vehicle-mounted hands-free are for further study.

8.2.3.1 Sending loudness rating (SLR)

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence. The broadband signal level then is adjusted to –28,7 dBPa at the HFRP or the HATS HFRP (as defined in ITU-Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as references to determine the sending sensitivity S_{mJ}.

b) The hands-free terminal is setup as described in clause 5. The sending sensitivity shall be calculated from each band of the 20 frequencies given in table A.2 of ITU-T Recommendation P.79 Annex A [16], 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.

c) The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23b), over bands 1 to 20, using m = 0,175 and the sending weighting factors from ITU-T Recommendation P.79 Annex A [16], table A.2.

8.2.3.2 Receiving loudness rating (RLR)

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.

b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. The receiving sensitivity shall be calculated from each band of the 20 frequencies given in table A.2 of ITU-T Recommendation P.79 Annex A [16], 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.

c) The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23c), over bands 1 to 20, using m = 0,175 and the receiving weighting factors from table A.2 of ITU-T Recommendation P.79 Annex A [16].

d) No leakage correction shall be applied. The hands-free correction as described in ITU-T Recommendation P.340 shall be applied. To compute the receiving loudness rating (RLR) for a hands-free terminal (see also ITU-T Recommendation P.340), when using the combination of left and right artificial ear signals from the HATS, the HFL_E has to be 8 dB instead of 14 dB. For further information see ITU-T Recommendation P.581.

8.2.4 Connections with hand-held hands-free UE

8.2.4.1 Sending loudness rating (SLR)

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence. The broadband signal level then is adjusted to –28,7 dBPa at the HFRP or the HATS HFRP (as defined in P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as reference to determine the sending sensitivity S_{mJ}.

b) The hands-free terminal is setup as described in clause 5. The sending sensitivity shall be calculated from each band of the 20 frequencies given in table A.2 of ITU-T Recommendation P.79 Annex A [16], 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.

c) The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23b), over bands 1 to 20, using m = 0,175 and the sending weighting factors from ITU-T Recommendation P.79 Annex A [16], table A.2.

8.2.4.2 Receiving loudness rating (RLR)

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.

b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. The receiving sensitivity shall be calculated from each band of the 20 frequencies given in table A.2 of ITU-T Recommendation P.79 Annex A [16], 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.

c) The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23c), over bands 1 to 20, using m = 0,175 and the receiving weighting factors from table A.2 of ITU-T Recommendation P.79 Annex A [16].

d) No leakage correction shall be applied. The hands-free correction as described in ITU-T Recommendation P.340 shall be applied. To compute the receiving loudness rating (RLR) for hands-free terminals (see also ITU-T Recommendation P.340) when using the combination of left and right artificial ear signals from the HATS the HFL_E has to be 8 dB, instead of 14 dB. For further information see ITU-T Recommendation P.581.

8.2.5 Connections with headset UE

Same as for handset.

8.3 Idle channel noise (handset and headset UE)

For idle noise measurements in sending and receiving directions, care should be taken that only the noise is windowed out by the analysis and the result is not impaired by any remaining reverberation or by noise and/or interference from various other sources. Some examples are air-conducted or vibration-conducted noise from sources inside or outside the test chamber, disturbances from lights and regulators, mains supply induced noise including grounding issues, test system and system simulator inherent noise as well as radio interference from the UE to test equipment such as ear simulators, microphone amplifiers, etc.

8.3.1 Sending

The terminal should be configured to the test equipment as described in subclause 5.1.

The environment shall comply with the conditions described in subclause 6.1.

The noise level at the output of the SS is measured with A-weighting. The A-weighting filter is described in IEC 60651.

A test signal may have to be intermittently applied to prevent "silent mode" operation of the MS. This is for further study.

The measured part of the noise shall be 170,667 ms (which equals 8192 samples in a 48 kHz sample rate test system). The spectral distribution of the noise is analyzed with an 8k FFT using windowing with ≤ 0,1 dB leakage for non bin-centered signals. This can be achieved with a window function commonly known as a 'flat top window'. Within the specified frequency range, the FFT bin that has the highest level is searched for; the level of this bin is the maximum level of a single frequency disturbance.

To improve repeatability, the test sequence (optional activation followed by the noise level measurement) may be contiguously repeated one or more times.

The total noise powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

The single frequency maximum powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

8.3.2 Receiving

The terminal should be configured to the test equipment as described in subclause 5.1.

The environment shall comply with the conditions described in subclause 6.1.

A test signal may have to be intermittently applied to prevent "silent mode" operation of the MS. This is for further study.

The noise shall be measured with A-weighting at the DRP with diffuse-field correction. The A-weighting filter is described in IEC 60651.

The measured part of the noise shall be 170,667 ms (which equals 8192 samples in a 48 kHz sample rate test system). The spectral distribution of the noise is analyzed with an 8k FFT using windowing with ≤ 0,1 dB leakage for non bin-centered signals. This can be achieved with a window function commonly known as a 'flat top window'. Within the specified frequency range the FFT bin that has the highest level is searched for; the level of this bin is the maximum level of a single frequency disturbance.

To improve repeatability, the test sequence (optional activation followed by the noise level measurement) may be contiguously repeated one or more times.

The total noise powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

The single frequency maximum powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

8.4 Sensitivity/frequency characteristics

8.4.1 Handset and headset UE sending

The headset case is similar to the handset one, except for the application force.

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence.

b) The handset terminal is setup as described in clause 5. Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 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.

c) The sensitivity is expressed in terms of dBV/Pa.

8.4.2 Handset and headset UE receiving

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.

b) The handset terminal is setup as described in clause 5. Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 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.

c) The HATS is diffuse-field equalized. The sensitivity is expressed in terms of dBPa/V. Information about correction factors is available in ITU-T Recommendation P.57 [14].

Optionally, the measurements may be repeated with 2 N and 13 N application force. For these test cases no normative values apply.

8.4.3 Desktop and vehicle-mounted hands-free UE sending

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence. The broadband signal level is then adjusted to –28,7 dBPa at the HFRP or the HATS HFRP (as defined in ITU-T Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as references to determine the sending sensitivity S_{mJ}.

b) The hands-free terminal is setup as described in clause 5. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 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.

c) The sensitivity is expressed in terms of dBV/Pa.

8.4.4 Desktop and vehicle-mounted hands-free UE receiving

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.

b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 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.

c) The sensitivity is expressed in terms of dBPa/V.

8.4.5 Hand-held hands-free UE sending

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence. The broadband signal level is then adjusted to –24,3 dBPa at the HFRP or the HATS HFRP (as defined in subclause 8.2.3.1) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as reference to determine the sending sensitivity S_{mJ}.

b) The hands-free terminal is setup as described in clause 5.1.3.3. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 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.

c) The sensitivity is expressed in terms of dBV/Pa.

8.4.6 Hand-held hands-free UE receiving

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.

b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave band frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 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.

c) The sensitivity is expressed in terms of dBPa/V.

8.5 Sidetone characteristics

8.5.1 Connections with handset UE

The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal shall be produced by the HATS. The test signal level shall be -4,7 dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence.

The handset UE is set up as described in clause 5. The application force shall be 13 N on the Type 3.3 artificial ear.

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 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 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 [16], table 4, 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 Rating (STMR), expressed in dB, shall be calculated from formula 5-1 of ITU-T Recommendation P.79 [16], using m = 0.225 and the weighting factors in table B2 (unsealed condition) of ITU-T Recommendation P.79 [16]. No leakage correction (L_E) shall be applied. DRP-ERP correction is used.

8.5.2 Headset UE

The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. 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 calculated over the complete test signal sequence.

 Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 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 [16], table 4, bands 1 to 20) is referred to the averaged test signal level measured in each frequency band.

The sidetone path loss (L_{meST}), as expressed in dB, shall be calculated from each band of the 20 frequencies given in table G.1 of ITU-T Recommendation P.79 Annex A [16], bands 1 to 20. The STMR (in dB) shall be calculated from formula B-4 of ITU-T Recommendation P.79 [16], using m = 0.225 and the weighting factors in table B.2 (unsealed condition) of ITU-T Recommendation P.79 [16]. No leakage correction (L_E) shall be applied. DRP-ERP correction is used.

8.5.3 Hands-free UE (all categories)

No requirement other than echo control.

8.5.4 Sidetone delay for handset or headset

The handset or headset terminal is setup as described in clause 5.

The test signal is a CS-signal complying with ITU-T Recommendation P.501 using a PN-sequence with a length, T, of 4 096 points (for a 48 kHz sample rate test system). The duration of the complete test signal is as specified in ITU-T Recommendation P.501. The level of the signal shall be -4,7 dBPa at the MRP.

The cross-correlation function $\Phi xy(\tau)$ 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:

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

The measurement window, *T*, shall be identical to the test signal period, T, with the measurement window synchronized to the PN-sequence of the test signal.

The sidetone delay is calculated from the envelope $E(\tau)$ of the cross-correlation function $\Phi xy(\tau)$. 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 $E(\tau)$ is calculated by the Hilbert transformation H {xy(τ)} of the cross-correlation:

$$H\{xy(\tau)\} = \sum_{u=-\infty}^{+\infty} \frac{\Phi_{xy}(u)}{\pi(\tau-u)} \tag{2}$$

$$E(\tau) = \sqrt{\left[\Phi_{xy}(\tau)\right]^2 + \left[H\{xy(\tau)\}\right]^2} \tag{3}$$

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

8.6 Stability loss

Where a user-controlled volume control is provided it is set to maximum.

Handset UE: The handset is placed on a hard plane surface with the earpiece facing the surface.

Headset UE: The requirement applies for the closest possible position between microphone and headset receiver within the intended wearing position.

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

Hands-free UE (all categories): No requirement other than echo loss.

Before the actual test a training sequence consisting of the British-English single talk sequence described in ITU-T Recommendation P.501 [22] is applied. The training sequence level shall be -16 dBm0 in order to not overload the codec.

