ETSI TS 102 924 V1.2.1 (2018-03)

TECHNICAL SPECIFICATION

Speech and multimedia Transmission Quality (STQ); Transmission requirements for Super-Wideband / Fullband handset and headset terminals from a QoS perspective as perceived by the user

Reference

RTS/STQ-208-1

Keywords

QoS, terminal

ETSI

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

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

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

Contents

Intellectual Property Rights

Essential patents

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

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

Trademarks

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

Foreword

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

Modal verbs terminology

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

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

Introduction

Speech terminals are currently implementing narrowband and wideband bandwidth. Terminal equipment may offer wider bandwidth, due to features already available in these terminals. Such equipment may implement conversational features that may benefit of the electroacoustic equipment already available in the terminal and may provide wider quality for the end users.

The present document is intended to provide initial requirements and test methods for such type of equipment.

1 Scope

The present document provides speech & audio transmission performance requirements and measurement methods for handset and headset functions of super-wideband/fullband terminals. The present document provides requirements in order to optimize the end to end quality perceived by users.

Users become more sensitive to voice and music quality (for music used in conversational services) when using ICT/terminal equipment and so are more demanding for further enhancement especially further extension of the audio coded bandwidth.

For instance, this is the case for high quality conferencing services with music on hold, better background environment rendering and longer duration than normal point to point calls.

Standardized super-wideband and fullband coders are now available, some being also compatible with wideband coders.

The present document will consider only conversational services (that may be mixed with other services) and does not cover the streaming-only services.

Such applications include:

- Speech and audio communication including conferencing.
- Bandwidth extension which may allow usage for some mixed content.
- Super-wideband enhancement coupled with stereo/dichotic.

The send path it can be characterized in two ways:

- The signal picked up by microphone may combine speech, music and every type of environmental signal.
- Direct insertion of any type of signal.

For receive path, signal may be combine two types:

- Communication signals such as described for send path.
- Signal coming from distributed applications (e.g. advertisement, music on hold, etc.).

2 References

2.1 Normative references

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

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

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

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

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

[2] Recommendation ITU-T P.10/G.100: "Vocabulary for performance and quality of service".

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

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

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

[6] Recommendation G.711.1 (2008) Amendment 4 (11/10): "Wideband embedded extension for G.711 pulse code modulation".

[7] Recommendation ITU-T G.722.1 (annex C): "Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss".

[8] Recommendation G.729.1 (05/06): "G.729-based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729".

[9] Recommendation ITU-T G.718 (06/08)": "Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s".

[10] Recommendation ITU-T G.719: "Low-complexity, full-band audio coding for high-quality, conversational applications".

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

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

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

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

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

[16] IEC 61260-1: "Electroacoustics - Octave-band and fractional-octave-band filters - Part 1: Specifications".

[17] Void.

[18] Void.

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

[20] Void.

[21] Recommendation ITU-T G.711.1 (annex F): "Wideband embedded extension for G.711 pulse code modulation".

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

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

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

[25] ETSI TR 126 952: "Universal Mobile Telecommunications System (UMTS); LTE; Codec for Enhanced Voice Services (EVS); Performance characterization (3GPP TR 26.952 version 12.2.0 Release 12)".

[26] ETSI TS 126 441: "Universal Mobile Telecommunications System (UMTS); LTE; Codec for Enhanced Voice Services (EVS); General overview (3GPP TS 26.441)".

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

[28] ETSI TS 103 281: "Speech and multimedia Transmission Quality (STQ); Speech quality in the presence of background noise: Objective test methods for super-wideband and fullband terminals".

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

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

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

[32] Recommendation ITU-T P.863.1: "Application Guide for Recommendation ITU-T P.863".

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

[34] IETF RFC 3550: "RTP: A Transport Protocol for Real-Time Applications".

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

[36] IETF RFC 6716: "Definition of the Opus Audio Codec".

2.2 Informative references

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

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

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

[i.1] ISO 532: "Acoustics -- Method for calculating loudness level".

[i.2] NIST Net™.

NOTE: Available at https://www-x.antd.nist.gov/itg/nistnet/.

[i.3] Netem™.

NOTE: Available at http://www.linuxfoundation.org/en/Net:Netem.

[i.4] Trace Control for Netem (TCN) (2006): "Trace Control for Netem, Semester Thesis SA-2006-15", ETH Zürich, A. Keller.

[i.5] ETSI EG 202 425: "Speech Processing, Transmission and Quality Aspects (STQ); Definition and implementation of VoIP reference point".

[i.6] STQ(15)48-0309: "Objective Codec Evaluation of EVS. HEAD acoustics GmbH".

3 Definitions and abbreviations

3.1 Definitions

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

binaural listening: both ears are involved for the perception of sound

dichotic: relating to or involving the presentation of a stimulus to one ear that differs in some respect (as pitch, loudness, frequency, or energy) from a stimulus presented to the other ear

diotic: pertaining to or affecting both ears (same signal in both ears)

dual channel mode: audio mode, in which two audio channels with independent programme contents (e.g. bilingual) are encoded within one audio bit stream

fullband bandwidth: transmission of speech with a nominal bandwidth of 20 Hz - 20 kHz

stereo mode: audio mode in which two channels forming a stereo pair (left and right) are encoded within one bit stream and for which the coding process is the same as for the Dual channel mode

super-wideband: transmission with supre-wideband bandwith which may cover at least mono capabilities. Stereo capabilities may be possible

super-wideband bandwidth: transmission of speech with a nominal pass-band wider than 100 Hz to 7 000 Hz, usually understood to be 50 Hz - 14 000 Hz (definition from Recommendation ITU-T P.10 /G.100 [2])

3.2 Abbreviations

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

ACR	Absolute Category Rating
DRP	ear Drum Reference Point
ERP	Ear reference Point
EVS	Enhanced Voice Services
FB	FullBand
GAT	Group Audio Terminal
G-MOS-LQO$_F$	Overall Quality Mean Opinion Score, Listening Quality Objective, fullband
HATS	Head and Torso Simulator
MCU	Multiplexing Control Unit
MRP	Mouth Reference Point
MS	Mid-sized Stereo
N-MOS-LQO$_F$	Noise Quality Mean Opinion Score, Listening Quality Objective, fullband
POI	Point Of Interconnection
SLR	Send Loudness Rating
S-MOS-LQO$_F$	Speech Quality Mean Opinion Score, Listening Quality Objective, fullband
SWB	Super-WideBand
TCL	Terminal Echo Loss

4 Applications and coder considerations

4.1 Applications

The following applications are within the scope of the present document:

- Speech and audio communication including conferencing using high quality hands free systems, for which super-wideband/fullband coding can better reproduce the audio environment and provide improved quality and audio immersion. These applications cover also GATs (Group Audio Terminals) and teleconference systems such as "Telepresence".

- Bandwidth extension which may allow usage for some mixed content applications where wider bandwidth could bring a significant added value for the customer (support of 14 kHz and 20 kHz bandwidth and stereo/multichannel capability).

- Super-wideband enhancement coupled with stereo/multichannel to maximize the quality enhancement for the customer when the terminal device can support this capability.

The send path can be characterized in two ways:

- The signal picked up by microphone(s) may combine speech, music and every type of environmental signal.

NOTE: For some applications (e.g. journalist reporting) the user should have the possibility to cancel the noise environment or to transmit it without degradation.

- Direct insertion of any type of signal.

For receive path, signal may combine the two following types:

- Communication signal such as described for send path.
- Signal coming from distributed applications (e.g. advertisement, music on hold, etc.).

4.2 Coder considerations

4.2.0 Premise

As indicated in the scope only coders supporting conversational SWB and FB services are applicable to the present document.

4.2.1 Super-wideband (SWB)

Table 0: Use cases for coders

Coder Reference	Speech	Other signals	Stereo	Remark
VoLTE (IMS) ETSI TS 126 441 [26]	X	X Music	X	
Recommendation ITU-T G.722.1 [7] annex C	X	X Music		For low frame loss
Recommendation ITU-T G.729.1 [8] annex E (extension SWB)	X	X background noise (X) Music		
Recommendation ITU-T G.718 [9] annex B	X	X Music		
Recommendation ITU-T G.711.1 annexes D [6] and F [21]	X	X	X (annex F)	
Recommendation ITU-T G.722 [19] annexes B and D	X	X	X (annex D)	
OPUS [36]	X	X	X	
NOTE: G 722.1 [7] is intended to be used for hand-free application. It is referenced here considering that a terminal using this coder may implement a handset or headset function.				

When X is in brackets, it means that the coder is not optimized for this application.

The following coders are recommended for Super-wideband:

- Recommendation ITU-T G.722.1 [7] Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss. Annex C 14 kHz mode at 24, 32 and 48 kbit/s.
  The algorithm is recommended for use in hands-free applications such as conferencing where there is a low probability of frame loss. It may be used with speech or music inputs. The bit rate may be changed at any 20 ms frame boundary. New annex C contains the description of a low-complexity extension mode to G.722.1, which doubles the algorithm to permit 14-kHz audio bandwidth using a 32-kHz audio sample rate, at 24, 32, and 48 kbit/s.
  Annex C of [7]: this annex provides a description of the 14-kHz mode at 24, 32 and 48 kbit/s for this Recommendation.

- Recommendation ITU-T G.729.1 [8] annex E (extension SWB for G.729.1 [8]).
  This annex provides the high-level description of the higher bit-rate extension of G.729 designed to accommodate a wide range of input signals, such as speech, with background noise and even music.

- Recommendation ITU-T G.718 [9] annex B Super-wideband scalable (extension for Recommendation ITU-T G.718 [9]). This annex describes a scalable super-wideband (SWB, 50 to 14 000 Hz) speech and audio coding algorithm operating from 36 to 48 kbit/s and interoperable with Recommendation ITU-T G.718 [9].

- Recommendation ITU-T G.711.1 [6] annex D defines the super-wideband extension
  Annex F defines the Stereo embedded extension for Recommendation ITU-T G.711.1 [6].
  "Annex F is intended as a stereo extension to the G.711.1 [6] wideband coding algorithm and its super-wideband annex D. Compared to discrete two-channel (dual-mono) audio transmission, this stereo extension G.711.1 [6] annex F saves valuable bandwidth for stereo transmission. It is specified to offer the stereo capability while providing backward compatibility with the monaural core in an embedded scalable way. The annex provides very good quality for stereo speech contents (clean speech and noisy speech with various stereo sound pickup systems: binaural, MS, etc.), and for most of the conditions it provides significantly higher quality than low bitrate dual-mono. For some music contents, e.g. highly reverberated and/or with diffuse sound, the algorithm may have some performance limitations and may not perform as good as dual-mono codecs, however it achieves the quality of state-of-the-art parametric stereo codecs".

- Recommendation ITU-T G.722 [19] annex B defines the super-wideband extension
  and annex D defines the Stereo embedded extension for Recommendation ITU-T G.722 [19].
  "Annex B describes a scalable super-wideband (SWB, 50 to 14 000 Hz) speech and audio coding algorithm operating at 64, 80 and 96 kbit/s. The Recommendation ITU-T G.722 [19] super-wideband extension codec is interoperable with Recommendation ITU-T G.722 [19]. The output of the Recommendation ITU-T G.722 [19] SWB coder has a bandwidth of 50 to 14 000 Hz".
  "Annex D describes a stereo extension of the wideband codec G.722 and its super-wideband extension, G.722 annex B. It is optimized for the transmission of stereo signals with limited additional bitrate, while keeping full compatibility with both codecs. Annex D operates from 64 to 128 kbit/s with four super-wideband stereo bitrates at 80, 96, 112 and 128 kbit/s and two wideband stereo bitrates at 64 and 80 kbit/s".

- 3GPP VoLTE (IMS) ETSI TS 126 441 [26]. The Enhanced Voice Services coder consists of the multi-rate audio coder optimized for operation with voice and music/mixed content signals, a source controlled rate scheme including a voice/sound activity detector and a comfort noise generation system, and an error concealment mechanism to combat the effects of transmission errors and lost packets.

  Coder EVS (Enhanced Voice Services) is defined in ETSI TS 126 441 [26] and ETSI TR 126 952 [25]. The tests conducted on codec implementations, e.g. [i.6] show that the requirements and test methods for SWB terminals as defined in the present document apply.

  EVS is designed for packet-switched and circuit-switched networks/Mobile VoIP and VoLTE is a key target application.
  The key features of EVS are Super-wideband speech (32 kHz sampling) with improved speech quality and improved music performance.

4.2.2 Fullband (FB)

The following codecs are recommended for fullband:

- Recommendation ITU-T G.719 [10] Low-complexity, fullband audio coding for high-quality, conversational applications.
  "Recommendation ITU-T G.719 [10] describes the G.719 [10] coding algorithm for low-complexity fullband conversational speech and audio, operating from 32 kbit/s up to 128 kbit/s".

  The encoder input and decoder output are sampled at 48 kHz. The codec enables full bandwidth, from 20 Hz to 20 kHz, encoding of speech, music and general audio content. The codec operates on 20-ms frames and has an algorithmic delay of 40 ms.

