# ETSI TS 102 925 V1.1.1 (2013-03)

**Technical Specification**

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

Reference

DTS/STQ-152-2

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

*Important notice*

Individual copies of the present document can be downloaded from:
http://www.etsi.org

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

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

*Copyright Notification*

*ETSI*

# Contents

# Intellectual Property Rights

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

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

# Foreword

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

# 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 equipments 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 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

# 1     Scope

The present document provides speech & audio transmission performance requirements and measurement methods for handsfree functions of superwideband/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 superwideband and fullband coders are now available, some being also compatible with wideband coders.

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

Such applications include:

- Speech and audio communication including conferencing using high quality handsfree systems.

- Bandwidth extension which may allow usage for some mixed content applications.

- Superwideband enhancement coupled with stereo/multichannel.

# 2     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 reference document (including any amendments) applies.

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

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

## 2.1     Normative references

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

[1]        Recommendation ITU-T P.501 Amendment 1: "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 202 396-1: "Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database".

[12]     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".

[13]     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".

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

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

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

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

[18]     ETSI TS 102 924: "Speech and multimediaTransmission Quality (STQ); Transmission requirements for Superwideband/Fullband headset terminals from a QoS perspective as perceived by the user".

[19]     Recommendation ITU-T P.800: "Methods for subjective determination of transmission quality".

[20]     Recommendation ITU-T P.830: "Subjective performance assessment of telephone-band and wideband digital codecs".

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

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

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

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

## 2.2      Informative references

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]     STQ(13)42-30: "Superwideband and fullband testing. Performance characteristics of the Head Acoustics HMS II.3 Artificial Head".

[i.4]     STQ(13)42-029: "Loudness depending on bandwidth and coder".

[i.5]     STQ(12)40-26: "Comparison between loudness ratings and loudness".

# 3 Definitions and abbreviations

## 3.1 Definitions

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

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

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

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

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

**fullband telephony:** transmission of speech with a nominal pass-band wider than 50 Hz to 14 000 Hz, usually understood to be 20 Hz to 20 000 Hz (definition from Recommendation ITU-T P.10/G.100 [2])

**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

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

NOTE: Superwideband covers at least moAno and stereo capabilities.

## 3.2 Abbreviations

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

| | |
|---|---|
| ACR | Absolute Category Rating |
| CSS | Composite Source Signal |
| EVS | Enhanced Voice Services |
| FB | Fullband |
| GAT | Group Audio Terminal |
| HATS | Head and Torso Simulator |
| HFRP | HandsFree Reference Point |
| MCU | Multiplexing Control Unit |
| MRP | Mouth Reference Point |
| PDA | Personal Digital Assistant |
| RLR | Receive Loudness Rating |
| SLR | Send Loudness Rating |
| SWB | Superwideband |

# 4 Applications and Coder considerations

## 4.1 Applications

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

- Speech and audio communication including conferencing using high quality handsfree systems, for which superwideband/fullband coding can better reproduce the audio environment and provides 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).

- Superwideband 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 Coder considerations

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

### 4.2.1 Superwideband (SWB)

| Coder Reference | Speech | Other signals | Stereo | Remark |
|---|---|---|---|---|
| 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 [16] Annexes D and F | X | X | X (Annex F) | |
| Recommendation ITU-T G.722 [21] Annexes B and D | X | X | X (Annex D) | |

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

The following coders are recommended for superwideband:

- 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 Superwideband scalable extension for Recommendation ITU-T G.718 [9]) *"This annex describes a scalable superwideband (SWB, 50-14000 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 [16], Annex D defines the superwideband extension.

  - Annex F defines the Stereo embedded extension for Recommendation ITU-T G.711.1 [16]

  - *"The Annex F is intended as a stereo extension to the G.711.1 wideband coding algorithm and its superwideband 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 [21], Annex B defines the superwideband extension and Annex D defines the Stereo embedded extension for Recommendation ITU-T G.722 [21].

  - *"Annex B describes a scalable superwideband (SWB, 50-14000 Hz) speech and audio coding algorithm operating at 64, 80 and 96 kbit/s. The Recommendation ITU-T G.722 [21] superwideband extension codec is interoperable with Recommendation ITU-T G.722 [21]. The output of the Recommendation ITU-T G.722 [21] SWB coder has a bandwidth of 50-14000 Hz."*

  - *"Annex D describes a stereo extension of the wideband codec G.722 and its superwideband 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 superwideband stereo bitrates at 80, 96, 112 and 128 kbit/s and two wideband stereo bitrates at 64 and 80 kbit/s".*

NOTE:     The potential future mobile coder EVS (Enhanced Voice Services) should be also considered when available. It will be relevant to reconsider the contents of the present document to consider the implications of the EVS coder implementation in terminals within the scope of the present document.

EVS is designed for packet-switched networks/Mobile VoIP and VoLTE is a key target application.

The key features of AVS are Superwideband speech (32 kHz sampling) with improved speech quality and improved music performance.

A future version of the present document will take into account this coder when available.

## 4.2.2    Fullband (FB)

The following coder is 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 [10] 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:     Amendment 1 adds new Annex A that 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).

# 5        Test considerations

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 Superwideband 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 equipments.

HATS is used to test narrowband and wideband speech terminals but has not been initially designed for applications with bandwidth above 10 kHz nor for lower frequency than 100 Hz.

Following the principles defined in TS 102 924 [18], HATS could be used for testing superwideband terminals, as indicated in [i.3].

To test the full bandwidth for fullband terminals, the alternative arrangements using a microphone and a loudspeaker, as defined in clause 5.1, should be used.

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

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

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

## 5.1      Test Set-ups

For handsfree and conferencing terminals an alternative to the use of 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 will be flat over 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 will 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:    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 (at least implementing the bit rate offering the best quality for the coder) or using the direct signal processing approach or acoustically.

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

# 5.1.1      Setup for terminals

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 5.1.1.1A and 5.1.1.1B.

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

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 5.1.1.1C.

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 5.1.1.1C.
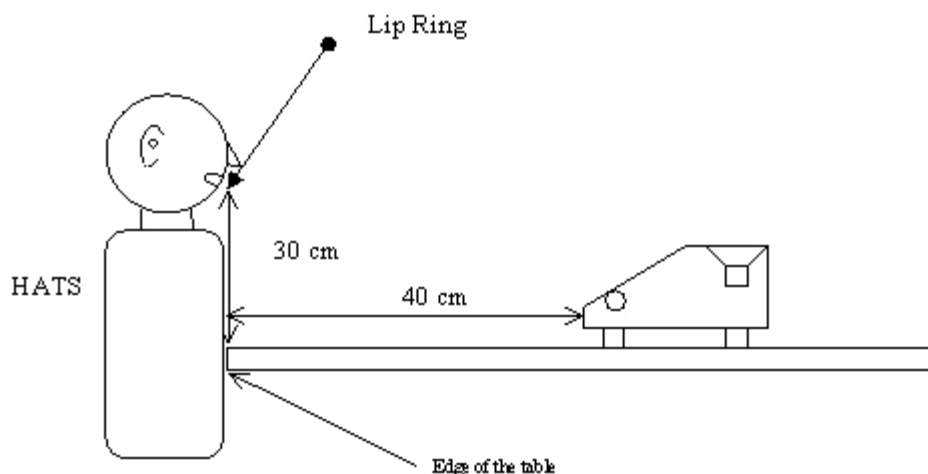