The test signal is a PN-sequence complying with ITU-T Recommendation P.501 with a length of 4 096 points (for a 48 kHz sampling rate system) and a crest factor of 6 dB instead of 11 dB. The PN-sequence is generated as described in P.501 with W(k) constant within the frequency range 100-8000 Hz and zero outside this range. The duration of the test signal is 250 ms. With an input signal of -3 dBm0, the attenuation from input to output of the system simulator shall be measured under the following conditions:

a) The handset or the headset, with the transmission circuit fully active, shall be positioned on a hard plane surface with at least 400 mm free space in all directions. The earpiece shall face towards the surface as shown in figure 17c;

b) The headset microphone is positioned as close as possible to the receiver(s) within the intended wearing position;

c) For a binaural headset, the receivers are placed symmetrically around the microphone.
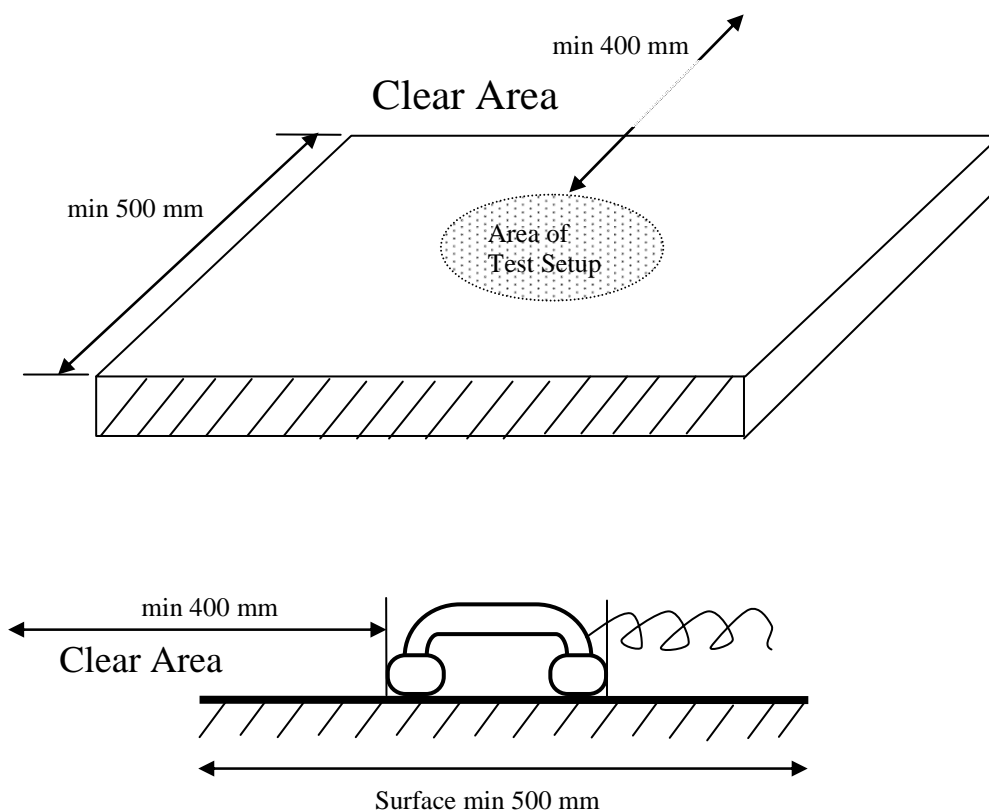
NOTE: All dimensions in mm.
Figure 17c. Test configuration for stability loss measurement on handset or headset UE

The attenuation from input to output shall be measured in the frequency range from 100 Hz to 8 kHz. The spectral distribution of the output signal is analysed with a 4k FFT (for a 48 kHz sample rate test system), thus the measured part of the output signal is 85,333 ms. To avoid leakage effects the frequency resolution of the FFT must be the same as the frequency spacing of the PN-sequence.

8.7 Acoustic echo control

8.7.1 General

The echo loss (EL) presented by the GSM or 3G networks at the POI should be at least 46 dB during single talk. This value takes into account the fact that UE is likely to be used in a wide range of noise environments.

8.7.2 Acoustic echo control in a hands-free UE

The hands-free UE is setup in a room with acoustic properties similar to a typical 'office-type' room; a vehicle-mounted hands-free UE should be tested in a vehicle or vehicle simulator, as specified by the UE manufacturer (see also 3GPP TS 03.58 [11]). The ambient noise level shall be ≤ -70 dBPa(A). The attenuation from reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, 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. 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).

The test signal level shall be -10 dBm0.

8.7.3 Acoustic echo control in a handset UE

The handset is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A). The attenuation from the reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, 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. 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).

The test signal level shall be -10 dBm0.

8.7.4 Acoustic echo control in a headset UE

The headset is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A). The attenuation from the reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, 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. 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).

The test signal level shall be -10 dBm0.

8.8 Distortion

8.8.1 Sending distortion

The handset, headset, or hands-free UE is setup as described in clause 5.

The signal used is a sine-wave signal with frequencies specified in clause 6.8 of 3GPP TS 26.131. The sine-wave signal level shall be calibrated to -4,7 dBPa at the MRP for all frequencies, except for the sine-wave with a frequency 1020 Hz

which shall be applied at the following levels at the MRP: 5, 0, -4,7, -10, -15, -20 dBPa. The test signals have to be applied in this sequence, i.e., from high levels down to low levels.

The duration of the sine-wave signal is recommended to be 360 ms. The manufacturer shall be allowed to request tone lengths up to 1 s. The measured part of the signal shall be 170,667 ms (which equals 2 * 4096 samples in a 48 kHz sample rate test system). The times are selected to be relatively short in order to reduce the risk that the test tone is treated as a stationary signal.

It is recommended that an optional activation signal be presented immediately preceding each test signal to ensure that the UE is in a typical state during measurement (see Note 1.). An appropriate speech or speech-like activation signal shall be chosen from ITU-T Recommendations P.501 or P.50 [10]. A recommendation for the use of an activation signal as part of the measurement is defined in figure 18. The RMS level of the active parts of this activation signal is recommended to be equal to the subsequent test tone RMS level. In practice, certain types of processing may be impacted due to the introduction of the activation signal. The manufacturer shall be allowed to specify disabling of the activation signal. It shall be reported whether an activation signal was used or not, along with the characteristics of the activation signal, as specified by the manufacturer.

The ratio of the signal to total distortion power of the signal output of the SS shall be measured with the psophometric noise weighting (see ITU-T Recommendations G.712, O.41 and O.132). The psophometric filter shall be normalized (0 dB gain) at 800 Hz as specified in ITU-T Recommendation O.41. The weighting function shall be applied to the total distortion component only (not to the signal component).

For measurement of the total distortion component an octave-wide band-stop filter shall be applied to the signal to suppress the sine-wave signal and associated coding artefacts. The filter shall have a lower passband ending at $0.7071 * f_S$, and an upper passband starting at $1.4142 * f_S$, where f_S is the frequency of the sine-wave signal. The passband ripple of the filter shall be ≤ 0,2 dB. The attenuation of the band-stop filter at the sine-wave frequency shall be ≥ 60 dB. Alternatively, the described characteristics can be implemented by an appropriate weighting on the spectrum obtained from an FFT. The total distortion component is defined as the measured signal within the frequency range 100 Hz to 6 kHz, after applying psophometric and stop filters (hence no correction for the lost power due to the stop filter, known as 'bandwidth correction', shall be applied).

To improve repeatability, considering the variability introduced by speech coding and voice processing, the test sequence (activation signal followed by the test signal) may be contiguously repeated one or more times. The single signal-to-total-distortion power ratios obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

Figure 18: Recommended activation sequence and test signal.

The activation signal consists of a 'Bandlimited composite source signal with speech-like power density spectrum' signal according to ITU-T Recommendation P.501 with 48,62 ms voiced part (1), 200 ms unvoiced part (2) and 101,38 ms pause (3), followed by the same signal but polarity inverted (4, 5, 6), followed by the voiced part only (7). The pure test tone is applied and after 50 ms settling time (8), the analysis is made over the following 170,667 ms (9).

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

NOTE 2: Void.

NOTE 3: Void.

NOTE 4: In order to ensure that the correct part of the signal is analyzed, the total delay of the terminal and SS may have to be determined prior to the measurement.

NOTE 5: For hands-free terminals tested in environments defined in subclause 6.1.2, care should be taken that the reverberation in the test room, caused by the activation signal, does not affect the test results to an unacceptable degree, referring to subclause 5.3.

8.8.2   Receiving

The handset, headset, or hands-free UE is setup as described in clause 5.

The signal used is a sine-wave signal with frequencies specified in clause 6.8 of 3GPP TS 26.131. The signal level shall be -16 dBm0, except for the sine-wave signal with a frequency 1020 Hz that shall be applied at the signal input of the SS at the following levels: 0, -3, -10, -16, -20, -30, -40, -45 dBm0. The test signals have to be applied in this sequence, i.e., from high levels down to low levels.
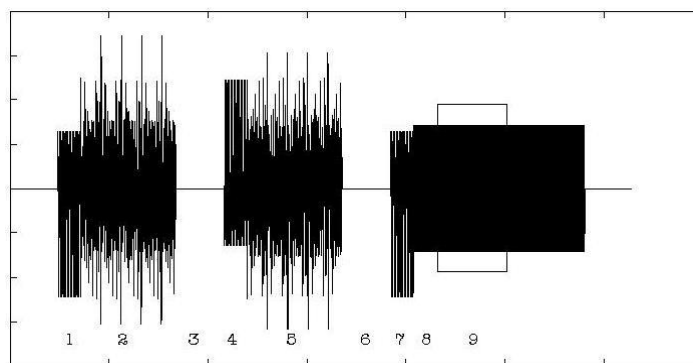
The duration of the sine-wave signal is recommended to be 360 ms. The manufacturer shall be allowed to request tone lengths up to 1 s. The measured part of the signal shall be 170,667 ms (which equals 2 * 4096 samples in a 48 kHz sample rate test system). The times are selected to be relatively short in order to reduce the risk that the test tone is treated as a stationary signal.

It is recommended that an optional activation signal be presented immediately preceding each test signal to ensure that the UE is in a typical state during measurement (see Note 1.). An appropriate speech or speech-like activation signal shall be chosen from ITU-T Recommendations P.501 or P.50 [10]. A recommendation for the use of an activation signal as part of the measurement is defined in figure 19. The RMS level of the active parts of this activation signal is recommended to be equal to the subsequent test tone RMS level for low and medium test levels. To avoid saturation of the SS speech encoder, it is recommended for high test levels that the activation signal level is adjusted so that its peak level equals the peak level of the test tone. In practice, certain types of processing may be impacted due to the introduction of the activation signal. The manufacturer shall be allowed to specify disabling of the activation signal. It shall be reported whether an activation signal was used or not, along with the characteristics of the activation signal, as specified by the manufacturer.

The ratio of the signal to total distortion power shall be measured at the applicable acoustic measurement point (DRP with diffuse-field correction for handset and headset modes; free field for hands-free modes) with the psophometric noise weighting (see ITU-T Recommendations G.712, O.41 and O.132). The psophometric filter shall be normalized to have 0 dB gain at 800 Hz as specified in ITU-T Recommendation O.41. The weighting function shall be applied to the total distortion component only (not to the signal component).