NOTE: Amendment 1 adds new annex A that specifies the use of the ISO base media file format as container for the G.719 [10] bitstream addresses non-conversational use cases of the codec (e.g. call waiting music playback and recording of teleconferencing sessions, voice mail messages, online "jam"-sessions).

- 3GPP VoLTE (IMS) ETSI TS 126 441 [26]. The Enhanced Voice Services coder consists of the multi-rate audio coder optimized for operation with voice and music/mixed content signals, a source controlled rate scheme including a voice/sound activity detector and a comfort noise generation system, and an error concealment mechanism to combat the effects of transmission errors and lost packets.

Coder EVS (Enhanced Voice Services) is defined in ETSI TS 126 441 [26] and ETSI TR 126 952 [25]. The tests conducted on codec implementations, e.g. [i.6] show that the requirements and test methods for FB terminals as defined in the present document apply.

EVS is designed for packet-switched and circuit-switched networks/Mobile VoIP and VoLTE is a key target application. The key features of EVS are Fullband speech with improved speech quality and improved music performance.

5 Test considerations and test equipment

5.0 Introduction

The terminals within the scope of the present document are not only dedicated to speech communication but are also mixing speech and audio contents and may implement stereo and multichannel transmissions. As a consequence there is a need to define new parameters, such as:

- **Loudness:** Loudness Rating is determined only for speech or speech-like signals. Loudness may be calculated over any types of signals (audio sequences, speech sequences and mix of these sequences). Moreover it is not intended to define Loudness Rating algorithms for Super-wideband and fullband speech. To be consistent with transmission planning, the loudness rating shall be determined using wideband calculation and loudness shall be measurement for all the bandwidths. Clause 5.4.1.2 details the measurement principles.
- **Binaural listening:** The most of the test assessment methods and requirements for speech terminals are based on monaural listening, Even if some of them (e.g. for Handsfree Loudness rating) are intended to take into account binaural listening, the basic methods and requirements are only taking into account correction factors. The plan is to adapt test methods to effective binaural listening.

As a consequence, the present document takes into account test arrangements that are defined for speech terminals or for audio equipment.

Recommendation ITU-T P.58 [3] give information about use of HATS only from 100 Hz to 10 kHz, but new designs offer wider bandwidths.

For send the HATS can be used between 50 Hz and 16 kHz. Until the development of new systems with larger bandwidth, send measurement will be limited to those frequencies.

NOTE 1: With some measurement equipment the use of such of bandwidth is not possible and should be limited to 100 Hz to 14 kHz.

For receive, a correction factor (given, in annex B) allows measurement at DRP until 16 kHz.

NOTE 2: It is not the intention of the present document to define new requirements to adapt HATS for super-wideband and fullband. However when terminals implement Super-wideband or Fullband within terminals support also WideBand and/or NarrowBand speech, it is intended to use as far as possible test methods defined for wideband terminals and consequently to use HATS for parameters measured in wideband bandwidth.

5.1 IP half channel measurement adaptor

The IP half channel measurement adaptor is described in ETSI EG 202 425 [i.5].

5.2 Environmental conditions for tests

The following conditions shall apply for the testing environment:

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

b) relative humidity: 5 % to 85 %;

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

5.3 Accuracy of measurements and test signal generation

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

Table 1: Measurement accuracy

Item	Accuracy
Electrical signal level	±0,2 dB for levels ≥ -50 dBV ±0,4 dB for levels < -50 dBV
Sound pressure	±0,7 dB
Frequency	±0,2 %
Time	±0,2 %
Application force	±2 N
Measured maximum frequency	20 kHz

NOTE: The measured maximum frequency is due to Recommendation ITU-T P.58 limitations [3].

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

Table 2: Accuracy of test signal generation

Quantity	Accuracy
Sound pressure level	+2/-6 dB for frequencies from 50 Hz to100 Hz±2 dB for 100 Hz (see note 2) ±1 dB for frequencies from 200 Hz to 8 000 Hz ±3 dB for frequencies from 8 000 Hz to 16 000 Hz
Electrical excitation levels	±0,4 dB across the whole frequency range
Frequency generation	±2 %
Time	±0,2 %
Specified component values	±1 %
NOTE 1: This tolerance may be used to avoid measurements at critical frequencies, e.g. those due to sampling operations within the terminal under test. NOTE 2: The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) - logarithmic (Hz) scale.	

NOTE: With some measurement equipment the use of such of bandwidth is not possible and should be limited to 100 Hz to 14 kHz.

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

5.4 Network impairment simulation

At least one set of requirements is based on the assumption of an error free packet network, and at least one other set of requirements is based on a defined simulated malperformance of the packet network.

An appropriate network simulator has to be used, for example NIST Net™ [i.2] (https://www-x.antd.nist.gov/itg/nistnet/) or Netem [i.3].

Based on the positive experience STQ have made during the ETSI Speech Quality Test Events with "NIST Net™" this will be taken as a basis to express and describe the variations of packet network parameters for the appropriate tests.

Here is a brief blurb about NIST Net™:

- The NIST Net™ network emulator is a general-purpose tool for emulating performance dynamics in IP networks. The tool is designed to allow controlled, reproducible experiments with network performance sensitive/adaptive applications and control protocols in a simple laboratory setting. By operating at the IP level, NIST Net can emulate the critical end-to-end performance characteristics imposed by various wide area network situations (e.g. congestion loss) or by various underlying subnetwork technologies (e.g. asymmetric bandwidth situations of xDSL and cable modems).

- NIST Net™ is implemented as a kernel module extension to the Linux™ operating system and an X Window System-based user interface application. In use, the tool allows an inexpensive PC-based router to emulate numerous complex performance scenarios, including: tunable packet delay distributions, congestion and background loss, bandwidth limitation, and packet reordering/duplication. The X interface allows the user to select and monitor specific traffic streams passing through the router and to apply selected performance "effects" to the IP packets of the stream. In addition to the interactive interface, NIST Net™ can be driven by traces produced from measurements of actual network conditions. NIST Net also provides support for user defined packet handlers to be added to the system. Examples of the use of such packet handlers include: time stamping/data collection, interception and diversion of selected flows, generation of protocol responses from emulated clients.

The key points of Netem™ can be summarized as follows:

- Netem™ is nowadays part of most Linux™ distributions, it only has to be switched on, when compiling a kernel. With Netem, there are the same possibilities as with NIST Net™, there can be generated loss, duplication, delay and jitter (and the distribution can be chosen during runtime). Netem can be run on a Linux™-PC running as a bridge or a router (NIST Net™ only runs on routers).

- With an amendment of Netem™, Trace Control for Netem (TCN) [i.4] which was developed by ETH Zurich, it is even possible, to control the behaviour of single packets via a trace file. So it is for example possible to generate a single packet loss, or a specific delay pattern. This amendment is planned to be included in new Linux™ kernels, nowadays it is available as a patch to a specific kernel and to the iproute2 tool (iproute2 contains Netem™).

- It is not advised to define specific distortion patterns for testing in standards, because it will be easy to adapt devices to these patterns (as it is already done for test signals). But if a pattern is unknown to a manufacturer, the same pattern can be used by a test lab for different devices and gives comparable results. It is also possible to take a trace of NIST Net distortions, generate a file out of this and playback exact the same distortions with Netem.

NOTE: NIST Net™, Netem™, Linux™ and X Window System™ are examples of suitable products available commercially. This information is given for the convenience of users of the present document and does not constitute an endorsement by ETSI of these product(s).

5.5 Acoustic environment

NOTE: The acoustic environment may influence more significantly the results in low and high frequencies. It should be adapted to the terminal bandwidth.

Unless stated otherwise measurements shall be conducted under quiet and "anechoic" conditions. Depending on the distance of the transducers from mouth and ear a quiet office room may be sufficient e.g. for handsets where artificial mouth and artificial ear are located close to the acoustical transducers.

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

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

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

Unless specified otherwise, the background noise level shall be less than -64 dBPa(A) in conjunction with NC30 (ISO 3745 [24]).

For specified tests, it is desirable to have a background noise level of less than -74 dBPa(A) in conjunction with NC20, but the background noise level of -64 dBPa(A) in conjunction with NC30 shall never be exceeded.

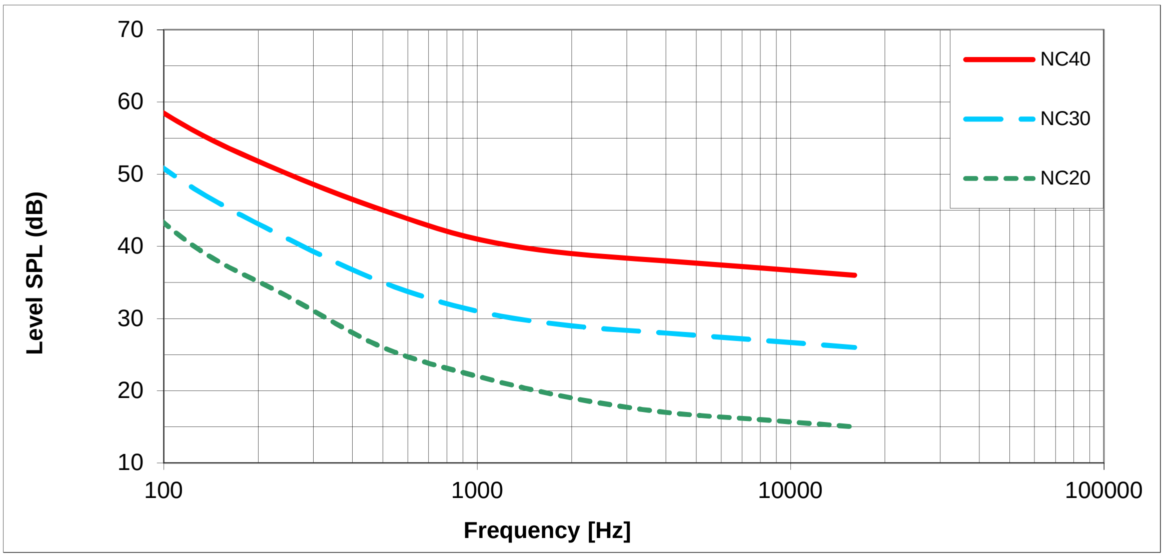


Figure 0

5.6 Verification of the environmental conditions

This test is not a mandatory test. This test is intended to be used in order to verify the environmental conditions as defined in the present document.

For the measurements no test signal is used.

A free-field measurement microphone is positioned in the test room.

The room noise is measured in the frequency range between 20 Hz and 20 kHz. The measurement duration is 5 seconds which is the averaging time for the idle channel noise.

The power density spectrum of the noise signal is determined using FFT (8 k samples/48 kHz sampling rate or equivalent). A Hanning window is used.

For checking the room noise level the measured spectrum is A-weighted.

For checking the NC-criteria the octave levels of the room noise are determined from 63 Hz to 16 kHz.

5.7 Influence of terminal delay on measurements

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

5.8 Specific test considerations

5.8.0 Premise

Even if the present document is dedicated to conversational services, the signals that are transmitted may combine speech and audio.

5.8.1 Loudness rating and Loudness

5.8.1.1 Loudness Rating

Loudness Rating, as defined in Recommendation ITU-T P.79 [5], applies for narrowband and wideband and is specific to telecommunications transmission systems. So, when terminals implement wideband speech or are intended to communicate with wideband or narrowband terminals, the terminals shall be calibrated for SLR and RLR values.

Due to the current bandwidth limitation of loudness rating's calculation it is not possible to calculate superwidband or fullband loudness ratings.

NOTE: For RLR and SLR, values are similar or derived from those defined in ETSI ES 202 739 [12] and ETSI TS 103 739 [13].

5.8.1.2 Loudness

Loudness quantifies the level perceived by the user and should be more relevant when the signal combines speech and audio sequences. ISO 532 [i.1] method B defines a standardized way to determine the loudness of a steady-state complex signal.

This assessment method takes into account the level, the spectrum of the signals and takes into account binaural listening. Loudness may be calculated for any type of signal (speech, music, noise) and mixed signals.

Standardized audio and speech signals (possibly based on combination of sequences) are defined in Recommendation ITU-T P.501 [1] and in ETSI TS 103 224 [11].

When the terminal provides super-wideband or fullband in addition with wideband or narrowband, the reference loudness value (expressed in phones) shall be determined for Narrowband or Wideband transmission.

The loudness measured in super-wideband or fullband should be equal and preferably higher than the loudness value measured for narrowband or wideband.

If the super-wideband and fullband terminal does not support wideband transmission, standardized loudness levels have to be defined (for further study).

5.8.2 Binaural listening

The scope of the present document includes terminals that may have two earpieces, distant sound pick-up using two or more microphones.

The terminal may also provide stereo listening or binaural rendering built from MCU.

So it should be relevant to consider binaural listening (for further study).

6 Requirements considerations and associated measurement Methodologies

6.1 Considerations

When possible, parameter requirements are derived from requirements defined for the wideband terminals. The recommended test method is also provided in the same clause as requirements.

NOTE 1: In general the test methods as described in the present document apply. If alternative methods exist they may be used if they have been proven to give the same result as the method described in the present document. This will be indicated in the test report.

