ETSI TS 102 925 V1.2.1 (2018-10)

TECHNICAL SPECIFICATION

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

Reference
RTS/STQ-208-2

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. Nowadays, terminal equipment may offer wider bandwidth, due to features already available in these terminals. Such equipment may implement conversational features that may be to the benefit of the electro acoustic equipment's already available in the terminal and may provide wider quality for the end users. High quality conferencing systems may also implement wider bandwidth in order to reach quality and behaviour close to normal face to face conditions.

The present document is intended to provide initial requirements and test methods for such equipment. The present document also provides materials for a further update of ETSI SR 002 959 [i.2]: Electronic Working Tools; Roadmap including recommendations for the deployment and usage of electronic working tools in the ETSI standardization process.

The present document complements the ETSI TS 102 924 [17] Handset and Headset mode specifications.

1 Scope

The present document provides speech & audio transmission performance requirements and measurement methods for handsfree functions of super-wideband/fullband terminals, including conferencing 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 codecs are now available, some being also compatible with wideband codecs.

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 using high quality handsfree systems.
- Bandwidth extension which may allow usage for some mixed content applications.
- Super-wideband enhancement coupled with stereo/dichotic/multichannel.

In the send path the signal may combine speech, music and environmental signals. The signal may be:

- acoustically captured by a microphone; or
- directly inserted through a digital or analog connection.

In the receive path, the signal may combine:

- communication signals such as described for send path, including environmental signals; and
- signals 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 (HATS) for hands-free and handset terminal testing".

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

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

[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 ITU-T G.729.1 (Annex E): "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 (Annex B): "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 ES 202 740: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user".

[12] ETSI TS 103 740: "Speech and multimedia Transmission Quality (STQ);Transmission requirements for wideband wireless terminals (handsfree) from a QoS perspective as perceived by the user".

[13] ETSI ETS 300 807: "Integrated Services Digital Network (ISDN); Audio characteristics of terminals designed to support conference services in the ISDN".

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

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

[16] Recommendation ITU-T P.1301: "Subjective quality evaluation of audio and audiovisual multiparty telemeetings".

[17] ETSI TS 102 924: "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".

[18] Void.

[19] Void.

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

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

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

[23] 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".

[24] Void.

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

[26] 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".

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

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

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

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

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

[32] Recommendation ITU-T P.1010: "Objective test methods for speech communication systems using complex test signals".

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] ITU-T Supplement P16: "Guidelines for placement of microphones and loudspeakers in telephone conference rooms and Group Audio Terminals (GATs)".

[i.2] ETSI SR 002 959: "Electronic Working Tools; Roadmap including recommendations for the deployment and usage of electronic working tools in the ETSI standardization process".

[i.3] ISO 532: "Acoustics - Method for calculating loudness level".

[i.4] ETSI EG 202 425: "Speech Processing, Transmission and Quality Aspects (STQ); Definition and implementation of VoIP reference point".

[i.5] NIST Net™.

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

[i.6] Netem™.

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

[i.7] STQ(15)48_039: "Objective Codec Evaluation of EVS. HEAD acoustics GmbH".

[i.8] ETSI TR 126 952: "Universal Mobile Telecommunications System (UMTS); LTE; Codec for Enhanced Voice Services (EVS); Performance characterization (3GPP TR 26.952)".

[i.9] STQ (12)40_32: "Proposal for correction factor when measuring receive part of super wide band and full band headset terminals".

3 Definitions of terms and abbreviations

3.1 Terms

For the purposes of the present document, the following terms 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: audio transmission bandwidth with a nominal pass-band wider than 50 Hz to 14 000 Hz, usually understood to be 20 Hz to 20 000 Hz

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: audio transmission bandwidth with a nominal pass-band wider than 100 Hz to 7 000 Hz, usually understood to be 50 Hz to 14 000 Hz

NOTE: Bandwidth definitions are adapted from Recommendation ITU-T P.10/G.100 [2].

3.2 Abbreviations

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

AM-FM	Amplitude Modulation - Frequency Modulation
CSS	Composite Source Signal
DRP	ear Drum Reference Point
DUT	Device Under Test
EC	Echo Cancellation
EL	Echo Loss
EVS	Enhanced Voice Services
FB	FullBand
FFT	Fast Fourier Transform
GAT	Group Audio Terminal
G-MOS-LQO$_F$	Overall Quality Mean Opinion Score, Listening Quality Objective, fullband
HATS	Head And Torso Simulator
HFRP	HandsFree Reference Point
IEC	International Electrotechnical Comission
IP	Internet Protocol
IPDV	IP Packet Delay Variation
L_E	Earcap Leakage
MCU	Multiplexing Control Unit
MOS	Mean Opinion Score
MRP	Mouth Reference Point
MS	Mid-sized Stereo
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
N-MOS-LQO$_F$	Noise Quality Mean Opinion Score, Listening Quality Objective, fullband
PC	Personal Computer
PDA	Personal Digital Assistant
POI	Point Of Interconnection
RLR	Receive Loudness Rating
SLR	Send Loudness Rating

S-MOS-LQO$_F$	Speech Quality Mean Opinion Score, Listening Quality Objective, fullband
SWB	Super-WideBand
TBD	To Be Determined
TCL	Terminal echo Coupling Loss
VAD	Voice Activity Detector

4 Applications and Codec considerations

4.1 Applications

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

- Speech and audio communication, including conferencing using high quality handsfree systems, for which super-wideband/fullband coding can better reproduce the audio environment and provide an improved sound quality, user's experience 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, the 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 Codec considerations

4.2.0 Introduction

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 4.2.1-1: List of super-wideband codecs covered by the present document

Coder Reference	Speech	Other signals	Stereo	Remark
ETSI TS 126 441 [25]	X	X Music	(X)	EVS codec. Stereo supported in a dual mono configuration
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 [15] Annexes D and F	X	X	X (Annex F)	
Recommendation ITU-T G.722 [20] Annexes B and D	X	X	X (Annex D)	
Opus [30]	X	X	X	

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

The following codecs are recommended for super-wideband:

- Recommendation ITU-T G.722.1 [7] Low-complexity coding at 24 kbit/s and 32 kbit/s for handsfree operation in systems with low frame loss. Annex C 14 kHz mode at 24 kbit/s, 32 kbit/s and 48 kbit/s.
  - The algorithm is recommended for use in handsfree 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 kbit/s, 32 kbit/s and 48 kbit/s.
  - Annex C. This annex provides a description of the 14 kHz mode at 24 kbit/s, 32 kbit/s 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 - 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 [15], Annex D defines the Super-wideband extension.
  - Annex F defines the Stereo embedded extension for Recommendation ITU-T G.711.1 [15].
  - *"Annex F is intended as a stereo extension to the G.711.1 wideband coding algorithm and its Super-wideband Annex D. Compared to discrete two-channel (dual-mono) audio transmission, this stereo extension G.711.1, 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 [20], Annex B defines the Super-wideband extension and Annex D defines the Stereo embedded extension for Recommendation ITU-T G.722 [20].
  - *"Annex B describes a scalable Super-wideband (SWB, 50-14 000 Hz) speech and audio coding algorithm operating at 64, 80 and 96 kbit/s. The Recommendation ITU-T G.722 Super-wideband extension codec is interoperable with Recommendation ITU-T G.722. The output of the Recommendation ITU-T G.722 SWB coder has a bandwidth of 50-14000 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".*
- ETSI TS 126 441 [25]. The Enhanced Voice Services (EVS) codec 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.
  - EVS is defined in ETSI TS 126 441 [25] and ETSI TR 126 952 [24]. The tests conducted on codec implementations, e.g. [i.7] show that the requirements and test methods for SWB terminals as defined in the present document apply.
  - EVS is designed for packet-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, full-band audio coding for high-quality, conversational applications.
  - "Recommendation ITU-T G.719 [10] describes the G.719 coding algorithm for low-complexity full-band 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: Recommendation ITU-T P.501 [1] Annex A specifies the use of the ISO base media file format as container for the G.719 bitstream addresses non-conversational use cases of the codec (e.g. call waiting music playback and recording of teleconferencing sessions, voice mail messages and online "jam"-sessions).

- ETSI TS 126 441 [25]. The Enhanced Voice Services (EVS) codec 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.
  - EVS is defined in ETSI TS 126 441 [25] and ETSI TR 126 952 [24]. The tests conducted on codec implementations, e.g. [i.7] show that the requirements and test methods for FB terminals as defined in this TS apply.
  - EVS is designed for packet-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 equipment and associated considerations

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 type of signal (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 for wideband calculation and loudness shall be calculated. 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.

5.1 Test Set-up

5.1.0 Introduction

Recommendation ITU-T P.58 [3] indicates:

> *"The artificial ears ... support super-wideband as well as full-band applications. It should be noted that the acoustical impedance of the artificial ears has some limitations in realistically simulating human ears".*
>
> *"The artificial mouth supports super-wideband applications, however it should be noted that the directionality of the artificial mouth is limited in its ability to simulate the human mouth in the super-wideband frequency range."*

For terminals supporting SWB or FB a HATS (Head And Torso Simulator) should be used. For terminals supporting SWB or FB in combination with Narrowband/Wideband functions a HATS (Head And Torso Simulator) shall be used for parameters defined for limited bandwidth such as RLR and SLR.

For send path the HATS shall be used for super-wideband. Until the development of new systems with larger bandwidth, send path measurements will be limited to super-wideband.

NOTE 1: Some HATS may provide a higher bandwidth. If a lab wants to apply the HATS for fullband testing, the lab should check if the HATS used for the tests has been developed and calibrated over the full bandwidth. For receive path, a correction factor (given in Annex B of STQ(12)40_32 [i.9]) allows measurement at DRP up to 16 kHz.

For handsfree and conferencing terminals an alternative to HATS is the use of a combination including a free field microphone (for receive measurements) and a loudspeaker (for send measurements). The frequency response of these equipments should cover the bandwidth of the terminal under test (at least from 50 Hz to 14 kHz for SWB and from 20 Hz to 20 kHz for FB). The characteristics of the free-field microphone and the loudspeaker shall be recorded in the test report.

The "lip ring" as defined for the artificial mouth of HATS will be defined as the centre of the front face of the loudspeaker and the acoustic centre of the free field microphone.

NOTE 2: The "centre" of the loudspeaker and the "equivalent lip ring" should be defined in more detail.

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.

When a codec with variable bite rate is used, one should adopt, for testing terminal electro acoustical parameters, the highest bit rate which is recognized as providing the best characteristics.

5.1.1 Setup for terminals

5.1.1.0 Introduction

As the scope of the present document includes all the potential types of handsfree terminals this clause defines the set up for each type of terminal.

5.1.1.1 Desktop operated handsfree terminal

The desktop operated handsfree terminal is intended to be placed on a table and the user is located close to the edge of this tables.

When HATS is used in the test equipment, the setups can be found in Recommendation ITU-T P.581 [4], and is placed according to figures 5.1.1.1-1 and 5.1.1.1-2.

When HATS is not used, it is replaced by free-field microphone for receive measurements and loudspeaker (called "artificial mouth" in figure 5.1.1.1-3) for send measurements, the arrangement defined in Recommendation ITU-T P.340 [6] applies (see figure 5.1.1.1-3).

When using a free-field microphone instead of the artificial ears of HATS the centre of the microphone is placed at the point "C" on figure 5.1.1.1-3.

When using a loudspeaker instead of the artificial mouth of HATS the centre of the front plane is placed at the point "C" on figure 5.1.1.1-3.