For measurement of the total distortion component an octave-wide band-stop filter shall be applied to the signal to suppress the sine-wave signal and associated coding artefacts. The filter shall have a lower passband ending at $0,7071 * f_S$, and an upper passband starting at $1,4142 * f_S$, where f_S is the frequency of the sine-wave signal. The passband ripple of the filter shall be $\leq 0,2$ dB. The attenuation of the band stop filter at the sine-wave frequency shall be ≥ 60 dB. Alternatively the described characteristics can be implemented by an appropriate weighting on the spectrum obtained from an FFT. The total distortion component is defined as the measured signal within the frequency range 100 Hz to 6 kHz, after applying psophometric and stop filters (hence no correction for the lost power due to the stop filter, known as 'bandwidth correction', shall be applied).

To improve repeatability, considering the variability introduced by speech coding and voice processing, the test sequence (activation signal followed by the test signal) may be contiguously repeated one or more times. The single signal-to-total-distortion power ratios obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

Figure 19: Recommended activation sequence and test signal.

The activation signal consists of a 'Bandlimited composite source signal with speech-like power density spectrum' signal according to ITU-T Recommendation P.501 with 48,62 ms voiced part (1), 200 ms unvoiced part (2) and 101,38 ms pause (3), followed by the same signal but polarity inverted (4, 5, 6), followed by the voiced part only (7). The pure test tone is applied and after 50 ms settling time (8), the analysis is made over the following 170,667 ms (9).

NOTE 1: Void.

NOTE 2: Void.

NOTE 3: In order to ensure that the correct part of the signal is analyzed, the total delay of the terminal and SS may have to be determined prior to the measurement.

NOTE 4: For hands-free terminals tested in environments defined in subclause 6.1.2, care should be taken that the reverberation in the test room, caused by the activation signal, does not affect the test results to an unacceptable degree, referring to subclause 5.3.

8.9 Ambient noise rejection

Handset and headset UE:

NOTE: The ambient noise rejection method has been updated to support basic testing of UEs that include noise suppression. Other methods are being investigated that may be more suitable for proper characterization of ambient noise rejection by UEs.

a) A 1/2-inch pressure microphone is calibrated using a known sound source and mounted at the MRP, without the HATS present. The measurement system is calibrated to enable the sound pressure levels at the microphone to be determined in 1/3-octave bands.

b) For calibration of the background noise generation system, flood the room in which the measurement is to be made with a band-limited (100 Hz to 8 kHz) pink-noise, accurate within ± 3 dB. The level at the MRP shall be adjusted to 70 dBSPL(A) (-24 dBPa(A)). The tolerance on this level is ± 1 dB. The resulting sound spectrum is P_{rn} (dBPa), measured in 1/3-octave bands.

To ensure that the sound field is sufficiently diffuse, the following applies:

The diffuse sound-field is calibrated in the absence of any local obstacles. The averaged field shall be uniform to within ± 3 dB within a radius of 0,15 m of the MRP, when measured in 1/3-octave bands from 100 Hz to 6,3 kHz.

NOTE 1: The pressure intensity index, as defined in ISO 9614, may prove to be a suitable method for assessing the diffuse field.

NOTE 2: Where more than one loudspeaker is used to produce the desired sound field, the loudspeakers must be fed with non-coherent electrical signals to eliminate standing waves and other interference effects.

c) The speech test signal to be used for measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is

calibrated at the MRP. The test signal level shall be +1,3 dBPa, measured at the MRP. The resulting sound spectrum is P_o (dBPa), measured in 1/3-octave bands.

d) Mount the MS under test, according to subclause 5.1.1.

e) Set up a send speech path between the MS and the SS and reproduce the artificial speech signal and pink noise signal simultaneously according to the sequence in figure 19a:

Figure 19a: Conditioning and test sequence for ambient noise rejection.

NOTE: Windows I, II, III are used only for conditioning of the MS. Window IV is used for the actual measurement of the MS speech send sensitivity and window V for the measurement of the MS noise send sensitivity.

f) To determine the send noise sensitivity, measure as a function of frequency in 1/3-octave bands (index j), the electrical output, V_{jrn}, (expressed as dB rel. 1 V) at the audio output of the SS for the applied noise acoustic pressure P_{jrn} (expressed as dB rel. 1 Pa) at the MRP. This measurement is performed during the period determined by window V shown in figure 19a. Since the MS sending sensitivity is not defined above 6,3 kHz, the measurement shall be cut off at 6,3 kHz. For the bands below 315 Hz, the send noise sensitivity shall be referenced to the send speech sensitivity at 315 Hz to yield the DELSM.

The send noise sensitivity is expressed as:

$$Sm_{jrn} = V_{jrn} \text{ (dBV)} - P_{jrn} \text{ (dBPa)}.$$

g) To determine the send speech sensitivity, measure as a function of frequency in 1/3-octave bands (index j) the electrical output, V_j, (expressed as dB rel. 1 V), at the audio output of the SS for the applied speech acoustic pressure, P_{j0}, (expressed as dB rel. 1 Pa), at the MRP. This measurement is performed during the period determined by window IV shown in figure 19a.

The send speech sensitivity is expressed as:

$$Sm_{js} \text{ (dB)} = V_j \text{ (dBV)} - P_{jo} \text{ (dBPa)} \text{ dB rel. 1 V/Pa.}$$

h) The difference between the room noise sensitivity and the speech sending sensitivity DELSM (Δ_{jSM}) in each 1/3-octave band for the MS is determined as:

$$\text{DELSM } (\Delta_{jSM}) = Sm_{jrn} - Sm_{js} \text{ (dB)} \quad \text{(for j = 1 to 2, } Sm_{js} = Sm_{3s}).$$

i) The Ambient noise rejection ANR is calculated as a single figure value according to the following formula:

$$ANR = -\frac{4}{5}\sum_{i=1}^{16}\Delta_{jSM}\cdot 10^{-0.0175W_{jsi}}$$

j = The index of 1/3-octave bands centered at frequencies from 200 Hz to 6300 Hz inclusive.

W_{jsi} = the sending weighting factors from ITU-T Recommendation P.79 Annex G [16], table G.1 for the jth 1/3-octave band centre frequency.

Hands-free UE (all categories):

No test method for hands-free operations.

8.10 Delay

8.10.1 Delay in sending direction (handset UE)

The handset terminal is setup as described in clause 5.1.1.

The delay shall include all entities in sending direction from MRP to the POI, but shall exclude the delays introduced by the test equipment.
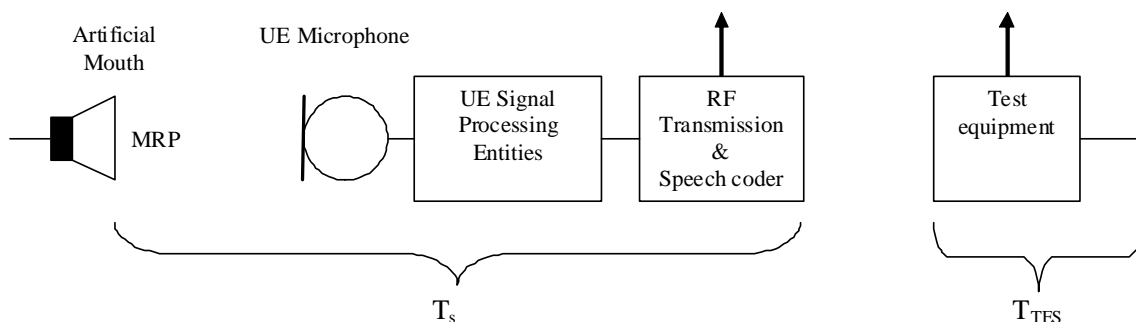
Figure 19b1: Different entities contributing to the delay in sending direction

The delay in sending direction, measured from MRP to POI, is T_s + T_{TES}.

All test equipment delays, for the network type, codec type and bitrate used according to clause 5, (including radio access, speech codec, A/D and D/A conversions etc.) are included in T_{TES}. The values used for testing (typical value considering variations due to interleaving etc.) as declared by the test equipment manufacturers shall be reported along with the measurement results.

1. For the measurements, a Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] 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 32 k samples (with 48 kHz sampling rate). The test signal level is -4,7 dBPa at the MRP.

2   The reference signal is the original signal (test signal). The setup of the handset/headset terminal is made corresponding to clause 5.1.

3. 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 subtracting the test equipment delay T_{TES}.

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

8.10.1a Delay in sending direction (headset UE)

The delay shall include all entities in sending direction from MRP to the POI, but shall exclude the delays introduced by the test equipment.
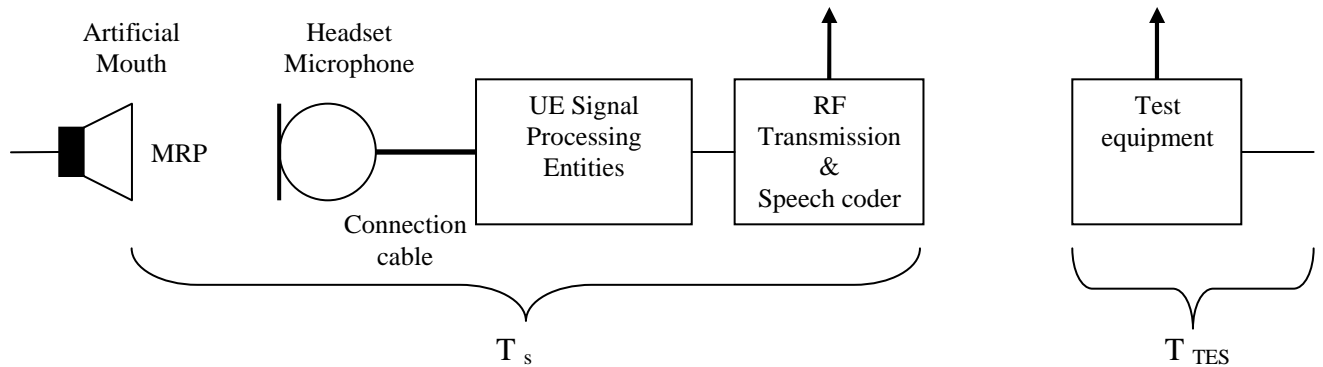
Figure 19b2: Different entities contributing to the delay in sending direction with a headset connected via cable

Note: The test setup only applies to headsets connected by wire. Wireless headsets (e.g. connected by Bluetooth) are currently out of scope.

The test method is the same as for handset UE (clause 8.10.1).

8.10.2 Delay in receiving direction (handset UE)

The handset terminal is setup as described in clause 5.