**Figure 5.1.1.1A: Position for test of desktop handsfree terminal with HATS, side view**
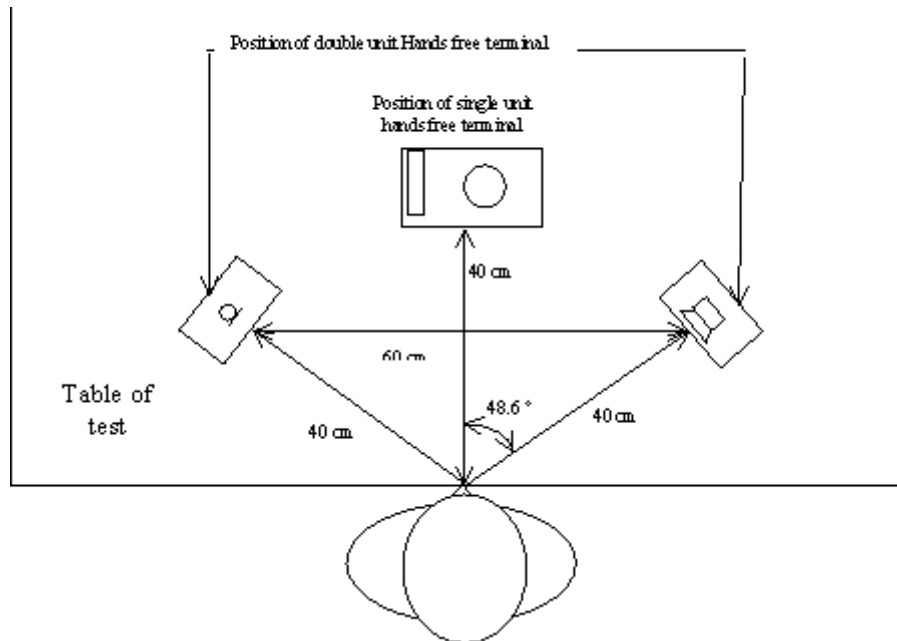
**Figure 5.1.1.1B: Position for test of desktop handsfree terminal with HATS, top view**
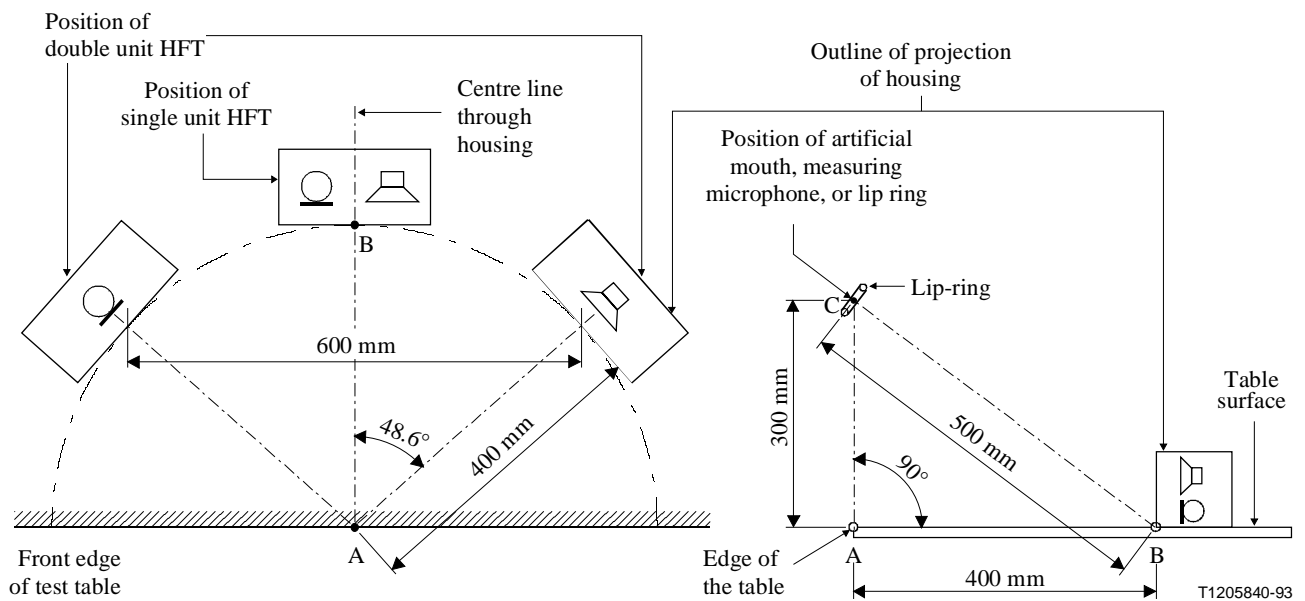


**Figure 5.1.1.1C: Position for test of desktop handsfree terminal with free-field microphone or with reference loudspeaker (from Recommendation ITU-T P.340 [6]), top and side views**

### 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.
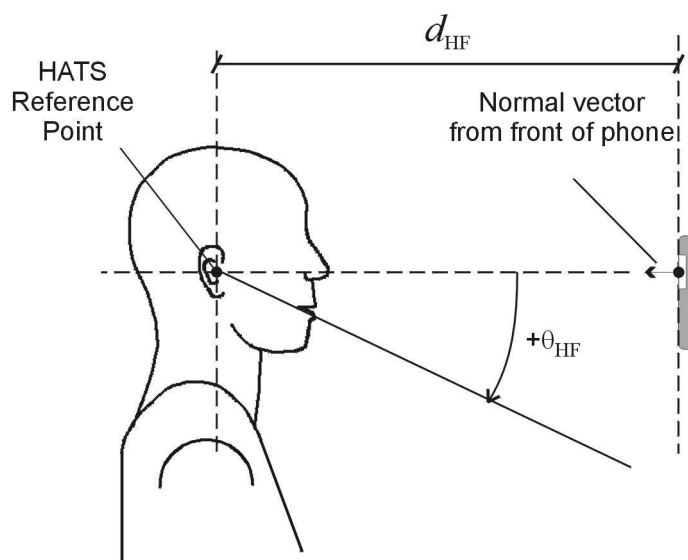
**Figure 5.1.1.2: Configuration of Hand-Held loudspeaker relative to the HATS side view**

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

### 5.1.1.3      Softphone (computer-based terminals)

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

If no other requirement is given by manufacturer 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 5.1.1.3A and 5.1.1.3B 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 5.1.1.3A.

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 5.1.1.3A.
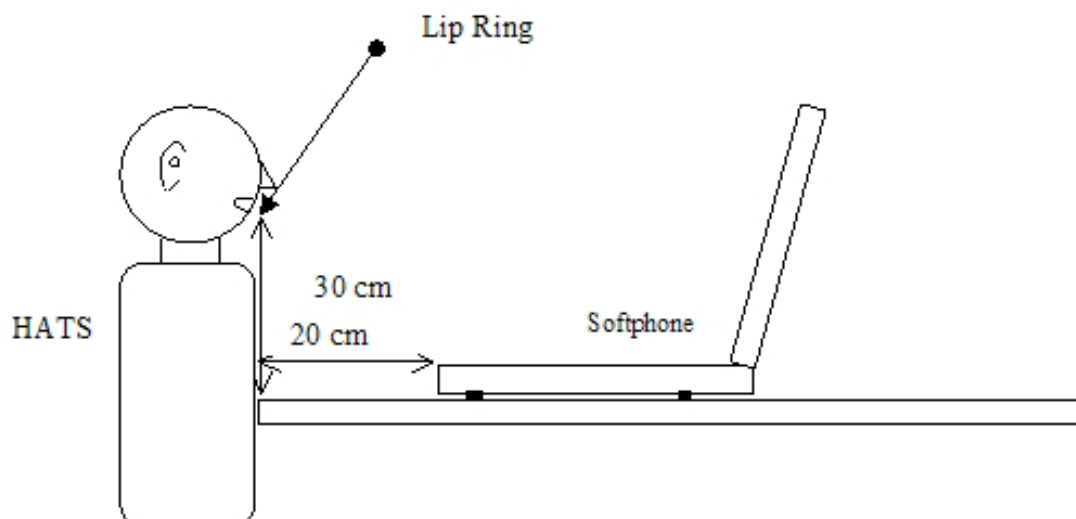
**Figure 5.1.1.3A: Configuration of softphone relative to the HATS side view**

When 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 5.1.1.3B: Configuration of softphone relative to the HATS top view**

When 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 F, 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 5.1.1.3C.