NOTE 2: Due to the time variant nature of IP connections delay variation may impair the measurements. In such cases the measurement should be repeated until a valid measurement result is achieved.

6.2 Test setup

6.2.1 General

The preferred acoustical access to terminals is the most realistic simulation of the "average" subscriber. This can be made by using Head And Torso Simulator (HATS) with appropriate ear simulation and appropriate means to fix handset and headset terminals in a realistic and reproducible way to the HATS. HATS is described in Recommendation ITU-T P.58 [3], appropriate ears are described in Recommendation ITU-T P.57 [22] (type 3.3 and type 3.4 ear), a proper positioning of handsets under realistic conditions is to be found in Recommendation ITU-T P.64 [23].

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

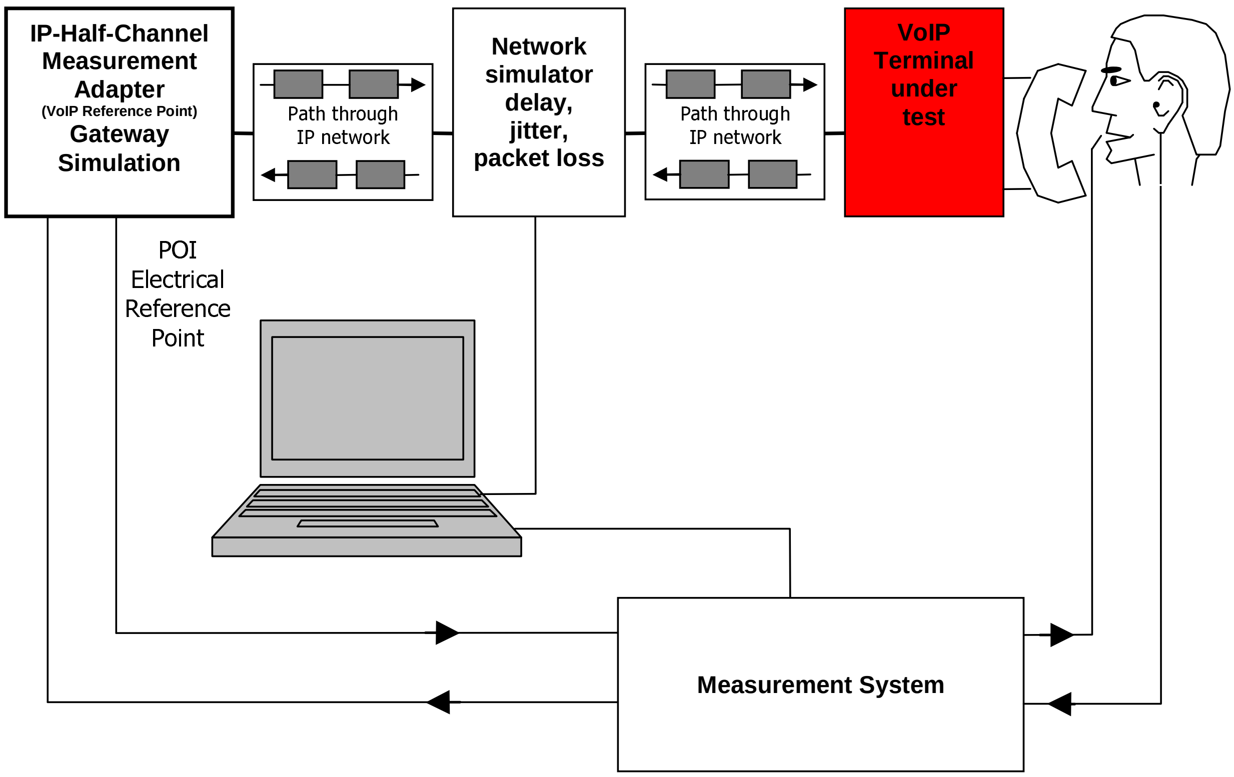


Figure 1: Half channel terminal measurement

6.2.2 Setup for handsets and headsets

When using a handset telephone the handset is placed in the HATS position as described in Recommendation ITU-T P.64 [23]. The artificial mouth shall be conform with Recommendation ITU-T P.58 [3]. The artificial ear shall conform with Recommendation ITU-T P.57 [22], type 3.3 or type 3.4 ears shall be used.

Recommendations for positioning headsets are given in Recommendation ITU-T P.380 [15]. If not stated otherwise headsets shall be placed in their recommended wearing position. Further information about setup and the use of HATS can be found in Recommendation ITU-T P.380 [15].

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

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

6.2.3 Position and calibration of HATS

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

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

The HATS shall be equipped with two type 3.3 or type 3.4 artificial ears. For binaural headsets two artificial ears are required. The type 3.3 or type 3.4 artificial ears as specified in Recommendation ITU-T P. 57 [22] shall be used. The artificial ear shall be positioned on HATS according to Recommendation ITU-T P. 58 [3].

The exact calibration and equalization can be found in Recommendation ITU-T P. 581 [4].

For calibration of mouth, equalization has to be limited between 50 Hz and 16 kHz.

With some measurement equipment the use of such of bandwidth is not possible and shall be limited to 100 Hz - 14 kHz.

If not stated otherwise, the HATS shall be diffuse-field equalized. The reverse nominal inverse field curve as found in table 3 of Recommendation ITU-T P.58 [3] shall be used.

NOTE: The inverse average diffuse field response characteristics of HATS as found in Recommendation ITU-T P.58 [3] is used and not the specific one corresponding to the HATS used. Instead of using the individual diffuse field correction, the average correction function is used because, for handset and headset measurements, mostly the artificial ear, ear canal and ear impedance simulations are effective. The individual diffuse-field correction function of HATS includes all diffraction and reflection effects of the complete individual HATS which are not effective in the measurement and potentially would lead to bigger measurement uncertainties than using the average correction.

6.2.4 Test signal and test signal levels

The test signals are defined according to Recommendation ITU-T P.501 [1].

As the bandwidth of the speech signals defined in Recommendation ITU-T P.501 [1] is fullband, these test signals shall be used in the present document:

- The test signal to be used for measurements such as Frequency response, Loudness Rating, shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [1].
- The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [1], shall be used as activation signal for measurements such as distortion, send noise.
- The compressed real speech signal described in clause 7.3.3 of Recommendation ITU-T P.501 [1], shall be used for measurements such as TCLw.

For double-talk performance:

A "double-talk" sequence representing typical double talk is provided in Recommendation ITU-T P.501 [1]. This uses the single-talk sequence described in section 7.3.1 of Recommendation ITU-T P.501 [1], shown in the lower pane, as the main speech and an additional competing speaker sequence, shown in the upper pane.

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

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

6.2.5 Setup of background noise simulation

A setup for simulating realistic background noises in a lab-type environment is described in ETSI TS 103 224 [11].

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

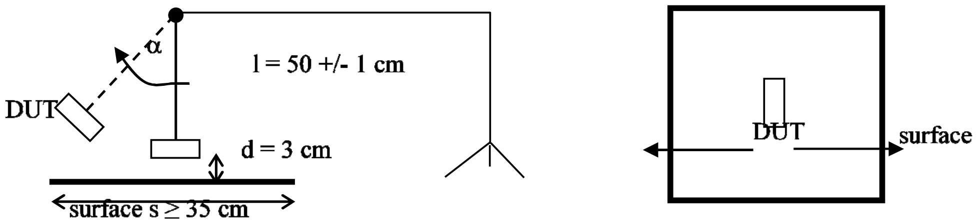
The following noises of ETSI TS 103 224 [11] shall be used:

Table 2a

Pub Noise (Pub)	HATS and microphone array in a pub	30 s	1: 77,2 dB 2: 76,6 dB 3: 75,7 dB 4: 76,0 dB 5: 76,0 dB 6: 76,3 dB 7: 76,0 dB 8: 76,4 dB
Sales Counter (SalesCounter)	HATS and microphone array in a supermarket	30 s	1: 66,6 dB 2: 66,1 dB 3: 65,7 dB 4: 66,5 dB 5: 66,3 dB 6: 66,8 dB 7: 66,6 dB 8: 67,1 dB
Callcenter 2 (Callcenter)	HATS and microphone array in business office	30 s	1: 60,2 dB 2: 60,0 dB 3: 60,1 dB 4: 60,8 dB 5: 60,2 dB 6: 60,6 dB 7: 60,2 dB 8: 60,7 dB

6.2.6 Setup of variable echo path

The handset is positioned d = 3 cm above a horizontal hard surface, facing the surface with speaker and microphone. The surface shall be at least 35 cm × 35 cm. The handset is fixed like a pendulum with a non-elastic cord 3 cm above the centre of the horizontal surface, see figure 2. The pivot is 55 ± 1 cm above the hard plate.



side view **top view**

Figure 2: Positioning of handset under test

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

For headsets the same measurement arrangement is used as described above. However, it has to be assured that the echo path (audio path between speaker output and microphone input) changes significantly. If the pendular motion across the base surface is not producing a sufficient change in echo path, another hard surface perpendicular to the base surface can be added. The dimension and position of the additional surface should be chosen such that it is positioned within the echo path when crossed by the pendulous headset but not within the echo path when the headset reaches the turning point of the pendulous motion. At the lowest point of pendular motion, the headset speaker and microphone should not exceed a distance of 3 cm from either of the surfaces.

NOTE: Depending on the geometry of the headset (monaural / binaural, microphone integrated into earpiece / earplug with microphone on short arm / microphone on long arm) a stable pendular motion has to be established. This may require two cords fixed with respect to the headset's balance point in order to avoid tumbling motion. Alternatively, the headset may be attached to a fixed radial arm to achieve a stable pendular motion.

Figure 3 shows an exemplary setup for a binaural headset with long microphone arm and vertical surface to increase echo path variation by changing the coupling between speaker and microphone during pendular motion. During one pendular period, the DUT is exposed to four sudden changes in echo path when passing the vertical surface.

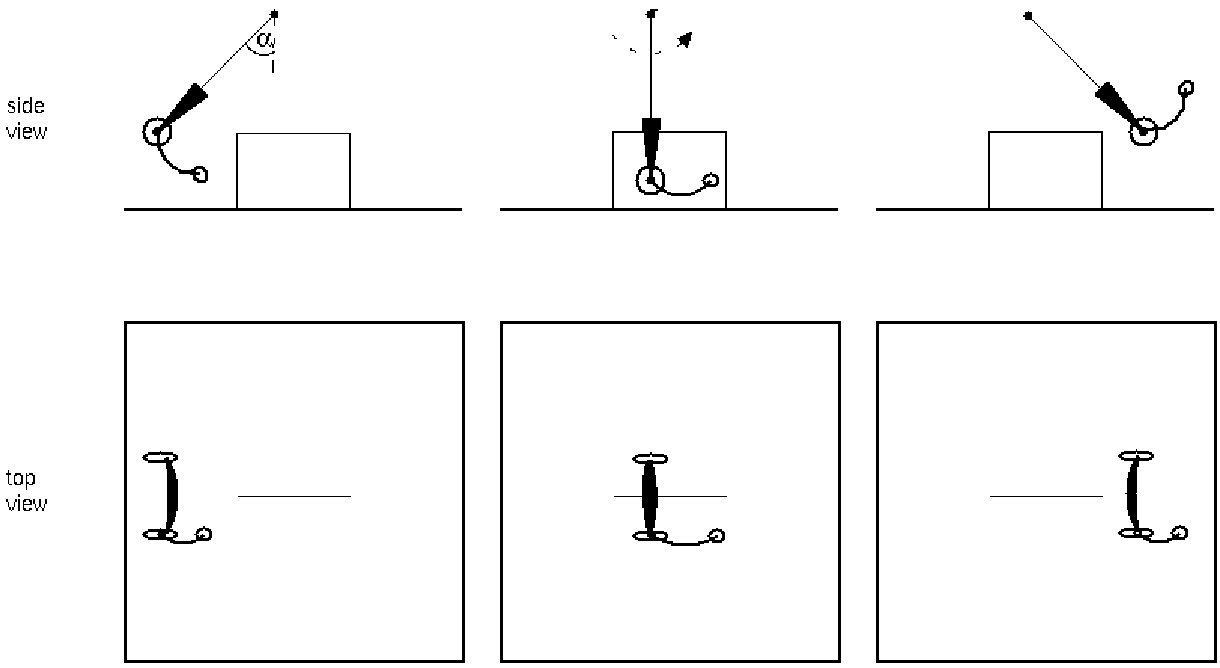


Figure 3: Example for positioning of a headset under test

6.3 Coding independent parameters

6.3.1 Send frequency response

Requirement

The send frequency response of the handset or headset shall be within a mask as defined in table 3 for SWB and table 4 for FB, and shown in figure 4 for SWB and figure 5 for FB. This mask shall be applicable for all types of handsets and headsets.

Table 3: Super-wideband send frequency response limits

Frequency	Upper Limit	Lower Limit
50 Hz	**0 dB**	
100 Hz	5 dB	-5 dB
12 500 Hz	5 dB	-5 dB
14 000 Hz	5 dB	-10 dB
NOTE: The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. The requirement is based on 1/12th octave measurement.		