The delay shall include all entities in receiving direction from the POI to the DRP, but shall exclude the delays introduced by the test equipment.
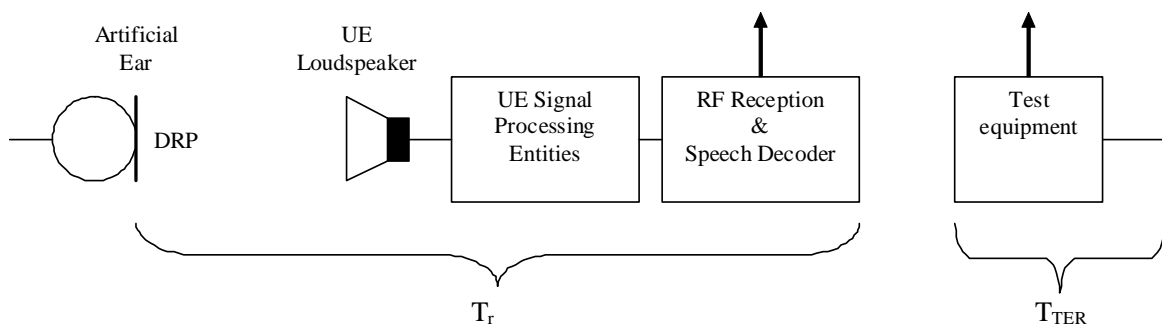
Figure 19b3: Different entities contributing to the delay in receiving direction

The delay in receiving direction, measured from POI to DRP, is $T_r + T_{TER}$.

All test equipment delays, for the network type, codec type and bitrate used according to clause 5, (including radio access, speech codec, A/D and D/A conversions etc.) are included in T_{TER}. The values used for testing (typical value considering variations due to interleaving etc.) as declared by the test equipment manufacturers shall be reported along with the measurement results.

1. For the measurements, a Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] 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 32 k samples (with 48 kHz sampling rate). The test signal level is -16 dBm0 measured at the digital reference point or the equivalent analogue point.

2 The reference signal is the original signal (test signal). The setup of the handset/headset terminal is in correspondence to clause 5.1.

3. 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 subtracting the test equipment delay T_{TER}.

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

8.10.2a Delay in receiving direction (headset UE)

The delay shall include all entities in receiving direction from the POI to the DRP, but shall exclude the delays introduced by the test equipment.
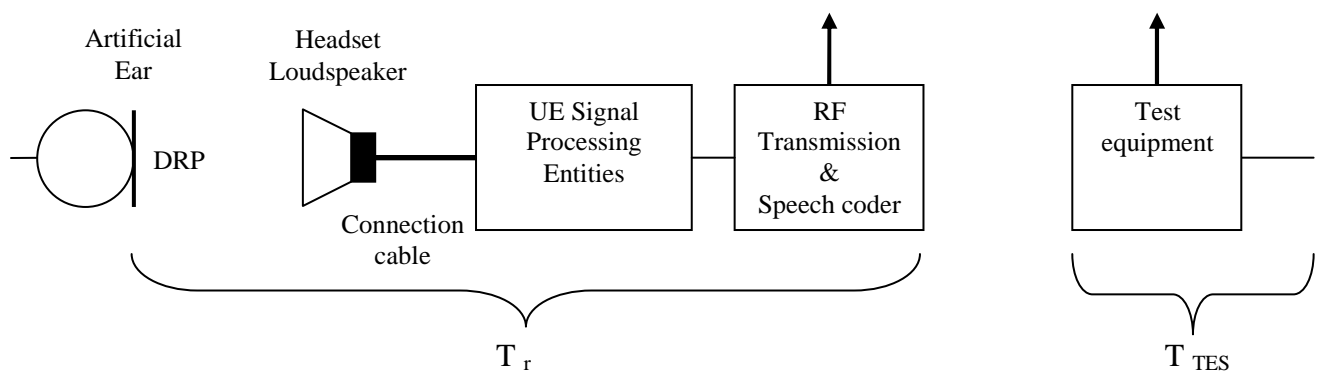
Figure 19b4: Different entities contributing to the delay in receiving direction with a headset connected via cable

Note: The test setup only applies to headsets connected by wire. Wireless headsets (e.g. connected by Bluetooth) are currently out of scope.

The test method is the same as for handset UE (subclause 8.10.2).

8.11   Echo control characteristics

8.11.1   Test set-up and test signals

The device is set up according to clause 5. The ambient noise level shall be $\leq$ -64 dBPa(A).

The test shall be performed with the British-English 'long' double-talk and conditioning speech sequences from ITU-T Recommendation P.501 [22], with the signals in the receiving direction band limited according to clause 5.4.

A description of the test stimuli is presented in Table 2e and Table 2f. The test sequence is composed of an initial conditioning sequence of 23,5 s and a double talk sequence of 35 s. For the analysis, the double talk sequence is divided into two segments, a first double-talk sequence with single short near-end words (0 – 20 s), and a second double-talk sequence with continuous double talk (20-35 s).

The sending speech during double-talk and the 'near-end speech only' are recorded individually, with the 'near-end speech only' sequence recorded with silence in the receiving direction. The time-alignment of the two recorded sequences is performed off-line during the analysis.

Table 2e: Test stimuli for recording of Echo Canceller operation

	Conditioning	Single words (segment 1) and full sentence (segment 2) double talk
Far-end signal	`FB_female_conditioning_seq_long.wav`	`FB_male_female_single-talk_seq.wav`
Artificial mouth signal	`FB_male_conditioning_seq_long.wav`	`FB_male_female_double-talk_seq.wav`

Table 2f: Test stimuli for reference "near-end speech only" recording.

	Conditioning	Single words (segment 1) and full sentence (segment 2) double talk
Far-end signal	`FB_female_conditioning_seq_long.wav`	`silence`
Artificial mouth signal	`FB_male_conditioning_seq_long.wav`	`FB_male_female_double-talk_seq.wav`

The level of the signal of the artificial mouth shall be - 4.7 dBPa measured at the MRP. In order to obtain a reproducible time alignment as seen by the UE, the artificial mouth signal shall be delayed by the amount of the receiving direction delay. For the purpose of this alignment, the receiving direction delay for handset and headset modes is defined from the system simulator input to the artificial ear. For handsfree modes, the downlink delay is defined from the system simulator input to the acoustic output from the UE loudspeaker.

The level of the downlink signal shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point.

8.11.2   Test method

The test method measures the duration of any level difference between the sending signal of a double-talk sequence (where the echo canceller has been exposed to simultaneous echo and near-end speech) and the sending signal of the same near-end speech only. The level difference is classified into eight categories according to Figure 19b5 and Table 2g, representing various degrees of 'Full duplex operation', 'Near-end clipping', and 'Residual echo'.

   NOTE: The limits for specifying the categories in Figure 19b5 and Table 2g are provisional pending further analysis and validation.

   NOTE: The categories in Figure 19b5 and Table 2g are labelled in a functional order and the subjective impression of the respective categories is for further study.

   NOTE: To reduce potential issues associated with low-frequency test room noise, a [4th]-order high-pass filter with a cut-off frequency of [100] Hz can be applied before the level computation.
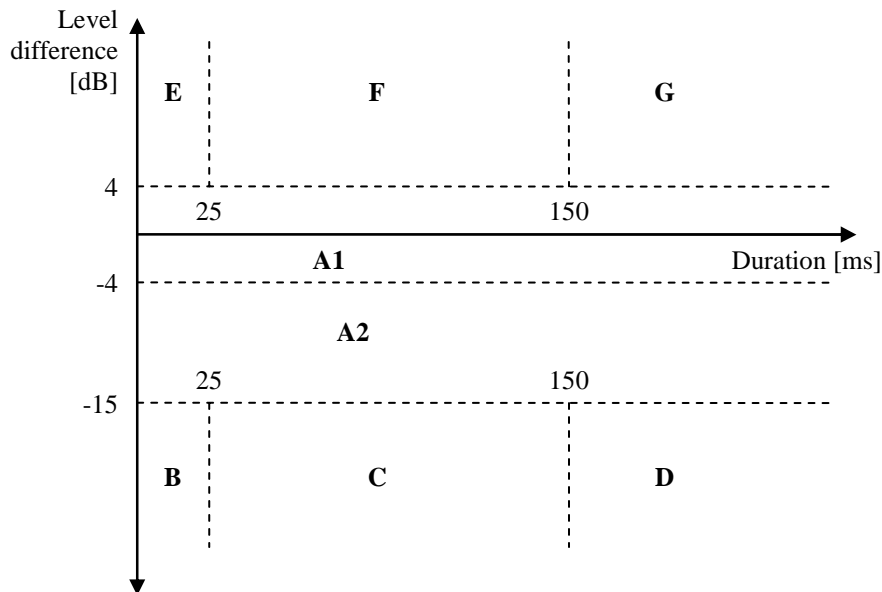
Figure 19b5: Classification of echo canceller performance

Table 2g: Categories for echo canceller performance classification

Category	Level difference (ΔL)	Duration (D)	Description
A1	-4 dB ≤ ΔL < 4 dB		Full-duplex and full transparency
A2	-15 dB ≤ ΔL < -4 dB		Full-duplex with level loss in Tx
B	ΔL < -15 dB	D < 25 ms	Very short clipping
C	ΔL < -15 dB	25 ms ≤ D < 150 ms	Short clipping resulting in loss of syllables
D	ΔL < -15 dB	D ≥ 150 ms	Clipping resulting in loss of words
E	ΔL ≥ 4 dB	D < 25 ms	Very short residual echo
F	ΔL ≥ 4 dB	25 ms ≤ D < 150 ms	Echo bursts
G	ΔL ≥ 4 dB	D ≥ 150 ms	Continuous echo

A pseudo-code reference of the test method including test scripts and test-vectors is presented in Clause C.3 and outlined in the following sub clauses.

8.11.2.1   Signal alignment

For the analysis of the signal level difference, the send signal during double-talk and the near-end only signal are aligned using a correlation analysis as described in Clause C.3.2.

8.11.2.2   Signal level computation and frame classification

The analysis is based on the digital level measured with a meter according to IEC 61672 [38] with a time constant of 12.5 ms, sampled at 5 ms intervals corresponding to the evaluated frames.

The 'double-talk' frames are defined as the frames were both the far-end (receiving direction) signal includes active speech (extended with a hang-over period of 200 ms) and the near-end signal is composed of active speech. Active speech is defined to be detected using a speech level meter according to ITU-T P.56, and frames within -15.9 dB from the active speech level are classified as active speech frames.

The 'far-end single-talk adjacent to double-talk' frames are similarly defined using a speech level meter according to ITU-T P.56 as the frames with active far-end speech (extended with a hang-over period of 200 ms) and no active near-end speech (extended with a hang-over period of 200 ms).

A reference implementation of the signal level computation and frame classification is presented in Clause C.3.3.

8.11.2.3   Classification into categories

The analysis and classification into the categories according to Figure 19b5 and Table 2g is performed according to the reference implementation described in Clause C.3.4 and C.3.4.