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 5.1.1.3C.

**Figure 5.1.1.3C: Configuration of softphone using external speakers
relative to the HATS top sight**

When free-field microphone or reference loudspeaker 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 5.1.1.3D, when using HATS.
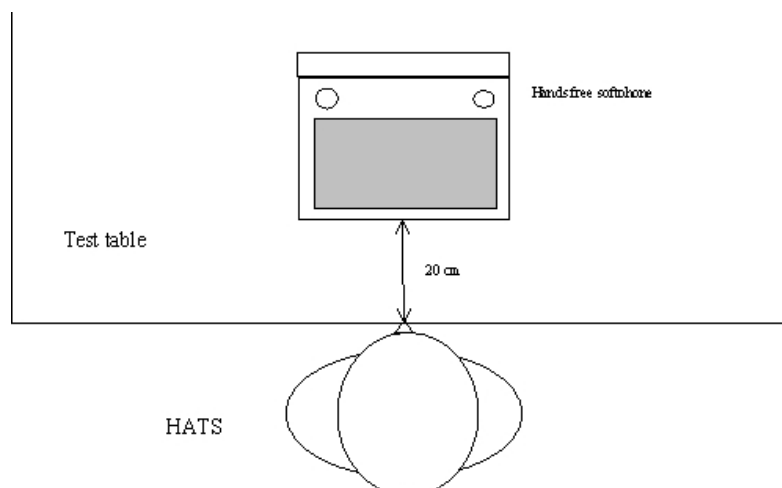
When using a free-field microphone instead of the artificial ears of HATS the centre of the microphone is placed at the point "lip ring" on figure 5.1.1.3D.

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 5.1.1.3D.

  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 5.1.1.3D applies.
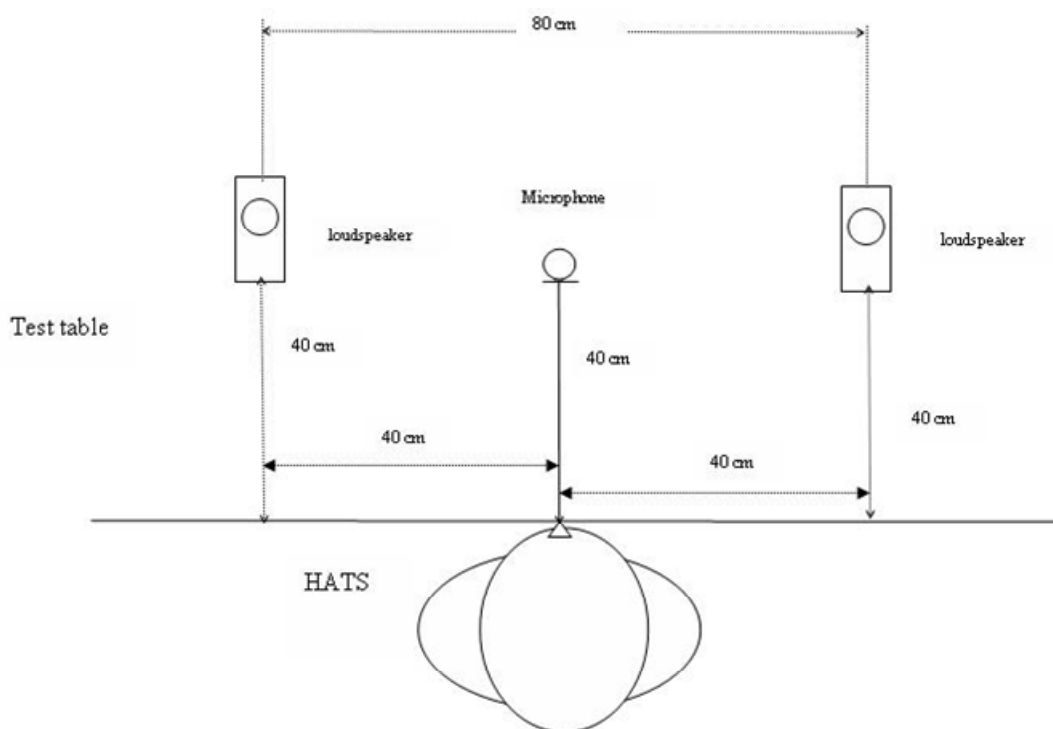
**Figure 5.1.1.3D: Configuration of softphone using
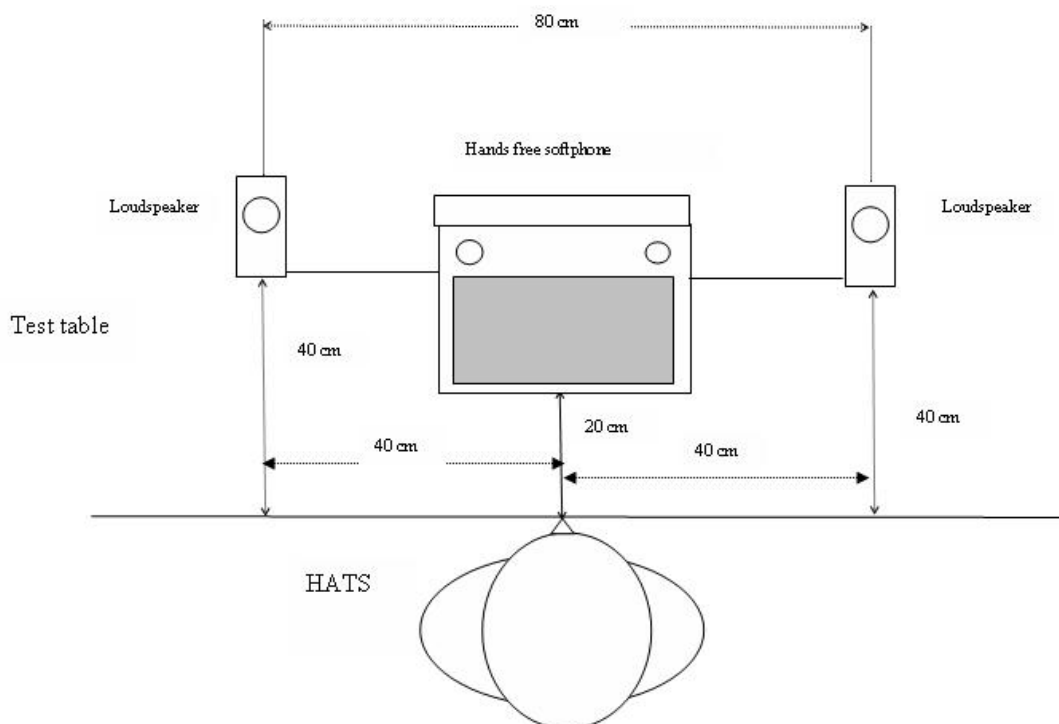external speakers and microphone relative to the HATS top sight**

When free-field microphone or reference loudspeaker 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 Superwideband/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 5.1.1.4A and 5.1.1.4B 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 5.1.1.4A and 5.1.1.4B.