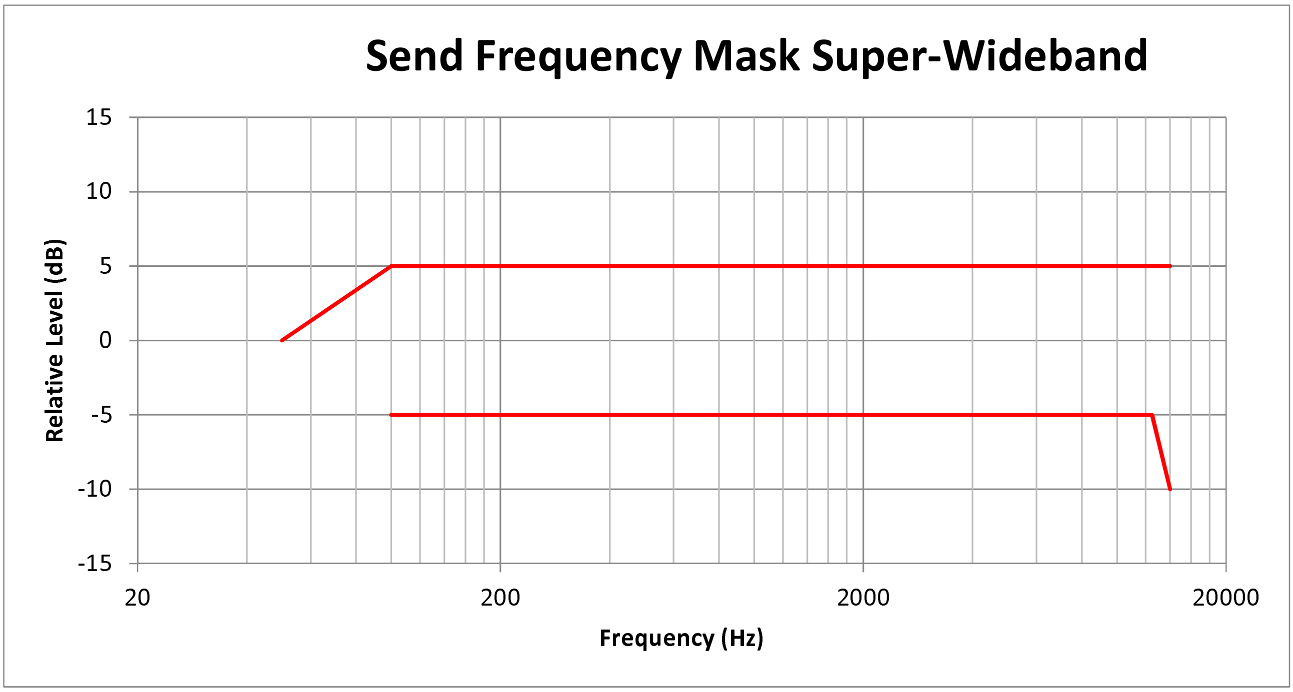


Figure 4: Send frequency response mask for super-wideband

Table 4: Fullband send frequency response limits

Frequency	Upper Limit	Lower Limit
20 Hz	**0 dB**	
50 Hz	**0 dB**	**-10 dB**
100 Hz	5 dB	-5 dB
12 500 Hz	5 dB	-5 dB
16 000 Hz	5 dB	-5 dB
20 000Hz	5 dB	
NOTE: The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) - logarithmic (Hz) scale. The requirement is based on 1/12th octave measurement.		

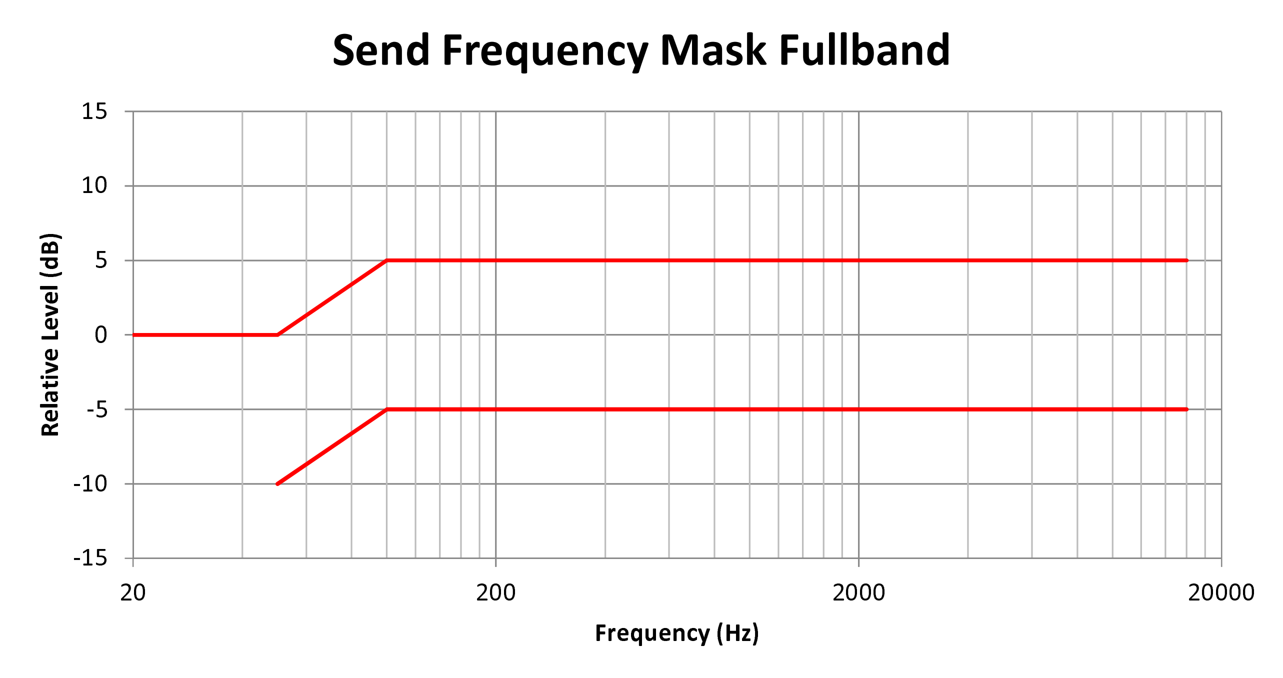


Figure 5: Send frequency response mask for Fullband

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

NOTE 2: A "balanced" frequency response is preferable from the perception point of view. If frequency components in the low frequency domain are attenuated in a similar way frequency components in the high frequency domain should be attenuated.

Measurement Method

The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [1]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free field conditions at the MRP. The test signal level shall be -4,7 dBPa, measured at the MRP. The test signal level is averaged over the complete test signal sequence.

The handset or headset terminal is setup as described in clause 6.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [23]). The application force used to apply the handset against the artificial ear is noted in the test report.

In case of headset measurements the tests are repeated 5 times, in conformance with Recommendation ITU-T P.380 [15]. The results are averaged (averaged value in dB, for each frequency).

For super-wideband terminals, measurements shall be made at one twelfth-octave intervals as given by the IEC 61260-1 [16] for frequencies from 50 Hz to 14 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 at the MRP.

For fullband terminals, measurements shall be made at one twelfth-octave intervals as given by the IEC 61260-1 [16] for frequencies from 20 Hz to 20 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 at the MRP.

The sensitivity is expressed in terms of dBV/Pa.

6.3.2 Send Loudness Rating (SLR)

Requirement:

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

- SLR(set) = 8 dB ± 3 dB

Measurement Method:

The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [1] shall be used. The spectrum of acoustic signal produced by the artificial mouth is calibrated under free field conditions at the MRP. The test signal level shall be -4,7 dBPa, measured at the MRP. The test signal level is averaged over the complete test signal sequence.

The handset or headset terminal is setup as described in clause 6.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [23]). The application force used to apply the handset against the artificial ear is noted in the test report.

In case of headset measurements the tests are repeated 5 times, in conformance with Recommendation ITU-T P.380 [15]. The results are averaged (averaged value in dB, for each frequency).

The sending sensitivity shall be calculated from each band of the 20 frequencies given in table 1 of Recommendation ITU-T P.79 [5], bands 1 to 20. For the calculation the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.

The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to Recommendation ITU-T P.79 [5] see annex A.

6.3.3 Mic mute

Requirement:

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

Measurement Method:

The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [1] shall be used. The spectrum of acoustic signal produced by the artificial mouth is calibrated under free field conditions at the MRP. The test signal level shall be -4,7 dBPa, measured at the MRP. The test signal level is averaged over the complete test signal sequence.

The handset or headset terminal is setup as described in clause 6.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [23]). The application force used to apply the handset against the artificial ear is noted in the test report.

In case of headset measurements the tests are repeated 5 times, in conformance with Recommendation ITU-T P.380 [15]. The results are averaged (averaged value in dB, for each frequency).

The sending sensitivity shall be calculated from each band of the 20 frequencies given in table 1 of Recommendation ITU-T P.79 [5], bands 1 to 20. For the calculation the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.

The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to Recommendation ITU-T P.79 [5] see annex A.

6.3.4 Linearity range of SLR

Requirement:

The sensitivity determined with input sound pressure levels between -24,7 dBPa and 5,3 dBPa shall not differ by more than ±2 dB from the sensitivity determined with an input sound pressure level of -4,7 dBPa. For the input sound pressure level of 5,3 dBPa a limit of +4 dB to -2 dB applies.

The sensitivity determined with input sound pressure levels between -24,7 dBPa and 5,3 dBPa shall not differ by more than ±2 dB from the sensitivity determined with an input sound pressure level of -4,7 dBPa. For the input sound pressure level of 5,3 dBPa a limit of +4 dB to -2 dB applies.

Table 5

Linearity range of SLR: ΔSLR = SLR - SLR@-4,7 dBPa			
Input Level	**Target ΔSLR**	**Upper limit**	**Lower limit**
-24,7 dBPa	0	2 dB	-2 dB
-19,7 dBPa	0	2 dB	-2 dB
-14,7 dBPa	0	2 dB	-2 dB
-9,7 dBPa	0	2 dB	-2 dB
-4,9 dBPa	0	2 dB	-2 dB
-4,7 dBPa	0	0 dB	0 dB
-4,5 dBPa	0	2 dB	-2 dB
0,3 dBPa	0	2 dB	-2 dB
5,3 dBPa	0	4 dB	-4 dB
NOTE: It is assumed that the variation of gain is mostly codec independent. In case codec specific requirements are needed, they are found in clause 6.4.			

Measurement Method:

The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [1] shall be used. The spectrum of acoustic signal produced by the artificial mouth is calibrated under free field conditions at the MRP. The test signal level shall be -4,7 dBPa, measured at the MRP. The test signal level is averaged over the complete test signal sequence.

The handset terminal is setup as described in clause 6.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [23]). The application force used to apply the handset against the artificial ear is noted in the test report.

The sending sensitivity shall be calculated from each band of the 20 frequencies given in table 1 of Recommendation ITU-T P.79 [5], bands 1 to 20. For the calculation the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.

The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to Recommendation ITU-T P.79 [5] see annex A.

6.3.5 Send Distortion

6.3.5.1 Signal to harmonic distortion versus frequency

Requirement:

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

Table 6: Send distortion for super-wideband

Frequency	Ratio
100 Hz	24 dB
200 Hz	26 dB
400 Hz	30 dB
1 kHz	30 dB
2 kHz	30 dB
3,15 kHz	30 dB
5 kHz	30 dB
NOTE: Limits at intermediate frequencies lie on a straight line drawn between the given values on a linear (dB ratio) - logarithmic (frequency) scale.	

Table 7: Send distortion for fullband

Frequency	Ratio
100 Hz	26 dB
200 Hz	30 dB
400 Hz	30 dB
1 kHz	30 dB
2 kHz	30 dB
3,15 kHz	30 dB
8 kHz	30 dB
NOTE: Limits at intermediate frequencies lie on a straight line drawn between the given values on a linear (dB ratio) - logarithmic (frequency) scale.	

Measurement method:

The handset or headset terminal is set-up as described in clause 6.2.

The signal used is an activation signal followed by a sine-wave signal with a frequency at 100 Hz, 315 Hz, 400 Hz, 500 Hz, 630 Hz, 800 Hz, 1 000 Hz, 2 000 Hz, 3 150 Hz and 7 000 Hz for super-wideband and 100 Hz, 315 Hz, 400 Hz, 500 Hz, 630 Hz, 800 Hz, 1 000 Hz, 2 000 Hz, 3 150 Hz and 8 000 Hz for fullband. The duration of the sine wave shall be less than 1 s. The sinusoidal signal level shall be calibrated to -4,7 dBPa at the MRP.

The signal to harmonic distortion ratio is measured selectively up to 14 kHz for super-wideband and 20 kHz for fullband.

The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [1] shall be used for activation. The level of this activation signal is -4,7 dBPa at the MRP.

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

6.3.5.2 Signal to harmonic distortion for higher input level

Requirement:

For the signal defined in the measurement method, the signal to harmonic distortion ratio shall be ≥ 30 dB.

Measurement method:

The handset or headset terminal is set-up as described in clause 6.2.