The frames are first categorized according to the level categories defined in Table 2g. To determine the durations, the amount of adjacent frames falling into the same level category is determined.

The classification is then performed individually for the following situations:

- frames classified as 'double-talk' from segment 1 of the double-talk sequence (see 8.11.1)

- frames classified as 'far-end single-talk adjacent to double-talk' from segment 1 of the double-talk sequence

- frames classified as 'double-talk' from segment 2 of the double-talk sequence

- frames classified as 'far-end single-talk adjacent to double-talk' from segment 2 of the double-talk sequence

To determine the percentage values for each category (A1, A2, B, C, D, E, F, and G) within each situation, the number of frames falling into the respective category is divided by the total number of frames within the situation in question.

To determine the averaged level difference of the frames for each category (A1, A2, B, C, D, E, F, and G) within each situation, the sum of the level difference (in dB) of the frames falling into the respective category is divided by the total number of frames within the situation in question.

8.12   Quality (speech quality, noise intrusiveness) in the presence of ambient noise

The speech quality in sending for narrowband systems is tested based on ETSI TS 103 106 [34]. This test method leads to three MOS-LQOw 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

The test arrangement is given in clause 5.1.5. The measurement is conducted for 8 noise conditions as described in Table 2h. The measurements should be made in the same unique and dedicated call. The noise types shall be presented according to the order specified in Table 2h.

Table 2h: Noise conditions used for ambient noise simulation

Description	File name	Duration	Level	Type
Recording in pub	Pub_Noise_binaural_V2	30 s	L: 75,0 dB(A) R: 73,0 dB(A)	Binaural
Recording at pavement	Outside_Traffic_Road_binaural	30 s	L: 74,9 dB(A) R: 73,9 dB(A)	Binaural
Recording at pavement	Outside_Traffic_Crossroads_binaural	20 s	L: 69,1 dB(A) R: 69,6 dB(A)	Binaural
Recording at departure platform	Train_Station_binaural	30 s	L: 68,2 dB(A) R: 69,8 dB(A)	Binaural
Recording at the drivers position	Fullsize_Car1_130Kmh_binaural	30 s	L: 69,1 dB(A) R: 68,1 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 a cafeteria	Mensa_binaural	22 s	L: 63,4 dB(A) R: 61,9 dB(A)	Binaural
Recording in business office	Work_Noise_Office_Callcenter_binaural	30 s	L: 56,6 dB(A) R: 57,8 dB(A)	Binaural

1) Before starting the measurements a proper conditioning sequence shall be used. The conditioning sequence shall be comprised of the four additional sentences 1-4 described in ETSI TS 103 106 [34], applied to the beginning of the 16-sentence test sequence. The conditioning signal level is - 1.7 dBPa at the MRP, measured as active speech level according to ITU-T P.56 [37].

NOTE: The sequence of speech samples concatenated for the test signal, consisting of alternating talkers in the sending direction, reduces the overall test time but may represent an unrealistic behaviour for certain voice enhancement technologies. Alternative concatenations are for further study.

2) The send speech signal consists of the 16 sentences of speech as described in ETSI TS 103 106 [34] The test signal level is - 1.7 dBPa at the MRP, measured as active speech level according to ITU-T P.56 [37]. Three signals are required for the tests:

   – The clean speech signal is used as the undisturbed reference (see ETSI TS 103 106 [34], ETSI EG 202 396-3 [36]).

   – The speech plus undisturbed background noise signal is recorded at the terminal's microphone position using an omnidirectional measurement microphone with a linear frequency response between 50 Hz and 12 kHz.

   – The send signal is recorded at the POI.

3) N-MOS-LQOw, S-MOS-LQOw and G-MOS-LQOw are calculated as described in ETSI TS 103 106 [34] on a per sentence basis and averaged over all 16 sentences. The results shall be reported as average and standard deviation.

4) The measurement is repeated for each ambient noise condition described in Table 2h.

5) The average of the results derived from all ambient noise types is calculated.

Annex A: Interpolation method for diffuse-field correction (normative)

Interpolated values for 1/12-octave bands shall be calculated from 1/3-octave band values using table 3.

For measurements requiring diffuse-field correction values for closer frequency spacing than 1/12-octave bands, linear interpolation on a log scale from the 1/12-octave band interpolated values in table 3 shall be used.

Table 3: Interpolation parameters on 1/12-octave bands.

Frequency (Hz)	Interpolated value (dB)	Frequency (Hz)	Interpolated value (dB)
95	0,000	1 000	5,000
100	0,000	1 060	5,375
106	0,000	1 120	5,750
112	0,000	1 180	6,125
118	0,000	1 250	6,500
125	0,000	1 320	6,800
132	0,000	1 400	7,150
140	0,000	1 500	7,550
150	0,000	1 600	8,000
160	0,000	1 700	8,550
170	0,000	1 800	9,175
180	0,000	1 900	9,850
190	0,000	2 000	10,500
200	0,000	2 120	11,500
212	0,125	2 240	12,550
224	0,250	2 360	13,500
236	0,390	2 500	14,050
250	0,500	2 650	13,850
265	0,525	2 800	13,250
280	0,500	3 000	12,400
300	0,480	3 150	12,000
315	0,500	3 350	11,750
335	0,600	3 550	11,650
355	0,725	3 750	11,600
375	0,875	4 000	11,500
400	1,000	4 250	11,425
425	1,135	4 500	11,375
450	1,275	4 750	11,275
475	1,375	5 000	11,000
500	1,500	5 300	10,400
530	1,625	5 600	9,550
560	1,650	6 000	8,600
600	1,800	6 300	8,000
630	2,000	6 700	7,375
670	2,450	7 100	6,800
710	3,000	7 500	6,450
750	3,500	8 000	6,500
800	4,000	8 500	7,150
850	4,325	9 000	8,250
900	4,550	9 500	9,450
950	4,750	10 000	10,450

Interpolated values for 1/12-octave bands can be also calculated from 1/3-octave band values using table 4 when frequencies are defined according to IEC 1260.

Table 4: Interpolation parameters on 1/12-octave bands with frequencies according to IEC 1260 (informative).

Frequency (Hz)	Interpolated value (dB)	Frequency (Hz)	Interpolated value (dB)
92	0,000	972	4,850
97	0,000	1029	5,180
103	0,000	1090	5,555
109	0,000	1155	5,969
115	0,000	1223	6,353
122	0,000	1296	6,720
130	0,000	1372	7,025
137	0,000	1454	7,345
145	0,000	1540	7,720
154	0,000	1631	8,165
163	0,000	1728	8,740
173	0,000	1830	9,370
183	0,000	1939	10,100
194	0,000	2054	10,900
205	0,055	2175	12,000
218	0,193	2304	13,080
230	0,330	2441	13,860
244	0,470	2585	13,985
259	0,520	2738	13,525
274	0,520	2901	12,810
290	0,490	3073	12,175
307	0,490	3255	11,850
325	0,550	3447	11,700
345	0,650	3652	11,625
365	0,790	3868	11,560
387	0,931	4097	11,460
410	1,055	4340	11,420
434	1,183	4597	11,375
460	1,313	4870	11,170
487	1,441	5158	10,700
516	1,560	5464	9,950
546	1,635	5788	9,070
579	1,720	6131	8,300
613	1,875	6494	7,700
649	2,180	6879	7,100
688	2,675	7286	6,610
729	3,222	7718	6,410
772	3,750	8175	6,655
818	4,140	8660	7,477
866	4,400	9173	8,680
917	4,620	9716	9,950

Annex B (informative):
Reference algorithm for echo control characteristics evaluation.

B.1 General

In this annex, a reference algorithm for evaluation of the echo control characteristics is described in pseudo code. The output of an implementation of the test method with the stimuli from the file 'echo_control_reference_files.zip' should equal the results presented in Table 3a and Table 3b. To run the verification, the additional file named 'p501-downlink_WB.pcm' in the pseudo code shall be created from the concatenated full band speech samples FB_female_conditioning_seq_long.wav and FB_male_female_single-talk_seq.wav from ITU-T Recommendation P.501, and processed with the following set of commands based on ITU-T Recommendation G.191:

```
filter -down HQ3 far_end_signal_48k.pcm far_end_signal_16k.pcm
filter P341 far_end_signal_16k.pcm p501-downlink_WB.pcm
```

Table 3a: Characterization of segment 1.

Category	Double talk		Single talk	
	Activity	Av. Level [dB]	Activity	Av. Level [dB]
A1	61,4%	-1,2	95,1%	0,1
A2	38,6%	-5,1	1,4%	-4,8
B	0,0%	0	0,0%	0
C	0,0%	0	0,0%	0
D	0,0%	0	0,0%	0
E	0,0%	0	0,3%	9,4
F	0,0%	0	3,2%	8,7
G	0,0%	0	0,0%	0

Table 3b: Characterization of segment 2.

Category	Double talk		Single talk	
	Activity	Av. Level [dB]	Activity	Av. Level [dB]
A1	51,6%	-1,2	93,8%	0,2
A2	39,7%	-7,1	0,3%	-6,1
B	1,1%	-16,9	0,0%	0
C	6,9%	-17,2	0,0%	0
D	0,0%	0	0,0%	0
E	0,0%	5,2	0,5%	9,5
F	0,7%	4	5,4%	6,2
G	0,0%	0	0,0%	0

The pseudo-code reference algorithm produces a text file output, and the implementation of the test method may be tested with the test script on the data in the file 'echo_control_reference_files.zip' for which the result shall equal

```
ms01-rec2; segm. 1; Processed signal;
active speech level [dBovl]; -45.8; RMS level [dBovl]; -51.5; speech activity; 0.269
ms01-rec2; segm. 1; Near end signal;
active speech level [dBovl]; -42.6; RMS level [dBovl]; -49.1; speech activity; 0.225
ms01-rec2; segm. 1; Downlink signal;
active speech level [dBovl]; -26.6; RMS level [dBovl]; -27.4; speech activity; 0.823
ms01-rec2; segm. 1; delay 0; DL delay 0;
DT activity 0.102; 0.614; 0.386; 0.000; 0.000; 0.000; 0.000; 0.000; 0.000;
ms01-rec2; segm. 1; delay 0; DL delay 0;
DT level diff; -1.2; -5.1; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0;
ms01-rec2; segm. 1; delay 0; DL delay 0;
ST activity 0.667; 0.951; 0.014; 0.000; 0.000; 0.000; 0.003; 0.032; 0.000;
ms01-rec2; segm. 1; delay 0; DL delay 0;
ST level diff; 0.1; -4.8; 0.0; 0.0; 0.0; 9.4; 8.7; 0.0;
```

```
ms01-rec2; segm. 2; Processed signal;
active speech level [dBovl]; -42.0; RMS level [dBovl]; -44.4; speech activity; 0.581
ms01-rec2; segm. 2; Near end signal;
active speech level [dBovl]; -40.6; RMS level [dBovl]; -42.7; speech activity; 0.625
ms01-rec2; segm. 2; Downlink signal;
active speech level [dBovl]; -26.5; RMS level [dBovl]; -27.2; speech activity; 0.841
ms01-rec2; segm. 2; delay -1; DL delay 0;
DT activity 0.359; 0.516; 0.397; 0.011; 0.069; 0.000; 0.000; 0.007; 0.000;
ms01-rec2; segm. 2; delay -1; DL delay 0;
DT level diff; -1.2; -7.1; -16.9; -17.2; 0.0; 5.2; 4.0; 0.0;
ms01-rec2; segm. 2; delay -1; DL delay 0;
ST activity 0.364; 0.938; 0.003; 0.000; 0.000; 0.000; 0.005; 0.054; 0.000;
ms01-rec2; segm. 2; delay -1; DL delay 0;
ST level diff; 0.2; -6.1; 0.0; 0.0; 0.0; 9.5; 6.2; 0.0;
```