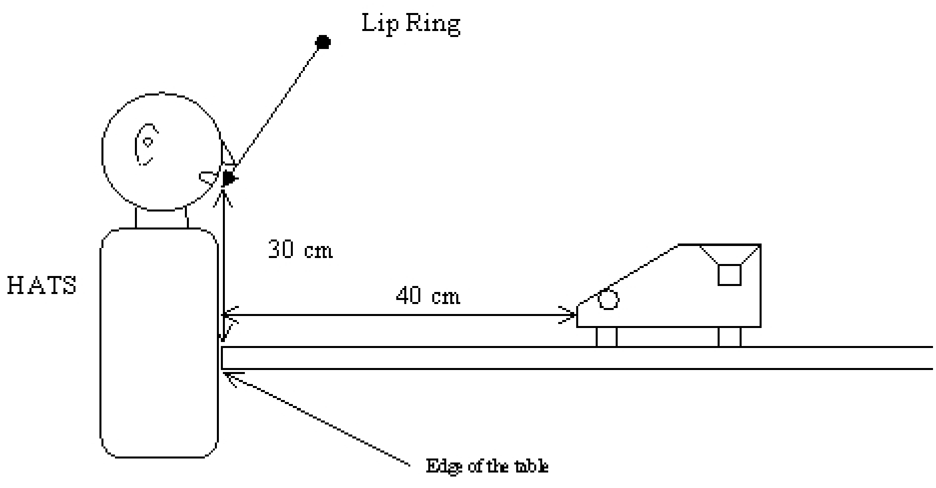

Figure 5.1.1.1-1: Position for test of desktop handsfree terminal with HATS, side view

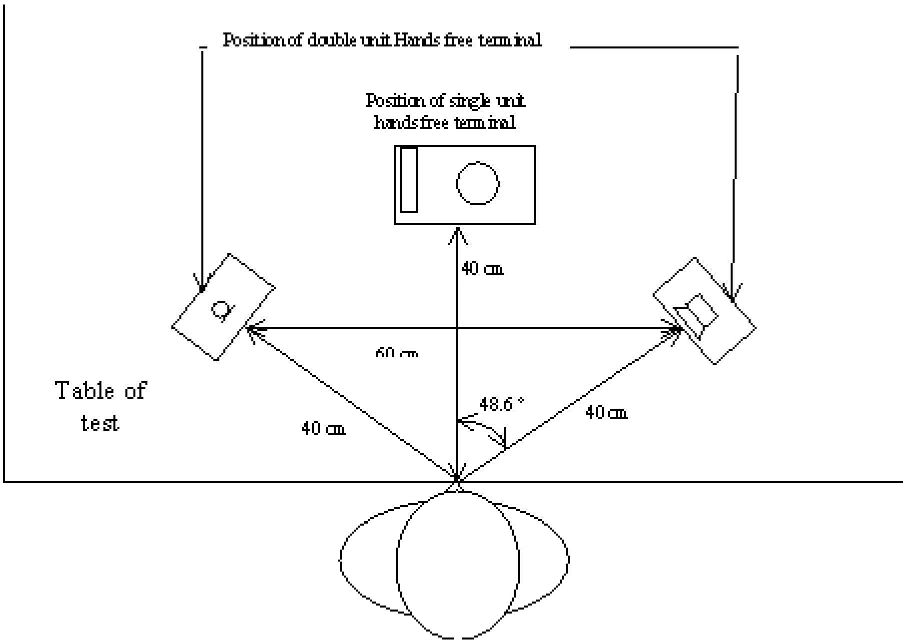

Figure 5.1.1.1-2: Position for test of desktop handsfree terminal with HATS, top view

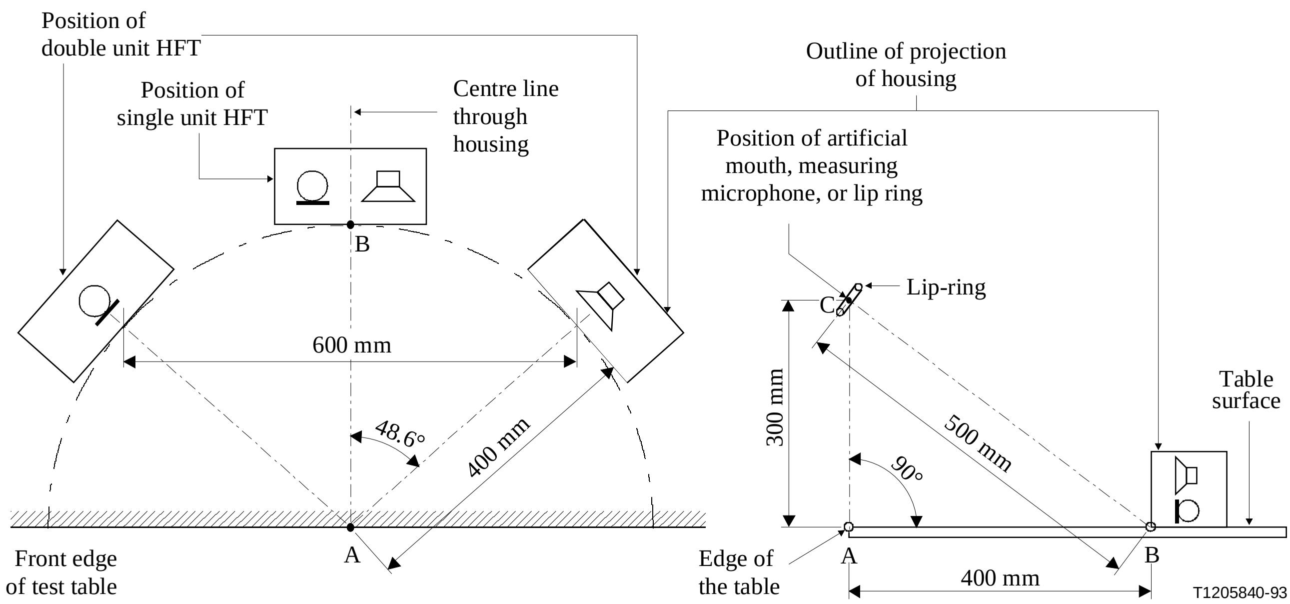

Figure 5.1.1.1-3: Position for test of desktop handsfree terminal with free-field microphone or with reference loudspeaker (from Recommendation ITU-T P.340 [6]), top and side views

5.1.1.2 Handheld handsfree terminal

This kind of terminal could implement SWB or FB; The test configuration is defined on figure 5.1.1.2-1.

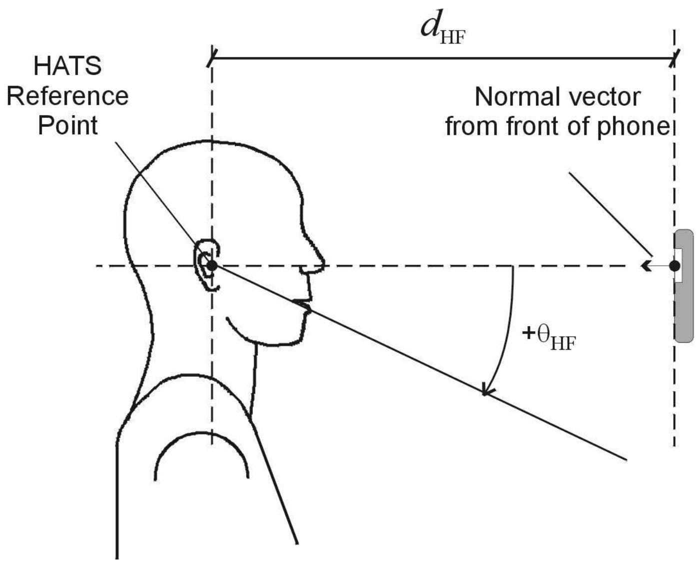

Figure 5.1.1.2-1: Configuration of Hand-Held loudspeaker relative to the HATS side view

NOTE: For a hand-held terminal using external microphone(s) the test set-up defined in clause 5.1.1.3 applies (the handheld terminal being placed at one of the locations of the loudspeaker as defined in figure 5.1.1.3-4).

5.1.1.3 Softphone (computer-based terminals)

When a manufacturer gives conditions of use, they will apply for test.

If no other requirement is given by a manufacturer, the softphone will be positioned according to the following conditions:

Softphone including loudspeakers and microphone

Two types of softphones are to be considered:

- Type 1 is to be used as a desktop type (e.g. notebook).
- Type 2 is to be used as a handheld type (e.g. PDA).

For Type 1 the configurations (side and top views) are defined in figures 5.1.1.3-1 and 5.1.1.3-2 when using HATS.

When using a free-field microphone instead of the artificial ears of HATS the centre of the microphone is placed at the point "lip ring" on figure 5.1.1.3-1.

When using a loudspeaker instead of the artificial mouth of HATS the centre of the front plane is placed at the point "lip ring" on figure 5.1.1.3-1.

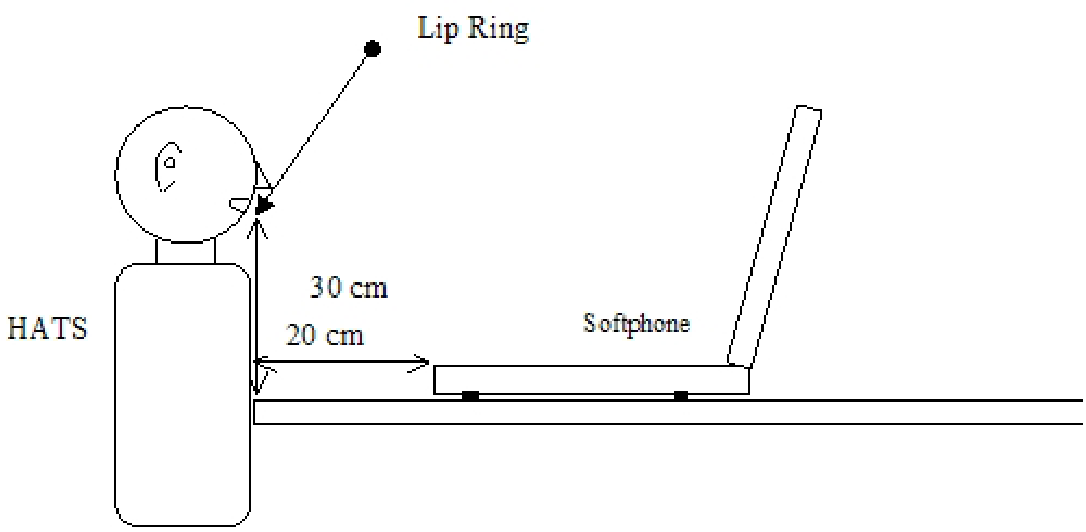

Figure 5.1.1.3-1: Configuration of softphone relative to the HATS side view

When a free-field microphone or reference loudspeaker is used instead of HATS, the microphone centre or the centre of the loudspeaker plane are positioned at the point defined as the lip ring position.

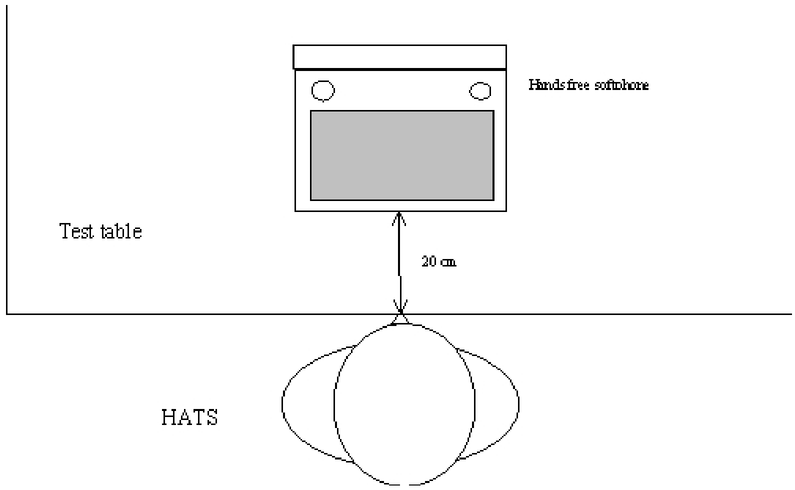

Figure 5.1.1.3-2: Configuration of softphone relative to the HATS top view

When a free-field microphone or reference loudspeaker is used instead of HATS, the microphone centre or the centre of the loudspeaker plane are positioned at the point defined as the lip ring position.