The signal used is an activation signal followed by a 1 kHz sine wave. The signal to harmonic distortion ratio is measured selectively up to 14 kHz for Super-wideband terminals and up to 20 kHz for fullband terminal.

The duration of the sine wave shall be ≤ 1 s. The sinusoidal signal level shall be calibrated to +10 dBPa at the MRP.

The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [1] shall be used for activation. The level of this activation signal is -4,7 dBPa at the MRP.

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

6.3.6 Send Noise

Requirement:

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

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

Measurement Method:

For the actual measurement no test signal is used. In order to reliably activate the terminal an activation signal is introduced before the actual measurement.

The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [1] shall be used for activation. The level of this activation signal is -4,7 dBPa at the MRP.

The handset or headset terminal is set-up as described in clause 6.2.

The send noise is measured at the POI in the frequency range from 50 Hz to 14 kHz for Super-wideband and 20 Hz to 20 kHz for Fullband. The analysis window is applied directly after stopping the activation signal but taking into account the influence of all acoustical components (reverberations). The averaging time is 1 s. The test house has to ensure (e.g. by monitoring the time signal) that during the test the terminal remains in activated condition. If the terminal is deactivated during the measurement, the measurement time has to be reduced to the period where the terminal remains in activated condition.

The noise level is measured in dBm0(A).

Spectral peaks are measured in the frequency domain in the frequency range from 100 Hz to 14 kHz for super-wideband and from 50 Hz to 16 kHz in fullband. The frequency spectrum of the idle channel noise is measured by a spectral analysis having a noise bandwidth of 8,79 Hz (determined using FFT 8 k samples/48 kHz sampling rate with Hanning window or equivalent). The idle channel noise spectrum is stated in dB. A smoothed average idle channel noise spectrum is calculated by a moving average (arithmetic mean) 1/3rd octave wide across the idle noise channel spectrum stated in dB (linear average in dB of all FFT bins in the range from 2^(-1/6)f to 2^(+1/6)f). Peaks in the idle channel noise spectrum are compared against a smoothed average idle channel noise spectrum.

NOTE: In case spectral peaks higher than 10 dB above the average noise floor are produced by the terminal, but which are considered to be inaudible due to the very low noise floor produced by the terminal on average, it is the responsibility of the test lab to demonstrate the desired performance subjectively.

6.3.7 Sidetone Masking Rating STMR (Mouth to ear)

Requirement:

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

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

NOTE 1: It is preferable to have a constant STMR independent of the volume control setting.

NOTE 2: STMR measurement in Recommendation ITU-T P.79 [5] is not defined above 8 kHz, but sidetone signal is not supposed to have such limitation.

Measurement Method:

The test signal is defined in clause 6.2. The test signal level shall be -4,7 dBPa, measured at the MRP. The handset or headset terminal is setup as described in clause 6.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [23]) and the application force shall be 13 N on the artificial ear type 3.3 or type 3.4.

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

Measurements shall be made at one twelfth-octave intervals as given by the IEC 61260-1 [16] for frequencies from 100 Hz to 8 kHz inclusive. For the calculation the averaged measured level at each frequency band (Recommendation ITU-T P.79 [5] table 3, bands 1 to 20) is referred to the averaged test signal level measured in each frequency band.

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

6.3.8 Sidetone delay

Requirement:

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

Measurement Method:

The handset or headset terminal is setup as described in clause 6.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [23]).

The test signal is a CS-signal complying with Recommendation ITU-T P.501 [1] using a PN sequence with a length of 4 096 points (for the 48 kHz sampling rate) which equals to the period T. The duration of the complete test signal is as specified in Recommendation ITU-T P.501 [1].

The level of the signal shall be -4,7 dBPa at the MRP.

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

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

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

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

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

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

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

6.3.9 Terminal Coupling Loss (TCL)

Requirement:

The TCL measured as unweighted Echo Loss shall be ≥ 46 dB for all positions of the volume control (if supplied).

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

Measurement method:

The handset or headset terminal is setup as described in clause 6.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [23]) and the application force shall be 2 N on the artificial ear type 3.3 or type 3.4 as specified in Recommendation ITU-T P.57 [22]. The ambient noise level shall be < -64 dBPa(A). The attenuation from electrical reference point input to electrical reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of Recommendation ITU-T P.501 [1].

TCL is calculated as unweighted echo loss from 100 Hz to 8 kHz. For the calculation the averaged measured echo level at each frequency band is referred to the averaged test signal level measured in each frequency band. The first 17,0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences).

For the measurement, a time window has to be applied which is adapted to the duration of the actual test signal. The echo loss is calculated by the equations:

$$L_e = C - 10\log_{10}\sum_{i=1}^{N}(A_i + A_{i-1})(\log_{10} f_i - \log_{10} f_{i-1}) \qquad (4)$$

and

$$C = 10\log_{10}(2\,(\log_{10} f_N - \log_{10} f_0) \qquad (5)$$

where:

- A_0 is the output/input power ratio at frequency f_0 = 100 Hz;
- A_1 the ratio at frequency f_i; and
- A_N the ratio at frequency f_N = 8 000 Hz.

Equation (2) is a generalized form of the equation defined in Recommendation ITU-T G.122 [29], clause B.4, for calculating echo loss based on tabulated data, which allows the calculation of echo loss within any frequency range between f_0 and f_N.

6.3.10 Stability loss

Requirement

With the handset lying on and the transducers facing a hard surface, the attenuation from the digital input to the digital output shall be at least 6 dB at all frequencies in the range of 50 Hz to 16 kHz for super-wideband and 20 to 20 kHz for Fullband. In case of headsets the requirement applies for the closest possible position between microphone and headset receiver.

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

Measurement method:

Before the actual test a training sequence of the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [1] is applied. The training sequence level shall be -16 dBm0 in order not to overload the codec.

The test signal is a PN sequence complying with Recommendation ITU-T P.501 [1] with a length of 4 096 points (for the 48 kHz sampling rate) and a crest factor of 6 dB. The duration of the test signal is 250 ms. With an input signal of -3 dBm0, the attenuation from digital input to digital output shall be measured for frequencies from 100 Hz to 8 kHz under the following conditions:

a) the headset, with the transmission circuit fully active, shall be positioned on one inside surface that is of three perpendicular planes, smooth, hard surfaces forming a corner. Each surface shall extend 0,5 m from the apex of the corner. One surface shall be marked with a diagonal line, extending from the corner formed by the three surfaces, and a reference position 250 mm from the corner, as shown in figure 4;

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

  1) the mouthpiece and ear cup shall face towards the surface;

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

3) the extremity of the handset shall coincide with the normal to the reference point, as shown in figure 6;

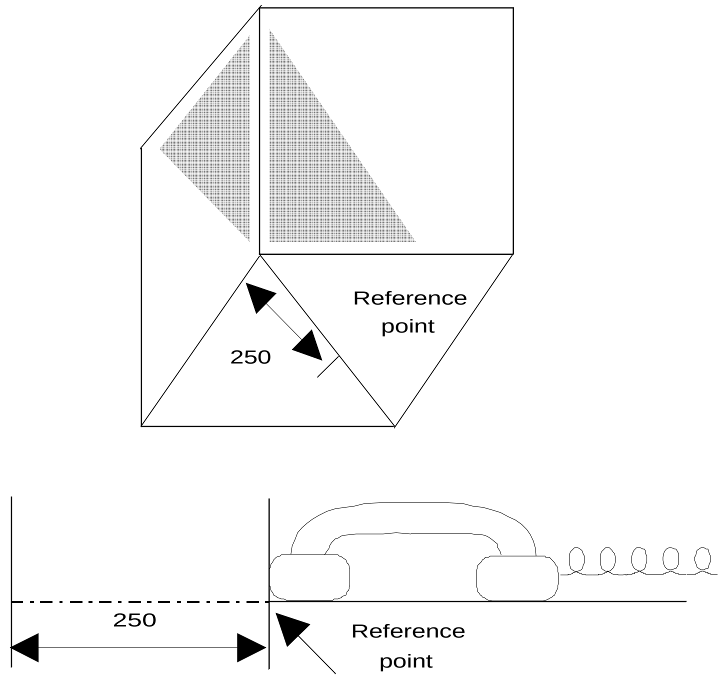
b2) the headset, with the transmission circuit fully active, shall be positioned on the defined surface as follows:

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

2) the headset receiver shall be placed centrally at the reference point as shown in figure 6;

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

4)



NOTE: All dimensions in mm.

Figure 6

6.4 Receive parameters

6.4.1 Equalization

This type of terminal may be used for reproduction of signals other than pure speech (e.g. music) for which user's preference may be different in term of sound signature.

So the terminal may implement an equalization function adjusting frequency response according to user's preferences.

When such a function is available the requirements in receive as defined in the present document shall be met for at least the default equalization setting.

6.4.2 Receive Frequency response

Requirement:

The receive frequency response of the handset or the headset shall be within a mask as defined in table 8 for SWB and table 9 for FB, and shown in figure 7 for SWB and figure 8 for FB.

Table 8: Super-wideband receive frequency response limits

Frequency	Upper Limit	Lower Limit
50 Hz	3 dB	-5 dB
400 Hz	3 dB	-5 dB
1010 Hz	(see note)	-5 dB
1 200 Hz	(see note)	-8 dB
1 500 Hz	(see note)	-8 dB
2 000 Hz	9 dB	-3 dB
3 200 Hz	9 dB	-3 dB
14 000 Hz	9 dB	-13 dB
NOTE: The limit curves shall be determined by straight lines joining successive co-ordinates given in the table, where frequency response is plotted on a linear dB scale against frequency on a logarithmic scale. It is a floating or "best fit" mask. The requirement is based on 1/12th octave measurement.		

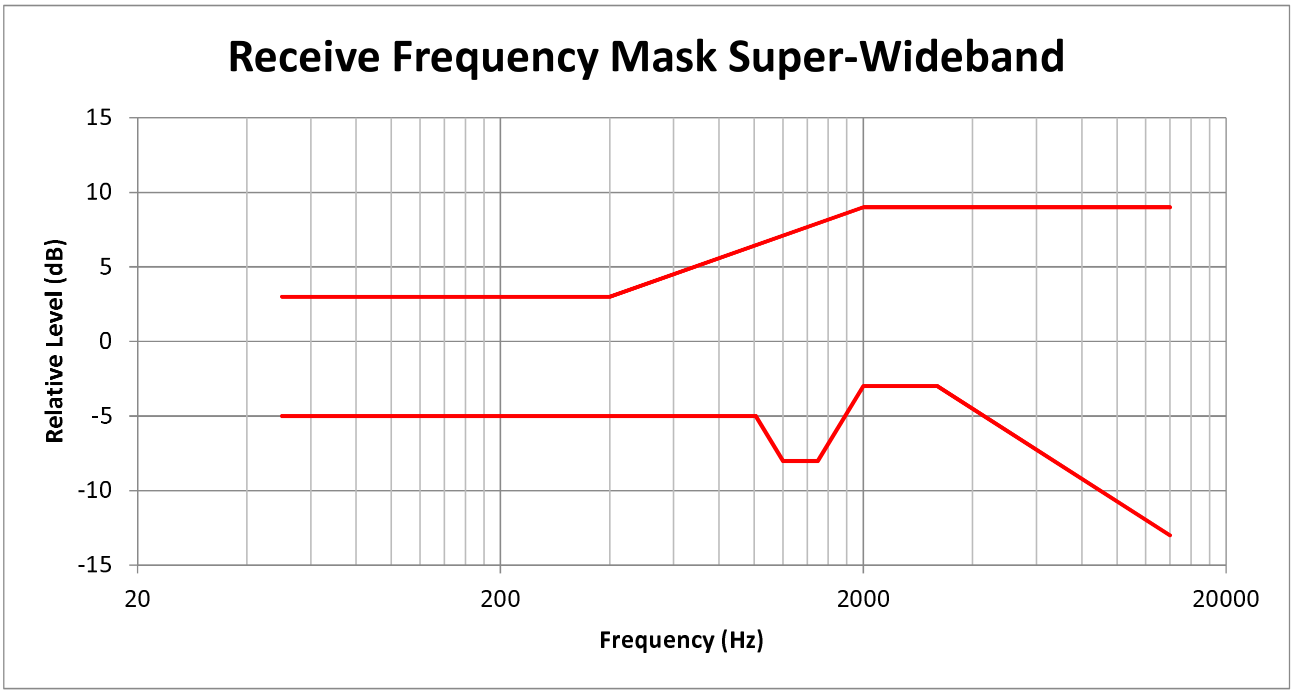


Figure 7: Receive frequency response mask for Super-Wideband

NOTE 1: This requirement applies to headphones not primarily designed for super-wideband communication but rather for music audition. It is the reason of rather open limits. In the next future, new limits will be discussed to apply when specially designed for super-wideband headphones will be available.