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 5.1.1.4A and 5.1.1.4B.
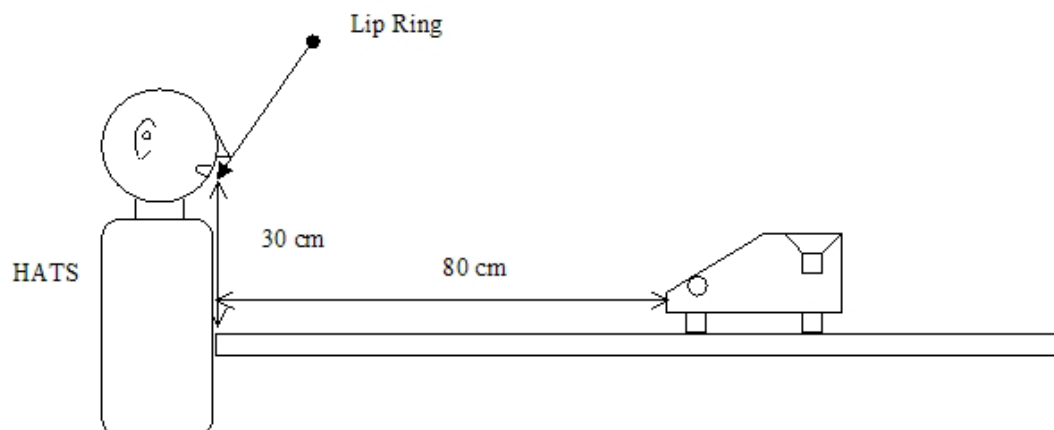
**Figure 5.1.1.4A: Configuration of group audio terminal relative to the HATS side view**

When free-field microphone or reference loudspeaker 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.
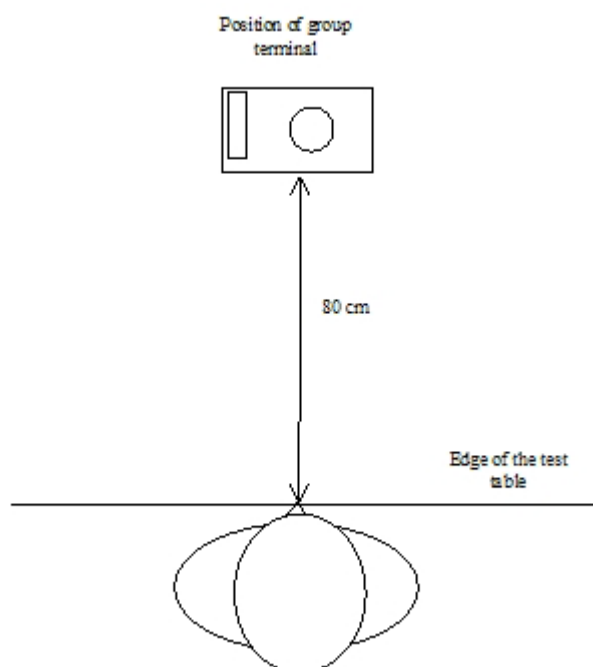


**Figure 5.1.1.4B: Configuration of group audio terminal relative to the HATS top sight**

When free-field microphone or reference loudspeaker 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, test laboratory can use a different position more representative of real use. The conditions of test will be given in the test report.

NOTE 2:   Experiences show that it should be needed to ensure that the quality is not too much 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 equipments may be used.

As there is no unique implementation, there is no standardized position(s) for free field microphone/loudspeaker(s). However, these test equipments 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 postions, 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        Systems such as "telepresence"

These systems are Teleconference systems with complementary features and functions (e.g. one microphone of one end terminal is coupled with a distant loudspeaker).

> NOTE:     ITU-T Study Group16 is currently producing a new Recommendation F.TPS-Reqs "Definitions, requirements, and use cases for Telepresence Systems" that will be taken into account in a future version of the present document.

## 5.1.2    Test signals

The test signals are defined according to Recommendation ITU-T P.501 Amendment 1 [1] for test made with speech signals. For some parameters it is needed 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 ES 202 396-1 [11].

As the bandwidth of the speech signals defined in Recommendation ITU-T P.501 Amendment 1 [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, Amendment 1 [1]

- A The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501, Amendment 1 [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.50, Amendment 1 [1], shall be used for measurements such as TCLw, switching characteristics.

For double-talk performance:

- A "double-talk" sequence representing typical double talk scenarios in real conversations is shown in figure 6.3.4. This uses the single-talk sequence described in section 7.3.1 of Recommendation ITU-T P.501, Amendment [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

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 calibrated at HFRP.

When using HATS it is positioned according to figure 5.1.3.1.

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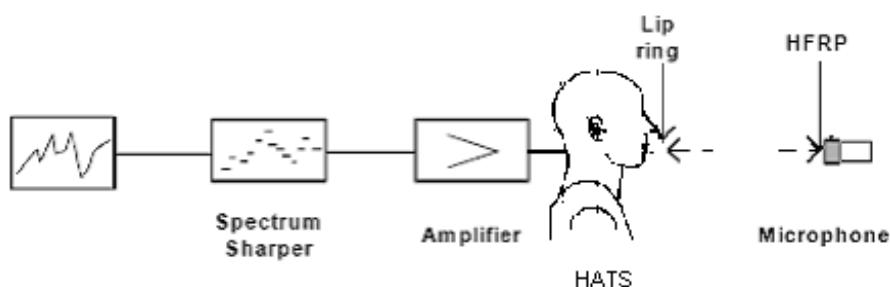


**Figure 5.1.3.1: Calibration at HFRP (with $d_{HFS}$ = 50 cm)**

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 ETS 300 807 [14], 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:

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

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.

### 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 ES 202 396-1 [11].

The signals attached to ES 202 396-1 [11] are fullband signals and should be used for background noise simulation.

## 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 are implemented the systems

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 for measurements

As delay is introduced by the terminal, care shall be taken for all measurements using an activation signal. It shall be checked that the test is performed on the test signal and not on the activation 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 [24]).

For specified tests, it is desirable to have a background noise level of less than -74 dBPa(A) in conjunction with NC20, but the background noise level of -64 dBPa(A) in conjunction with NC30 shall never be exceeded.
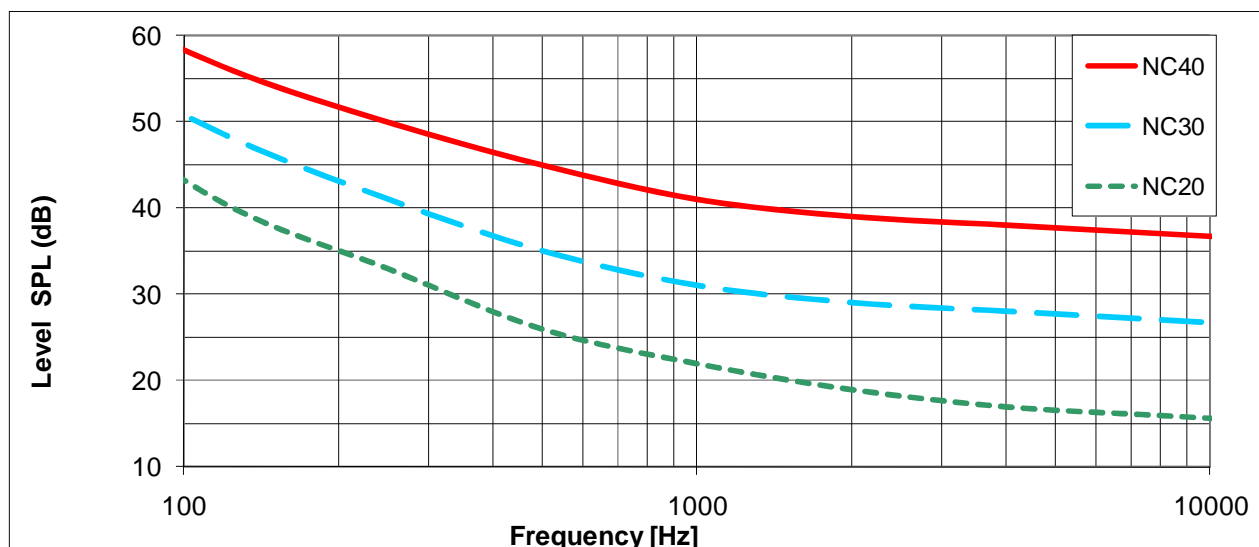
**Figure 5.2: NC-criteria for test environment**

# 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 5.3A: 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 % |

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

**Table 5.3B: 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:      This tolerance may be used to avoid measurements at critical frequencies, e.g. those due to sampling operations within the terminal under test. | |

> NOTE:      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

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 superwideband or fullband functions, or is intended to communicate with wideband terminals, the terminal shall be calibrated for SLR and RLR values for wideband/Narrowband bandwidth.