Softphone with separate loudspeakers

When separate loudspeakers are used, these loudspeakers will be positioned as in figure 5.1.1.3-3, when using HATS.

When using a free-field microphone instead of the artificial ears of HATS the centre of the microphone is placed at the point "lip ring" on figure 5.1.1.3-3.

When using a loudspeaker instead of the artificial mouth of HATS the centre of the front plane is placed at the point "lip ring" on figure 5.1.1.3-3.

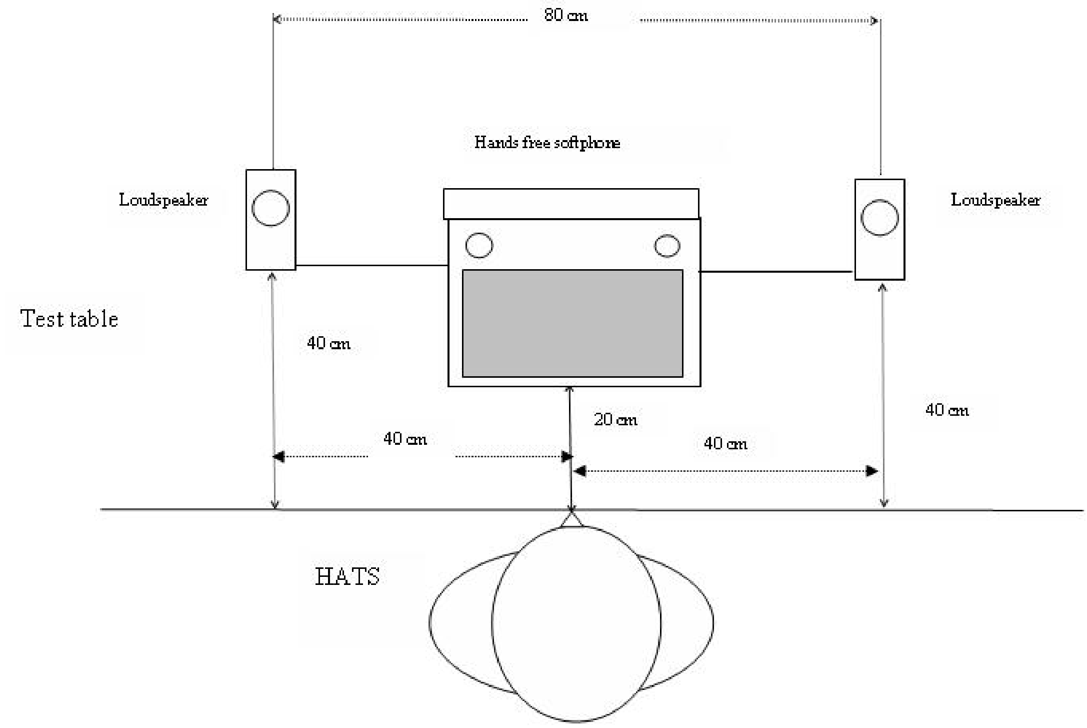

Figure 5.1.1.3-3: Configuration of softphone using external speakers relative to the HATS top sight

When a free-field microphone or reference loudspeaker is used instead of HATS, the microphone centre or the centre of the loudspeaker plane are positioned at the point defined as the lip ring position.

Softphone with separate loudspeakers and external microphone

When external microphone and loudspeakers are used, they are positioned as in figure 5.1.1.3-4, when using HATS.

When using a free-field microphone instead of the artificial ears of HATS the centre of the microphone is placed at the point "lip ring" on figure 5.1.1.3-4.

When using a loudspeaker instead of the artificial mouth of HATS the centre of the front plane is placed at the point "lip ring" on figure 5.1.1.3-4.

NOTE: For some specific applications (e.g. sound pick-up, journalist reporting), the terminal may be used with an external microphone (monaural or stereo). The test set-up as defined in figure 5.1.1.3-4 applies.

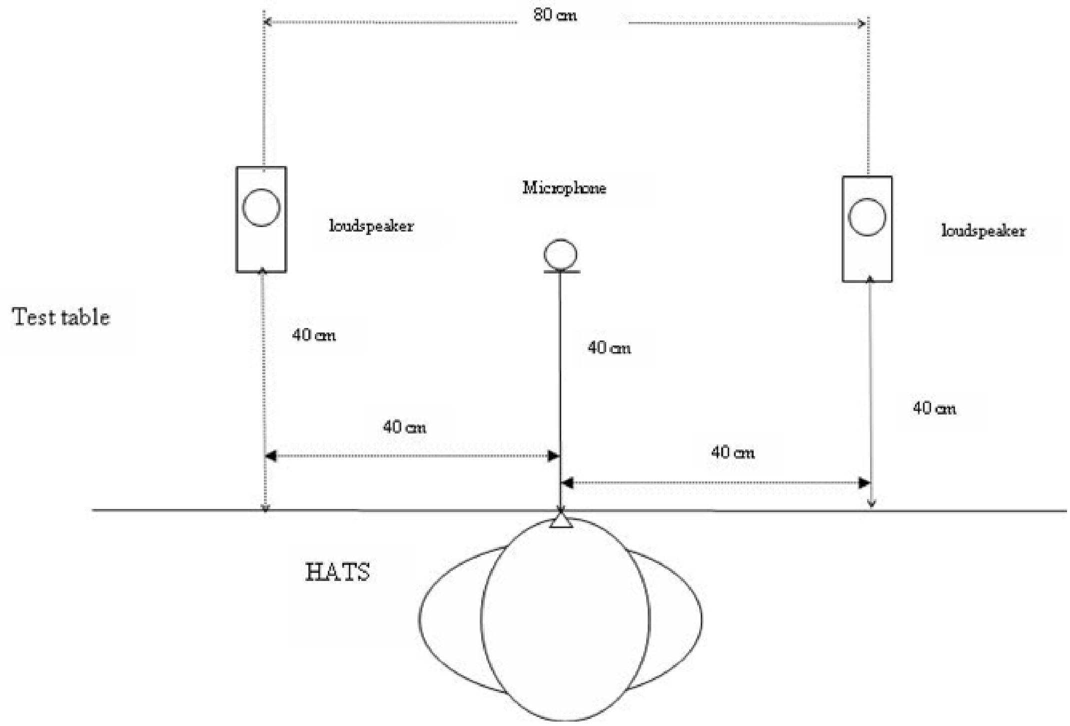

Figure 5.1.1.3-4: Configuration of softphone using external speakers and microphone relative to the HATS top sight

When a free-field microphone or reference loudspeaker is used instead of HATS, the microphone centre or the centre of the loudspeaker plane are positioned at the point defined as the lip ring position.

5.1.1.4 Group audio terminal (GAT)

The Group audio terminal as defined in the present document is considered as a "one-piece" terminal including loudspeaker/microphone in the same "box".

When supplementary microphones/loudspeakers may be added to the Group Audio Terminal, the test set-up "teleconference" should be used; as defined below.

When manufacturer's guidance defines conditions for use, these conditions apply for the test.

When no requirement from manufacturer is available, the following conditions will be used by the test laboratory.

When the Super-wideband/Fullband Group Audio terminal also implements Wideband coders, some parameters may be tested using a HATS test equipment.

Other parameters should be tested using free-field microphone and a reference loudspeaker.

Figures 5.1.1.4-1 and 5.1.1.4-2 define the test positions to be used when using HATS.

When using a free-field microphone instead of the artificial ears of HATS, the centre of the microphone is placed at the point "lip ring" on figures 5.1.1.4-1 and 5.1.1.4-2.

When using a loudspeaker instead of the artificial mouth of HATS, the centre of the front plane is placed at the point "lip ring" on figures 5.1.1.4-1 and 5.1.1.4-2.

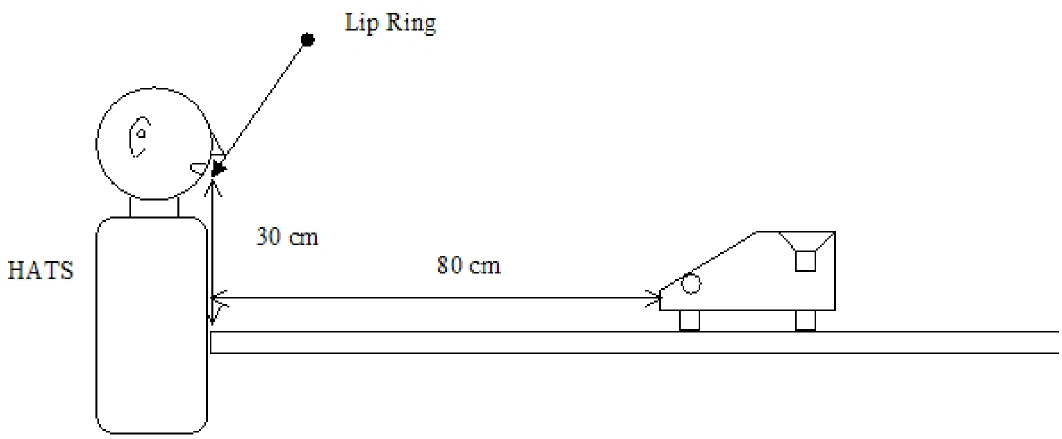

Figure 5.1.1.4-1: Configuration of group audio terminal relative to the HATS side view

When a free-field microphone or reference loudspeaker is used instead of HATS, the microphone centre or the centre of the loudspeaker plane are positioned at the point defined as the lip ring position.

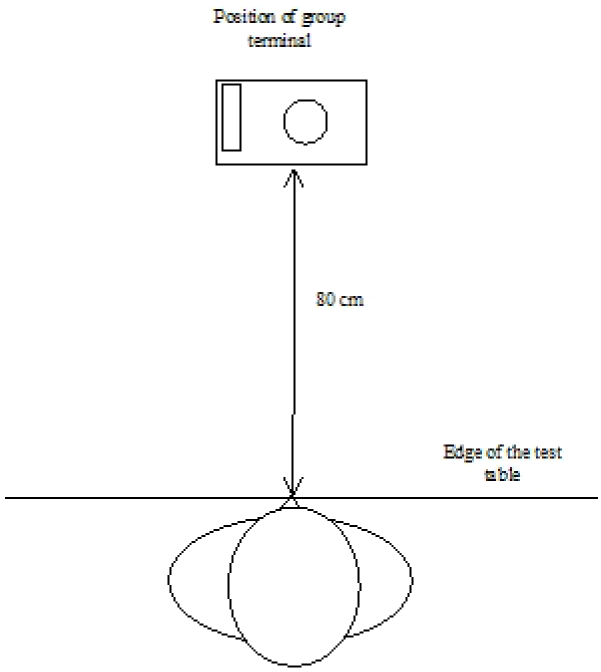

Figure 5.1.1.4-2: Configuration of group audio terminal relative to the HATS top sight

When a free-field microphone or reference loudspeaker is used instead of HATS, the microphone centre or the centre of the loudspeaker plane are positioned at the point defined as the lip ring position.

NOTE 1: In case of special casing where those conditions are not realistic, the test laboratory can use a different position more representative of real use. The conditions of test should be given in the test report.

NOTE 2: Experiences show that it should be ensured that the quality is not substantially affected when the speaker moves in front of the group audio terminal or if he turns his head. Specific arrangements should be defined to check these practical conditions.

NOTE 3: For a terminal using external microphone(s) the test set-up defined in clause 5.1.1.3 applies.