Table 9: Fullband receive frequency response limits

Frequency	Upper Limit	Lower Limit
20 Hz	3 dB	-15 dB
50 Hz	3 dB	-5 dB
400 Hz	3 dB	-5 dB
1010 Hz	(see note)	-5 dB
1 200 Hz	(see note)	-8 dB
1 500 Hz	(see note)	-8 dB
2 000 Hz	9 dB	-3 dB
3 200 Hz	9 dB	-3 dB
20 000 Hz	9 dB	-15 dB
NOTE: The limit curves shall be determined by straight lines joining successive co-ordinates given in the table, where frequency response is plotted on a linear dB scale against frequency on a logarithmic scale. It is a floating or "best fit" mask. The requirement is based on 1/12th octave measurement.		

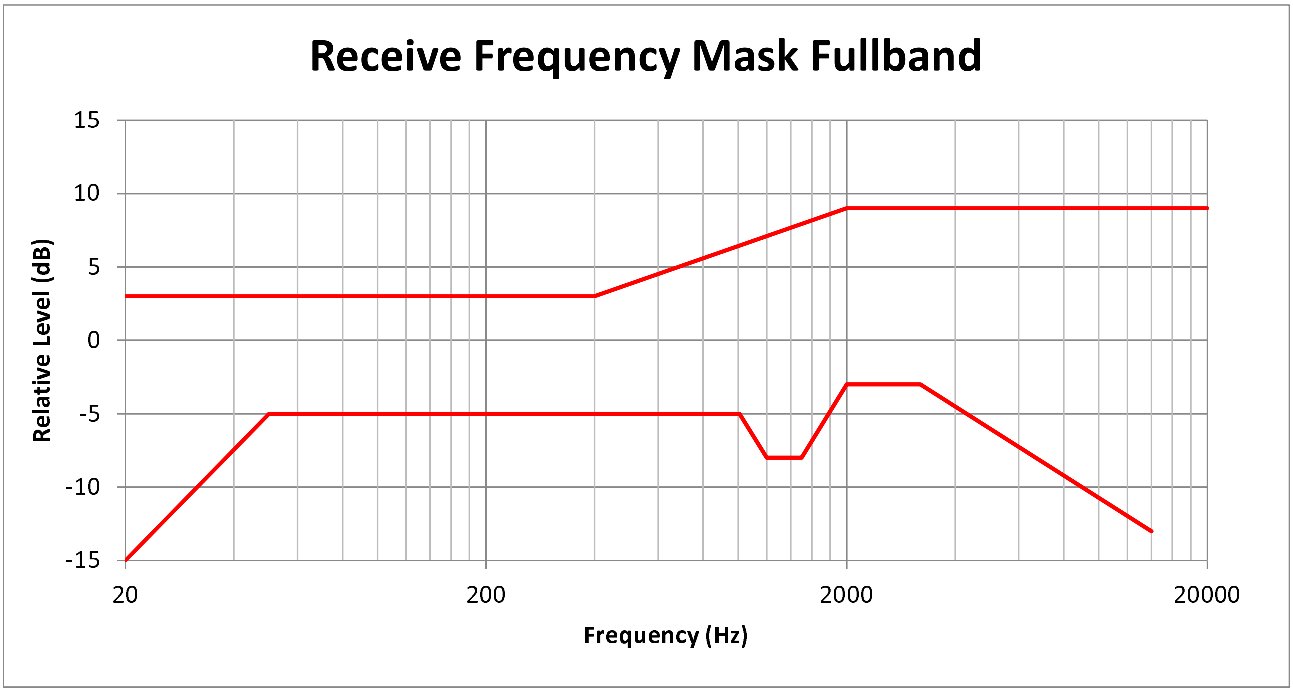


Figure 8: Receive frequency response mask for Fullband

NOTE 2: This requirement applies to headphones not primarily designed for fullband communication but rather for music audition. It is the reason of rather open limits. In the next future, new limits will be discussed to apply when specially designed for fullband headphones will be available.

Measurement Method:

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

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

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

The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [1]. The test signal level shall be -16 dBm0, measured according to Recommendation ITU-T P.56 [27] at the digital reference point or the equivalent analogue point.

The handset or headset terminal is setup as described in clause 6.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [23]). The application forces used to apply the handset against the artificial ear is 2 N, 8 N and 13 N.

The sound pressure level is measured at the DRP of the HATS for each 1/12th octave band.

In case of headset measurements the tests are repeated 5 times, in conformance with Recommendation ITU-T P.380 [15]. The results are averaged (averaged value in dB, for each frequency).

The HATS is diffuse-field equalized as described in Recommendation ITU-T P.581 [4]. The diffuse-field correction as defined in Recommendation ITU-T P.58 [3] is applied.

For super-wideband terminals, measurements shall be made at one twelfth-octave intervals as given by the IEC 61260-1 [16] for frequencies from 50 Hz to 14 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.

For fullband terminals, measurements shall be made at one twelfth-octave intervals as given by the IEC 61260-1 [16] for frequencies from 20 Hz to 20 kHz inclusive. For the calculation the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

The sensitivity is expressed in terms of dBPa/V.

6.4.3 Receive Loudness Rating (monaural reproduction)

Requirement:

When terminal implements Wideband speech functions or when the super-wideband/fullband functions may interact with wideband terminals, the terminal shall fulfil the requirements on RLR as defined in ETSI ES 202 739 [12], clause 7.1.7.

Measurement Method:

The test signal to be used for the measurements shall be British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [1] shall be used. The test signal level shall be -16 dBm0, measured at the digital reference point or the equivalent analogue point. The test signal level is averaged over the complete test signal sequence.

The handset or headset terminal is setup as described in clause 6.2. The handset is mounted in the HATS position (see Recommendation ITU-T P.64 [23]). The application force used to apply the handset against the artificial ear is noted in the test report. The HATS is *NOT* diffuse field equalized as described in Recommendation ITU-T P.581 [4]. The DRP-ERP correction as defined in Recommendation ITU-T P.57 [22] is applied. The application force used to apply the handset against the artificial ear is noted in the test report. By default, 8 N is used.

In case of headset measurements the tests are repeated 5 times, in conformance with Recommendation ITU-T P.380 [15]. The results are averaged (averaged value in dB, for each frequency).

The receiving sensitivity shall be calculated from each band of the 20 frequencies given in table 1 of Recommendation ITU-T P.79 [5], bands 1 to 20. For the calculation the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to Recommendation ITU-T P.79 [5], see annex A. No leakage correction shall be applied for the measurement.

6.4.4 RLR for stereo/dichotic reproduction

For further study.

6.4.5 Loudness

For further study.

6.4.6 Receive Distortion

Requirement:

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

Table 10: Receive distortion for super-wideband

Frequency	Signal to distortion ratio limit, receiving
100 Hz	24 dB
315 Hz	26 dB
400 Hz	30 dB
500 Hz	30 dB
800 Hz	30 dB
1 kHz	30 dB
2 kHz	30 dB
3,15 kHz	30 dB
5 kHz	30 dB
NOTE: Limits at intermediate frequencies lie on a straight line drawn between the given values on a linear (dB ratio) - logarithmic (frequency) scale.	

Table 11: Receive distortion for fullband

Frequency	Signal to distortion ratio limit, receiving
50 Hz	24 dB
100 Hz	26 dB
315 Hz	30 dB
400 Hz	30 dB
500 Hz	30 dB
800 Hz	30 dB
1 kHz	30 dB
2 kHz	30 dB
3,15 kHz	30 dB
8 kHz	30 dB
NOTE: Limits at intermediate frequencies lie on a straight line drawn between the given values on a linear (dB ratio) - logarithmic (frequency) scale.	

Measurement Method:

The handset or headset terminal is positioned as described in clause 6.2.

The signal used is an activation signal followed by a sine-wave signal with a frequency at 100 Hz, 315 Hz, 400 Hz, 500 Hz, 630 Hz, 800 Hz, 1 000 Hz, 2 000 Hz, 3 150 Hz and 7 000 Hz for super-wideband and 50 Hz, 100 Hz, 315 Hz, 400 Hz, 500 Hz, 630 Hz, 800 Hz, 1 000 Hz, 2 000 Hz, 3 150 Hz and 8 000 Hz for fullband.

The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [1] shall be used for activation. The signal level shall be -16 dBm0.

The signal to harmonic distortion ratio is measured selectively up to 14 kHz for super-wideband and 20 kHz for fullband.

The ratio of signal to harmonic distortion shall be measured at the DRP of the artificial ear with a correction by the curve of reference microphone.

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

6.4.7 Minimum activation level and sensitivity in Receive direction

For further study.

6.4.8 Receive Noise

Requirement:

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

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

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

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

Measurement Method:

The handset or headset terminal is setup as described in clause 6.2.

The A-weighted noise level shall be measured at DRP of the artificial ear with the diffuse field equalization active. The noise level is measured until 20 kHz.

The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [1] shall be used for activation. The activation signal level shall be -16 dBm0.

The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [1] shall be used for activation. The activation signal level shall be -16 dBm0.

Spectral peaks are measured in the frequency domain in the frequency range from 50 Hz to 14 kHz for Super-wideband and 50 Hz to 18 kHz for Fullband. The frequency spectrum of the idle channel noise is measured by a spectral analysis having a noise bandwidth of 8,79 Hz (determined using FFT 8 k samples/48 kHz sampling rate with Hanning window or equivalent). The idle channel noise spectrum is stated in dB. A smoothed average idle channel noise spectrum is calculated by a moving average (arithmetic mean) $1/3^{rd}$ octave wide across the idle noise channel spectrum stated in dB (linear average in dB of all FFT bins in the range from 2^(-1/6)f to 2^(+1/6)f). Peaks in the idle channel noise spectrum are compared against a smoothed average idle channel noise spectrum.

NOTE: In case spectral peaks higher than 10 dB above the average noise floor are produced by the terminal, but which are considered to be inaudible due to the very low noise floor produced by the terminal on average, it is the responsibility of the test lab to demonstrate the desired performance subjectively.

6.4.9 Automatic level control in receiving

For further study.

6.4.10 Double talk performance

6.4.10.1 General

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

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

The most important parameters determining the speech quality during double talk are (see Recommendations ITU-T P.340 [30] and P.502 [31]):

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

6.4.10.2 Attenuation range in send direction during double talk $A_{H,S,dt}$

Requirement

Based on the level variation in send direction during double talk $A_{H,S,dt}$ the behaviour of the terminal can be classified according to table 12.

Table 12

Category (according to Rec. ITU-T P.340 [30])	**1**	**2a**	**2b**	**2c**	**3**
	Full Duplex Capability	Partial Duplex Capability			No Duplex Capability
$A_{H,S,dt}$ [dB]	≤ 3	≤ 6	≤ 9	≤ 12	> 12

In general table 12 provides a quality classification of terminals regarding double talk performance. However, this does not mean that a terminal which is category 1 based on the double talk performance is of high quality concerning the overall quality as well.

Measurement method

The long conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [1] shall be used for conditioning the terminal, with the female speaker in the receive direction. The test signal to determine the attenuation range during double talk is the double talk speech sequence as defined in clause 7.3.5 of Recommendation ITU-T P.501 [1] as shown in figure 9. The competing speaker is always inserted as the double talk sequence sdt(t) either in send or receive and is used for analysis.

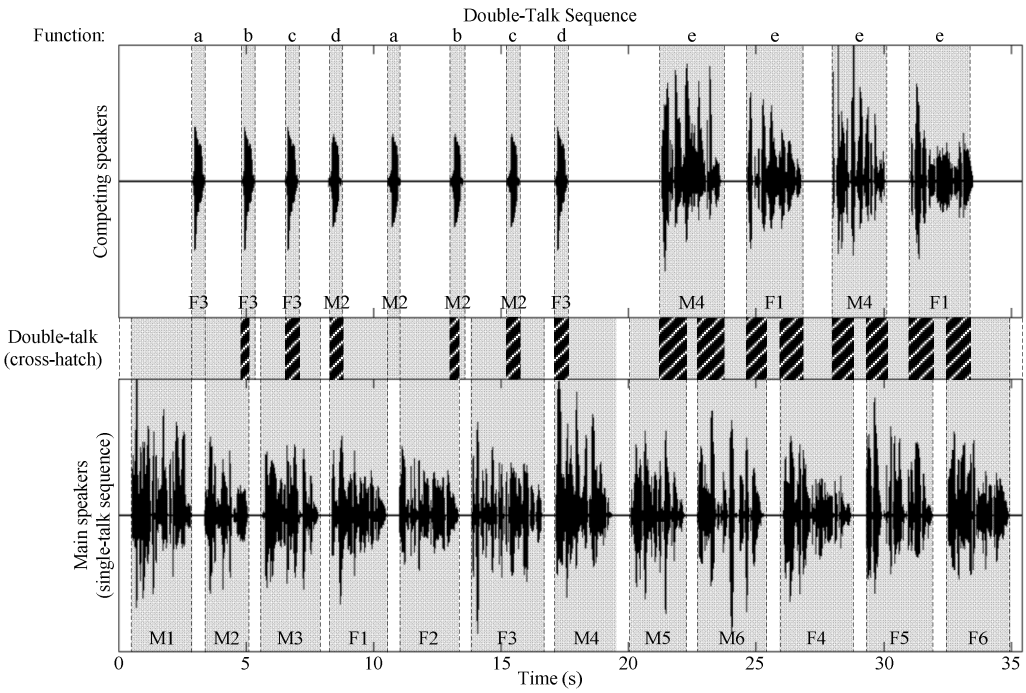


Figure 9: Double talk test sequence with overlapping speech sequences in send and receive direction

The attenuation range during double talk is determined as described in Appendix III of Recommendation Recommendation ITU-T P.502 [31]. The double talk performance is analysed for each word and sentence produced by the competing speaker. The requirement has to be met for each word and sentence produced by the competing speaker.