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

> NOTE:     RLR and SLR, values are based on those defined in ES 202 740 [12] and TS 103 740 [13].

### 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 and for superwideband and fullband.

The assessment method takes into account the level, the spectrum of the signals and may also take 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, Amendment 1 [1] and in ES 202 396-1 [11].

When the terminal provides superwideband or fullband in addition with wideband or narrowband the reference loudness value (expressed in phons) shall 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.

Preliminary measurement methods and requirements are available in [i.4].

The loudness measured in superwideband 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 [17] 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 superwideband and fullband terminals compared to wideband terminals.

# 6        Requirement considerations and test methods

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.

## 6.1      Send

All the types of terminals within the scope of the present document shall fulfil 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 superwideband 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 superwideband shall fulfil the mask as defined in table 6.1.1 and figure 6.1.1.

**Table 6.1.1A: Frequency mask for superwideband 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. | | |

## Send frequency response mask for SWB



**Figure 6.1.1A: Frequency mask for superwideband terminals - Send**
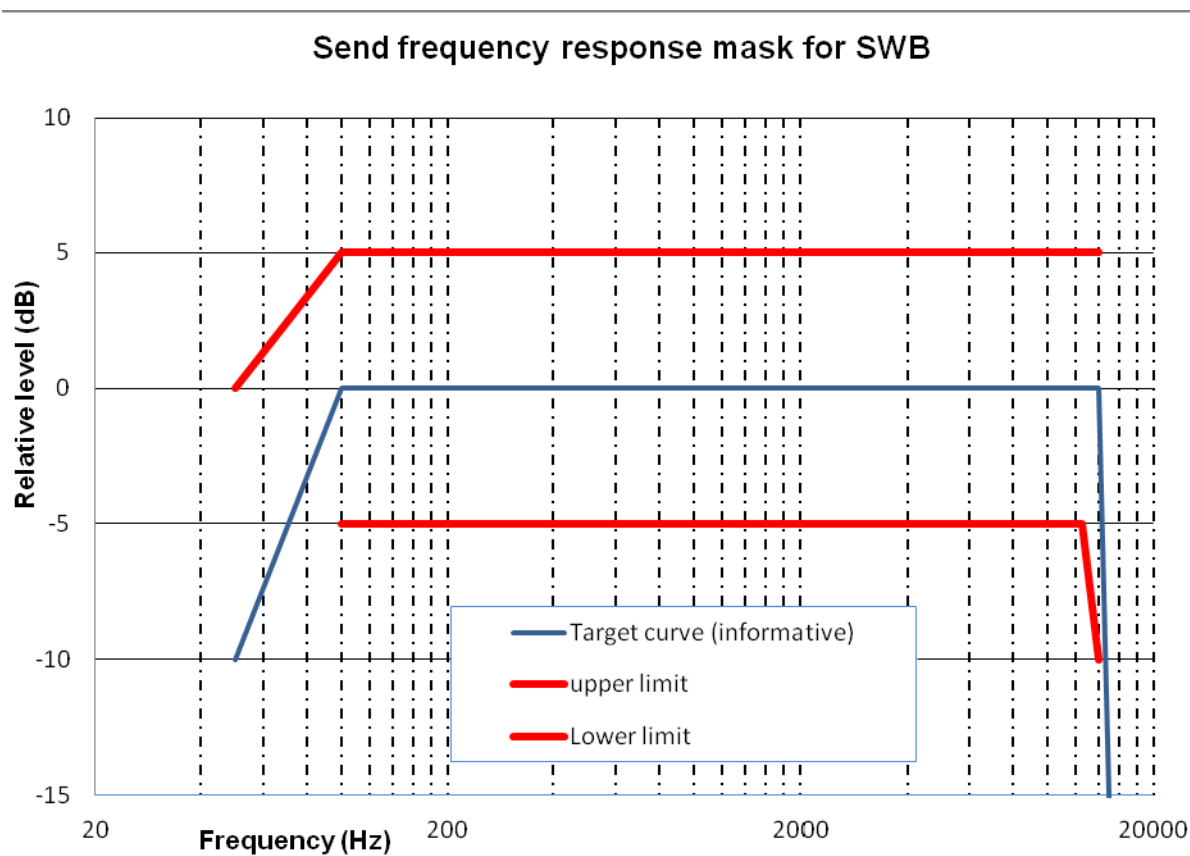
Fullband

**Table 6.1.1B: Frequency mask for fullband terminals - Send**

| Frequency (Hz) | Upper limit (dB) | Lower limit (dB) |
|---|---|---|
| 50 | 0 | -10 |
| 100 | 5 | -5 |
| 200 | 5 | -5 |
| 12 500 | 5 | -5 |
| 14 000 | 5 | -5 |
| 16 000 | 5 | -5 |
| NOTE:     All sensitivity values are expressed in dB on an arbitrary scale. | | |
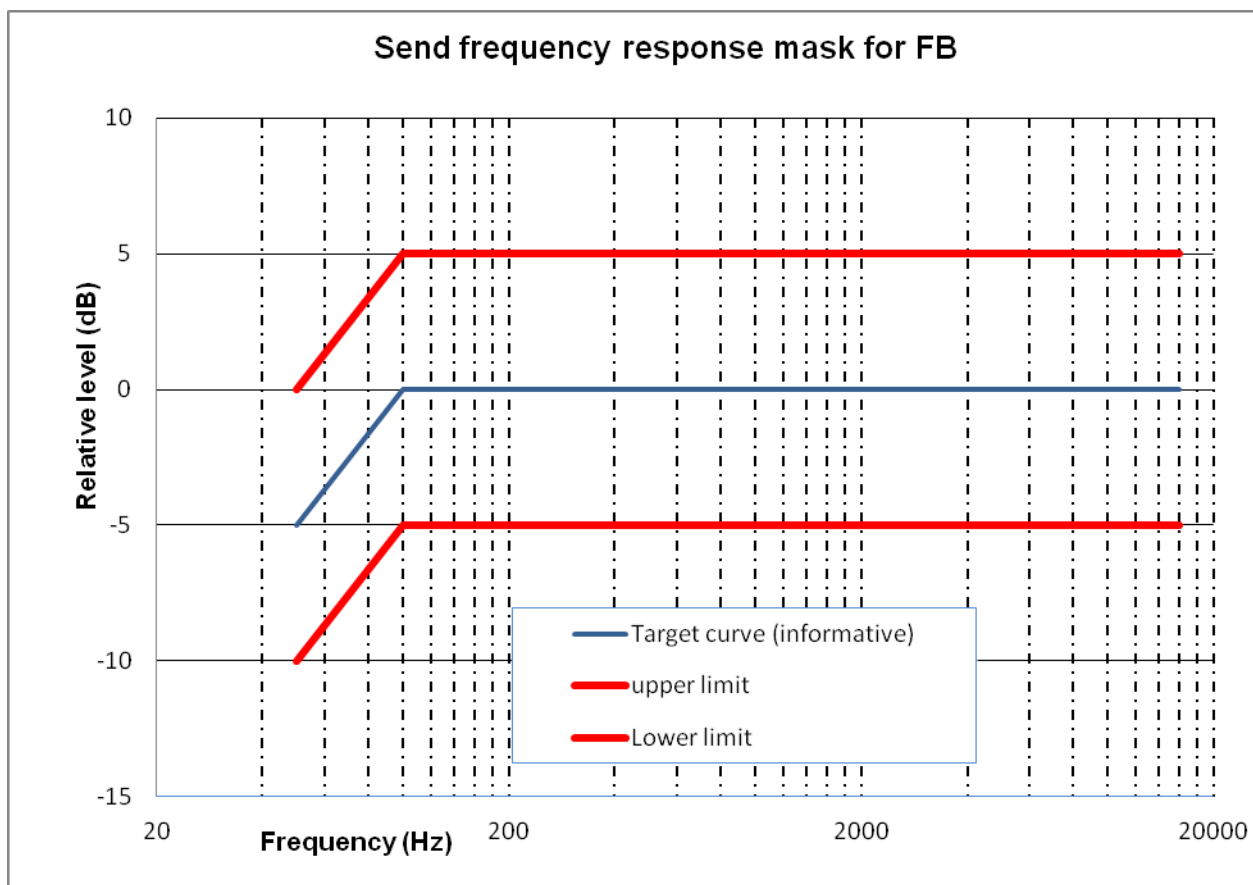
**Figure 6.1.1B: Frequency mask for fullband terminals - Send**