5.1.1.5 Teleconference systems

Teleconference systems may implement video and currently use multi-microphone systems and/or multi-loudspeaker systems.

For SWB teleconference systems, HATS may be used for some tests. For FB teleconference systems and for other tests of SWB teleconference systems, additional tests are conducted using freefield microphone and a high quality loudspeaker. For some specific tests, several test equipment may be used.

As there is no unique implementation, there is no standardized position(s) for free field microphone/loudspeaker(s). However, these test equipment are placed as close as possible to the users positions recommended by the manufacturers.

NOTE 1: Special cases to be considered: multichannel implementations.

NOTE 2: From the experience, it appears that one very important request for video communication is to ensure the eye-to-eye contact. This principle should be taken into account when defining the measurement positions and conditions for audiovisual communications, such as "telepresence".

If the room is designed with microphone arrangements, HATS will be placed at the users positions.

When the terminal is intended to be used for different users positions, the test is to be done at least at two or three positions (to be defined by the manufacturer or, by default, the test laboratory).

5.1.1.6 Void

5.1.2 Test signals

The test signals are defined according to Recommendation ITU-T P.501 [1] for tests made with speech signals. For some parameters it is necessary to combine speech signals with other types of signals (e.g. music, background noise) or the test signal may be an audio signal mixing any type of materials. Such signals are defined in ETSI TS 103 224 [28].

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 and 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 and 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 and switching characteristics.

For double-talk performance:

- A "double-talk" sequence representing typical double talk scenarios in real conversations is shown in table 6.3.4.2-1 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.

5.1.3 Test signal levels

5.1.3.0 General

The level dependency should be considered and consequently tests should also be done with signal levels lower and higher than the reference level defined in the following clauses.

5.1.3.1 Send

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

When using HATS it is positioned according to figure 5.1.3.1-1.

When using a reference loudspeaker its centre is positioned at the lip ring position defined in figure 5.1.3.1. The loudspeaker is intended to be free-field equalized.

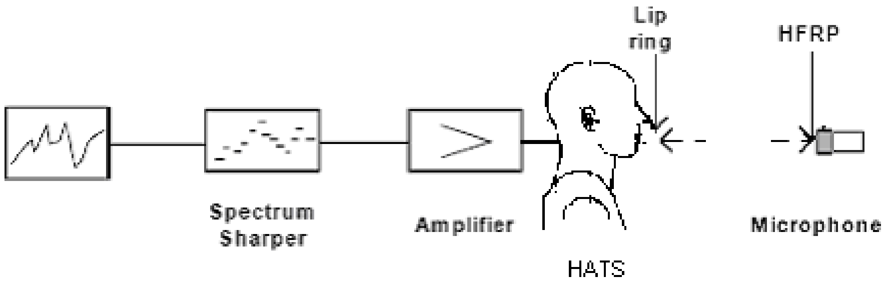

Figure 5.1.3.1-1: Calibration at HFRP (with d_{HFS} = 50 cm)

NOTE 1: The distance used for level calibration corresponds to the following values:

- Desktop terminal: 50 cm and level to adjust -28,7 dBPa.
- Handheld terminal: 30 cm with -24,3 dBPa.
- Softphone: 36 cm with -25,8 dBPa.
- Group audio terminal: 85 cm with -33,3 dBPa. (85 cm correspond to a distance of 80 cm between the table edge and the front part of the GAT).
- Teleconference systems: 100 cm with -34,7 dBPa.

Telepresence systems. The distance(s) and users position(s) have to be defined by the manufacturer.

NOTE 2: As defined in ETSI ETS 300 807 [13], in order to take into account the difference between the reference test positioning and the actual microphone-talker operating distance (d_S) for which the terminal is adjusted, the following correction factor F_S is defined:

$$F_S \text{ (dB)} = 20 \text{ Log } (d_S/0,5) \text{ } (d_S \text{ in meters}) \qquad (1)$$

The formula may be used to define the relevant level calibration for telepresence systems when using the reference signal level defined for desktop terminal.

In the formula, 0,5 meter is equal to d_{HFs} in figure 5.1.3.1-1.

5.1.3.2 Receive

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

5.1.4 Setup of background noise simulation

A setup for simulating realistic background noises in a lab-type environment is described in ETSI TS 103 224 [28].

The signals attached to ETSI TS 103 224 [28] are fullband signals and should be used for background noise simulation.

The following noises of ETSI TS 103 224 [28] shall be used.

Table 5.1.4-1: Noises used for background noise simulation

Name	Description	Length	Handsfree Levels
Pub Noise (Pub)	HATS and microphone array in a pub	30 s	1: 75,2 dB 2: 75,1 dB 3: 74,9 dB 4: 75,1 dB 5: 74,8 dB 6: 74,8 dB 7: 74,8 dB 8: 75,0 dB
Full-size car 130 km/h (FullSizeCar_130)	HATS and microphone array at co-drivers position	30 s	1: 69,5 dB 2: 68,6 dB 3: 68,6 dB 4: 68,7 dB 5: 68,8 dB 6: 68,8 dB 7: 69,2 dB 8: 69,7 dB
Cafeteria (Cafeteria)	HATS and microphone array inside a cafeteria	30 s	1: 69,0 dB 2: 69,7 dB 3: 69,6 dB 4: 69,8 dB 5: 69,5 dB 6: 69,5 dB 7: 69,7 dB 8: 70,0 dB
Roadnoise (*Roadnoise*)	HATS and microphone array standing outside near a road	30 s	1: 69,9 dB 2: 70,7 dB 3: 70,9 dB 4: 71,0 dB 5: 70,8 dB 6: 70,8 dB 7: 70,9 dB 8: 71,0 dB
Recording in airport hallway (*Airport*)	HATS and microphone array in an airport hallway with overhead public address announcement	30 s	1: 77,2 dB 2: 77,4 dB 3: 77,6 dB 4: 77,7 dB 5: 78,1 dB 6: 77,9 dB 7: 77,8 dB 8: 77,9 dB

5.1.5 Acoustic environment

5.1.5.1 Measurement environment

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

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

Unless stated otherwise measurements shall be conducted under quiet and "anechoic" conditions.

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

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

5.1.5.2 Acoustic environment for the rooms where systems are implemented

The acoustic environment may have an important influence on the quality, in particular for group audio terminals and conference systems.

Information is available in Annex A.

5.1.6 Influence of terminal delay issue 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.

5.2 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).

d) Unless specified otherwise, the background noise level shall be less than -64 dBPa(A) in conjunction with NC30 (ISO 3745 [23]).

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.

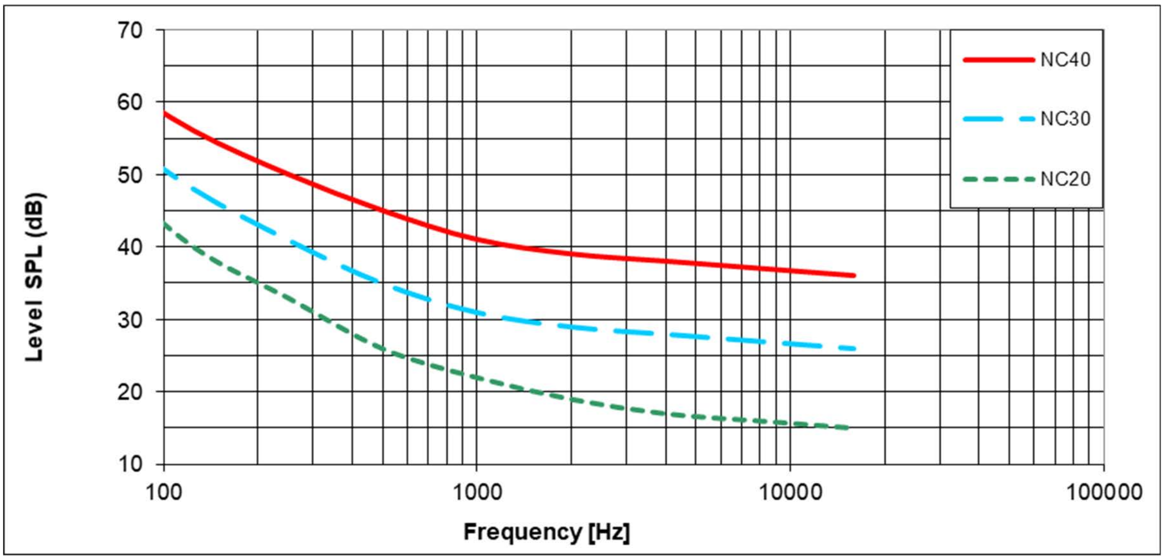

Figure 5.2-1: NC-criteria for test environment

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 5.3-1: 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 %

The values in table 5.3-1 shall apply up to the measured maximum frequency of 20 kHz.

NOTE 1: 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 5.3-2: Accuracy of test signal generation

Quantity	Accuracy
Sound pressure level at HandsFree Reference Point (HFRP)	0 to -6 dB for frequencies from 50 Hz to 100 Hz ±1 dB for frequencies from 100 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 2: With some measurement equipment the use of such a 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 Specific test considerations

5.4.0 Introduction

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

5.4.1 Loudness Rating and Loudness

5.4.1.1 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 a terminal implements wideband speech in addition with Super-wideband or fullband functions, or is intended to communicate with wideband terminals, the terminal 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 Super-wideband or fullband loudness ratings.

NOTE: RLR and SLR, values are similar or derived from those defined in ETSI ES 202 740 [11] and ETSI TS 103 740 [12].

5.4.1.2 Loudness

Loudness quantifies the level as perceived by the user and should be more relevant when the signal combines speech and audio sequences for Super-wideband and fullband. ISO 532 [i.3] 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 also takes into account binaural listening. Loudness may be calculated for any type of signal (speech, music and noise) and mixed signals.

Standardized audio and speech signals are defined in Recommendation ITU-T P.501 [1], ETSI TS 103 224 [28].

When the terminal provides Super-wideband or fullband in addition with wideband or narrowband the reference loudness value (expressed in phons) should be determined for narrowband or wideband transmission.

If the superwideband and fullband terminals do not support wideband transmissions, standardized loudness levels have to be defined. This is for further study.

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

5.4.2 Binaural listening

The scope of the present document includes terminals that may have two or more microphones and two or more loudspeakers.

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

NOTE: Loudness calculation should be based on binaural listening.

5.4.3 Subjective considerations

Recommendation ITU-T P.1301 [16] defines the subjective quality evaluation of audio and audiovisual multiparty telemeetings:

"This recommendation concerns subjective quality assessment of telemeeting systems that provide multiparty communication between distant locations, using audio-only, video-only, audiovisual, text-based or graphical means as communication modes. The term multiparty refers to more than two meeting participants who can be located at two or more than two locations.

Evaluation of those systems can focus on audio-only, video-only or audiovisual quality aspects and non-interactive or conversational quality can be assessed.

This recommendation gives an overview of relevant aspects that need to be considered for subjective quality evaluation of multiparty telemeetings and it provides guidance to recommendations describing the details of applicable methods and procedures. Aspects in this recommendation are also applicable to two-party telemeetings".

In addition to this methodology, it should be needed to add some new perceptual criteria, such as Intelligibility, naturalness, etc. that should be improved for Super-wideband and fullband terminals compared to wideband terminals.

5.4.4 Setup of variable echo path

Test setup for desktop hands free terminals: A notebook is positioned at least 20 cm in front of the device (or devices) with the transducers, as shown in figure 5.4.4-1. The notebook lid is moved during the measurement.