6.4.10.3 Attenuation range in receive direction during double talk $A_{H,R,dt}$

Requirement

Based on the level variation in receive direction during double talk $A_{H,R,dt}$ the behaviour of the terminal can be classified according to table 13.

Table 13

Category (according to Rec. ITU-T P.340 [30])	1	2a	2b	2c	3
	Full duplex capability	Partial duplex capability			Full duplex capability
$A_{H,R,dt}$ **[dB]**	≤ 3	≤ 5	≤ 8	≤ 10	> 10

In general table 13 provides a quality classification of terminals regarding double talk performance. However, this does not mean that a terminal which is category 1 based on the double talk performance is of high quality concerning the overall quality as well.

Measurement method

The long conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [1] shall be used for conditioning the terminal, with the female speaker in the receive direction. The test signal to determine the attenuation range during double talk is shown in figure 7. A sequence of speech signals is used which is inserted in parallel in send and receive direction. The test signals are synchronized in time at the acoustical interface. The delay of the test arrangement should be constant during the measurement.

The test arrangement is according to clause 6.2.

The attenuation range during double talk is determined as described in Appendix III of Recommendation Recommendation ITU-T P.502 [31]. The double talk performance is analysed for each word and sentence produced by the competing speaker. The requirement has to be met for each word and sentence produced by the competing speaker.

6.4.10.4 Detection of echo components during double talk

Requirement

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

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

Under these conditions the requirements given in table 14 are applicable (more information can be found in annex A of Recommendation ITU-T P.340 [30]).

Table 14

Category (according to Rec. ITU-T P.340 [30])	1	2a	2b	2c	3
	Full Duplex Capability	Partial Duplex Capability			No Duplex Capability
Echo Loss [dB]	≥ 27	≥ 3	≥ 17	≥ 11	< 11

Measurement method

The test arrangement is according to clause 6.2.

The double talk signal consists of a sequence of orthogonal signals which are realized by voice-like modulated sine waves spectrally shaped similar to speech. A detailed description can be found in Recommendation ITU-T P.501 [1].

The signals are fed simultaneously in send and receive direction. The level in send direction is -4,7 dBPa at the MRP (nominal level), the level in receive direction is -16 dBm0 at the electrical reference point (nominal level).

The settings for the signals are as follows.

Table 15: Parameters of the two test signals for double talk measurement based on AM-FM modulated sine waves

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

The test signal is measured at the electrical reference point (send direction). The measured signal consists of the double talk signal which was fed in by the artificial mouth and the echo signal. The echo signal is filtered by comb filter using mid-frequencies and bandwidth according to the signal components of the signal in receive direction (see Recommendation ITU-T P.501 [1]). The filter will suppress frequency components of the double talk signal.

In each frequency band which is used in receive direction the echo attenuation can be measured separately. The requirement for category 1 is fulfilled if in any frequency band the echo signal is either below the signal noise or below the required limit. If echo components are detectable, the classification is based on table 15. The echo attenuation is to be achieved for **each individual frequency band** according to the different categories.

6.4.11 Switching characteristics

6.4.11.1 Note

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

6.4.11.2 Activation in send direction

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

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

Requirement

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

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

Measurement method

The test signal is the "short words for activation" sequence described in clause 7.3.4 of Recommendation ITU-T P.501 [1] with increasing level for each single word.

Table 16

	Single word duration / pause duration	Level of the first single word (at the MRP)	Level difference between two periods of the test signal
single word to determine switching characteristic in send direction	~600 ms / ~400 ms	-24 dBPa (see note)	1 dB
NOTE: The signal level is determined for each utterance individually according to Recommendation ITU-T P.56 [27].			

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

The test arrangement is described in clause 6.2.

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

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

6.4.11.3 Silence suppression and comfort noise generation

Requirements and measurement methods are for further study.

NOTE: In general it is not recommended to use silence suppression at all.

6.4.12 Speech and audio quality in presence of noise

6.4.12.1 Performance in send in the presence of background noise

Requirement

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

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

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

The spectral difference between comfort noise and original (transmitted) background noise shall be within the mask given through straight lines between the breaking points on a logarithmic (frequency) - linear (dB sensitivity) scale as given in table 17.

Table 17: Requirements for spectral adjustment of comfort noise (mask)

Frequency	Upper limit	Lower limit
200 Hz	12 dB	-12 dB
800 Hz	12 dB	-12 dB
800 Hz	10 dB	-10 dB
2 000 Hz	10 dB	-10 dB
2 000 Hz	6 dB	-6 dB
4 000 Hz	6 dB	-6 dB
14 000 Hz	6 dB	-6 dB
20 000 Hz*	6 dB	-6 dB
NOTE 1: All sensitivity values are expressed in dB on an arbitrary scale. NOTE 2: For fullband terminals only.		

Measurement method

The background noise simulation as described in clause 6.2 is used.

The handset or headset terminal is set-up as described in clause 6.2. The handset is mounted at the HATS position (see Recommendation ITU-T P.64 [23]).

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

In a second step a test signal is applied in receive direction consisting of an initial pause of 10 s and a periodical repetition of the female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [1] in receive direction (duration 10 s) with nominal level to enable comfort noise injection simultaneously with the background noise. For the measurement the background noise sequence has to be started at the same point as it was started in the previous measurement.

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

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

6.4.12.2 Speech quality in the presence of background noise

Requirement

Speech Quality for super-wideband and fullband systems can be tested based on ETSI TS 103 281 [28].

For the background noises defined in clause 6.2 the following requirements apply:

- N-MOS-LQO$_f \geq$ 3,5.
- S-MOS-LQO$_f \geq$ 4.
- G-MOS-LQO$_f \geq$ 3,5.

NOTE: It is recommended to test the terminal performance with other types of background noises if the terminal is likely to be exposed to other noises than specified in clause 6.2.

Measurement method

The background noise simulation as described in clause 6.2 is used. The handset or headset terminal is set-up as described in clause 6.2. The handset is mounted at the HATS position (see Recommendation ITU-T P.64 [23]).

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

The near end speech signal consists of 16 sentences of speech (4 male and 4 female talkers, 2 sentences each). The American English speech samples from ETSI TS 103 281 [28] are used. The test signal level is -1,7 dBPa at the MRP.

The following signals are required for the tests:

1) The clean speech signal is used as the undisturbed reference (see ETSI TS 103 281 [28]).

2) The send signal is recorded at the electrical reference point.

N-MOS-LQO_f, S-MOS LQO_f and G-MOS LQO_f are calculated as described in ETSI TS 103 281 [28]. Either Model A or Model B can be used. The model chosen shall be documented in the test report.

When using model A the following mapping functions apply:

$$\text{S-MOS}'_{\text{LQOf}} = 1{,}418 \cdot \text{S-MOS}_{\text{LQOf}} - 1{,}145$$
$$\text{N-MOS}'_{\text{LQOf}} = 1{,}346 \cdot \text{N-MOS}_{\text{LQOf}} - 1{,}584$$
$$\text{G-MOS}'_{\text{LQOf}} = 1{,}279 \cdot \text{G-MOS}_{\text{LQOf}} - 0{,}7364$$

6.4.12.3 Quality of background noise transmission (with far end speech)

Requirement

The test is carried out applying a speech signal in receive direction. During and after the end of the speech signal the signal level in send direction should not vary more than 10 dB (during transition to transmission of background noise without far end speech). The measurement is conducted for all types of background noise as defined in clause 6.2.

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

Measurement method

The test arrangement is according to clause 6.2.

The background noises are generated as described in clause 6.2.

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

In a second step the same measurement is conducted but with inserting the speech signal at the far end. The exactly identical background noise signal is applied. The background noise signal shall start at the same point in time which was used for the measurement without far end signal. The background noise should be applied for at least 5 s in order to allow adaptation of the noise reduction algorithms. After at least 5 s a series of the female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [1] is applied in receive direction with duration of at least 10 s. The test signal level is -16 dBm0 at the electrical reference point.

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

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

6.4.13 Quality of echo cancellation

6.4.13.1 Temporal echo effects

Requirement

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

Measurement method

The test arrangement is according to clause 6.2.

The test signal consists of periodically repeated Composite Source Signal according to Recommendation ITU-T P.501 [1] with an average level of -5 dBm0 as well as an average level of -25 dBm0. The echo signal is analysed during a period of at least 2,8 s which represents 8 periods of the CS signal. The integration time for the level analysis shall be 35 ms, the analysis is referred to the level analysis of the reference signal.

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

The difference between the maximum attenuation and the minimum attenuation is measured.

NOTE 1: In addition tests with more speech like signals should be made, e.g. Recommendation ITU-T P.501 [1] to see time variant behaviour of EC. However, for such tests the simple broadband attenuation based test principle as described above cannot be applied due to the time varying spectral content of the speech like signals.

NOTE 2: The analysis is conducted only during the active signal part, the pauses between the Composite Source Signals are not analysed. The analysis time is reduced by the integration time (35 ms) of the level analysis taking into account the exponential character of the integration time in any tolerance scheme.

NOTE 3: Care should be taken not to confuse noise or comfort noise with residual echo. In cases of doubt the measured echo signal should be compared to the residual noise signal measured under the same conditions without inserting the receive signal. If the level vs. time analysis leads to the identical result it can be assumed that no echo but just comfort noise is present.

6.4.13.2 Spectral echo attenuation

Requirement

The echo attenuation versus frequency shall be below the tolerance mask given in table 18.

Table 18: Echo attenuation limits

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

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

NOTE: The measurement is performed in wideband since no super-wideband or fullband data are available. The extension of the method to super-wideband and fullband is for further study.

Measurement method

The test arrangement is according to clause 6.2.

Before the actual measurement a training sequence consisting of the compressed real speech signal described in clause 7.3.3 of Recommendation ITU-T P.501 [1] is fed. The level of the training sequence is -16 dBm0.

The test signal is the compressed real speech signal described in clause 7.3.3 of Recommendation ITU-T P.501 [1]. The measurement is carried out under steady-state conditions. The average test signal level is -16 dBm0, averaged over the complete test signal. The power density spectrum of the measured echo signal is referred to the power density spectrum of the original test signal. The analysis is conducted using FFT analysis with 8 k points (48 kHz sampling rate, Hanning window).

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

6.4.13.3 Occurrence of artefacts

For further study.

6.4.13.4 Variable echo path

Requirement

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

Measurement method

The test arrangement is according to clause 6.2.

As test signal the compressed real speech signal described in clause 7.3.3 of Recommendation ITU-T P.501 [1] is used. The signal level shall be -10 dBm0. The terminal volume control is set to nominal RLR. The first 4 sentences of the test signal are used to allow full convergence of the echo canceller. The next 4 sentences (from 10,75 s to 22,5 s) are used for the analysis. The echo signal level is analysed over time. The echo signal level is analysed for 11,75 s, using a time constant of 35 ms.

The measurement result is displayed as echo level versus time.

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

6.4.14 Variant impairments; network dependant

6.4.14.1 Clock accuracy send

Requirement

The clock accuracy in send direction between the VoIP-Terminal and the IP reference interface shall be less than 150 ppm under ideal network conditions.

NOTE: The clock accuracy does not cover all possible network configurations. Especially it is not sufficient for data transmission or distributed TDM PBX where synchronization is required.

Measurement method

A sequence of CS signals (active signal length = 250 ms) is repeated for 120 s in order to analyse clock accuracy and any other time-variant delay. The pause length between two CS bursts is 100 ms and 1,2 s after every fourth burst in order to simulate a speech pause, which may lead to buffer adjustments. The test signal level shall be -4,7 dBPa at the MRP.

A cross correlation analysis versus time is carried out over the whole 120 s sequence between the received and the original test signal. The duration of the measurement (120 s) is indicated on the x-axis, the result of the cross correlation analysis (delay) is plotted on the y-axis.

The resulting clock accuracy within an analysis time range of at least 60 s is calculated as follows:

$$clock\ accuracy\left[ppm\right] = \frac{delay\ change\left[s\right]}{analysis\ duration\left[s\right]} \cdot 1 \cdot 10^{6} \qquad (6)$$

6.4.14.2 Clock accuracy receive

Requirement

The clock accuracy in receive direction between the IP reference interface and the VoIP-Terminal shall be less than 150 ppm under ideal network conditions.

Measurement method

A sequence of CS signals (active signal length = 250 ms) is repeated for 120 s in order to analyse clock accuracy and any other time-variant delay. The pause length between two CS bursts is 100 ms and 1,2 s after every fourth burst in order to simulate a speech pause, which may lead to buffer adjustments. The test signal level at the IP reference interface shall be -16 dBm0.