B.2   Test script

```
%
% Set data format
%
fs = 16000;
conditioningTime = 23.5;
downlinkSystemDelay = 0;


%
% Segment the data
%
offsetDoubleTalk = conditioningTime;
offsetNearEnd    = conditioningTime;

segmentDoubleTalkIndex(1) = {[0, 20]};
segmentNearEndIndex(1) = {[0, 20]};

segmentDoubleTalkIndex(2) = {[20, 35]};
segmentNearEndIndex(2) = {[20, 35]};

lengthDoubleTalk = max(cell2mat(segmentDoubleTalkIndex(end)));
lengthNearEnd = max(cell2mat(segmentNearEndIndex(end)));

firstSampleDoubleTalk = round(fs*offsetDoubleTalk) + 1;
firstSampleNearEnd    = round(fs*offsetNearEnd) + 1;

lastSampleDoubleTalk = round(fs*(offsetDoubleTalk+lengthDoubleTalk));
lastSampleNearEnd    = round(fs*(offsetNearEnd+lengthNearEnd));

indexDoubleTalk = [firstSampleDoubleTalk, lastSampleDoubleTalk];
indexNearEnd    = [firstSampleNearEnd,    lastSampleNearEnd];

%
% Read data from file
%
fid = fopen('ms01_WB_rec2.pcm', 'r');
fseek(fid, 2*round(fs*offsetDoubleTalk), 'bof');
processedData = fread(fid, round(fs*lengthDoubleTalk), 'int16');
fclose(fid);

fid = fopen('ms01_WB_ref.pcm', 'r');
fseek(fid, 2*round(fs*offsetNearEnd), 'bof');
nearendData = fread(fid, round(fs*lengthNearEnd), 'int16');
fclose(fid);

fid = fopen('p501-downlink_WB.pcm', 'r');
fseek(fid, 2*round(fs*offsetDoubleTalk), 'bof');
downlinkData = fread(fid, round(fs*lengthDoubleTalk), 'int16');
fclose(fid);


%
% Evaluate
%
ecEvaluation(processedData, nearendData, downlinkData, ...
            segmentDoubleTalkIndex, segmentNearEndIndex, ...
            'ms01-rec2', downlinkSystemDelay, ...
            fs, 'bitExactTest.txt');
```

B.3 Reference algorithm

B.3.1 Main algorithm

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% processedData:          processed samples
% originalData:           near-end-only samples
% downlinkData:           down-link (loudspeaker) samples
% processedSegmentSet:    set of indices to processed data segments
% originalSegmentSet:     set of indices to original data segments
% PROC_FILE:              name shown in diagrams
% downlinkSystemDelayInMs: delay in DL signal from data to acoustic out
% sampleRate:             sampling frequency of the data
% resultsFile:            output file
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function ecEvaluation(...
  processedData, ...
  nearendData, ...
  downlinkData, ...
  indexProcessed, ...
  indexNearend, ...
  PROC_FILE, ...
  downlinkSystemDelayInMs, ...
  sampleRate, ...
  resultFile)


fid = fopen(resultFile, 'a');

% Define the categories
global D1 D2 D3 D4 L1 L2 L3;
D1 = 25;
D2 = 150;
D3 = 25;
D4 = 150;
L1 = 4;
L2 = -4;
L3 = -15;

global FRAME_LENGTH_MS ...
       MAX_DURATION_MS ...
       MAX_DURATION_FRAMES ...
       MAX_LEVEL_DIFFERENCE ...
       MIN_LEVEL_DIFFERENCE ...
       HISTOGRAM_RESOLUTION_MS

FRAME_LENGTH_MS = 5;
MAX_DURATION_MS = 200;
MAX_DURATION_FRAMES = MAX_DURATION_MS/FRAME_LENGTH_MS;
MAX_LEVEL_DIFFERENCE = 40;
MIN_LEVEL_DIFFERENCE = -40;
HISTOGRAM_RESOLUTION_MS = FRAME_LENGTH_MS;


% Main processing loop
frameLengthInSamples = FRAME_LENGTH_MS*sampleRate/1000; % 5ms frames

for segment = 1:length(indexProcessed)
  % Get the data samples for the segment
  segmentDataProcessed = cell2mat(indexProcessed(segment));
  segmentDataNearend = cell2mat(indexNearend(segment));

  index = (sampleRate*segmentDataProcessed(1)+1):sampleRate*segmentDataProcessed(2);
  x = processedData(index);
  z = downlinkData(index);
  index = (sampleRate*segmentDataNearend(1)+1):sampleRate*segmentDataNearend(2);
  y = nearendData(index);

  % Estimate and compensate for delay between processed and near end
  [x, y, z, delay] = compensateDelay(x, y, z, 0.5*sampleRate);
```

```
  % Compute the signal levels and classify the frames
  [Rx, Ry, Rz, doubleTalkFrames, singleTalkFrames] = ...
  computeSignalLevels(x, y, z, ...
                      sampleRate, frameLengthInSamples, ...
                      downlinkSystemDelayInMs, ...
                      PROC_FILE, segment, fid);

  % Evaluate double-talk performance
  numberOfDoubleTalkFrames = length(doubleTalkFrames);
  [H_dt, ld_ax_dt, dur_ax_dt] = ...
  levelTimeStatistics(Rx(doubleTalkFrames), Ry(doubleTalkFrames));

  [C_dt, L_dt] = evaluateHistogram(H_dt, ld_ax_dt, dur_ax_dt, ...
                                   numberOfDoubleTalkFrames);
  activityFactorDoubleTalk = numberOfDoubleTalkFrames/length(Rx);

  % Evaluate single-talk performance
  numberOfSingleTalkFrames = length(singleTalkFrames);
  [H_st, ld_ax_st, dur_ax_st] = ...
  levelTimeStatistics(Rx(singleTalkFrames), Ry(singleTalkFrames));

  [C_st, L_st] = evaluateHistogram(H_st, ld_ax_st, dur_ax_st, ...
                                   numberOfSingleTalkFrames);
  activityFactorSingleTalk = numberOfSingleTalkFrames/length(Rx);


  % Save to result file
  writeResultsToFile(fid, ...
                     PROC_FILE, ...
                     segment, ...
                     delay, ...
                     round(downlinkSystemDelayInMs), ...
                     activityFactorDoubleTalk, ...
                     activityFactorSingleTalk, ...
                     C_dt, ...
                     C_st, ...
                     L_dt, ...
                     L_st);
end

fclose(fid);
```

B.3.2　Delay compensation

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Compensate for delay in processed file
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [x, y, z, delay] = ...
compensateDelay(...
  x, ...
  y, ...
  z, ...
  maxLag)

ii = 1:min(1000000, length(x));

r = xcorr(x(ii), y(ii), maxLag);
[~, delay] = max(abs(r));
delay = delay-maxLag-1;

if (delay > 0)
  x = x((delay+1):end);
  z = z((delay+1):end);
  y = y(1:(end-delay));
elseif (delay < 0)
  y = y((-delay+1):end);
  x = x(1:(end+delay));
  z = z(1:(end+delay));
end;
```

B.3.3   Signal level computation and frame classification

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Determine speech activity and signal levels
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [Rx, Ry, Rz, doubleTalkFrames, singleTalkFrames] = ...
computeSignalLevels(x, y, z, ...
                    sampleRate, frameLengthInSamples, ...
                    downlinkSystemDelayInMs, ...
                    PROC_FILE, segment, fid)

LEVEL_METER_INIT_TIME_MS = 100;
DOWNLINK_HANGOVER_FRAMES = 40;
NEAREND_HANGOVER_FRAMES = 40;

levelMeterInitTime = LEVEL_METER_INIT_TIME_MS*sampleRate/1000;

% Level according to IEC61672
Rx = IEC61672(x, sampleRate, 12.5);
Ry = IEC61672(y, sampleRate, 12.5);
Rz = IEC61672(z, sampleRate, 12.5);

% Correct for system delay
nRz = length(Rz);
minRz = min(Rz(levelMeterInitTime:end));
Rz = [minRz*ones(floor(downlinkSystemDelayInMs*sampleRate/1000), 1); Rz];
Rz = Rz(1:nRz);

% Sub-sample and avoid initialization period of level meter
Rx = Rx(levelMeterInitTime:frameLengthInSamples:end);
Ry = Ry(levelMeterInitTime:frameLengthInSamples:end);
Rz = Rz(levelMeterInitTime:frameLengthInSamples:end);

% Active speech level according to P.56
[activeSpeechLevelProcessed, ...
 longTermLevelProcessed, ...
 activityFactorProcessed] = ...
speechLevelMeter(x, sampleRate);

[activeSpeechLevelNearend, ...
 longTermLevelNearend, ...
 activityFactorNearend] = ...
speechLevelMeter(y, sampleRate);

[activeSpeechLevelDownlink, ...
 longTermLevelDownlink, ...
 activityFactorDownlink] = ...
speechLevelMeter(z, sampleRate);

% Write active speech levels to file
writeSpeechLevelsToFile(PROC_FILE, segment, fid, ...
  activeSpeechLevelProcessed, ...
  activeSpeechLevelNearend, ...
  activeSpeechLevelDownlink, ...
  longTermLevelProcessed, ...
  longTermLevelNearend, ...
  longTermLevelDownlink, ...
  activityFactorProcessed, ...
  activityFactorNearend, ...
  activityFactorDownlink);

%
% Only evaluate for active downlink/near-end speech including hang-over
%
activeRyFrames = find(Ry > activeSpeechLevelNearend-15.9);
activeRzFrames = find(Rz > activeSpeechLevelDownlink-15.9);