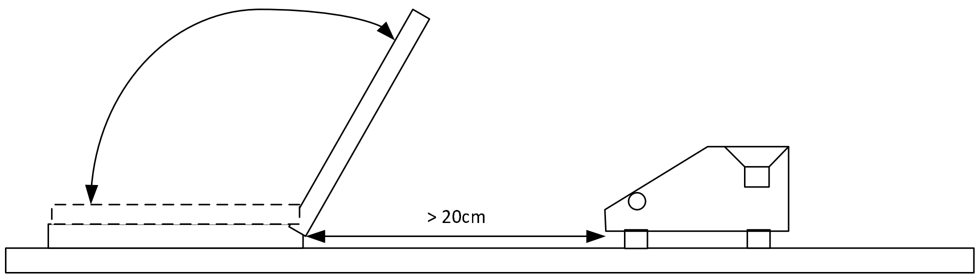

Figure 5.4.4-1: Positioning of DUT

Test setup for softphone: The test setup is described in clause 5.1.1.3. The notebook lid is moved during the measurement, as shown in figure 5.4.4-2. This setup is valid for all combinations of notebook with or without external speakers or microphone.

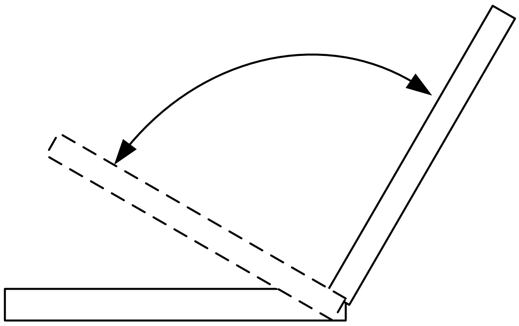

Figure 5.4.4-2: Positioning of DUT

Test setup for other handsfree devices is for further study.

5.5 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.5] (https://www-x.antd.nist.gov/itg/nistnet/) or NIST Netem™ [i.6].

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 description 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).

- 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).

5.6 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 clause 5.2.

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 per figure 5.2-1, the octave levels of the room noise are determined from 63 Hz to 16 kHz.

6 Requirements and associated measurement methodologies

6.0 Considerations

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

NOTE: 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.

6.1 Send

6.1.0 Introduction

All the types of terminals within the scope of the present document shall fulfill the requirements of this clause. Even if these terminals are rather different, the intention of the present document is to guarantee that all the terminals effectively transmit Super-wideband and/or fullband bandwidths.

6.1.1 Frequency response

Requirements

The objective is to define a flat frequency curve over the whole bandwidth.

The frequency response for Super-wideband shall fulfill the mask as defined in table 6.1.1-1 and figure 6.1.1-1.

Table 6.1.1-1: Frequency mask for Super-wideband terminals - Send

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.		

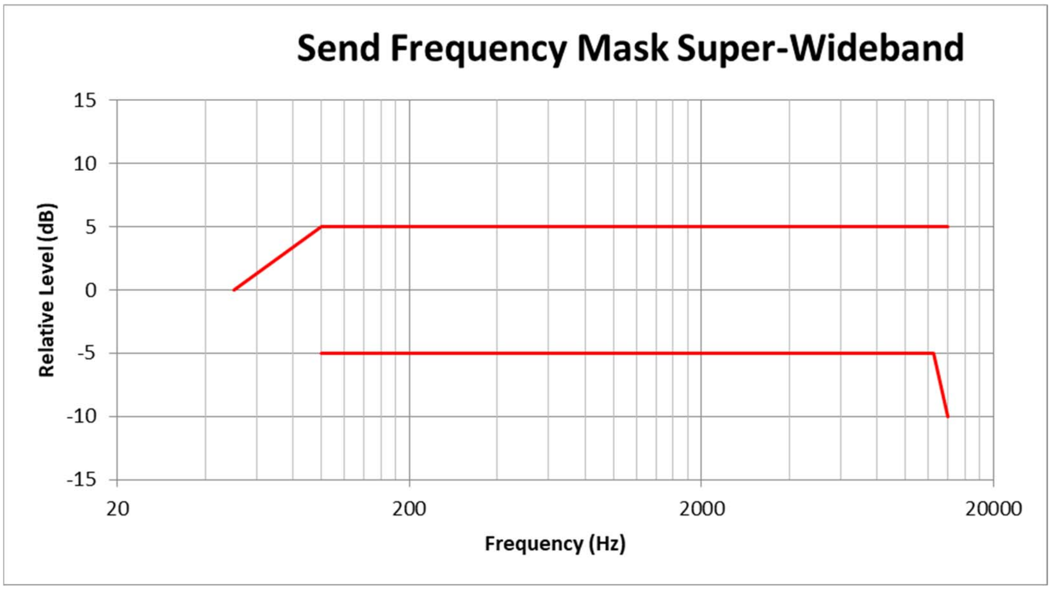

Figure 6.1.1-1: Send frequency response mask for super-wideband

The frequency response for fullband shall fulfill the mask as defined in table 6.1.1-2 and figure 6.1.1-2.

Table 6.1.1-2: Frequency mask for fullband terminals - Send

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
20	0	
50	0	-10
100	5	-5
12 500	5	-5
16 000	5	-5
20 000	5	
NOTE: All sensitivity values are expressed in dB on an arbitrary scale. The requirement is based on 1/12th octave measurement.		

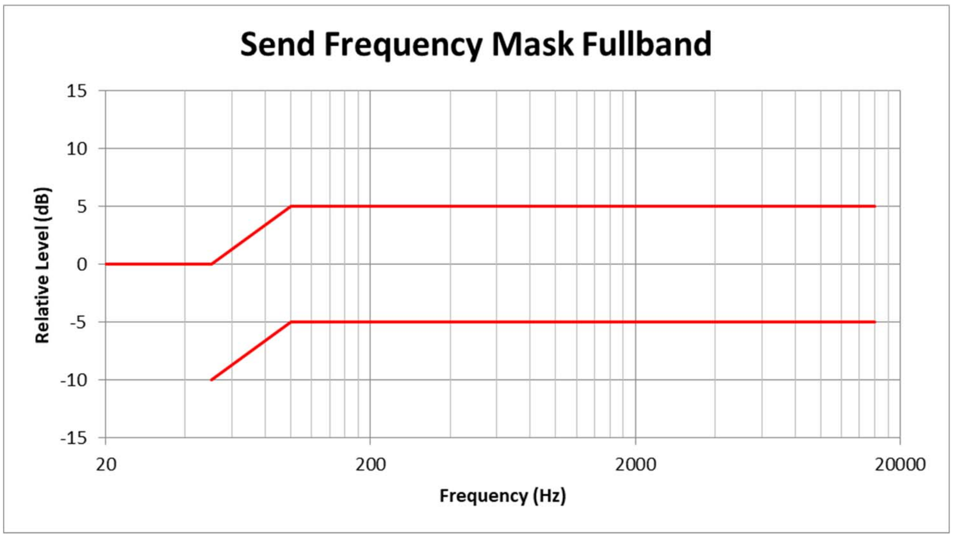

Figure 6.1.1-2: Send frequency mask for fullband

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

Additional requirements are for further study when the system is intended to be used by several users, when stereo features are made available or when microphone array(s) are used.

Measurement Method

The terminal is set according to clause 5.1.1. The test signal is defined in clause 5.1.2. The test signal level is defined according to clause 5.1.3.

Measurements shall be made at one twelfth-octave intervals as given by IEC 61260-1 [22] for frequencies from 50 Hz to 14 kHz inclusive for SWB and from 20 Hz to 20 kHz inclusive for FB.

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 HFRP.

The sensitivity is expressed in terms of dBV/Pa.

6.1.2 Send Loudness rating (SLR)

Requirements

To ensure the compatibility with other terminals or systems a reference SLR needs to be defined.

The requirements refer to wideband handsfree terminals, ETSI TS 103 740 [12].

Nominal value: +13dB ± 3 dB.

There is no specific requirement for SWB or FB bandwidth.

Measurement method (of Wideband Loudness rating):

The terminal is positioned as described in clause 5.1.1.

For a correct activation of the system, 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 acoustic signal produced by the artificial mouth is calibrated under free field conditions at the MRP. The test signal level shall be -4,7 dBPa, measured at the MRP. The test signal level is averaged over the complete test signal sequence.

Calibration is realized as explained in clause 5.1.3.

The send 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], Annex A.

6.1.3 Void

6.1.4 Send Noise

Requirements

The limit for the send noise is the following:

- send noise level maximum -64 dBm(A).

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

NOTE 1: Softphones with cooling devices (fans) can produce a rather high level of noise; largely dependent of the activity of system.

Measurement method

The terminal is set-up according to clause 5.1.1.

The female speaker 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 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 2: 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.

6.1.5 Send Distortion

6.1.5.1 Signal to harmonic distortion versus frequency

Requirements

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

The following draft requirements are defined for all the terminals within the scope of the present document, as it is needed to ensure that any terminal intended to be used in Super-wideband and fullband sends good quality signals. Care should be taken on the distortion of the HATS or of the loudspeaker used to test the send distortion of the terminal.

For Super-wideband

Table 6.1.5.1-1

Frequency	Ratio
100 Hz	25 dB
200 Hz	30 dB
400 Hz	30 dB
1 kHz	30 dB
2 kHz	30 dB
3,15 kHz	30 dB
5 kHz	30 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 signal to harmonic distortion ratio is measured selectively up to 16 kHz.

For Fullband

Table 6.1.5.1-2

Frequency	Ratio
100 Hz	25 dB
200 Hz	30 dB
400 Hz	30 dB
1 kHz	30 dB
2 kHz	30 dB
3,15 kHz	30 dB
5 kHz	30 dB
8 kHz	30 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 signal to harmonic distortion ratio is measured selectively up to 20 kHz.

Measurement Method

The terminal is set according to clause 5.1.1.

For Super-wideband terminal, the signal used is an activation signal followed by a series sine wave signal with a frequency at 100 Hz, 200 Hz, 400 Hz, 1 kHz, 2 kHz, 3,15 kHz and 5 kHz. The signal to harmonic distortion ratio is measured selectively up to 14 kHz.

For fullband terminal, the signal used is an activation signal followed by a series sine wave signal with a frequency at 100 Hz, 200 Hz, 400 Hz, 1 kHz, 2 kHz, 3,15 kHz, 5 kHz and 8 kHz. The signal to harmonic distortion ratio is measured selectively up to 18 kHz.

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.

For a correct activation of the system, 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 1: Depending on the type of codec or signal processing the test signal used may need to be adapted.

NOTE 2: When using HATS, due to the distortion limit of the artificial mouth at 100 Hz, as defined in table 10 of Recommendation ITU-T P.58 [3], the measurements with a frequency of 100 Hz and possibly 200 Hz have to take into account the actual distortion of the artificial mouth.

6.1.5.2 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 terminal is set according to clause 5.1.1.

For Super-wideband and fullband terminal, the signal used is an activation signal followed by a series sine wave signal with a frequency at 1 kHz. 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 less than 1 s. The sinusoidal signal level shall be calibrated to +10 dBPa at the MRP.

For a correct activation of the system, 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 or signal processing the test signal used may need to be adapted.

6.1.6 Microphone mute

Requirement

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

Measurement method

The terminal will be positioned as described in clause 5.1.

For a correct activation of the system, 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 acoustic signal produced by the artificial mouth is calibrated under freefield conditions at the MRP. The test signal level shall be -4,7 dBPa, measured at the MRP. The test signal level is averaged over the complete test signal sequence.

Calibration is realized as explained in clause 5.1.3.

The send 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], Annex A.

6.2 Receive

6.2.0 Introduction

The scope of the present document includes a lot of different types of handsfree terminals. The receive performance may significantly depend on the size and on the application of the terminal. In the following clause the requirements are defined for the different types of terminals within the scope of the present document.

6.2.1 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 terminals may implement an equalization function adjusting frequency response according to user's preference.

When such a function is available it is necessary that the receive frequency response conforms to requirements defined in clause 6.2.2 for at least the default equalization setting.

6.2.2 Frequency response

6.2.2.0 General

When using HATS (with the restrictions defined in clause 5) HATS shall be equalized according to Recommendation ITU-T P.581 [4].