A cross correlation analysis versus time is carried out over the whole 120 s sequence between the received and the original test signal. The duration of the measurement (120 s) is indicated on the x-axis, the result of the cross correlation analysis (delay) is plotted on the y-axis.

The resulting clock accuracy within an analysis time range of at least 60 s is calculated as follows:

$$clock\ accuracy\left[ppm\right] = \frac{delay\ change\left[s\right]}{analysis\ duration\left[s\right]} \cdot 1 \cdot 10^{6} \qquad (7)$$

6.4.14.3 Send packet delay variation

Requirement

The measured maximum delay variation of RTP packets in send direction of the VoIP-terminal under test should be less than 1 ms.

NOTE: Any delay variation of RTP packets introduced in send direction will lead to potentially increased delay due to increased de-jitter buffer at the far end terminal.

Measurement method

The RTP data stream in send direction should be monitored with a tap or a switch providing a monitoring port, positioned at the location of the network impairment simulator (see clause 6.1).
The test arrangement is according to clause 6.1.

The monitoring time should be 60 s. A signal like the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [1] is played back in send direction using a nominal level of -4,7 dBPa at the MRP. This speech signal is only necessary to make sure, RTP is played out, even in the case VAD is active.

The delay variation for each packet D(i) is evaluated according to IETF RFC 3550 [34]:

$$d(i) = \Delta t_{eff(i)} - \Delta t_{exp(i)}$$

$$D(i) = (15 \times D(i-1) + |\, d(i)\, |) / 16 \qquad (8)$$

With:

- $\Delta t_{exp(i)}$ = the expected time between packet i and packet i-1; and
- $\Delta t_{eff(i)}$ = the effective time between packet i and packet i-1.

Maximum delay variation = MAX(D(i)).

6.4.15 Send and receive delay - round trip delay

The roundtrip delay of a VoIP-terminal is defined as the sum of send and receive delays. In the following clauses the calculation of the requirements for send and receive delay are explained. For a telecommunication connection, only the roundtrip delay can be experienced. For this reason, also the requirement for VoIP-terminals is given only for the roundtrip delay. As long as the measured roundtrip delay fulfils the requirements, send or receive delays may be above the theoretical requirements.

Requirement

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

The roundtrip delay of the VoIP-terminal T_{rtd} (sum of receive and send delay) shall be less than 100 ms. (category B in Recommendation ITU-T P-1010 [33]). From the users perspective, a value less than 50 ms (category A in Recommendation ITU-T P-1010 [33]) is preferred.

NOTE 1: The limit for the roundtrip delay T_{rtd} of the VoIP-terminal is derived from the sum of the send and receive delay limits.

NOTE 2: This requirement is based on the lowest possible delay values which can be expected under ideal network conditions. Caution should be exercised to ensure that the terminal is operated under optimum conditions in order to avoid adverse effects, e.g. network conditions, settings and memory effects of the terminal jitter buffer.

Measurement method

- **Send direction**

The delay in send direction is measured from the MRP to POI. The delay measured in send direction is:

$$T_s + t_{System} \quad (9)$$

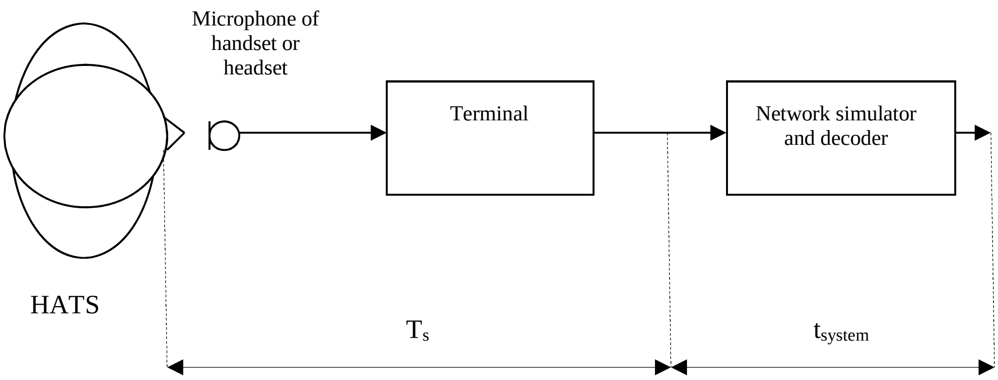


Figure 10: Different blocks contributing to the delay in send direction

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

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

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

   The setup of the handset/headset terminal is in correspondence to clause 6.2.

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

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

- **Receive direction**

The delay in receive direction is measured from POI to the Drum Reference Point (DRP). The delay measured in receive direction is:

$$T_r + t_{System} \quad (10)$$

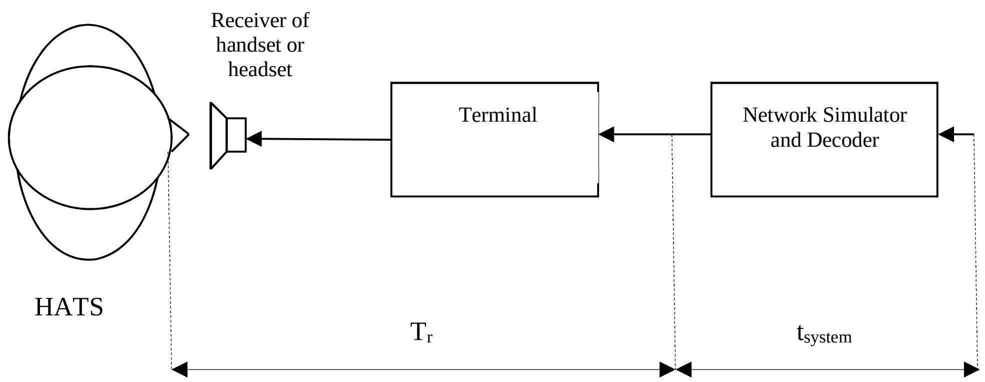


Figure 11: Different blocks contributing to the delay in receive direction

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

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

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

2) The test arrangement is according to clause 6.2.

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

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

NOTE 3: It is not necessary to know the delays T_s, T_r and t_{system} per direction. The roundtrip delay of the terminal is the sum of send and receive delays minus the roundtrip delay of the measurement equipment and (if applicable) the network.

6.5 Other parameters

6.5.1 Objective listening quality

Requirement

For further study.

Measurement method

Objective listening speech quality is measured using Recommendation ITU-T P.863 [14] in fullband mode.

The test signal to be used for the measurements shall be 4 sentence pairs (male/female) fulfilling the requirements of Recommendation ITU-T P.863.1 [32]. The 4 sentence pairs are taken from Recommendation ITU-T P.501 [1]. It shall be stated, which sentence pairs were used. The test signal level is averaged over all sentence pairs (4 sentence pairs). The measurement is done 4 times, every time using another pair of the speech sentences. The result of the measurement is the averaged value of all 4 measurements.

NOTE 1: For the use of P.863 the following applies (see Recommendation ITU-T P.863.1 [32]):

- fullband Context (MOS-LQO$_f$):
  - For the performance tests with network impairments the following settings are used.

Table 19: Network conditions for electrical-acoustical measurements (speech samples)

Condition	Packet Loss (Equal)	Delay Variation
0 (see note 2) (VAD)	0	No
1	0	No
2	0	20 ms (see note 1)
3	1 %	No
4	1 %	20 ms (see note 1)
5	3 %	No
NOTE 1: Delay variation produced with a Pareto-Distribution and r = 0,5. NOTE 2: VAD on, all other conditions (1 to 5) tested with VAD off. NOTE 3: For some network emulation tools, it is necessary to introduce a constant delay to offer the possibility to generate a delay variation distribution. This delay has to be subtracted from the measured delay before interpreting the results. NOTE 4: The settings are derived from the ones used in the ETSI Plugtest VoIP speech quality test events.		

NOTE 2: The delay requirements for conditions with network impairments are based on the measured roundtrip delay of the terminal in the absence of network impairments T_{rtd}clean (see clause 6.4.15). A small additional tolerance takes into account the variable behaviour of the delay.

6.5.2 Quality of jitter buffer adjustment

Requirements

The speech quality during and after inserted IP delay variation shall be as follows:

Table 20: Requirements for variant network impairments

Codec	MOS-LQO$_f$
VoLTE (IMS) ETSI TS 126 441 [26]	> 3,8
Recommendation ITU-T G.722.1 [7], annex C	> 3,8
Recommendation ITU-T G.729.1 [8], annex E (extension SWB)	> 3,8
Recommendation ITU-T G.718 [9] annex B	> 3,8
Recommendation ITU-T G.711.1 [6], annexes D and F [21]	> 3,8
Recommendation ITU-T G.722 [19], annexes B and D	> 3,8
Opus [36]	> 3,8

The delay measured 20 s after ending of the IP delay variation shall be maximum 10 ms higher than the delay measured before the IP delay variation.

Measurement method

The test signal consists of a CSS-signal, followed by 5 times the same speech sentence, fulfilling the requirements of Recommendation ITU-T P.863.1 [32], then again a CSS signal (20 s after the IP delay variation stops). The speech signal level is averaged over all used (original) sentences (8 sentences).

NOTE 1: The 8 used sentences consist of the 8 single sentences taken from the 4 sentence pairs used in clauses 6.5.2.

NOTE 2: For every new measurement a new call has to be setup to start with an initial delay. Depending on the algorithm used in the variable jitter buffer (e.g. jitter buffer starting with a high fill size), it may be necessary to let some time pass under clean conditions until the measurement is started.

The first CSS signal is used to measure the delay prior to the IP impairment (in clean network conditions). The second CSS signal is used to measure the delay 20 s after the IP impairment stops. The difference of the two delays is the measurement result for the variation of the jitter buffer per measurement. The overall result is the average of all 10 measurements.

The first sentence (during which IPDV of 50 ms is applied) is used to measure the speech quality during jitter buffer adaption (low to high). MOS-LQO$_f$ of the first sentence is measured using Recommendation ITU-T P.863 [14] in super-wideband mode. The overall result is the average MOS-LQO$_f$ of the 8 measurements.

The second to the fifth sentence (every 5 s a sentence) are used to measure the speech quality during jitter buffer adaption (high to low). MOS-LQO$_f$ is measured using Recommendation ITU-T P.863 [14] in super-wideband mode for each of these four sentences. The minimum MOS-LQO$_f$ of this four sentences is used for the averaging over all 8 measurements. The overall result for the speech quality during jitter buffer adaption (high to low) is the average of the minimum MOS-LQO$_f$-value of the 8 measurements.

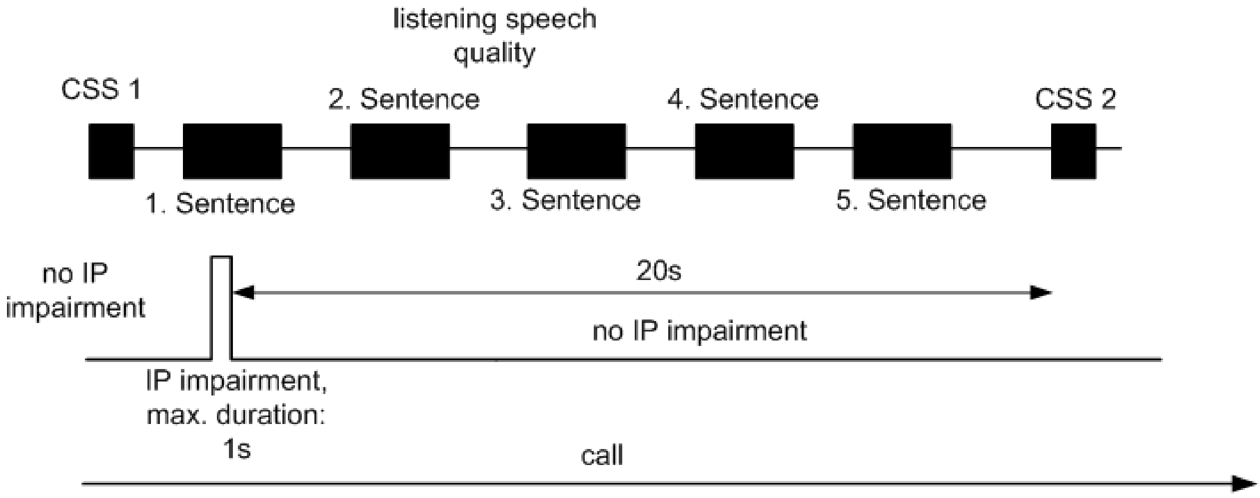


Figure 12: Test sequence to measure quality of jitter buffer adjustment (with 1 of 8 sentences)

The IP impairment consists of additional packet delay (IPDV) up to 50 ms, during max. 1 second. The impairment can be in form of jitter, but also with only some single packets delayed. An example for the impairment can be found in annex B of ETSI ES 202 737 [35].

NOTE 3: Care should be given, that no packet reordering occurs (this could happen if e.g. one packet is delayed by 50 ms and the next one is not delayed, they will change order, which will not happen in real networks except in a failover situation or with bad implementations of load balancing).

Annex A: Void

History

Document history		
V1.1.1	March 2013	Publication
V1.2.1	March 2018	Publication