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 the R.40 series of preferred numbers in ISO 3 [23] for frequencies from 100 Hz to 14 kHz inclusive for SWB and from 50 Hz to 18 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    Loudness rating (SLR),

**Requirement**

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

The requirements refer to wideband handsfree terminals, ES 202 740 [12].

Nominal value: +13dB ± 3 dB.

There is no specific requirement for SWB or FB bandwidth.

**Measurement method** of Wideband Loudness rating. The terminal will be 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, Amendment 1 [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    Level dependency

The loudness/loudness ratings are tested for different input levels (at least the nominal signal level, a 10 dB lower and a 5 dB higher).

Requirements are for further study.

> NOTE:    This parameter should also be checked for different positions of the HATS (see clasue 5.1.1.4) when the terminal is intended to be used simultaneously by several users located in the same room.

## 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:    Softphones with cooling devices (fans) can produce a rather high level of noise, furthermore largely dependent of activity of system.

**Measurement method**

The terminal is set 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, Amendment 1 [1] shall be used for activation. The level of this activation signal will be -4,7 dBPa at the MRP.

The level at the output of the test setup is measured with a A weighting, in the bandwidth from 50 Hz and 20 kHz.

## 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 superwideband 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 Superwideband

**Table 6.1.5.1A**

| Frequency | Ratio |
|-----------|-------|
| 100 Hz | 25 dB |
| 200 Hz | 30 dB |
| 400 Hz | 30 dB |
| 1 kHz | 30 dB |
| 2 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 6.1.51B**

| Frequency | Ratio |
|-----------|-------|
| 100 Hz | 25 dB |
| 200 Hz | 30 dB |
| 400 Hz | 30 dB |
| 1 kHz | 30 dB |
| 2 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 superwideband 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 et 5 kHz. The signal to harmonic distortion ratio is measured selectively up to 14 kHz.

For superwideband 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, 5 kHz et 8 kHz. The signal to harmonic distortion ratio is measured selectively up to 18 kHz.

The duration of the sine wave shall be of 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, Amendment 1 [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.

## 6.1.5.2    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 superwideband 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 Superwideband terminals and up to 18 kHz for fullband terminal.

The duration of the sine wave shall be of 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, Amendment 1 [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.

# 6.2     Receive

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 one setting.

For all settings the conformance with other parameter of the present shall be ensured.

## 6.2.2     Frequency response

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 6.2.2.1A: 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. | | |
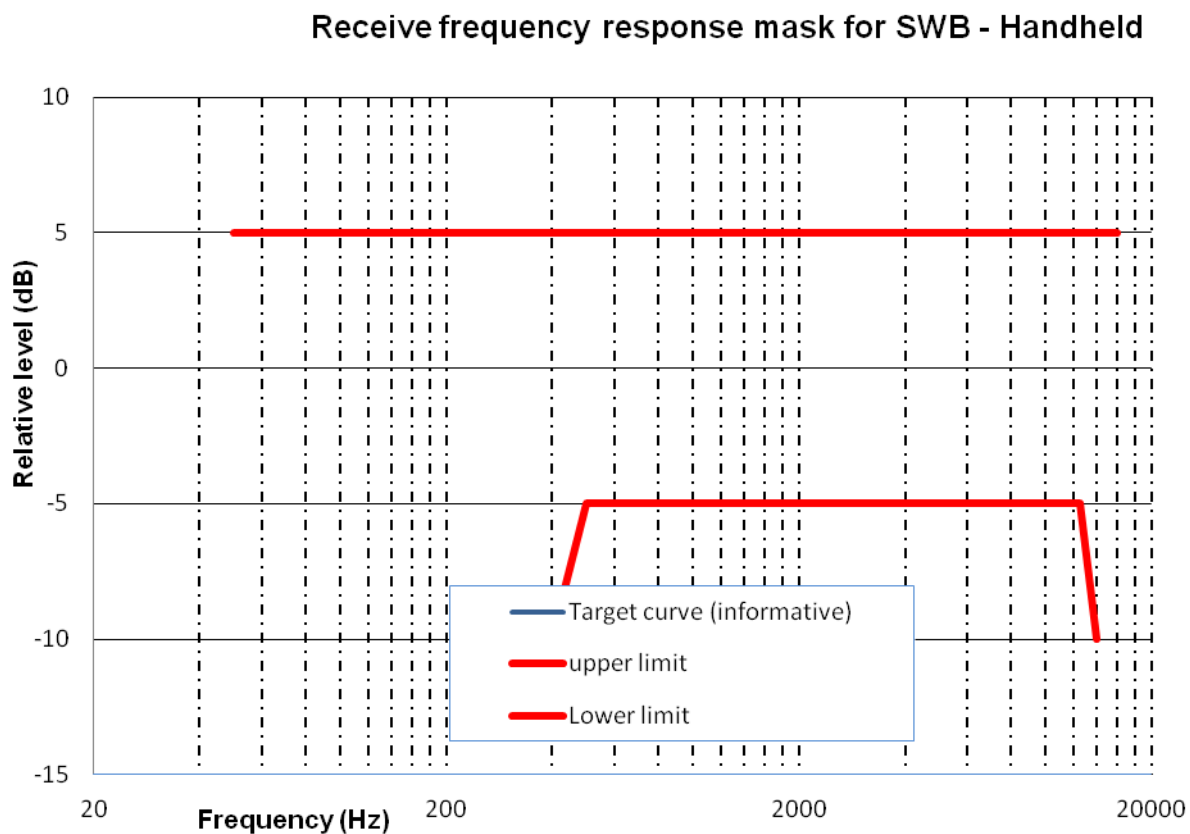
**Figure 6.2.2.1A: Frequency mask for superwideband handheld terminals - Receive**

Fullband

**Table 6.2.2.1B: 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. | | |

## Receive frequency response mask for FB - Handheld



**Figure 6.2.2.1B: 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 [22] 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 3 [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 3 [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**