However, at least for fullband terminals, it is recommended to use free-field microphones instead of the HATS.

6.2.2.1 Handheld terminal

Requirements

Superwideband

Table 6.2.2.1-1: Frequency mask for Superwideband handheld terminals - Receive

Frequency	Upper Limit	Lower Limit
50 Hz	5 dB	
400 Hz	5 dB	-5 dB
12 500 Hz	5 dB	-5 dB
14 000 Hz	5 dB	-10 dB
16 000 Hz	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.		

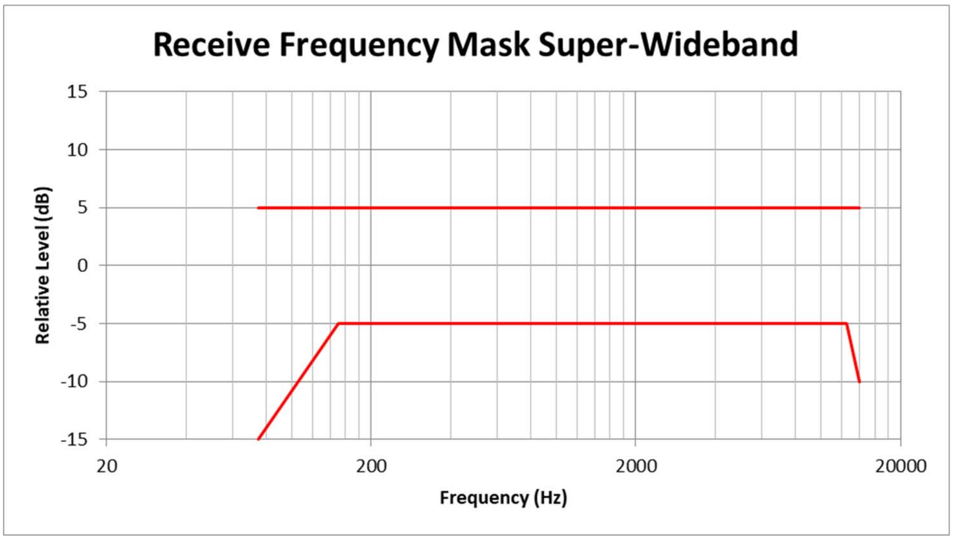

Figure 6.2.2.1-1: Frequency mask for superwideband handheld terminals - Receive

Fullband

Table 6.2.2.1-2: Frequency mask for fullband handheld terminals - Receive

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
20	5	
400	5	-10
500	5	-5
12 500	5	-5
14 000	5	-5
16 000	5	-5
20 000	5	
NOTE: All sensitivity values are expressed in dB on an arbitrary scale.		

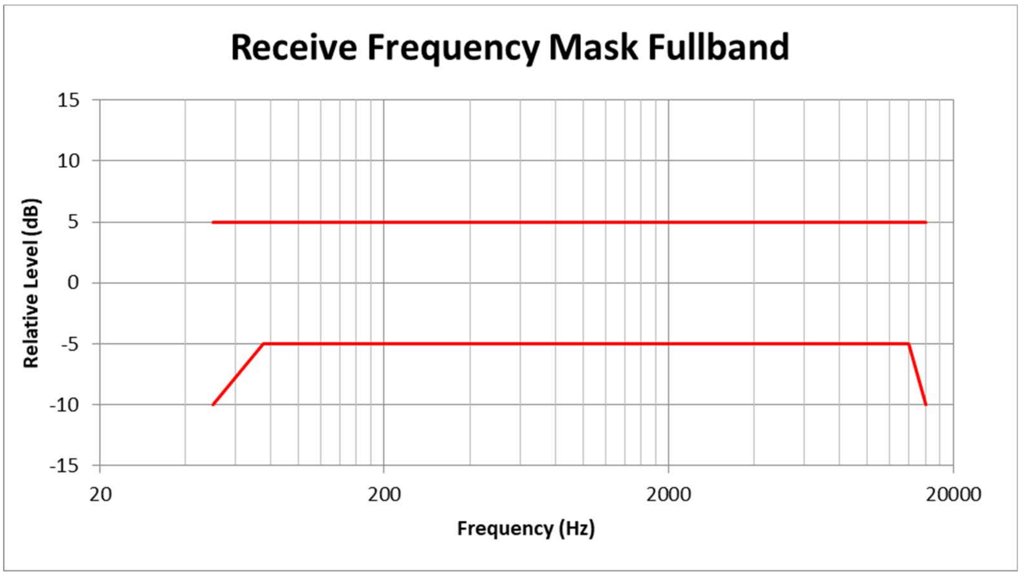

Figure 6.2.2.1-2: Frequency mask for fullband handheld terminals - Receive

Measurement methods

The terminal is set according to clause 5.1.1.

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, Amendment 1 [1]. The test signal level shall be -16 dBm0, measured according to Recommendation ITU-T P.56 [21] at the digital reference point or the equivalent analogue point.

The equalized output signal is power-averaged on the total time of analysis. The 1/3 octave band data are considered as the input signal to be used for calculations or measurements.

For superwideband terminals measurements shall be made at one third-octave intervals as given by the R.40 series of preferred numbers in ISO 3745 [23] for frequencies from 400 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 third-octave intervals as given by the R.40 series of preferred numbers in ISO 3745 [23] for frequencies from 400 Hz to 16 kHz inclusive. For the calculation the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

The sensitivity is expressed in terms of dBPa/V.

6.2.2.2 Desktop terminal

Requirements

Super-wideband

Table 6.2.2.2-1: Frequency mask for Super-wideband desktop terminals - Receive

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
75	5	-15
150	5	-5
12 500	5	-5
14 000	5	-10
NOTE: All sensitivity values are expressed in dB on an arbitrary scale.		

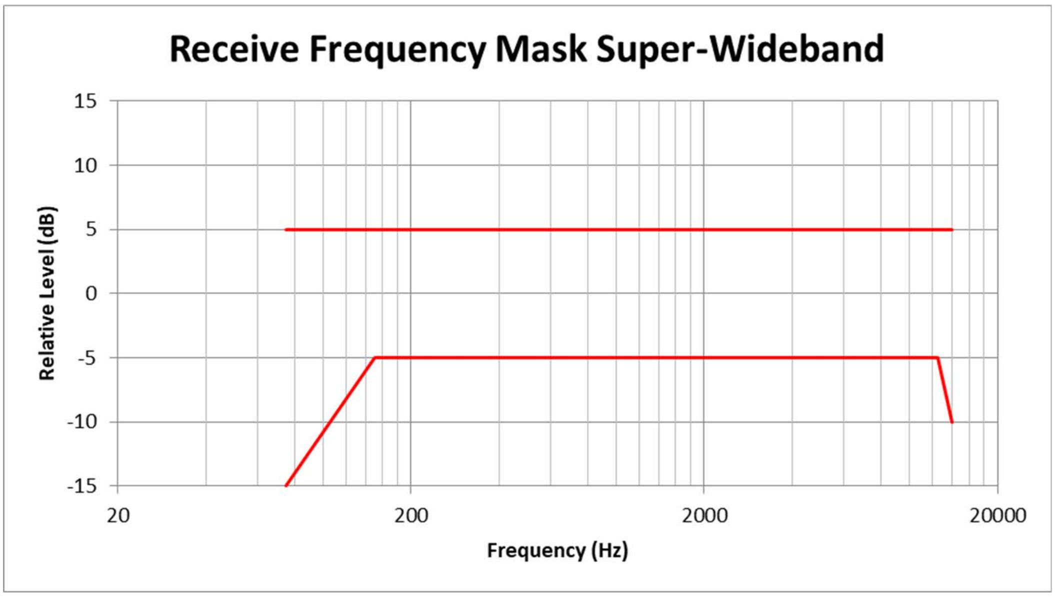

Figure 6.2.2.2-1: Frequency mask for Super-wideband desktop terminals - Receive

Fullband

Table 6.2.2.2-2: Frequency mask for fullband desktop terminals - Receive

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
50	5	-10
75	5	-5
14 000	5	-5
16 000	5	-10
NOTE: All sensitivity values are expressed in dB on an arbitrary scale.		

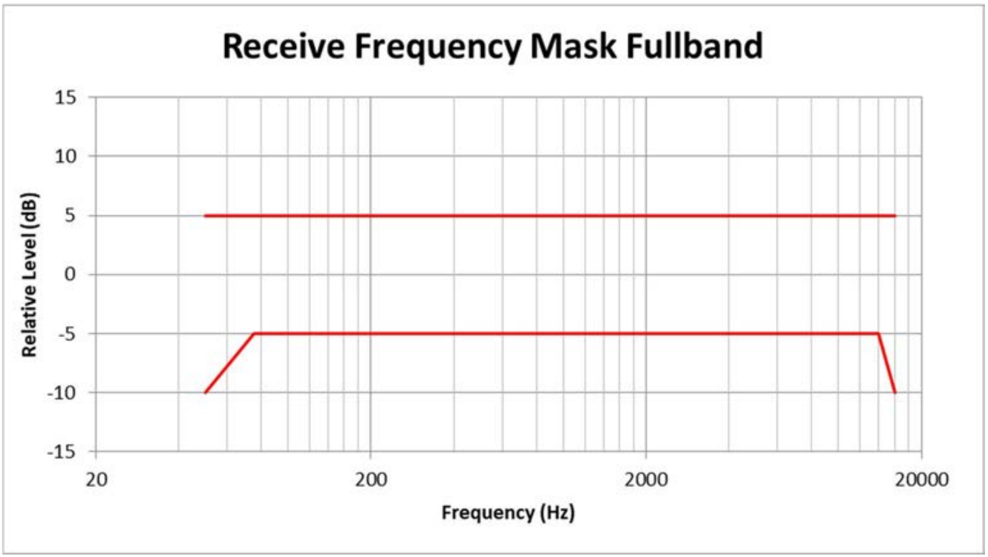

Figure 6.2.2.2-2: Frequency mask for fullband terminals – Receive

Measurement methods

The terminal is set according to clause 5.1.1.

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

$$S_{Jeff} = 20 \log (p_{e_{ff}} / v_{RCV}) \text{ dB rel 1 Pa / V} \qquad (2)$$

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

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]. The test signal level shall be -16 dBm0, measured according to Recommendation ITU-T P.56 [21] at the digital reference point or the equivalent analogue point.

The equalized output signal is power-averaged on the total time of analysis. The 1/3 octave band data are considered as the input signal to be used for calculations or measurements.

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

For Super-wideband terminals, measurements shall be made at one third-octave intervals as given by IEC 61260-1 [22] for frequencies from 75 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 third-octave intervals as given by IEC 61260-1 [22] for frequencies from 50 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.

6.2.2.3 Terminals intended to be used simultaneously by several users

Additional requirements to be defined:

- when the terminal is intended to be used by several users;

- when stereo features are made available.

For all the testing positions the frequency curve shall fulfil the requirements defined for desktop terminals.

6.2.3 Receive Loudness Rating (RLR) and Loudness

6.2.3.1 Receive Loudness Rating

Requirements

When terminal implements wideband speech functions or when the super-wideband/fullband functions may interact with wideband terminals, the handsfree terminal shall fulfill the requirements on RLR as defined in ETSI TS 103 740 [12], clause 7.1.7:

- Desktop terminal:
    - Nominal value of RLR will be 5 dB ± 3 dB. This value has to be fulfilled for one position of volume range.
    - The value of RLR at the upper part of the volume range shall be less than (louder) or equal to -2 dB: RLR ≤ -2 dB.
    - The range of volume control shall be ≥15 dB.
- Handheld terminal:
    - Nominal value of RLR will be 9 dB ± 3 dB. This value has to be fulfilled for one position of volume range.
    - Value of RLR at upper part of volume range shall be less than (louder) or equal to 5 dB: RLR ≤ 5 dB.
    - Range of volume control shall be equal or exceed 15 dB.
- Softphone (computer-based terminal):
    - Type 1 or softphone with external speakers: requirement as for desktop terminal.
    - Type 2 requirement as for handheld terminal.
- Group audio terminal:
    - Nominal value of RLR will be 5 dB ± 3 dB. This value has to be fulfilled for one position of volume range.
    - Value of RLR at upper part of volume range shall be less than (louder) or equal to -6 dB: RLR ≤ -6 dB.
    - Range of volume control shall be ≥19 dB.

NOTE 1: Due to the lack of experience in the application of wide band loudness rating calculation as defined in Annex G of Recommendation ITU-T P.79 [5] the loudness rating calculation as described in Annex A is used.

NOTE 2: Loudness Rating measurement corresponding to level with speech signal, it can be considered that a measurement in wideband may be sufficient. Indeed, energy of speech beyond bandwidth of wideband is rather small.

NOTE 3: Receive Loudness Rating for stereo/dichotic is for further study.

Measurement Method

The test set-up is described in clause 5.1

The measurement is conducted at nominal volume control setting.

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]. The test signal level shall be -16 dBm0, measured according to Recommendation ITU-T P.56 [21] at the digital reference point or the equivalent analogue point. The test signal level is averaged over the complete test signal sequence.

The HATS is free field equalized as described in Recommendation ITU-T P.581 [4]. The equalized output signal is power-averaged on the total time of analysis. The 1/3 octave band data are considered as the input signal to be used for calculations or measurements.

Measurements shall be made at one third-octave intervals as given by IEC 61260-1 [22] for frequencies from 100 Hz to 8 kHz inclusive. 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], Annex A. The RLR shall then be corrected as RLR minus 14 dB according to Recommendation ITU-T P.340 [6] and without the L_E factor.

6.2.3.2 Loudness

The implementation of loudness measurements is very important for the receive part of the terminal.

So, for the super-wideband and fullband transmissions it should be relevant to determine the loudness of the signals delivered by the terminal. When the terminal provides Super-wideband or fullband in addition with wideband the reference loudness value could be determined for Wideband transmission and to align the loudness in super-wideband or fullband with this reference. When the super-wideband and fullband terminals do not support wideband transmissions, it should be needed to define standardized loudness levels.

NOTE 1: The requirements and test methods are for further study.

NOTE 2: Preliminary results on Loudness measurements are available in STQ(12)40_26 [i.4], and detailed values are available in ISO 532 [i.3].

6.2.4 Receive Noise

Requirements

- A-weighted

The noise level measured until 10 kHz shall not exceed -54 dBPa(A) at nominal setting of the volume control.

- Third-octave band spectrum.

For SWB: The level in any 1/3-octave band, between 50 Hz and 12,5 kHz shall not exceed a value of -64 dBPa.

For FB: The level in any 1/3-octave band, between 50 Hz and 16 kHz shall not exceed a value of -64 dBPa.

NOTE 1: No peaks in the frequency domain higher than 10 dB above the average noise spectrum should occur.

NOTE 2: For softphone, fan noise should be avoided in order to fulfill this condition.

Measurement Method

The terminal is set according to clause 5.1.1.

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 for activation. Level of this activation signal will be -16 dBm0.

For the A-weighted noise level measurement the noise level is measured at DRP of the artificial ear until 14 kHz for Super-wideband terminal and until 18 kHz for fullband terminal. Freefield equalization shall be used.

For the 1/3 octave band spectrum the level is measured in all the 1/3-octave bands, between 50 Hz and 12,5 kHz in SWB and between 50 Hz and 16 kHz in FB.

The noise shall be measured just after interrupting the activation signal.

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) 1/3rd 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 3: 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.

6.2.5 Receive Distortion

Requirements

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

Table 6.2.5-1: Receive distortion for super-wideband and fullband

Frequency	Signal to distortion ratio limit, receive for desktop terminal (SWB)	Signal to distortion ratio limit, receive for desktop terminal (FB)
100 Hz		20 dB
200 Hz	20 dB	22 dB
315 Hz	26 dB	26 dB
400 Hz	30 dB	30 dB
500 Hz	30 dB	30 dB
800 Hz	30 dB	30 dB
1 kHz	30 dB	30 dB
2 kHz	30 dB	30 dB
5 kHz	30 dB	30 dB
8 kHz	30 dB	30 dB
NOTE: The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) - logarithmic (Hz) scale.		

In low frequencies, GAT and telemeeting terminals should have higher signal to noise ratio than desktop terminals, to guarantee a better use of these terminals.

The requirements defined above apply to speech transmission only. Higher values for signal to distortion ratio are for terminals, also intended to transmit audio signals, e.g. music. This is for further study.

Measurement Method

The terminal is set according to clause 5.1.1.

The signal used is an activation signal followed by a sine wave signal with a frequency at 100 Hz, 200 Hz, 315 Hz, 400 Hz, 500 Hz, 800 Hz, 1 kHz, 2 kHz, 5 kHz and 8 kHz. The duration of the sine wave shall be of less than 1 s. Appropriate signals for activation and signal combinations can be found in Recommendation ITU-T P.501 [1]. The sinusoidal 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. Level of this activation signal shall be -16 dBm0.

The signal to harmonic distortion ratio is measured selectively up to 14 kHz for Super-wideband terminal and up to 18 kHz for fullband terminal.

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.

6.3 Other parameters

6.3.0 Introduction

The interest of such types of terminals is to provide a very high quality. The parameters to be defined in this clause are intended to guarantee that this expected quality is effectively offered to the user(s).

6.3.1 Round-trip Delay

6.3.1.1 Round-trip Delay for VoIP terminals

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 fulfills 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 [32]). From the users perspective, a value less than 50 ms (category A in Recommendation ITU-T P-1010 [32]) 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:

$$T_s + t_{System} \qquad (3)$$

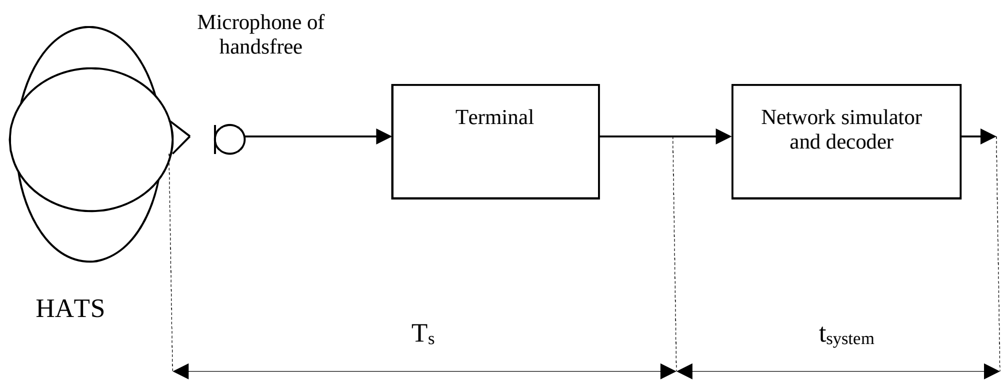

Figure 6.3.1.1-1: Different blocks contributing to the delay in send direction

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

1) For the measurements a Composite Source Signal (CSS) according to Recommendation ITU-T P.501 [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 5.1.

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:

$$T_r + t_{System} \tag{4}$$

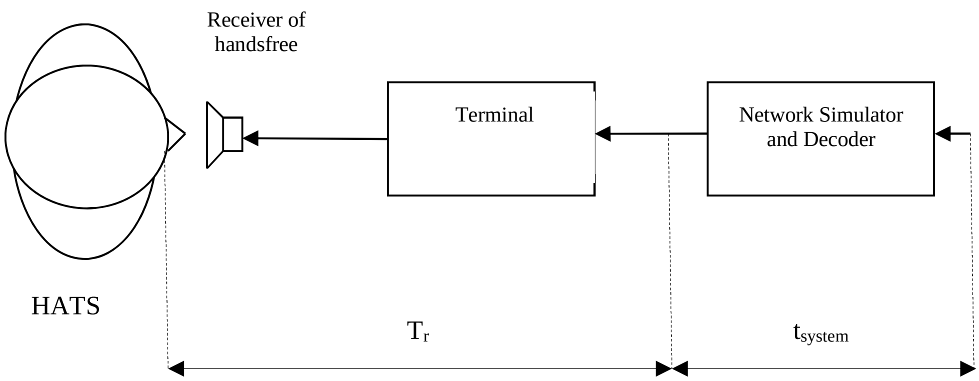

Figure 6.3.1.1-2: Different blocks contributing to the delay in receive direction

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

1) For the measurements a Composite Source Signal (CSS) according to 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 5.1.

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.

6.3.1.2 Quality of jitter buffer adjustment

Requirements

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

Table 6.3.1.2-1: Requirements for variant network impairments

Codec	MOS-LQO$_f$
ETSI TS 126 441 [25]	TBD
Recommendation ITU-T G.722.1 [7] Annex C	TBD
Recommendation ITU-T G.729.1 [8] Annex E (extension SWB)	TBD
Recommendation ITU-T G.718 [9] Annex B	TBD
Recommendation ITU-T G.711.1 [15], Annexes D and F	TBD
Recommendation ITU-T G.722 [20] Annexes B and D	TBD
Opus [30]	TBD

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 [27], 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.

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.

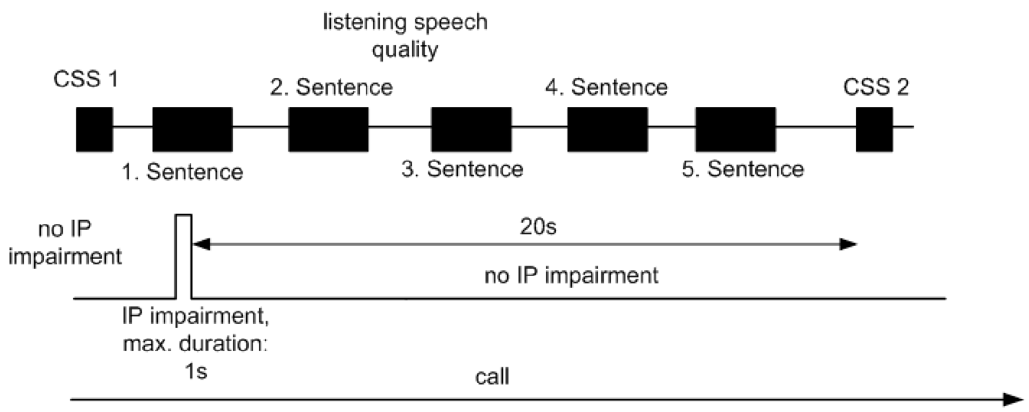

Figure 6.3.1.2-1: Test sequence to measure quality of jitter buffer adjustment (with 1 of 8 sentences)

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 202 740 [11].

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).

6.3.2 Terminal Echo Loss (TCL)

Requirement:

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

Measurement method:

The handsfree terminal is set-up as described in clause 5.1.

For hands-free measurement, HATS is positioned but not used.

For loudspeaking measurement, handset is positioned on HATS (right ear).

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:

$$L_e = C - 10\log_{10}\sum_{i=1}^{N}(A_i + A_{i-1})(\log_{10} f_i - \log_{10} f_{i-1}) \quad (5)$$

and

$$C = 10\log_{10}(2\,(\log_{10} f_N - \log_{10} f_0) \quad (6)$$

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 (5) 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.

6.3.3 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 [27]. 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 [27]):

- Fullband Context (MOS-LQO$_f$).

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

Table 6.3.3-1: 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 (1-5) 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 6.3.1). A small additional tolerance takes into account the variable behaviour of the delay.

6.3.4 Double talk performance

6.3.4.1 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 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 [6] 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.

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

Requirement

Based on the level variation in send direction during double talk $A_{H,S,dt}$ the behaviour of the terminal can be classified according to table 6.3.4.2-1.

Table 6.3.4.2-1

Category (according to Recommendation ITU-T P.340 [6])	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 6.3.4.2-1 provides a quality classification of terminals regarding double talk performance. However, this does not mean that a terminal which is category 1 based on the double talk performance is of high quality concerning the overall quality as well.

This measurement shall be done for the desktop hands free terminals and soft phones also with variable echo path.

Measurement Method

To assess double talk performance, the signals to be used are defined in Recommendation ITU-T P.501 [1]: 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 A "double-talk" sequence representing typical double talk scenarios in real conversations is shown in figure 6.3.4.2-1. 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.

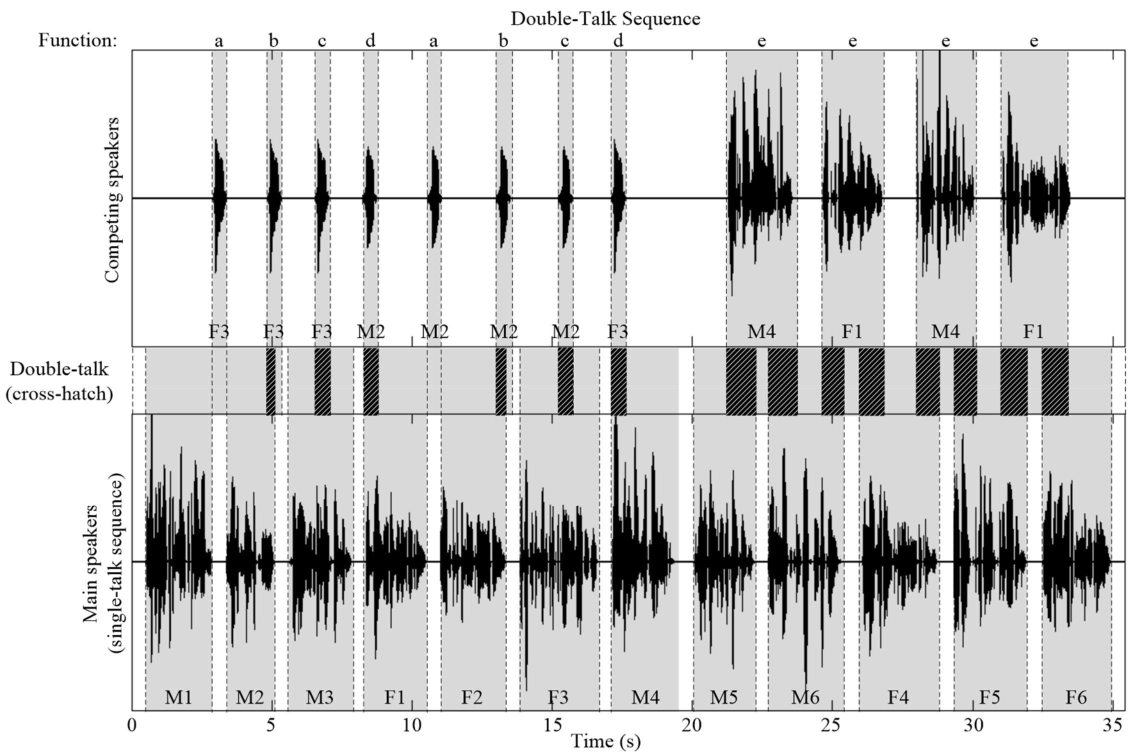

Figure 6.3.4.2-1: Double-talk test sequence using the single-talk sequence and competing speech serving different functions (a - e)

NOTE: Cross-hatched areas between the upper and lower panes show periods of double talk.

The competing-speaker sequence includes single words (the word "five") spoken by speakers F3 and M2 during the first half of the sequence followed by full sentences by speakers F1 and M4 during the second half of the sequence. No speaker is competing with themselves during the sequence.

The competing samples serve different double-talk functions, defined as functions "a" to "e" above the upper pane of figure 6.3.4.2-1. The functions are:

a) Competing word within a speech pause.

b) Competing word partially masked.

c) Competing word fully masked within a sentence.

d) Competing word fully masked coincident with the start of a sentence.

e) Sentence masking another sentence.

These are meant to represent possible double-talk situations in normal conversation. The area between the upper and lower pane of figure 6.3.4.2-1 shows the periods during which double-talk happens as cross-hatched patches. The competing sequence can be used either as a send signal or a receive signal in testing.

6.3.4.3 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 6.3.4.3-1.

Table 6.3.4.3-1

Category (according to Recommendation ITU-T P.340 [6])	**1**	**2a**	**2b**	**2c**	**3**
	Full Duplex Capability	*Partial Duplex Capability*			*No Duplex Capability*
$A_{H,R,dt}$ [dB]	*≤ 3*	*≤ 5*	*≤ 8*	*≤ 10*	*> 10*

In general table 6.3.4.3-1 provides a quality classification of terminals regarding double talk performance. However, this does not mean that a terminal which is category 1 based on the double talk performance is of high quality concerning the overall quality as well.

This measurement shall be done for the desktop hands free terminals and soft phones also with variable echo path.

Measurement method

The test setup is described in clause 5.1.

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 clause 6.3.4.2. 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 attenuation range during double talk is determined as described in Appendix III of 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.

6.3.4.4 Detection of echo components during double talk

Requirement

"Echo Loss" (EL) is the echo suppression provided by the terminal measured at the electrical reference point. Under these conditions the requirements given in table 6.3.4.4-1 are applicable (more information can be found in Annex A of Recommendation ITU-T P.340 [6].

Table 6.3.4.4-1

Category (according to Recommendation ITU-T P.340 [6])	**1**	**2a**	**2b**	**2c**	**3**
	Full Duplex Capability	*Partial Duplex Capability*			*No Duplex Capability*
Echo Loss [dB]	*≥ 27*	*≥ 23*	*≥ 17*	*≥ 11*	*< 11*

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

This measurement shall be done for the desktop hands free terminals and soft phones also with variable echo path.

Measurement method

The test setup is described in clause 5.1.

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 6.3.4.4-2: 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 6.3.4.4-2. The echo attenuation is to be achieved for **each individual frequency band** according to the different categories.

6.3.4.5 Minimum activation level and sensitivity of double talk detection

For further study.

6.3.5 Speech and audio quality in presence of noise

6.3.5.1 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 6.3.5.1-1.

Table 6.3.5.1-1: Requirements for spectral adjustment of comfort noise (mask)

Frequency	Upper limit	Lower limit
200 Hz	12 dB	-12 dB
800 Hz	12 dB	-12 dB
800 Hz	10 dB	-10 dB
2 000 Hz	10 dB	-10 dB
2 000 Hz	6 dB	-6 dB
4 000 Hz	6 dB	-6 dB
14 000 Hz	6 dB	-6 dB
20 000 Hz (see note 2)	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 5.1.4 is used.

The handsfree is set-up as described in clause 5.1. The handset is mounted at the HFRP position per clause 5.1.4.

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.

6.3.5.2 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 [26].

For the background noises defined in clause 5.1.4 the following requirements apply:

- N-MOS-LQO$_f$ $\geq$ 2.7.

- S-MOS-LQO$_f$ ≥ 3.3.
- G-MOS-LQO$_f$ ≥ 2.7.

NOTE: It is recommended to test the terminal performance with other types of background noises if the terminal is likely to be exposed to other noises than specified in clause 5.1.4.

Measurement method

The background noise simulation as described in clause 5.1.4 is used. The handsfree terminal is set-up as described in clause 5.1. The handset is mounted at the HFRP position per clause 5.1.4.

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 [26] 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 [26]).
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 [26]. 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:

$$\text{S-MOS}'_{\text{LQOf}} = 1{,}418 \cdot \text{S-MOS}_{\text{LQOf}} - 1{,}145$$

$$\text{N-MOS}'_{\text{LQOf}} = 1{,}346 \cdot \text{N-MOS}_{\text{LQOf}} - 1{,}584$$

$$\text{G-MOS}'_{\text{LQOf}} = 1{,}279 \cdot \text{G-MOS}_{\text{LQOf}} - 0{,}7364$$

6.3.5.3 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 5.1.4.

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

Measurement method

The test arrangement is according to clause 5.1.

The background noises are generated as described in clause 5.1.4.

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.

6.3.6 Potential other quality features

6.3.6.1 Sound localization and binaural performance

For further study.

6.3.6.2 Dereverberation performance

This feature is intended to reduce the reverberant signals due to the room where the terminal is installed.

For further study.

6.3.6.3 Switching characteristics between transducers

For further study.

6.3.7 Quality of echo cancellation

6.3.7.1 Temporal echo effects

Requirements

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

Measurement method

The test setup is described in clause 5.1.

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 seconds which represents 8 periods of the CSS signal. The integration time for the level analysis shall be 35 ms, the analysis is referred to the level analysis of the reference signal.

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

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.

6.3.7.2 Spectral echo attenuation

Requirements

The echo attenuation vs. frequency shall be below the tolerance mask given in table 6.3.7.2-1.

Table 6.3.7.2-1: Spectral echo loss limits

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

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

Measurement method

The test setup is described in clause 5.1.

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

6.3.7.3 Occurrence of artefacts

For further study.

6.3.7.4 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 setup is described in clause 5.1.

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.

6.3.8 Switching characteristics

6.3.8.1 Note

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

6.3.8.2 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).

Requirements

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 setup is described in clause 5.1.

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

The settings of the test signal are as follows.

Table 6.3.8.2-1: Test file settings

	Single word/ Pause Duration	**Level of the first single word (active Signal Part at the MRP)**	**Level Difference between two Periods of the Test Signal**
Single word to Determine Switching Characteristic **in Send Direction**	~600 ms / ~500 ms	-24 dBPa (see note)	1 dB
NOTE: The signal level is determined for each utterance individually according to Recommendation ITU-T P.56 [21].			

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

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

6.3.8.3 Silence suppression and comfort noise generation

For further study.

Annex A (normative): Room acoustics and electro acoustic equipment positioning

The positioning of transducers in the acoustic environment can strongly influence their effective performances and suitable installation criteria should be followed in order to maximize the signal-to-noise and signal-to-reverberation ratios.

In particular the main parameters to be taken into account when installing teleconference/videoconference systems are:

- room acoustics (e.g. reverberation);
- background noise;
- sound insulation (privacy), mainly for individual use.

Additional parameters to be taken into account are at least:

- A room suitable for a normal face-to-face conference shall be selected.
- Maximum talker to microphone distance shall be determined taking into account both the noise and reverberation dependencies.
- The microphones and loudspeakers shall be positioned in accordance with both these distances.
- The microphone type should be chosen according to the room environment.

More detailed information is available in ETSI ETS 300 807 [13]. Audio characteristics of terminals designed to support conference services ITU-T Supplement P16 [i.1]; Guidelines for placement of microphones and loudspeakers in telephone conference rooms and for Group Audio Terminals.

Annex B (informative): Bibliography

- Meunier S. and al.: "Calcul des indicateurs de sonie : revue des algorithmes et implémentation", 10ème Congrès Français d'Acoustique (2010).
- ANSI S3, 4-2007: "American National Standard Procedure for the Computation of Loudness of Steady Sound" (2007).
- Zwicker E. and Fastl H.: "Psychoacoustics: Facts and models", 2nd Edition, Springer-Verlag, Berlin (1999).
- Glasberg, B. R., and Moore, B. C. J.: "A model of loudness application to time-varying sounds", J. Audio Eng. Soc, Vol. 50, n°5, 331-342 (2002).
- STQ(12)40_31: "New results of Laboratory Test of reference headphones on HATS for defining correction factors".

History

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