% Downlink with added hangover
activeDownlinkSpeechFrames = zeros(size(Rz));
activeDownlinkSpeechFrames(activeRzFrames) = ones(size(activeRzFrames));

activeDownlinkSpeechFrames = conv(activeDownlinkSpeechFrames, ...
                                  ones(DOWNLINK_HANGOVER_FRAMES, 1));
```

```
activeDownlinkSpeechFrames = activeDownlinkSpeechFrames(1:length(Rz));

% Near-end
activeNearEndSpeechFrames = zeros(size(Ry));
activeNearEndSpeechFrames(activeRyFrames) = ones(size(activeRyFrames));
activeNearEndSpeechHtFrames = conv(activeNearEndSpeechFrames, ...
                                   ones(NEAREND_HANGOVER_FRAMES, 1));
activeNearEndSpeechHtFrames = activeNearEndSpeechHtFrames(1:length(Rz));

% Only evaluate double talk when both rx+hangover and near-end
doubleTalkSpeechFrames = (activeDownlinkSpeechFrames & ...
                          activeNearEndSpeechFrames);
doubleTalkFrames = find(doubleTalkSpeechFrames > 0);

% Single talk defined as rx and no near-end including 200 ms hangover
singleTalkSpeechFrames = (activeDownlinkSpeechFrames & ...
                          ~activeNearEndSpeechHtFrames);
singleTalkFrames = find(singleTalkSpeechFrames > 0);


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Average speech and noise levels
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [...
  activeSpeechLevel, ...
  longTermLevel, ...
  activityFactor ...
  ] = ...
speechLevelMeter(x, sampleRate)

SPEECH_LEVEL_HANGOVER_TIME_IN_MS = 200;

% Filter data
g = exp(-1/(0.03*sampleRate));
p = filter((1-g), [1, -g], abs(x));
q = filter((1-g), [1, -g], abs(p));

% Add 200ms hangover
hTimeInSamples = SPEECH_LEVEL_HANGOVER_TIME_IN_MS*sampleRate/1000;
qht = q;
for loop = 1:hTimeInSamples
  qht = max(qht, [zeros(loop, 1); q(1:end-loop)]);
end

% Compute cumulative histogram of signal power with hangover
nData = length(x);
cBins = 2.0.^(0:14)';
histogramCsum = zeros(size(cBins));

for loop = 1:length(cBins)
  histogramCsum(loop) = length(find(qht>cBins(loop)));
end

% Get the levels
sumSquare = sum(x.^2);
refdB = 20*log10(32768);

longTermLevel = 10*log10(sumSquare/nData) - refdB;
A = 10*log10(sumSquare./histogramCsum) - refdB;
C = 20*log10(cBins) - refdB;

Diff = A-C;
if ((A(1) == 0) || ((A(1) - C(1)) <= 15.9))
  activeSpeechLevel = -100;
else
  index = find(Diff <= 15.9, 1, 'first');

  if (Diff(index) == 15.9)
    activeSpeechLevel = A(index);
  else
    C_level = C(index) + ...
              (15.9 - Diff(index))* ...
              (C(index)-C(index-1))/(Diff(index)-Diff(index-1));
    activeSpeechLevel = A(index) + ...
```

```
                          (C_level - C(index))* ...
                          (A(index)-A(index-1))/(C(index)-C(index-1));
  end
end

activityFactor = 10.0^((longTermLevel-activeSpeechLevel)/10);


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Speech level meter according to IEC61672
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function Rx = IEC61672(x, sampleRate, tc)
%
%
% This functions computes the power of a sampled signal
% using a discrete filter with time constant equivalent to a first order
% continous time exponential averaging circuit,
%
%               1/tc
%      Rx = ----------  x^2
%            s + 1/tc
%
% according to IEC 61672 (1993, section 7.2).
%

T = 1/sampleRate;
tc = tc/1000;


%
% Design H by sampling of Hc
%
la = exp(-T/tc);
B = 1-la;
A = [1, -la];

Rx = filter(B, A, x.^2);


%
% Transform Rx to dBov (square wave),
%
% 0 dBov <=> power of maximum square wave signal, 32768
%
%         10^0 = 32768^2/X   => X = 32768^2
%
% Avoid log(0) by using log(max(eps, Rx))
%
Rx = 10*log10(max(eps, Rx)/32768/32768);
```

B.3.4   Level vs time computation

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Computation of level and time statistics
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [...
  levelVsDurationHistogram, ...
  levelDifferenceAxis, ...
  durationAxis] = ...
levelTimeStatistics(processedLevel, nearEndLevel)

global MAX_DURATION_FRAMES MAX_LEVEL_DIFFERENCE MIN_LEVEL_DIFFERENCE

FIRST_OCCURENCE = 1;


%
% Compute level difference
%
```

```
levelDifference = processedLevel - nearEndLevel;

%
% Only evaluate in integers (rounded towards 0) of dB and limit to max/min difference
%
levelDifference = fix(levelDifference);
levelDifference = min(levelDifference, MAX_LEVEL_DIFFERENCE);
levelDifference = max(levelDifference, MIN_LEVEL_DIFFERENCE);

%
% Produce axis
%
levelDifferenceAxis = MIN_LEVEL_DIFFERENCE:MAX_LEVEL_DIFFERENCE;
durationAxis = 1:(MAX_DURATION_FRAMES+1);

%
% Set initial values for computations and loop through all frames
%
numberOfEvaluatedFrames = length(levelDifference);

levelIncludedInEvaluation = (MAX_LEVEL_DIFFERENCE+1)*...
                            ones(numberOfEvaluatedFrames, 1);
levelAndRunLength = zeros(numberOfEvaluatedFrames, 4);
levelVsDurationHistogram = zeros(MAX_LEVEL_DIFFERENCE+ ...
                                (-MIN_LEVEL_DIFFERENCE)+1, ...
                                MAX_DURATION_FRAMES+1);

previousLevelDifference = 0;

for frame = 1:numberOfEvaluatedFrames-1;
  currentLevelDifference = levelDifference(frame);

  %
  % Evaluate all levels from the previous level up to the current level
  %
  if currentLevelDifference <= 0
    firstEvaluatedLevelDifference = max(min(0, previousLevelDifference), ...
                                        currentLevelDifference);
    step = -1;
  else
    firstEvaluatedLevelDifference = min(max(0, previousLevelDifference), ...
                                        currentLevelDifference);
    step = 1;
  end

  %
  % Loop the levels to be evaluated
  %
  for evaluatedLevelDifference = ...
        firstEvaluatedLevelDifference:step:currentLevelDifference
    %
    % Check that the current frame is not already included
    % in evaluation for earlier frames
    %
    if (evaluatedLevelDifference ~= levelIncludedInEvaluation(frame))
      if (evaluatedLevelDifference > 0)
        duration = find(levelDifference(frame+1:end) < ...
                        evaluatedLevelDifference, FIRST_OCCURENCE);
      else
        duration = find(levelDifference(frame+1:end) > ...
                        evaluatedLevelDifference, FIRST_OCCURENCE);
      end

      if (isempty(duration))
        duration = numberOfEvaluatedFrames-frame+1;
      end

      %
      % Set the frames during duration of the level difference
      % as being evaluated
      %
      if (duration > 1)
        levelIncludedInEvaluation(frame:(frame+duration-1)) = ...
            evaluatedLevelDifference*ones(duration, 1);
      end;

      %
      % Add the number of frames in the duration that have
```

```
      % absolute level diff greater or equal to evalutedLevel
      %
      durationIndex = min(duration, MAX_DURATION_FRAMES);
      levelIndex = evaluatedLevelDifference+(-MIN_LEVEL_DIFFERENCE)+1;
      levelVsDurationHistogram(levelIndex, durationIndex) = ...
         levelVsDurationHistogram(levelIndex, durationIndex) + duration;
    end
  end

  previousLevelDifference = currentLevelDifference;
end
```

B.3.5   Categorization

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Evaluate the histogram data
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [categories, averageLevelsInCategories] = ...
evaluateHistogram(...
  histogramData, ...
  levelDiff_ax, ...
  duration_ax, ...
  numberOfFrames)

global D1 D2 D3 D4 L1 L2 L3 HISTOGRAM_RESOLUTION_MS;

D1_scaled = D1/HISTOGRAM_RESOLUTION_MS;
D2_scaled = D2/HISTOGRAM_RESOLUTION_MS;
D3_scaled = D3/HISTOGRAM_RESOLUTION_MS;
D4_scaled = D4/HISTOGRAM_RESOLUTION_MS;

levelIndex_L1 = find(levelDiff_ax == L1);
levelIndex_L2 = levelDiff_ax == L2;
levelIndex_L3 = find(levelDiff_ax == L3);

duration_A2 = duration_ax;
duration_B = duration_ax<=D1_scaled;
duration_C = (D1_scaled<duration_ax)&(duration_ax<=D2_scaled);
duration_D = duration_ax>D2_scaled;
duration_E = duration_ax<=D3_scaled;
duration_F = (D3_scaled<duration_ax)&(duration_ax<=D4_scaled);
duration_G = duration_ax>D4_scaled;

framesInCategoryB = sum(histogramData(levelIndex_L3,  duration_B));
framesInCategoryC = sum(histogramData(levelIndex_L3,  duration_C));
framesInCategoryD = sum(histogramData(levelIndex_L3,  duration_D));
framesInCategoryE = sum(histogramData(levelIndex_L1,  duration_E));
framesInCategoryF = sum(histogramData(levelIndex_L1,  duration_F));
framesInCategoryG = sum(histogramData(levelIndex_L1,  duration_G));

framesInCategoryA2 = sum(histogramData(levelIndex_L2, duration_A2));
framesInCategoryA2 = framesInCategoryA2 - ...
                     framesInCategoryB - ...
                     framesInCategoryC - ...
                     framesInCategoryD;

framesInCategoryA1 = numberOfFrames - ...
                     framesInCategoryA2 - ...
                     framesInCategoryB - ...
                     framesInCategoryC - ...
                     framesInCategoryD - ...
                     framesInCategoryE - ...
                     framesInCategoryF - ...
                     framesInCategoryG;

categories = [framesInCategoryA1;
              framesInCategoryA2;
              framesInCategoryB;
              framesInCategoryC;
              framesInCategoryD;
              framesInCategoryE;
```

```
                framesInCategoryF;
                framesInCategoryG]/numberOfFrames;

averageLevelsInCategories = zeros(8, 1);

% Category A1
index = levelDiff_ax < L1;
index = levelDiff_ax(index) > L2;
weight = levelDiff_ax(index);

duration = duration_ax;
levelTimesDuration = (weight*histogramData(index, duration)).*duration;
nData = sum(histogramData(index, duration)*duration');
if (nData > 0)
  averageLevelsInCategories(1) = sum(levelTimesDuration)/nData;
end

% Category A2
index = levelDiff_ax <= L2;
index = levelDiff_ax(index) > L3;
weight = levelDiff_ax(index);

duration = duration_ax;
levelTimesDuration = (weight*histogramData(index, duration)).*duration;
nData = sum(histogramData(index, duration)*duration');
if (nData > 0)
  averageLevelsInCategories(2) = sum(levelTimesDuration)/nData;
end


% Category B, C, D
index = find(levelDiff_ax <= L3);
weight = levelDiff_ax(index);

duration = duration_ax(duration_B);
levelTimesDuration = (weight*histogramData(index, duration_B)).*duration;
nData = sum(histogramData(index, duration_B)*duration');
if (nData > 0)
  averageLevelsInCategories(3) = sum(levelTimesDuration)/nData;
end

duration = duration_ax(duration_C);
levelTimesDuration = (weight*histogramData(index, duration_C)).*duration;
nData = sum(histogramData(index, duration_C)*duration');
if (nData > 0)
  averageLevelsInCategories(4) = sum(levelTimesDuration)/nData;
end

duration = duration_ax(duration_D);
levelTimesDuration = (weight*histogramData(index, duration_D)).*duration;
nData = sum(histogramData(index, duration_D)*duration');
if (nData > 0)
  averageLevelsInCategories(5) = sum(levelTimesDuration)/nData;
end

% Category E, F, G
index = find(levelDiff_ax >= L1);
weight = levelDiff_ax(index);

duration = duration_ax(duration_E);
levelTimesDuration = (weight*histogramData(index, duration_E)).*duration;
nData = sum(histogramData(index, duration_E)*duration');
if (nData > 0)
  averageLevelsInCategories(6) = sum(levelTimesDuration)/nData;
end

duration = duration_ax(duration_F);
levelTimesDuration = (weight*histogramData(index, duration_F)).*duration;
nData = sum(histogramData(index, duration_F)*duration');
if (nData > 0)
  averageLevelsInCategories(7) = sum(levelTimesDuration)/nData;
end

duration = duration_ax(duration_G);
levelTimesDuration = (weight*histogramData(index, duration_G)).*duration;
nData = sum(histogramData(index, duration_G)*duration');
if (nData > 0)
  averageLevelsInCategories(8) = sum(levelTimesDuration)/nData;
```

```
end
```

B.3.6 Auxiliary functions for reporting data

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Write the classification to file
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function writeResultsToFile(fid, ...
                   PROC_FILE, ...
                   segment, ...
                   delay, ...
                   downlinkSystemDelay, ...
                   activityFactorDoubleTalk, ...
                   activityFactorSingleTalk, ...
                   C_dt, ...
                   C_st, ...
                   L_dt, ...
                   L_st)

str = sprintf('%s; segm. %d; delay %d; DL delay %d; DT activity %1.3f; %1.3f; %1.3f; %1.3f; %1.3f;
%1.3f; %1.3f; %1.3f; %1.3f;', ...
        PROC_FILE, segment, delay, downlinkSystemDelay, activityFactorDoubleTalk, ...
        C_dt(1), C_dt(2), C_dt(3), C_dt(4), ...
        C_dt(5), C_dt(6), C_dt(7), C_dt(8));
disp(str);
if (fid > -1)
  fprintf(fid, [str, '\n']);
end;

str = sprintf('%s; segm. %d; delay %d; DL delay %d; DT level diff; %1.1f; %1.1f; %1.1f; %1.1f;
%1.1f; %1.1f; %1.1f; %1.1f;', ...
        PROC_FILE, segment, delay, downlinkSystemDelay, ...
        L_dt(1), L_dt(2), L_dt(3), L_dt(4), L_dt(5), L_dt(6), L_dt(7), L_dt(8));
disp(str);
if (fid > -1)
  fprintf(fid, [str, '\n']);
end;

str = sprintf('%s; segm. %d; delay %d; DL delay %d; ST activity %1.3f; %1.3f; %1.3f; %1.3f;
%1.3f; %1.3f; %1.3f; %1.3f;', ...
        PROC_FILE, segment, delay, downlinkSystemDelay, activityFactorSingleTalk, ...
        C_st(1), C_st(2), C_st(3), C_st(4), ...
        C_st(5), C_st(6), C_st(7), C_st(8));
disp(str);
if (fid > -1)
  fprintf(fid, [str, '\n']);
end;

str = sprintf('%s; segm. %d; delay %d; DL delay %d; ST level diff; %1.1f; %1.1f; %1.1f; %1.1f;
%1.1f; %1.1f; %1.1f; %1.1f;', ...
        PROC_FILE, segment, delay, downlinkSystemDelay, ...
        L_st(1), L_st(2), L_st(3), L_st(4), L_st(5), L_st(6), L_st(7), L_st(8));
disp(str);
if (fid > -1)
  fprintf(fid, [str, '\n']);
end;



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Write the signal levels to file
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function writeSpeechLevelsToFile(PROC_FILE, segment, fid, ...
  activeSpeechLevelProcessed, ...
  activeSpeechLevelNearend, ...
  activeSpeechLevelDownlink, ...
  longTermLevelProcessed, ...
```

```
   longTermLevelNearend, ...
   longTermLevelDownlink, ...
   activityFactorProcessed, ...
   activityFactorNearend, ...
   activityFactorDownlink)

str = sprintf('%s; segm. %d; Processed signal; active speech level [dBovl]; %3.1f; RMS level
[dBovl]; %3.1f; speech activity; %1.3f', ...
   PROC_FILE, segment, activeSpeechLevelProcessed, ...
   longTermLevelProcessed, activityFactorProcessed);
disp(str);
if (fid > -1)
  fprintf(fid, [str, '\n']);
end;

str = sprintf('%s; segm. %d; Near end signal; active speech level [dBovl]; %3.1f; RMS level [dBovl];
%3.1f; speech activity; %1.3f', ...
   PROC_FILE, segment, activeSpeechLevelNearend, ...
   longTermLevelNearend, activityFactorNearend);
disp(str);
if (fid > -1)
  fprintf(fid, [str, '\n']);
end;

str = sprintf('%s; segm. %d; Downlink signal; active speech level [dBovl]; %3.1f; RMS level [dBovl];
%3.1f; speech activity; %1.3f', ...
   PROC_FILE, segment, activeSpeechLevelDownlink, ...
   longTermLevelDownlink, activityFactorDownlink);
disp(str);
if (fid > -1)
  fprintf(fid, [str, '\n']);
end;
```

Annex C (informative): Change history

TSG SA#	TSG doc	Spec	CR	Rev	Cat	Vers	New Vers	Subject
08							3.0.0	Approved
09	SP-000397	26.132	001		F	3.0.0	3.1.0	Handheld hands-free Test Setup
11	SP-010107	26.132	002	1	F	3.1.0	3.2.0	Harmonisation of test methods for acoustics between 3GPP and GSM
11		26.132				3.2.0	4.0.0	Release 4
11	SP-010107	26.132	003	1	B	4.0.0	5.0.0	Compatibility with testing wideband telephony transmission performance
13	SP-010454	26.132	004		B	5.0.0	5.1.0	Extended scope of test signals for Ambient Noise Rejection
13	SP-010454	26.132	006		F	5.0.0	5.1.0	Restricted scope of ITU-T P.501 test signals for 3G acoustic tests
13	SP-010454	26.132	008		A	5.0.0	5.1.0	Bandwidth of test signals for acoustic testing
15	SP-020080	26.132	011	1	A	5.1.0	5.2.0	Correction of references and editorial changes (wrong decimal separators)
16	SP-020435	26.132	016		F	5.2.0	5.3.0	Correction on ANR test for hands-free Ues
21	SP-030445	26.132	026		F	5.3.0	5.4.0	Loudness rating measurements at lower bit rates
25	SP-040649	26.132	028		C	5.4.0	6.0.0	Change of sending distortion test case
35	SP-070026	26.132	0030		F	6.0.0	6.1.0	Reference Update for ITU-T Recommendation P.57 'Artificial Ears'
35	SP-070026	26.132	0031	1	F	6.0.0	6.1.0	Update of reference [16] to P.79-2001 Annex G
35	SP-070026	26.132	0032	1	F	6.0.0	6.1.0	Distinction between narrow-band and wideband telephony in the frequency ranges and loudness rating and STMR weights, and in ANR calculation
36		26.132				6.1.0	7.0.0	Version for Release 7
38	SP-070759	26.132	0034	2	F	7.0.0	7.1.0	Changing the sidetone test to allow type 3.3 or 3.4 artificial ears
42	SP-080674	26.132	0035	1	F	7.1.0	7.2.0	Correction to allow wideband testing for GSM terminals
42	SP-080685	26.132	0037	3	C	7.2.0	8.0.0	Updated test methods for wideband terminal acoustics
43	SP-090015	26.132	0038		F	8.0.0	8.1.0	Clarification on Distortion with psophometric filter
43	SP-090018	26.132	0036	2	C	8.1.0	9.0.0	Speech and video telephony terminal acoustic test
45	SP-090568	26.132	0040	2	A	9.0.0	9.1.0	Correction of STMR calculation
45	SP-090573	26.132	0041	1	F	9.0.0	9.1.0	Handling Acoustic Testing with Noise Suppression Algorithms Employed
47	SP-100021	26.132	0042	1	F	9.1.0	9.2.0	Correction of distortion measurements
51	SP-110042	26.132	0043	5	B	9.2.0	10.0.0	Alignment of 3GPP Audio Test Case Specification
52	SP-110304	26.132	0045	3	C	10.0.0	10.1.0	Remaining modifications to EAAT WI
53	SP-110549	26.132	0047	1	F	10.1.0	10.2.0	Correction to Acoustic Echo Control
54	SP-110793	26.132	0048	1	F	10.2.0	10.3.0	Correction to Ambient Noise Rejection Test Procedure
54	SP-110793	26.132	0049		F	10.2.0	10.3.0	Clarification of Stability loss test signal
55	SP-120022	26.132	0050	1	F	10.3.0	10.4.0	Correction of receiving distortion
57	SP-120501	26.132	0053	2	F	10.4.0	10.5.0	Addition of 1/12 octave diffuse field table values with frequencies according to IEC 1260 (for information)
57	SP-120503	26.132	0052	3	B	10.5.0	11.0.0	Addition of UE delay test method
57	SP-120503	26.132	0054	1	B	10.5.0	11.0.0	Extension of Acoustic Tests
58	SP-120760	26.132	0055	1	F	11.0.0	11.1.0	Correction of references and levels

History

Document history		
V11.0.0	October 2012	Publication
V11.1.0	January 2013	Publication