Superwideband

**Table 6.2.2.2A: Frequency mask for superwideband desktop terminals - Receive**

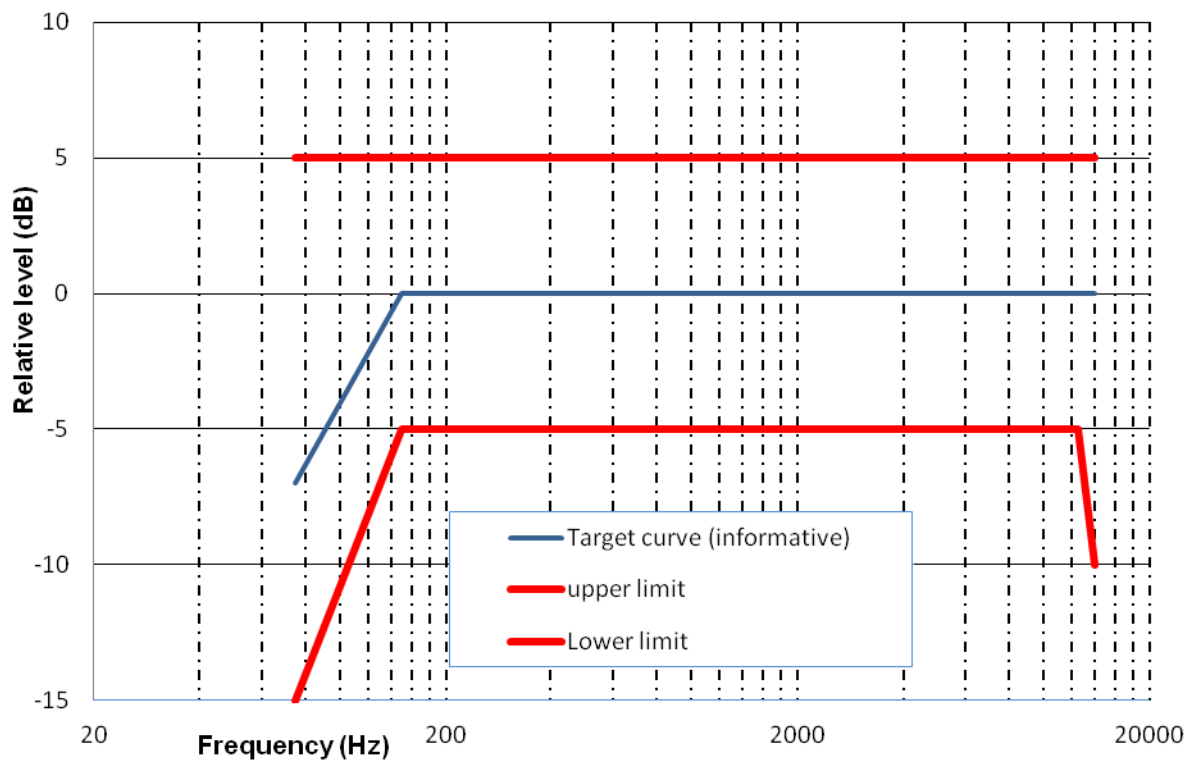| Frequency (Hz) | Upper limit (dB) | Lower limit (dB) |
|---|---|---|
| 75 | 5 | -15 |
| 150 | 5 | -5 |
| 200 | 5 | -5 |
| 400 | 5 | -5 |
| 500 | 5 | -5 |
| 12 500 | 5 | -5 |
| 14 000 | 5 | -10 |
| NOTE:    All sensitivity values are expressed in dB on an arbitrary scale. | | |



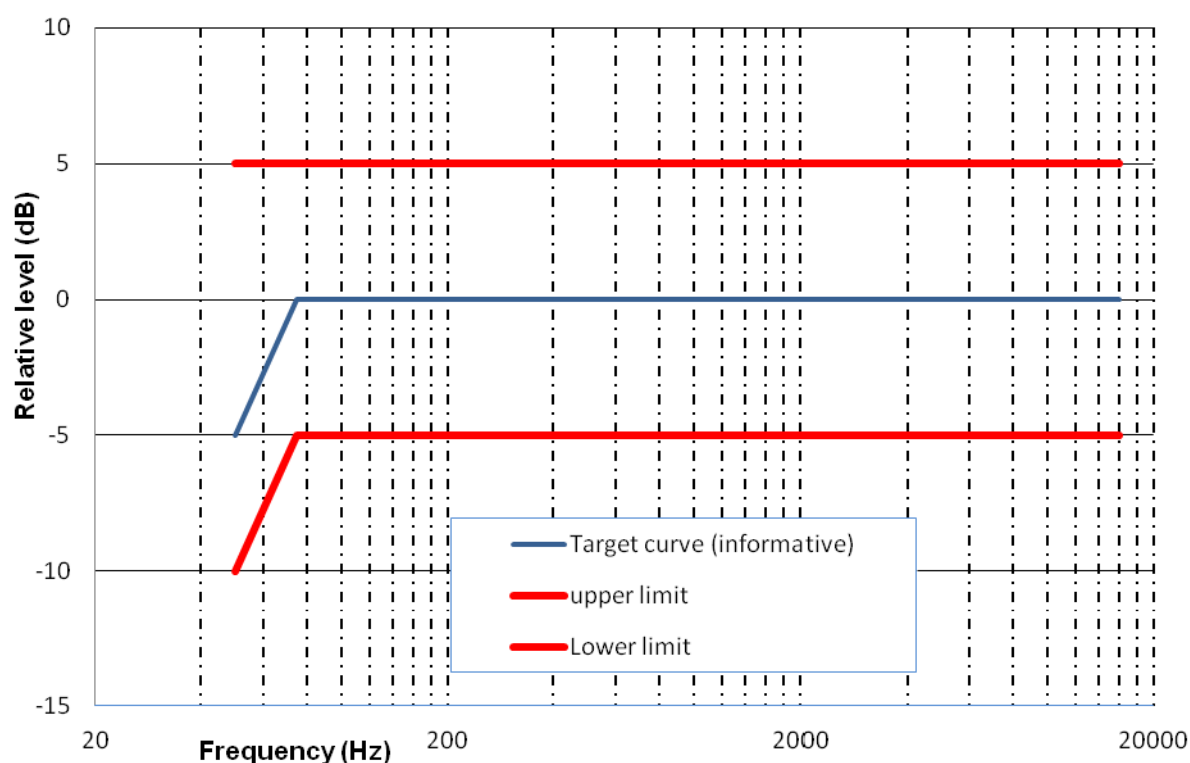**Figure 6.2.2.2A: Frequency mask for superwideband desktop terminals - Receive**

Fullband

**Table 6.2.2.2B: Frequency mask for fullband desktop terminals - Receive**

| Frequency (Hz) | Upper limit (dB) | Lower limit (dB) |
|---|---|---|
| 50 | 5 | -10 |
| 75 | 5 | -5 |
| 150 | 5 | -5 |
| 200 | 5 | -5 |
| 400 | 5 | -5 |
| 500 | 5 | -5 |
| 12 500 | 5 | -5 |
| 14 000 | 5 | -5 |
| 16 000 | 5 | -10 |
| NOTE: All sensitivity values are expressed in dB on an arbitrary scale. | | |



**Figure 6.2.2.2B: Frequency mask for fullband 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 [22] 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 3 [23] 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 the R.40 series of preferred numbers in ISO 3 [23] for frequencies from 50 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.

### 6.2.2.3        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        Loudness Rating (RLR) and Loudness

### 6.2.3.1        Loudness Rating

**Requirements**

When terminal implements wideband speech functions or when the superwideband/fullband functions may interact with wideband terminals, the handsfree terminal shall fulfil the requirements on RLR as defined in ES 202 740 [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 ≥15 dB.

- 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 setup is described in clause 5.1.

The measurement is conducted at nominal volume control setting.

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

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

$S_{Jeff}$     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, Amendment 1 [1]. The test signal level shall be -20 dBm0, measured according to Recommendation ITU-T P.56 [22] at the digital reference point or the equivalent analogue point.

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 the R.40 series of preferred numbers in ISO 3 [23] for frequencies from 100 Hz to 8 kHz inclusive. For the calculation the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

The sensitivity is expressed in terms of dBPa/V.

## 6.2.3.2 Loudness

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

So for the superwideband and fullband transmissions it should be relevant to determine the loudness of the signals delivered by the terminal. When the terminal provides superwideband or fullband in addition with wideband the reference loudness value could be determined for Wideband transmission and to align the loudness in superwideband or fullband with this reference. When the superwideband 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.5], and detailed values are available in [i.4].

## 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 fulfil 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, Amendment 1 [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 until 14 kHz for superwideband terminal and until 18 kHz for fullband terminal.

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.

## 6.2.5 Receive distortion

**Requirements**

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

**Table 6.2.5**

| Frequency | Signal to distortion ratio limit, receive for desktop terminal (SWB) | Signal to distortion ratio limit, receive for desktop terminal (FB) | Signal to distortion ratio limit, receive for handheld terminal |
|---|---|---|---|
| 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 | 15 dB |
| 800 Hz | 30 dB | 30 dB | 20 dB |
| 1 kHz | 30 dB | 30 dB | 25 dB |
| 2 kHz | 30 dB | 30 dB | 25 dB |
| 5 kHz | 30 dB | 30 dB | 30 dB |
| 8 kHz | 30 dB | 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 et 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, Amendment 1 [1]. The sinusoidal signal level shall be calibrated to -16 dBm0.

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

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

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

# 6.3        Other parameters

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

Send delay plus receive delay.

Codec delay is not included.

For conversational services, the delay shall be kept as small as possible to ensure a good quality and in particular a high level of interactivity between the users.

For further study.

## 6.3.2        Terminal Echo Loss

NOTE:        To ensure that the terminal does not provide annoying echo for speech communication, the TCLw requirement and test method defined in ES 202 740 [12] apply for the wideband bandwidth.

Recommendation ITU-T P.501, Amendment 1 [1] indicates that in general, high frequency echo components are more annoying than lower frequency echo components. This is especially important for wideband and fullband echo testing. The major impairment typically occurs if such high frequency echo components reach the users' ear unmasked. Especially for superwideband and fullband connections it is important that a test signal provides excitation energy in the high frequency range above 3,5 kHz.

For further study.

## 6.3.3        Objective listening quality

*"Recommendation ITU-T P.863 [15] describes an objective method for predicting overall listening speech quality from narrowband (300 to 3 400 Hz) to superwideband (50 to 14'000 Hz) telecommunication scenarios as perceived by the user in an Recommendation ITU-T P.800 [19] or Recommendation ITU-T P.830 [20] ACR listening only test"*

The speech sequences shall be carefully selected.

For terminals also intended to transmit audio signal, e.g. music, specific models should be used.

The requirements and the detailed measurement method are for further study.

## 6.3.4        Double talk performance

**Requirements** are for further study.

**Measurement Method**

To assess double talk performance, the signals to be used are defined in Recommendation ITU-T P.501, Amendment 1 [1]: A "double-talk" sequence representing typical double talk scenarios in real conversations is shown in figure 6.3.4. This uses the single-talk sequence described in section 7.3.1 of Recommendation ITU-T P.501, Amendment 1 [1], shown in the lower pane, as the main speech and an additional competing speaker sequence, shown in the upper pane.
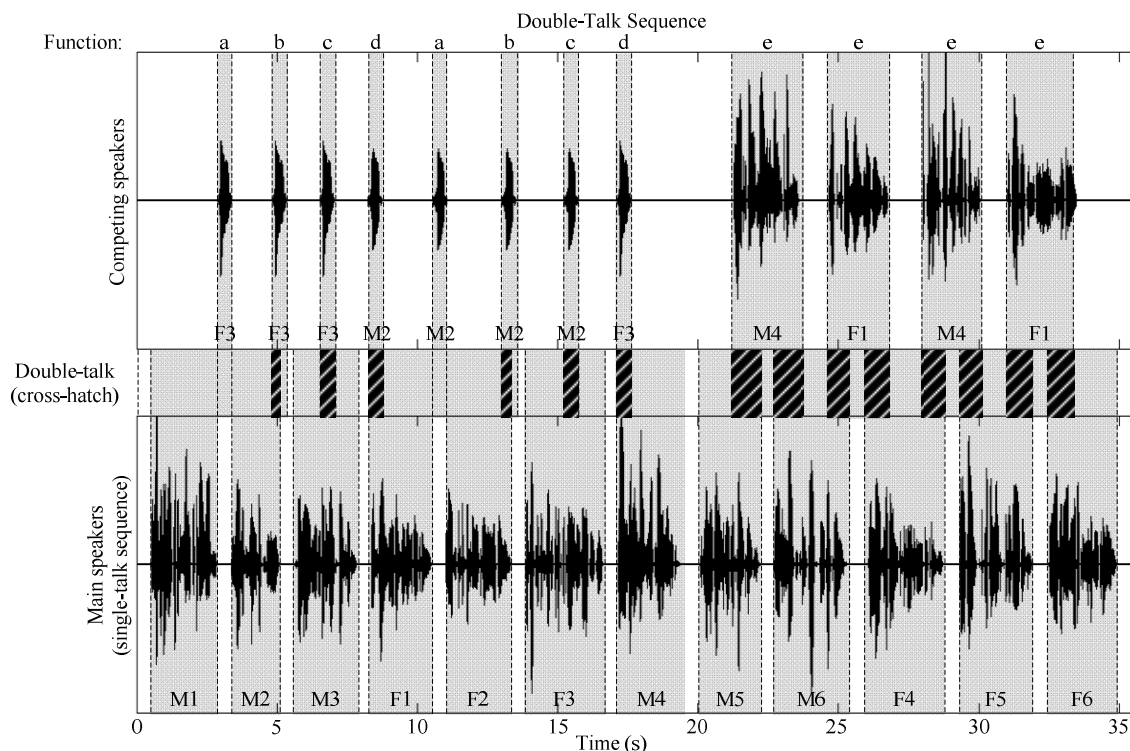
NOTE:     Cross-hatched areas between the upper and lower panes show periods of double talk.

**Figure 6.3.4: Double-talk test sequence using the single-talk sequence and
competing speech serving different functions (a - e)**

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. 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 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.

## 6.3.5     Speech and audio quality in presence of noise

For further study.

## 6.3.6     Potential other quality features

### 6.3.6.1     Sound localisation 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.

# 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 ETS 300 807 [14]. 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

[B.1]     Meunier S. and al., "Calcul des indicateurs de sonie : revue des algorithmes et implémentation", 10ème Congrès Français d'Acoustique (2010).

[B.2]     ISO 532 B: "Method for calculating loudness", International standard (1975).

[B.3]     ANSI S3, 4-2007: "American National Standard Procedure for the Computation of Loudness of Steady Sound" (2007).

[B.4]     Zwicker E. and Fastl H., "Psychoacoustics: Facts and models", 2nd Edition, Springer-Verlag, Berlin (1999).

[B.5]     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).

[B.6]     STQ(12)40_31: New results of Laboratory Test of reference headphones on HATS for defining correction factors.

[B.7]     STQ (12)40_32: Proposal for correction factor when measuring receive part of super wide band and full band headset terminals.

# History

| Document history | | |
|---|---|---|
| V1.1.1 | March 2013 | Publication |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |