ETSI ES 202 740 V1.4.1 (2015-01)

Speech and multimedia Transmission Quality (STQ);
Transmission requirements for wideband
VoIP loudspeaking and handsfree terminals
from a QoS perspective as perceived by the user

Reference
RES/STQ-205

Keywords
handsfree, loudspeaking, quality, speech,
terminal, VoIP, Wideband

ETSI

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

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

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

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://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 final draft ETSI Standard (ES) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ), and is now submitted for the ETSI standards Membership Approval Procedure.

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**may not**", "**need**", "**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

Traditionally, the analogue and digital telephones were interfacing switched-circuit 64 kbit/s PCM networks. With the fast growth of IP networks, wideband terminals providing higher audio-bandwidth and directly interfacing packet-switched networks (VoIP) are being rapidly introduced. Such IP network edge devices may include gateways, specifically designed IP phones, soft phones or other devices connected to the IP based networks and providing telephony service. Due to the unique characteristics of the IP networks including packet loss, delay, etc. new performance specification, as well as appropriate measuring methods, will have to be developed. Terminals are getting increasingly complex.

The advanced signal processing of terminals is targeted to speech signals. Therefore, wherever possible speech signals are used for testing in order to achieve mostly realistic test conditions and meaningful results.

The present document provides speech transmission performance requirements for wideband VoIP loudspeaking and hands-free terminals.

NOTE: Requirement limits are given in tables, the associated curve when provided is given for illustration.

1 Scope

The present document provides speech transmission performance requirements for 8 kHz wideband VoIP loudspeaking and hands-free terminals; it addresses all types of IP based terminals, including wireless, softphones and group audio terminals.

In contrast to other standards which define minimum performance requirements it is the intention of the present document to specify terminal equipment requirements which enable manufacturers and service providers to enable good quality end-to-end speech performance as perceived by the user.

In addition to basic testing procedures, the present document describes advanced testing procedures taking into account further quality parameters as perceived by the user.

NOTE: The present document does not concern headset terminals.

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 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.

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

[1] ETSI ETSI I-ETS 300 245-6: "Integrated Services Digital Network (ISDN); Technical characteristics of telephony terminals; Part 6: Wideband (7 kHz), loudspeaking and hands free telephony".

[2] ETSI TS 126 171: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR speech codec, wideband; General description (3GPP TS 26.171 version 6.0.0 Release 6)".

[3] Recommendation ITU-T G.108: "Application of the E-model: A planning guide".

[4] Recommendation ITU-T G.109: "Definition of categories of speech transmission quality".

[5] Void.

[6] Void.

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

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

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

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

[11] Recommendation ITU-T G.1020: "Performance parameter definitions for quality of speech and other voiceband applications utilizing IP networks".

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

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

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

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

[16] Recommendation ITU-T P.310: "Transmission characteristics for telephone band (300-3400 Hz) digital telephones".

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

[18] Recommendation ITU-T P.341: "Transmission characteristics for wideband (150-7000 Hz) digital hands-free telephony terminals".

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

NOTE: At the publication date of the present document, annex C to P.501 is available as P.501, Amendment 2.

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

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

[22] Recommendation ITU-T P.862: "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs".

[23] IEC 61260: "Electroacoustics- Octave-band and fractional-octave-band filters".

[24] Recommendation ITU-T P.800.1: "Mean Opinion Score (MOS) terminology".

[25] 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".

[26] Recommendation ITU-T P.863.1: "Application guide for Recommendation ITU-T P.863".

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

[28] ETSI ES 202 718: "Speech and multimedia Transmission Quality (STQ); Transmission Requirements for IP-based Narrowband and Wideband Home Gateways and Other Media Gateways from a QoS Perspective as Perceived by the User".

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

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

[i.3] ETSI TR 102 648-1: "Speech Processing, Transmission and Quality Aspects (STQ); Test Methodologies for ETSI Test Events and Results; Part 1: VoIP Speech Quality Testing".

[i.4] NIST net.

NOTE: Available at http://snad.ncsl.nist.gov/itg/nistnet/.

[i.5] Netem.

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

[i.6] Recommendation ITU-T G.729: "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)".

[i.7] Recommendation ITU-T G.723.1: "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s".

3 Definitions and abbreviations

3.1 Definitions

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

artificial ear: device for the calibration of earphones incorporating an acoustic coupler and a calibrated microphone for the measurement of the sound pressure and having an overall acoustic impedance similar to that of the median adult human ear over a given frequency band

codec: combination of an analogue-to-digital encoder and a digital-to-analogue decoder operating in opposite directions of transmission in the same equipment

ear-Drum Reference Point (DRP): point located at the end of the ear canal, corresponding to the ear-drum position

freefield equalization: artificial head is equalized in such a way that for frontal sound incidence in anechoic conditions the frequency response of the artificial head is flat

freefield reference point: point located in the free sound field, at least in 1,5 m distance from a sound source radiating in free air

NOTE: In case of a head and torso simulator (HATS) in the centre of the artificial head with no artificial head present.

group-audio terminal: handsfree terminal primarily designed for use by several users which will not be equipped with a handset

handsfree telephony terminal: telephony terminal using a loudspeaker associated with an amplifier as a telephone receiver and which can be used without a handset

HATS Hands-Free Reference Point (HATS HFRP): reference point "n" from Recommendation ITU-T P.58 [14] "n" is one of the points numbered from 11 to 17 and defined in table 6a of Recommendation ITU-T P.58 [14], (coordinates of far field front point)

NOTE: The HATS HFRP depends on the location(s) of the microphones of the terminal under test: the appropriate axis lip-ring/HATS HFRP is to be as close as possible to the axis lip-ring/HFT microphone under test.

Head And Torso Simulator (HATS) for telephonometry: manikin extending downward from the top of the head to the waist, designed to simulate the sound pick-up characteristics and the acoustic diffraction produced by a median human adult and to reproduce the acoustic field generated by the human mouth

loudspeaking function: function of a handset telephone using a loudspeaker associated with an amplifier as a telephone receiver

Mouth Reference Point (MRP): point located on axis and 25 mm in front of the lip plane of a mouth simulator

nominal setting of the volume control: setting which is closest to the nominal RLR

softphone: speech communication system based upon a computer

3.2 Abbreviations

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

AM-FM	Amplitude Modulation-Frequency Modulation
AMR	Adaptative Multi-Rate
CSS	Composite Source Signal
DRP	ear Drum Reference Point
EC	Echo Canceller
EL	Echo Loss
ERP	Ear Reference Point
ETH	Eidgenössische Technische Hochschule
FFT	Fast Fourrier Transform
GSM	Global System for Mobile Communications
HATS	Head And Torso Simulator
HFRP	Hands Free Reference Point
IEC	International Electrotechnical Commission
IP	Internet Protocol
IPDV	IP Packet Delay Variation
ITU-T	International Telecommunication Union –Telecommunication standardization sector
LAN	Local Area Network
LE	Earphone coupling Loss
MOS	Mean Opinion Score
MOS-LQOy	Mean Opinion Score - Listening Quality Objective, y being n for narrow-band, w for wideband, and M for mixed

NOTE: See Recommendation ITU-T P.800.1 [24].

MRP	Mouth Reference Point
NIST	National Institute of Standards and Technology
NLP	Non Linear Processor
PC	Personal Computer
PCM	Pulse Code Modulation
PLC	Packet Loss Concealment
POI	Point Of Interconnection
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RLR	Receive Loudness Rating
RLRmax	Receive Loudness Rating corresponding to the maximum setting of the volume control
RLRmin	Receive Loudness Rating corresponding to the minimum setting of the volume control
RMS	Root Mean Square
SLR	Send Loudness Rating
TCL	Terminal Coupling Loss
TCN	Trace Control for Netem
TELR	Talker Echo Loudness Rating
VoIP	Voice over Internet Protocol

4 General considerations

4.1 Coding Algorithm

The assumed coding algorithm is according to Recommendation ITU-T G.722 [8]. VoIP terminals may support other coding algorithms.

NOTE: Associated Packet Loss Concealment, e.g. as defined in Recommendation ITU-T G.722 [8], appendixes 3 and 4, should be used.

4.2 End-to-end considerations

In order to achieve a desired end-to-end speech transmission performance (mouth-to-ear) it is recommended that general rules of transmission planning tasks are carried out with the E-model taking into account that E-model does not directly address handsfree or loudspeaking terminals; this includes the a-priori determination of the desired category of speech transmission quality as defined in Recommendation ITU-T G.109 [4].

While, in general, the transmission characteristics of single circuit-oriented network elements, such as switches or terminals can be assumed to have a single input value for the planning tasks of Recommendation ITU-T G.108 [3], this approach is not applicable in packet based systems and thus there is a need for the transmission planner's specific attention.

In particular the decision as to which delay measured according to the present document should be acceptable or representative for the specific configuration is the responsibility of the individual transmission planner.

Recommendation ITU-T G.108 [3] with its amendments provides further guidance on this important issue.

The following optimum terminal parameters from a users' perspective need to be considered:

- Minimized delay in send and receive direction.
- Optimum loudness Rating (RLR, SLR).
- Compensation for network delay variation.
- Packet loss recovery performance.
- Maximized terminal coupling loss.
- Some more basic (ETSI I-ETS 300 245-6 [1]) parameters are applicable, if Recommendation ITU-T G.722 [8] is used.

4.3 Parameters to be investigated

4.3.1 Basic parameters

The basic parameters are given in ETSI I-ETS 300 245-6 [1], Recommendation ITU-T P.340 [17] and Recommendation ITU-T P.341 [18].

4.3.2 Further Parameters with respect to Speech Processing Devices

For VoIP terminals that contain non-linear speech processing devices, the following parameters require additional attention in the context of the present document.

The measurements for further parameters with respect to speech processing devices which are novelties to terminal requirement standards, have been successfully used in the ETSI Speech Quality Test Events (see ETSI TR 102 648-1 [i.3]):

- Objective evaluation of speech quality for VoIP terminals.

- Minimum activation level and sensitivity in Receive direction.
- Automatic Level Control in Receive.
- Double Talk Performance.
- Minimum activation level and sensitivity of double talk detection.
- Switching characteristics.
- Quality of echo cancellation.
- Variant Impairments; Network dependant, etc.

5 Test equipment

5.1 IP half channel measurement adaptor

The IP half channel measurement adaptor is described in ETSI EG 202 425 [i.1].

5.2 Environmental conditions for tests

The following conditions shall apply for the testing environment:

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

b) relative humidity: 5 % to 85 %;

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

5.3 Accuracy of measurements and test signal generation

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

Table 1: Measurement Accuracy

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

NOTE: The measured maximum frequency is due to Recommendation ITU-T P. 58 [14] limitations.

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

Table 2: Accuracy of test signal generation

Quantity	Accuracy
Sound pressure level at Mouth Reference Point (MRP)	±3 dB for frequencies from 100 Hz to 200 Hz ±1 dB for frequencies from 200 Hz to 4 000 Hz ±3 dB for frequencies from 4 000 Hz to 8 000 Hz
Electrical excitation levels	±0,4 dB across the whole frequency range
Frequency generation	±2 % (see note)
Time	±0,2 %
Specified component values	±1 %
NOTE: This tolerance may be used to avoid measurements at critical frequencies, e.g. those due to sampling operations within the terminal under test.	

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 alternate current, the test shall be conducted within ±4 % of the rated frequency.

5.4 Network impairment simulation

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

An appropriate network simulator has to be used, for example NISTnet [i.4] (http://snad.ncsl.nist.gov/itg/nistnet/) or Netem [i.5].

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

Here is a brief blurb about NIST Net:

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

- NIST Net is implemented as a kernel module extension to the Linux™ operating system and an X Window System-based user interface application. In use, the tool allows an inexpensive PC-based router to emulate numerous complex performance scenarios, including: tuneable 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 nistnet, 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 (Nistnet only runs on routers).

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

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

5.5 Acoustic environment

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. Depending on the distance of the transducers from mouth and ear a quiet office room may be sufficient e.g. for handsets where artificial mouth and artificial ear are located close to the acoustical transducers. But this is not applicable for handsfree and loudspeaking terminals.

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.

In case where an anechoic room is not available the test room has to be an acoustically treated room with few reflections and a low noise level.

Considering this, test laboratory, in the case where its test room does not conform to anechoic conditions as given in Recommendation ITU-T P.341 [18], has to present difference in results for measurements due to its test room.

5.6 Influence of terminal delay on measurements

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

6 Test Setup

In order to use a compatible test system for all types of speech terminals a HATS (Head And Torso Simulator) will be used instead of free field microphone (for receive measurement) and artificial mouth (for send measurement). HATS is described in Recommendation ITU-T P.58 [14].

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 coder with variable bite rate is used, it should be adopted, for testing terminal electro acoustical parameters, the bit rate recognized as giving the best characteristics is selected.

EXAMPLE:

- Recommendation ITU-T G.722 [8]: 64 kbit/s.
- ETSI TS 126 171 [2]: 19,85 kbit/s.
- Recommendation ITU-T G.729.1 [10]: 32 kbit/s.

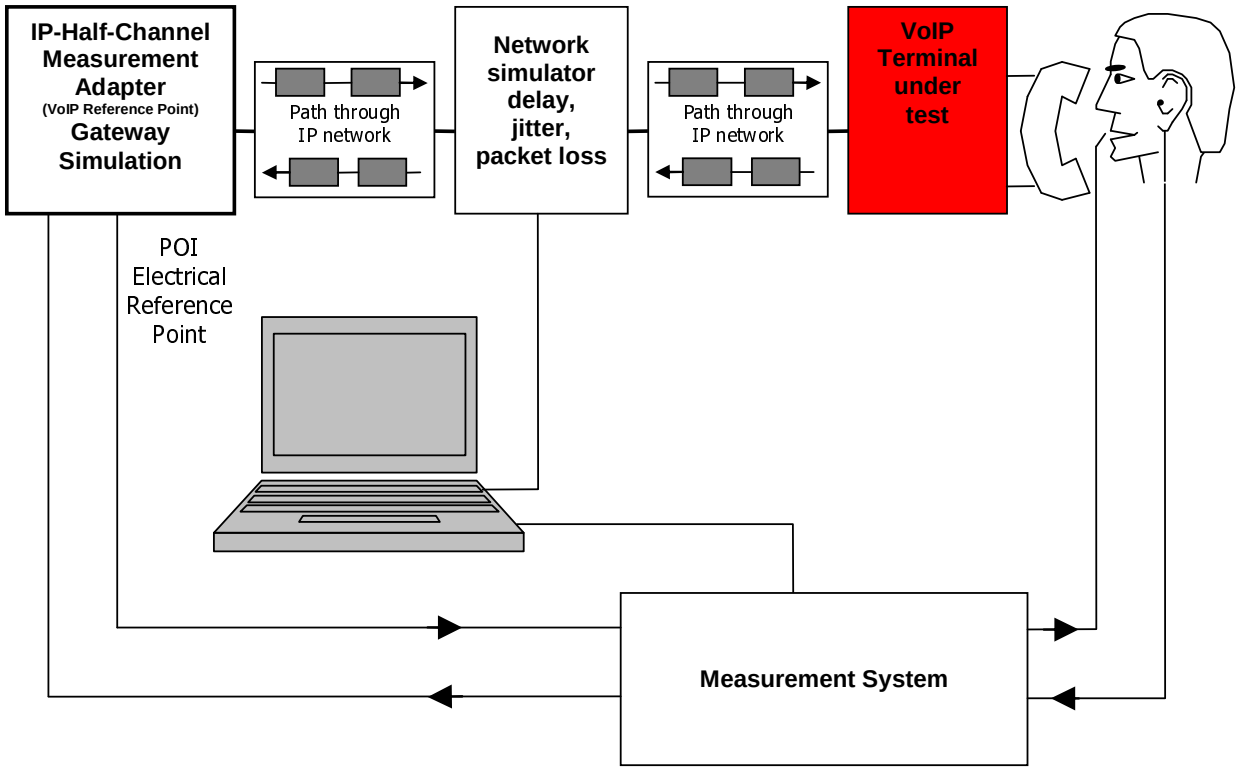

Figure 1: Half channel terminal measurement

6.1 Setup for terminals

6.1.1 Hands-free measurements

The ear used for measurement will be indicated in the test report.

Desktop operated hands-free terminal

For HATS test equipment, definition of hands-free terminal and setups for hands-free terminal can be found in Recommendation ITU-T P.581 [21].

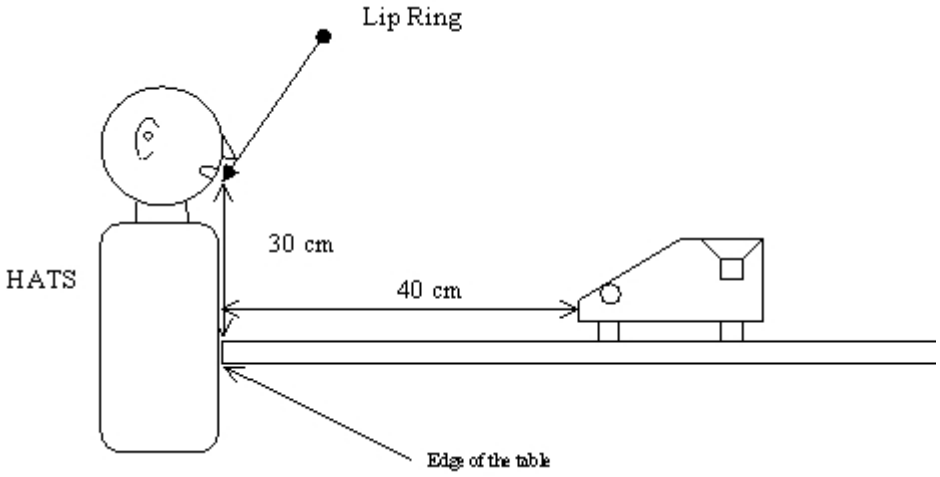

Figure 2: Position for test of desktop hands free terminal side view

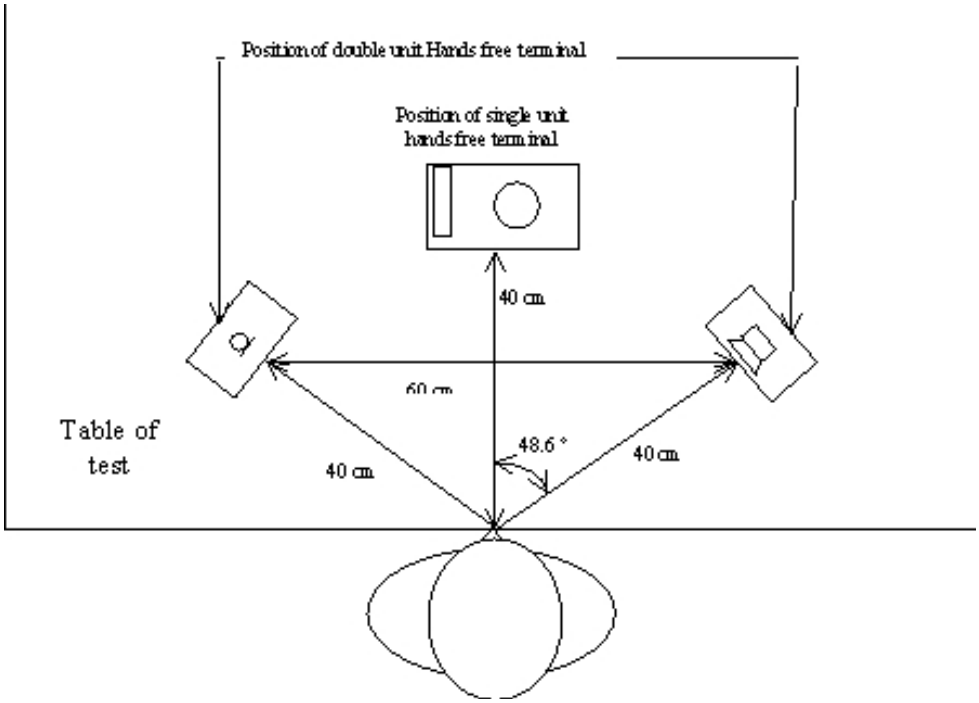

Figure 3: Position for test of desktop hands free terminal top sight

Handheld hands-free terminal

It should be placed in according to figure 4. The HATS should be positioned so that the HATS Reference Point is at a distance d_{HF} from the centre point of the visual display of the Mobile Station. The distance d_{HF} is specified by the manufacturer. A vertical angle θ_{HF} may be specified by the manufacturer.

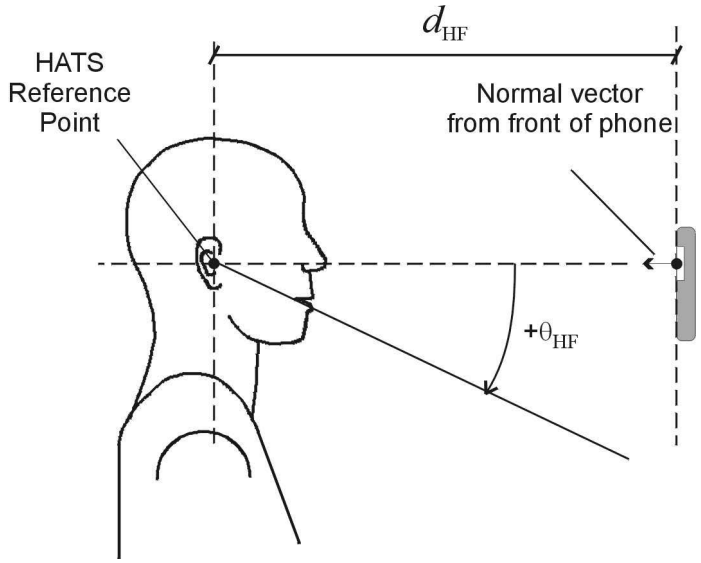

Figure 4: Configuration of Hand-Held loudspeaker relative to the HATS side view

The HATS reference point should be located at a distance d_{HF} from the centre of the visual display of the Mobile Station. The distance d_{HF} is specified by the manufacturer, $d_{HFR}=d_{HF}$, $d_{HFS}=d_{HF}-d_{EM}$, where d_{HFR} is the distance for receive measurement, d_{HFS} is the distance for send measurement, and d_{EM} is the distance from ERP to MRP.

When no operating distance is specified by manufacturer, value for d_{HFS} will be 30 cm. A calculation of d_{EM} for HATS gives 12 cm.

A value of 42 cm will be taken for d_{HF}.

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 the following conditions:

Softphone including speakers 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).

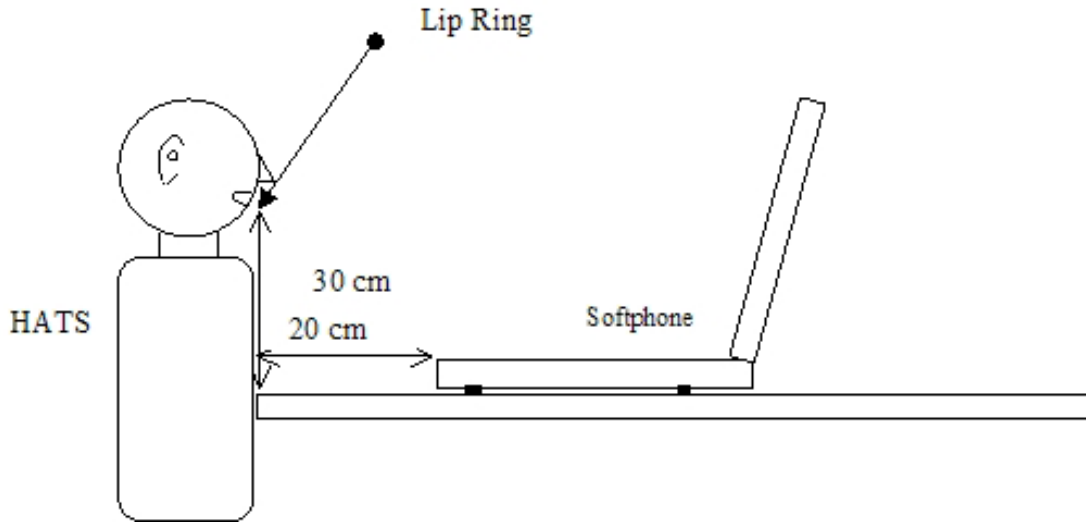

Figure 5: Configuration of softphone relative to the HATS side view

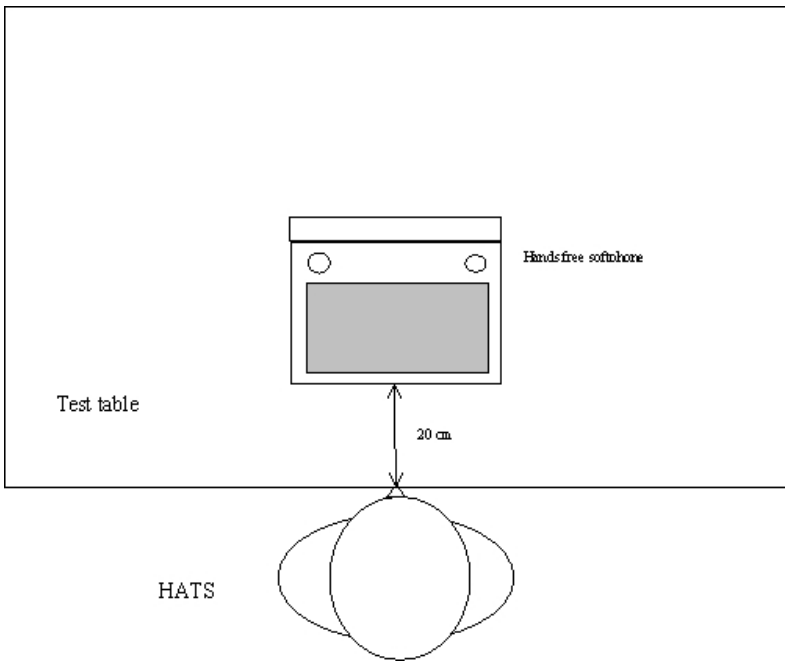

Figure 6: Configuration of softphone relative to the HATS top sight

Softphone with separate speakers

When separate loudspeakers are used, system will be positioned as in figure 7.

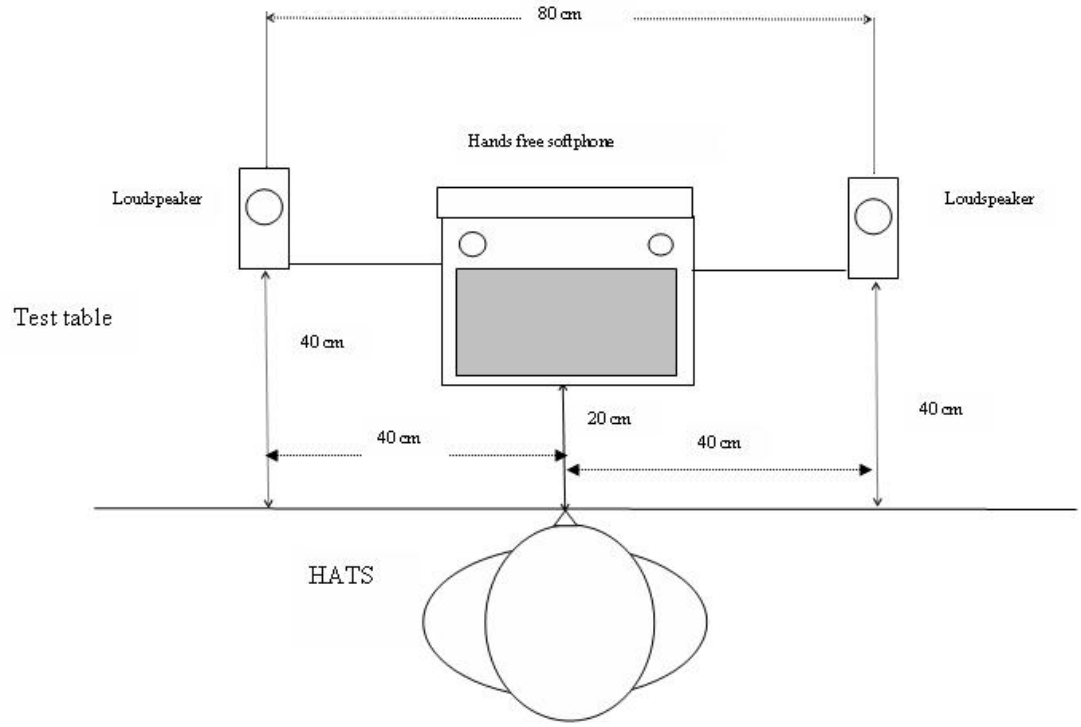

Figure 7: Configuration of softphone using external speakers relative to the HATS top sight

When external microphone and speakers are used, system will be positioned as in figure 8.

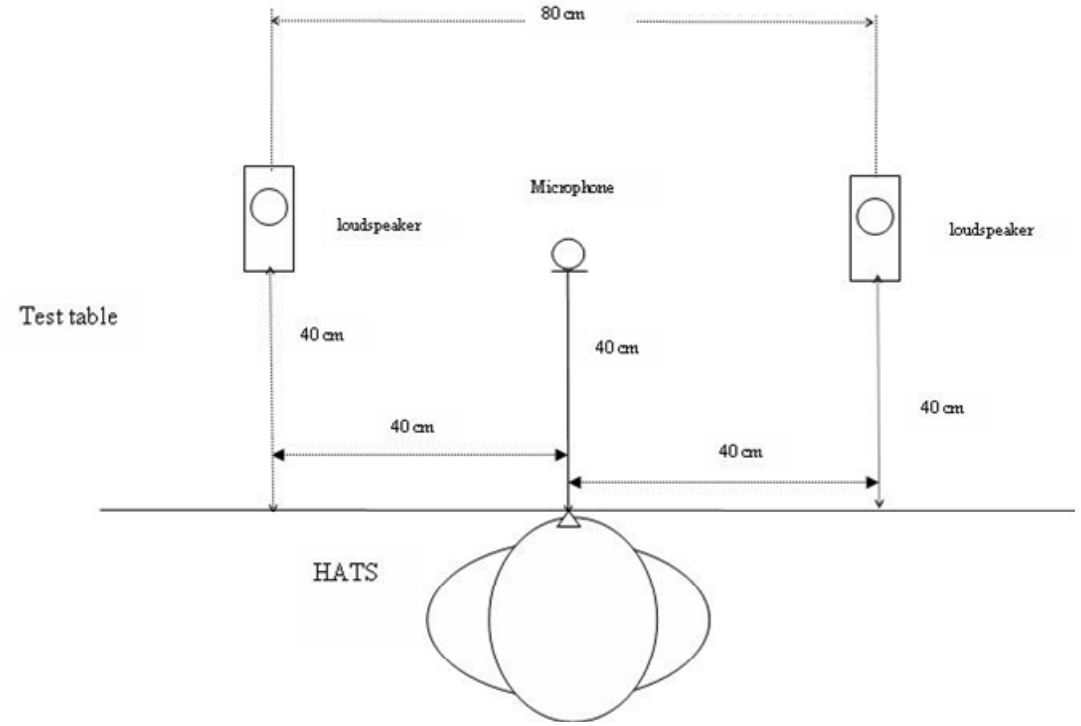

Figure 8: Configuration of softphone using external speakers and microphone relative to the HATS top sight

Group audio terminal

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

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

Measurement will be conducted by using a HATS test equipment.

The following test position will be used.

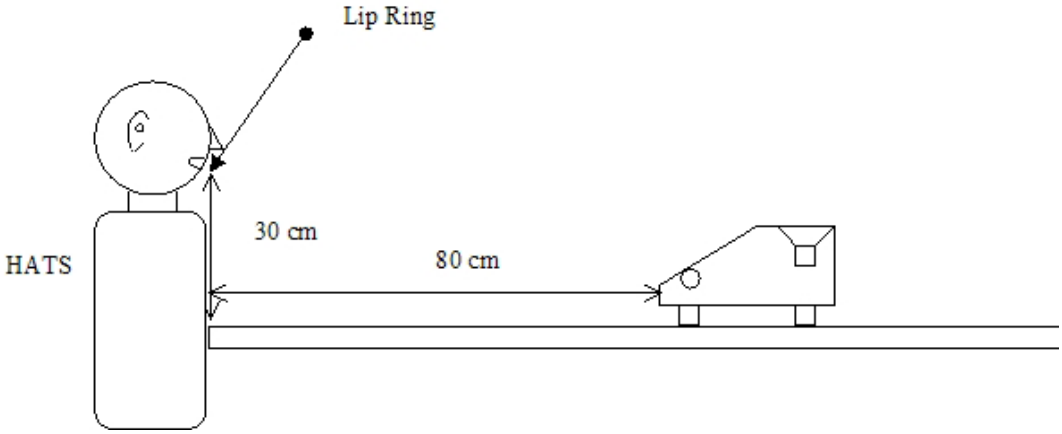

Figure 9: Configuration of group audio terminal relative to the HATS side view

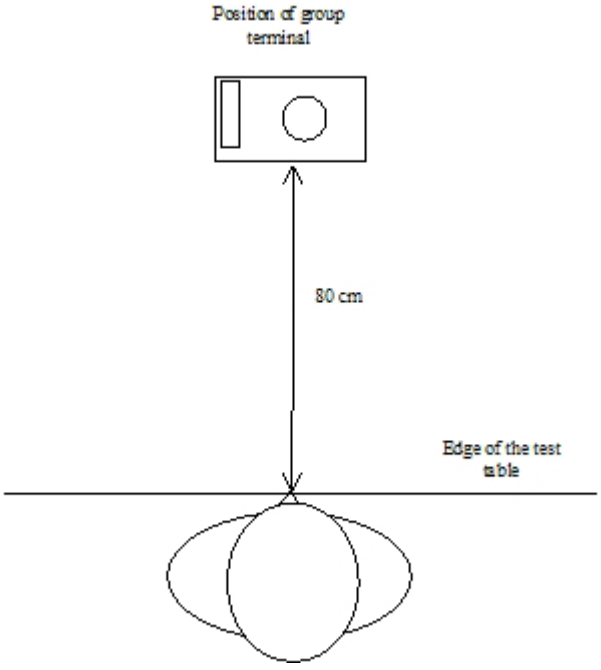

Figure 10: Configuration of group audio terminal relative to the HATS top sight

NOTE: 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.

6.1.2 Measurements in loudspeaking mode

For those measurements HATS will be used.

It will be positioned as defined in clause 6.1.1 measurement will be performed on one ear and handset will be placed on the other ear. The ear used for measurement will be specified in test report.

NOTE: Only desktop terminals are concerned by loudspeaking measurement.

6.2 Test signal levels

6.2.1 Send

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

The following procedure shall be used to perform the calibration of the artificial mouth of the HATS:

The input signal from the artificial mouth is first calibrated under free-field conditions at the MRP. The total level on the frequency range is set to -4,7 dBPa.

The spectrum at MRP is recorded.

This spectrum is used as a reference for the send characteristics.

The spectrum at the HATSHFRP is recorded.

The spectrum at HATSHFRP is calibrated to the nominal spectrum given for the relevant HATSHFRP in [14] table 7d and [14] table 7e.

Then the level is adjusted to the level given further in this text (depending of type of terminal tested (for example -24,3 dBPa at 30 cm for handheld terminal)).

The level at MRP (measured in third octave bands) adjusted at the first step (with total level of -4,7 dBPa) is used as the reference for send characteristics.

The test setup shall be in conformance with figure 11 but, depending on the type of terminal, the appropriate distance and level will be used. When using this calibration method, send sensitivity shall be calculated as follows:

- **SmJ** = 20 log **Vs** - 20 log **PMRP**

where:

- **Vs** is the measured voltage across the appropriate termination (unless stated otherwise, a 600 Ω termination).
- **PMRP** is the applied sound pressure at the MRP at the first step of calibration.

NOTE: Reason for this procedure of calibration in two steps is to take into account the different variation of signal with distance by using different implementations of HATS.

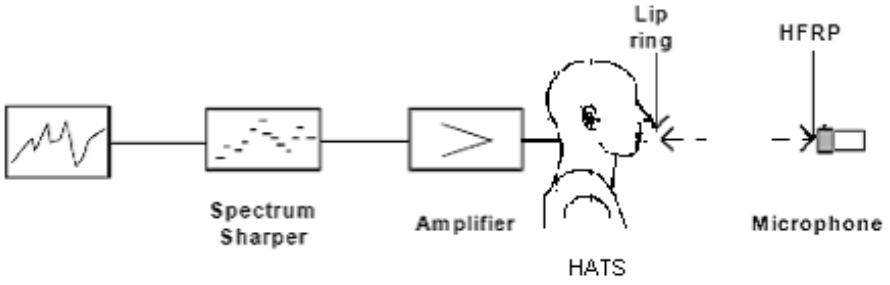

Figure 11: Calibration at HFRP (with d_{HFS} = 50 cm)

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.

6.2.2 Receive

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

All measurement values produced by HATS are intended to be free-field equalized.

6.3 Setup of background noise simulation

A setup for simulating realistic background noises in a lab-type environment is described in ETSI ES 202 396-1 [25].

The ETSI ES 202 396-1 [25] contains a description of the recording arrangement for realistic background noises, a description of the setup for a loudspeaker arrangement suitable to simulate a background noise field in a lab-type environment and a database of realistic background noises, which can be used for testing the terminal performance with a variety of different background noises.

The principle loudspeaker setup for the simulation arrangement is shown in figure 12.

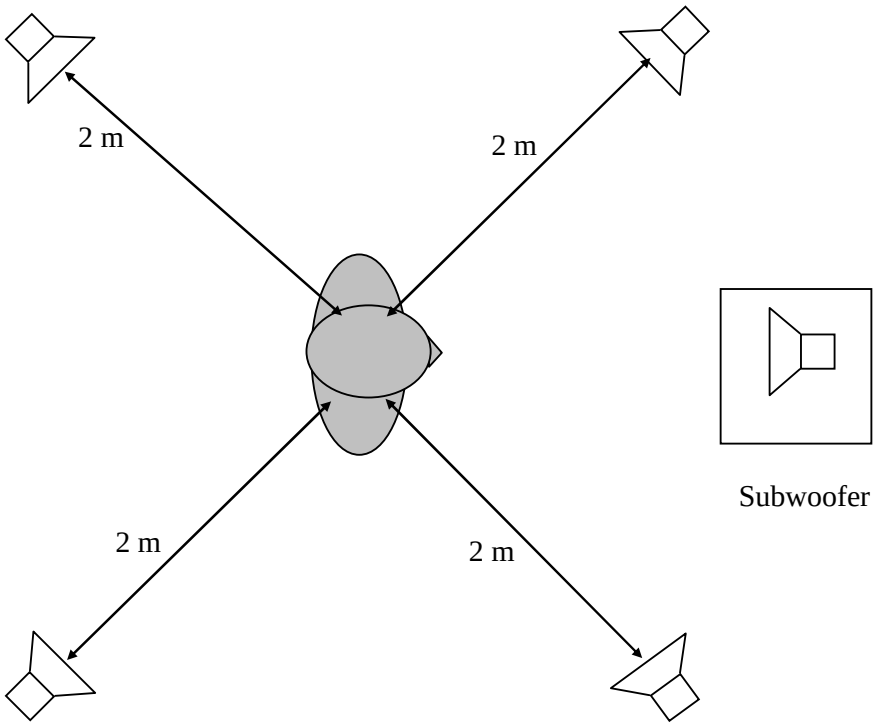

Figure 12: Loudspeaker arrangement for background noise simulation

The equalization and calibration procedure for the setup is described in detail in ETSI ES 202 396-1 [25].

If stated otherwise this setup is used in all measurements where background noise simulation is required.

The following noises of ETSI ES 202 396-1 [25] shall be used.

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

7 Measurements and Requirements for Basic Parameters

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

NOTE 2: Due to time variant nature of IP connection, delay variation may impair the measurement. In such case, the measurement has to be repeated until a valid measurement can be achieved.

7.1 Coding independent parameters

7.1.1 Send sensitivity/frequency response

7.1.1.1 Requirement

The send sensitivity/frequency response shall be within the limits given in table 3.

Table 3

Frequency	Upper limit	Lower limit
100 Hz	4 dB	
125 Hz	4 dB	-10 dB
200 Hz	4 dB	-4 dB
1 000 Hz	4 dB	-4 dB
5 000 Hz	(see note)	-4 dB
6 300 Hz	9 dB	-7 dB
8 000 Hz	9 dB	
NOTE: The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) - logarithmic (Hz) scale.		

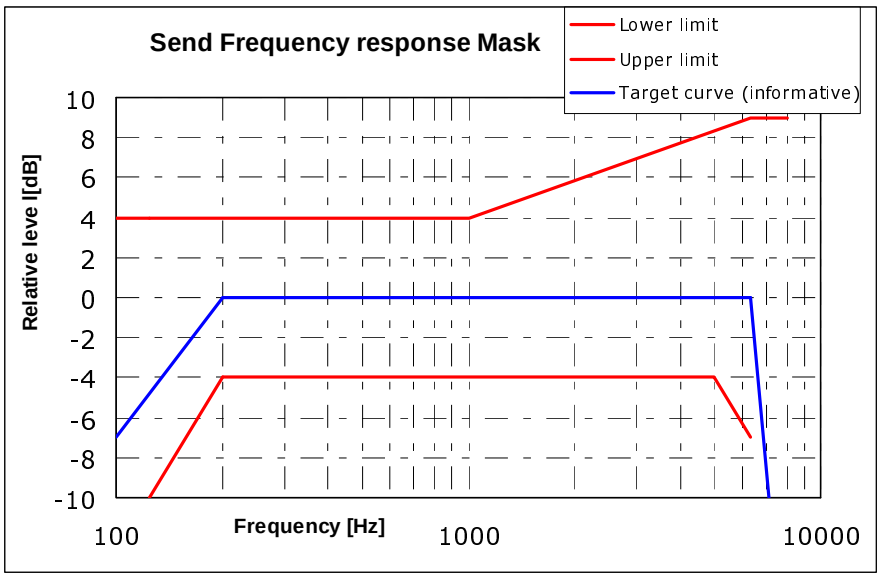

Figure 13: Send sensitivity/frequency mask for HFT

NOTE 1: Level at 125 Hz can be reduced (low limit at -10 dB, it can be useful for reduction of transmitted noise and obtaining a more well balanced response curve relative to high frequencies (see note 2)).

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

7.1.1.2 Measurement method

The terminal will be positioned as described in clause 6.1.

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 [19]. The spectrum of acoustic signal produced by the artificial mouth is calibrated under free field conditions at the MRP. The signal level is adjusted according to clause 6.2.1.

The spectrum at the MRP and the actual level at the MRP (measured in third octaves) is used as reference to determine the send sensitivity SmJ.

Measurements shall be made at one third-octave bands as given in IEC 61260 [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 dBV/Pa.

7.1.2 Send loudness rating

7.1.2.1 Requirement

The value of SLR shall be +13 dB ± 3 dB.

This value is derived from Recommendation ITU-T P.310 [16]. According to Recommendation ITU-T P.340 [17], the SLR of a hands-free telephone should be about 5 dB higher than the SLR of the corresponding handset telephone.

This value will be identical for all type of terminal (desktop, handheld, etc.). Difference in efficiency will be given by conditions for measurement (see clause 6.1).

NOTE: 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 [15] the loudness rating calculation as described in annex A is used.

7.1.2.2 Measurement method

The terminal will be positioned as described in clause 6.1.

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 [19]. 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 6.2.1.

The send sensitivity shall be calculated from each band of the 20 frequencies given in table 1 of Recommendation ITU-T P.79 [15], 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 [15], annex A.

7.1.3 Send distortion

7.1.3.1 Requirement

The terminal will be positioned as described in clause 6.1.

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

Table 4

Frequency	Ratio
200 Hz	25 dB
315 Hz	26 dB
400 Hz	30 dB
1 kHz	30 dB
2 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.	

7.1.3.2 Measurement method

The terminal will be positioned as described in clause 6.1.

The signal used is an activation signal followed by a series sine wave signal with a frequency at 200 Hz, 315 Hz, 400 Hz, 500 Hz, 630 Hz, 800 Hz, 1 000 Hz and 2 kHz. The duration of the sine wave shall be of less than 1 second. The sinusoidal signal level shall be calibrated to -4,7 dBPa at the MRP.

The signal to harmonic distortion ratio is measured selectively up to 6,3 kHz.

The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [19] shall be used for activation. The level of this activation signal shall be -4,7 dBPa at the MRP.

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

7.1.4 Out-of-band signals in send direction (informative)

7.1.4.1 Requirement

The level of any in-band image frequencies resulting from application of input signals at 8 kHz and above should be attenuated by at least 25 dB compared to the output level of a 1 kHz input signal.

7.1.4.2 Measurement method

The terminal will be positioned as described in clause 6.1.

The female speaker of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [19] shall be used for activation. The level of this activation signal shall be -4,7 dBPa at the MRP.

For the test, an out-of-band signal shall be provided as a frequency band signal centred on 8,5 kHz, 9 kHz and 10 kHz respectively. The level of any image frequencies at the digital interface shall be measured.

The levels of these signals shall be -4,7 dBpa at the MRP.

The complete test signal is constituted by t1 ms of in-band signal (reference signal), t2 ms of out-of-band signal and another time t1 ms of in-band signal (reference signal).

The observation of the output signal on the first and second in-band signals permits control if the set is correctly activated during the out-of-band measurement. This measurement shall be performed during t2 period:

- a value of 250 ms is suggested for t1;
- t2 depends on the integration time of the analyser, typically less than 150 ms.

NOTE: The frequency range of artificial mouth according to Recommendation ITU-T P.58 [14] is specified up to 8 kHz. The production of out-of-band frequencies up to 10 kHz however is possible. So the out-of-band test is limited up to 10 kHz.

7.1.5 Send noise

7.1.5.1 Requirement

The limit for the send noise is the following:

- send noise level maximum -64 dBm0(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.

7.1.5.2 Measurement method

The terminal will be positioned as described in clause 6.1.

The female speaker of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [19] shall be used for activation. The level of this activation signal shall be -4,7 dBPa at the MRP.

The level at the output of the test setup is measured with a A weighting.

Spectral peaks are measured in the frequency domain. The frequency spectrum of the A-weighted idle channel noise is measured by a spectral analysis having a noise bandwidth of 8,79Hz (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.

7.1.6 Receive Frequency Response

7.1.6.1 Requirement

The following masks are required for handsfree and loudspeaking terminals. The mask is drawn as straight lines between the breaking points in the table on a logarithmic (frequency) - linear (dB sensitivity) scale.

Desktop operated loudspeaker

Table 5: Receive frequency response mask-desktop

Frequency	Upper limit	Lower limit
125 Hz	8 dB	
200 Hz	8 dB	-12 dB
250 Hz	8 dB	-9 dB
315 Hz	7 dB	-6 dB
400 Hz	6 dB	-6 dB
5 000 Hz	6 dB	-6 dB
6 300 Hz	6 dB	-9 dB
8 000 Hz	6 dB	

NOTE 1: Referring to ETSI ETS 300 245-6 [1], low limit has been modified: no requirement at 160 Hz, -12 dB at 200 Hz and -9 dB at 250 Hz instead of -15 dB, -9 dB and -6 dB. Rationale: better balanced response curve and avoiding necessity in most case to introduce "bass boost" for amplification.

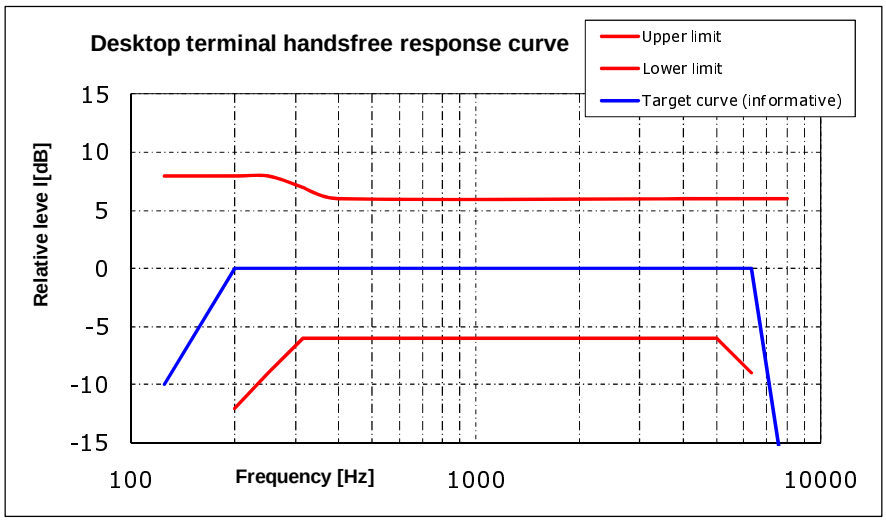

Figure 14: Receive frequency mask for Desktop HFT

Handheld hands-free terminal

Table 6: Receive frequency response mask-handheld

Frequency	Upper limit	Lower limit
125 Hz	6 dB	
400 Hz	6 dB	-12 dB
500 Hz	6 dB	-6 dB
4 000 Hz	6 dB	-6 dB
5 000 Hz	6 dB	-9 dB
6 300 Hz	6 dB	-12 dB
8 000 Hz	6 dB	
NOTE: The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) - logarithmic (Hz) scale.		

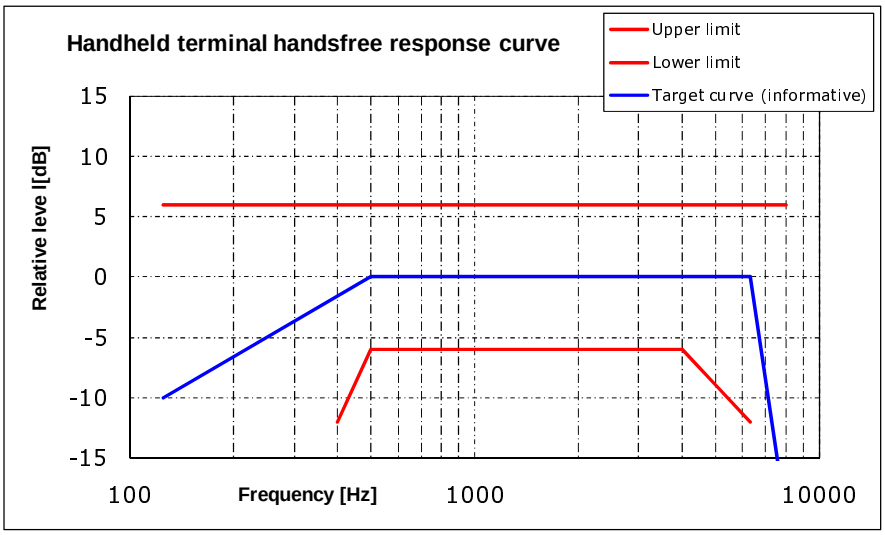

Figure 15: Receive frequency response mask for Hand-held HFT

NOTE 2: At high frequencies, low limit is relaxed. It is necessary to take into account that in most case measurement will be made facing to the opposite side of output of loudspeaker; see figure 4.

Softphone (computer-based terminals)

Type 1 or softphone with external speakers: requirement as for desktop terminal.

Type 2 requirement as for handheld terminal.

Group audio terminal

Same requirement as desktop terminals.

7.1.6.2 Measurement method

The test setup is described in clause 6.1.

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)

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

S_{Jeff} Receive Sensitivity; Junction to HATS Ear with free field correction.

pe_{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 [19]. The test signal level shall be -20 dBm0, measured according to Recommendation ITU-T P.56 [13] at the digital reference point or the equivalent analogue point.

The HATS is free field equalized as described in Recommendation ITU-T P.581 [21]. 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 bands as in IEC 61260 [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.

7.1.7 Receive Loudness Rating

7.1.7.1 Requirement

Desktop operated loudspeaker

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 equal or exceed 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 equal or exceed 19 dB.

NOTE: 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 [15] the loudness rating calculation as described in annex A is used.

7.1.7.2 Measurement method

The test setup is described in clause 6.1.

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 [19]. The test signal level shall be -20 dBm0, measured according to Recommendation ITU-T P.56 [13] at the digital reference point or the equivalent analogue point.

The receive sensitivity shall be calculated from each band of the 20 frequencies given in table 1 of Recommendation ITU-T P.79 [15], 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 [15], annex A. The RLR shall then be corrected as RLR minus 14 dB according to Recommendation ITU-T P.340 [17] and without the LE factor.

7.1.8 Receive distortion

7.1.8.1 Requirement

Desktop and Handheld terminals

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

Table 7

Frequency	Signal to distortion ratio limit, receive for desktop terminal	Signal to distortion ratio limit, receive for handheld terminal	Signal to distortion ratio limit, receive for all terminals at maximum volume
315 Hz	26 dB		
400 Hz	30 dB		
500 Hz	30 dB	20 dB	
800 Hz	30 dB	30 dB	20 dB
1 kHz	30 dB	30 dB	
2 kHz	30 dB	30 dB	
3 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.			

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

Same requirement as for desktop terminal.

7.1.8.2 Measurement method

The test setup is described in clause 6.1.

The signal used is an activation signal followed by a sine wave signal with a frequency at 315 Hz, 400 Hz, 500 Hz, 630 Hz, 800 Hz, 1 000 Hz, 2 000 Hz, 3 000 Hz. The duration of the sine wave shall be of less than 1 second. Appropriate signals for activation and signal combinations can be found in Recommendation ITU-T P.501 [19]. 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 [19] 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 10 kHz.

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

7.1.9 Out-of-band signals in receive direction (informative)

7.1.9.1 Requirement

Any spurious out-of-band image signals in the frequency range from 9 kHz to 12 kHz measured selectively shall be lower than the in-band level measured with a reference signal. The minimum level difference between the reference signal level and the out-of-band image signal level shall be as given in table 8.

Table 8

Frequency	Signal limit
9 kHz	50 dB
10 kHz	52 dB
NOTE: The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) - logarithmic (kHz) scale.	

7.1.9.2 Measurement Method

The test setup is described in clause 6.1.

Measurement is operated at nominal value of volume control.

The signal used is an activation signal followed by a sine wave signal. For input signals at the frequencies 6 kHz and 7 kHz applied at the level of -16 dBm0, the level of spurious out-of-band image signals at frequencies up to 10 kHz is measured selectively at measurement point.

The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [19] shall be used for activation. Level of this activation signal shall be -16 dBm0.

7.1.10 Receive noise

7.1.10.1 Requirement

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

The level in any 1/3-octave band, between 100 Hz and 10 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.

7.1.10.2 Measurement method

The test setup is described in clause 6.1.

The female speaker signal of the short conditioning sequence described in clause 7.3.7 of Recommendation ITU-T P.501 [19] shall be used for activation. Level of this activation signal shall be -16 dBm0.

The noise level is measured until 10 kHz.

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

7.1.11 Terminal Coupling Loss

7.1.11.1 Requirement

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

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

7.1.11.2 Measurement method

The setup for terminal is described in clause 6.1.

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

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

The test signal is the compressed real speech signal described in clause 7.3.3 of Recommendation ITU-T P.501 [19].

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 seconds of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences).

The ambient noise level shall be < -64 dBPa(A).

NOTE: It should be relevant to perform the test in a real room instead in an anechoic room.

7.1.12 Stability Loss

7.1.12.1 Requirement

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. It shall exceed 6 dB for all frequencies and for all settings of volume control.

7.1.12.2 Measurement method

For handsfree mode the test setup is identical as for TCL.

For loudspeaking mode handset is placed at 50 cm beside terminal with transducers facing the table (see figure 16).

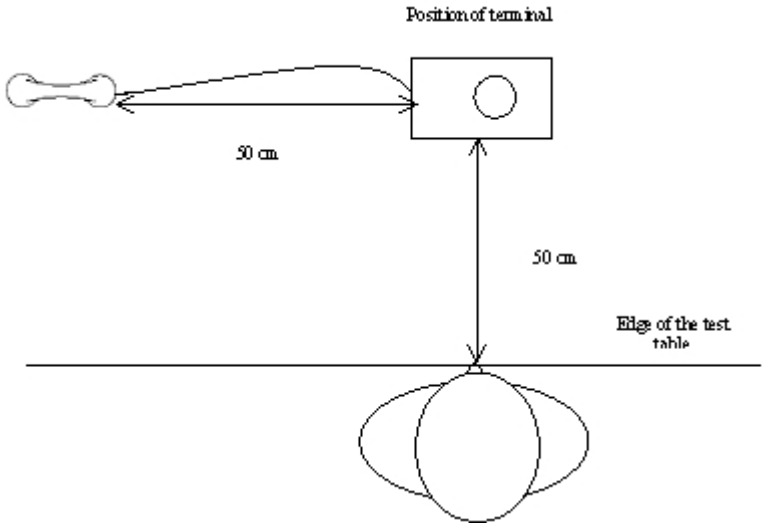

Figure 16: Stability loss position for loudspeaking function

Before the actual test a training sequence consisting of the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [19] shall be applied. The training sequence level shall be -16 dBm0 in order not to overload the codec.

The test signal is a PN sequence complying with Recommendation ITU-T P.501, annex C [19] with a length of 4 096 points (for the 48 kHz sampling rate) and a crest factor of 6 dB. The duration of the test signal is 250 ms. With an input signal of -3 dBm0, the attenuation from digital input to digital output shall be measured for frequencies from 100 Hz to 8 kHz.

7.2 Codec Specific Requirements

7.2.1 Send Delay

For a VoIP terminal, send delay is defined as the one-way delay from the acoustical input (mouthpiece) of this VoIP terminal to its interface to the packet based network. The total send delay is the upper bound on the mean delay and takes into account the delay contributions of all of the elements shown in figures 2 and A.1 of Recommendation ITU-T G.1020 [11], respectively.

The send delay T(s) is defined as follows:

$$T(s) = T(ps) + T(la) + T(rif) + T(asp) \quad (2)$$

Where:

- T(ps) = packet size = N × T(fs);
- N = number of frames per packet;

- T(fs) = frame size of encoder;
- T(la) = look-ahead of encoder;
- T(aif) = air interface framing;
- T(asp) = allowance for signal processing.

The additional delay required for IP packet assembly and presentation to the underlying link layer will depend on the link layer. When the link layer is a LAN (e.g. Ethernet), this additional time will usually be quite small. For the purposes of the present document it is assumed that in the test setup this delay can be neglected.

NOTE: The size of T(aif) is for further study.

7.2.1.1 Requirement

The allowance for signal processing shall be T(asp) < 40 ms (including processing time for handsfree).

NOTE: With the knowledge of the codec specific values for T(fs) and T(la) the requirements for send delay for any type of coder and any frame size T(fs) can easily be calculated by equation 2. Table 9 provides requirements calculated accordingly for frequently used codecs and packet sizes.

Table 9

Codec	N	T(fs) in ms	T(ps) in ms	T(la) in ms	T(aif) in ms	T(asp) in ms	T(s) Requirement in ms	T(s) Requirement in ms
Recommendation ITU-T G.722 [8]	80	0,0625	10	0	0	40	< 50,0625	< 51
	160	0,0625	20	0	0	40	< 60,0625	< 61
Recommendation ITU-T G.722.1 [9]	1	20	10	5	0	40	To be completed	To be completed
	2	20	20	5	0	40	To be completed	To be completed
L16-256	160	0,0625	10	0	0	40	< 60,0625	< 61

Further information about the different sources of delay for different codecs can be found in annex A.

7.2.1.2 Measurement Method

The test setup is described in clause 6.1.

The test signal to be used for the measurements shall be a Composite Source Signal (CSS) as described in Recommendation ITU-T P.501, annex C [19]. The test signal consists of the voiced part as described in Recommendation ITU-T P.501, annex C [19] followed by a pseudo random noise sequence with a periodicity of minimum 500 ms. 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.

NOTE 1: If the expected delay is higher than 500 ms a pseudo random sequence with a higher periodicity should be used.

The delay is calculated using the cross correlation function between the signal at the electrical test point and the signal at the MRP. The cross correlation analysis has to be chosen in such a way that the maximum delay of 500 ms can be analysed. The measurement is corrected by the delay introduced by the test equipment.

The delay is expressed in ms, determined from the maximum of the cross correlation function.

NOTE 2: Delay may be time variant. Therefore constant monitoring of the actual delay may be required when evaluating the range of delay which can be observed in a given connection. The test setup should take into account either real network conditions or the tools needed to simulate typical causes for time variant delay (e.g. packet loss) during the measurement period. Other methods like running cross correlation or delay estimation procedures e.g. used in PESQ (Recommendation ITU-T P.862 [22]) may be used.

7.2.2 Receive delay

For a VoIP terminal, receive delay is defined as the one-way delay from the interface to the packet based network of this VoIP terminal to its acoustical output (earpiece). The total receive delay is the upper bound on the mean delay and takes into account the delay contributions of all of the elements shown in figures 3 and A.2 of Recommendation ITU-T G.1020 [11], respectively.

The receive delay T(r) is defined as follows:

$$T(r) = T(fs) + T(fi) + T(aif) + T(jb) + T(plc) + T(asp) \quad (3)$$

Where:

- T(fs) = frame size of encoder;
- T(fi) = filter processing delay;
- T(aif) = air interface framing;
- T(jb) = jitter buffer size;
- T(plc) = PLC buffer size;
- T(asp) = allowance for signal processing.

The additional delay required for IP packet dis-assembly and presentation from the underlying link layer will depend on the link layer. When the link layer is a LAN (e.g. Ethernet), this additional time will usually be quite small. For the purposes of the present document it is assumed that in the test setup this delay can be neglected.

NOTE: The size of T(aif) is for further study.

7.2.2.1 Requirements

The allowance for signal processing shall be T(asp) < 10 ms.

The additional delay introduced by the jitter buffer shall be T(jb) ≤ 10 ms.

For Coders with integrated PLC the additional PLC buffer size shall be T(plc) = 0 ms.

For Coders with integrated PLC the additional PLC buffer size shall be T(plc) = 0 ms.

NOTE 1: With the knowledge of the codec specific values for T(fs) and T(la) the requirements for receive delay for any type of coder and any frame size T(fs) can easily be calculated by equation 3. Table 10 provides requirements calculated accordingly for some frequently used codecs and packet sizes as an example.

Table 10

Codec	N	T(fs)	T(fi)	T(aif)	T(jb)	T(plc)	T(asp)	T(r) Requirement	T(r) Requirement in ms
Recommendation ITU-T G.722 [8]	80	0,0625	0	0	10	10	10	< 30,0625	< 31
	160	0,0625	0	0	10	10	10	< 30,0625	< 31
Recommendation ITU-T G.722.1 [9]	1	20	0	0	10	0	10	< 40	< 40
	2	20	0	0	10	0	10	< 40	< 40
L 16-256	160	0,0625	0	0	10	10	10	< 30,0625	< 31
NOTE 1: T(ps) = packet size = N × T(fs). NOTE 2: N = number of frames per packet.									

NOTE 2: These requirements are 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.

7.2.2.2 Measurement Method

The test setup is described in clause 6.1.

The test signal to be used for the measurements shall be a Composite Source Signal (CSS) as described in Recommendation ITU-T P.501, annex C [19]. The test signal consists of the voiced part as described in Recommendation ITU-T P.501, annex C [19] followed by a pseudo random noise sequence with a periodicity of minimum 500 ms. The test signal level shall be -16 dBm0, measured at the electrical test point. The test signal level is averaged over the complete test signal sequence.

NOTE 1: If the expected delay is higher than 500 ms a pseudo random sequence with a higher periodicity should be used.

The delay is calculated using the cross correlation function between the signal at the electrical test point and the signal at the DRP. The cross correlation analysis has to be chosen in such a way that the maximum delay of 500 ms can be analysed. The measurement is corrected by the delay introduced by the test equipment.

The delay is expressed in ms, determined from the maximum of the cross correlation function.

NOTE 2: Delay may be time variant. Therefore constant monitoring of the actual delay may be required when evaluating the range of delay which can be observed in a given connection. The test setup should take into account either real network conditions or the tools needed to simulate typical causes for time variant delay (e.g. packet loss) during the measurement period. Other methods like running cross correlation or delay estimation procedures e.g. used in PESQ (Recommendation ITU-T P.862 [22]) may be used.

8 Measurements and Requirements for Parameters with respect to Speech Processing Devices

8.1 Objective Listening Speech Quality MOS-LQOM in Send direction

For further study.

8.2 Objective Listening Quality MOS-LQOM in Receive direction

For further study.

8.3 Minimum activation level and sensitivity in Receive direction

For further study.

8.4 Automatic Level Control in Receive

For further study.

8.5 Double Talk Performance

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

In order to guarantee sufficient quality under double talk conditions the Talker Echo Loudness Rating 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 [17] and P.502 [20]):

- 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.

8.5.1 Attenuation Range in Send Direction during Double Talk $A_{H,S,dt}$

8.5.1.1 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 11.

Table 11

Category (according to Recommendation ITU-T P.340 [17])	**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 11 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.

8.5.1.2 Measurement Method

The test signal to determine the attenuation range during double talk is the double talk speech sequence as defined in clause 7.3.5 of Recommendation ITU-T P.501 [19] as shown in figure 17. The competing speaker is always inserted as the double talk sequence sdt(t) either in send or receive and is used for analysis.

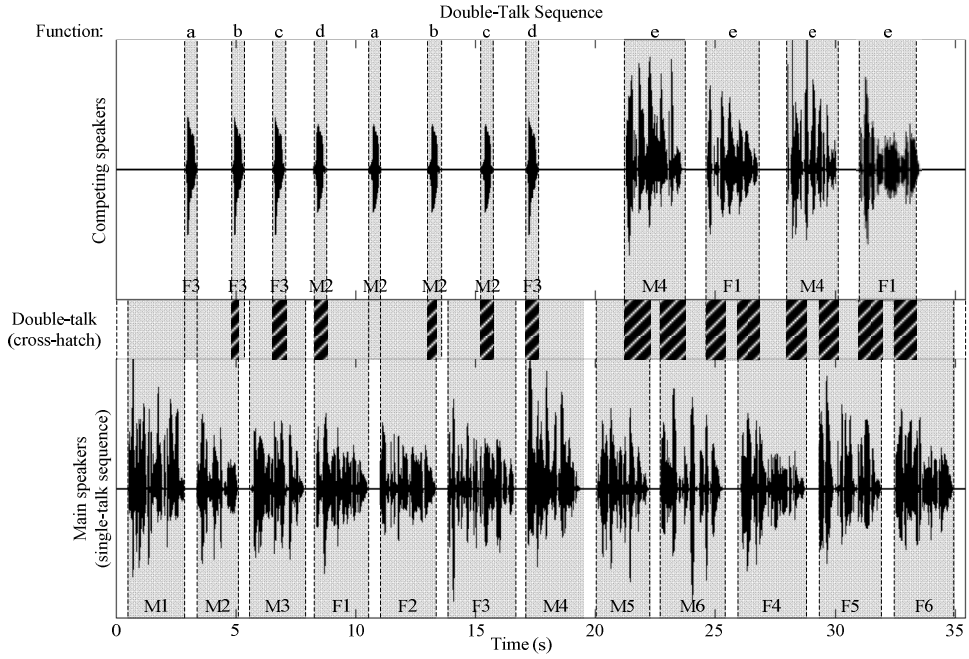

Figure 17: Double Talk Test Sequence with overlapping speech sequences in send and receive direction

Measurement method

The attenuation range during double talk is determined as described in Appendix III of Recommendation ITU-T P.502 [20]. 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.

Table 12: Void

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

8.5.2.1 Requirement

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

Table 13

Category (according to Recommendation ITU-T P.340 [17])	**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 13 provides a quality classification of terminals regarding double talk performance. However, this does not mean that a terminal which is category 1 based on the double talk performance is of high quality concerning the overall quality as well.

8.5.2.2 Measurement Method

The test setup is described in clause 6.1.

The test signal to determine the attenuation range during double talk is shown in figure 17. 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 [20]. 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.

Table 14: Void

8.5.3 Detection of Echo Components during Double Talk

8.5.3.1 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 15 are applicable (more information can be found in annex A of the Recommendation ITU-T P.340 [17].

Table 15

Category (according to Recommendation ITU-T P.340 [17])	**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).

8.5.3.2 Measurement Method

The test setup is described in clause 6.1.

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. The measurement signals used are shown in figure 17a. A detailed description can be found in Recommendation ITU-T P.501, annex C [19].

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

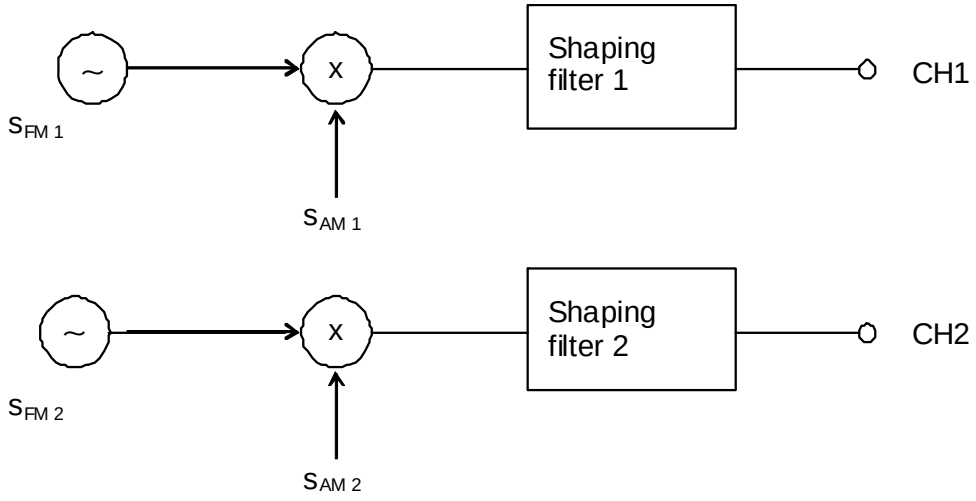

Figure 17a: Measurement signals

$$s_{FM1,2}(t) = \sum A_{FM1,2} * \cos(2\pi t\, n * F_{01,2}) \;;\; n = 1,2,... \quad (4)$$

$$s_{AM1,2}(t) = A_{AM1,2} * \cos(2\pi t F_{AM1,2}); \quad (5)$$

NOTE: A is determined by the required test signal level as found in the individual test cases.

The settings for the signals are as follows.

Table 16: Parameters of the two Test Signals for Double Talk Measurement based on AM-FM modulated sine waves

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

The test signal is measured at the electrical reference point (send direction). The measured signal consists of the double talk signal which was fed in by the artificial mouth and the echo signal. The echo signal is filtered by comb filter using mid-frequencies and bandwidth according to the signal components of the signal in receive direction (see Recommendation ITU-T P.501, annex C [19]). 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 16. The echo attenuation is to be achieved for **each individual frequency band** according to the different categories.

8.5.4 Minimum activation level and sensitivity of double talk detection

For further study.

8.5.5 Switching characteristics

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

8.5.5.1 Activation in Send Direction

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

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

8.5.5.1.1 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.

8.5.5.1.2 Measurement Method

The test setup is described in clause 6.1.

The test signal is the activation of the short conditioning sequence described in clause 7.3.4 of Recommendation ITU-T P.501 [19] with increasing level for each single word.

Figure 18: Void

The settings of the test signal are as follows.

Table 17

	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 notes)	1 dB
NOTE 1: The level of the active signal part corresponds to an average level of -24,7 dBPa at the MRP for the test signals according to Recommendation ITU-T P. 501, annex C [19] assuming a pause of about 400 ms. NOTE 2: The signal level is determined for each utterance individually according to Recommendation ITU-T P.56 [13].			

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.

8.5.5.2 Silence Suppression and Comfort Noise Generation

For further study.

8.5.5.3 Performance in send direction in the presence of background noise

8.5.5.3.1 Requirement

The level of comfort noise, if implemented, shall be within in a range of +2 dB and -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 18.

Table 18: Requirements for Spectral Adjustment of Comfort Noise (Mask)

Frequency	Upper Limit	Lower Limit
200 Hz	12 dB	-12 dB
800 Hz	12 dB	-12 dB
800 Hz	10 dB	-10 dB
2 000 Hz	10 dB	-10 dB
2 000 Hz	6 dB	-6 dB
4 000 Hz	6 dB	-6 dB
8 000 Hz	6 dB	-6 dB
NOTE: All sensitivity values are expressed in dB on an arbitrary scale.		

8.5.5.3.2 Measurement Method

The test setup is described in clause 6.1.

The background noise simulation as described in clause 6.3 is used.

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

In a second step a test signal is applied in receive direction consisting of an initial pause of 10 seconds and a periodical repetition of the Composite Source Signal in receive direction (duration 10 seconds) with nominal level to enable comfort noise injection simultaneously with the background noise. For the measurement the background noise sequence has to be started at the same point as it was started in the previous measurement. Alternatively other speech like test signals (e.g. artificial voice) with the same signal level can be used.

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

The power density spectra measured in send direction without far end speech simulation averaged between 10 seconds and 20 seconds 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 seconds and 20 seconds. Level and spectral differences between both power density spectra are analysed and compared to the requirements.

8.5.5.4 Speech Quality in the Presence of Background Noise

8.5.5.4.1 Requirement

Speech Quality for wideband systems can be tested based on ETSI EG 202 396-3 [i.2]. The test method is applicable for narrowband (100 Hz to 4 kHz) and wideband (100 Hz to 8 kHz) transmission systems. LQOn is used for narrowband and LQOw is used for wideband systems. The test method described leads to three MOS-LQO quality numbers:

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

For the background noises defined in clause 7.1 the following requirements apply:

- N-MOS-LQOw $\geq$ 3.0,
- S-MOS-LQOw $\geq$ 3.0,
- G-MOS-LQOw $\geq$ 3.0.

NOTE: It is recommended to test the terminal performance with other types of background noises if the terminal is likely to be exposed to other noises than specified in clause 6.1.

8.5.5.4.2 Measurement Method

The background noise simulation as described in clause 7.1 is used. The terminal is set-up as described in clause 6.1.

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

The near end speech signal consists of 8 sentences of speech (2 male and 2 female talkers, 2 sentences each). Appropriate speech samples can be found in Recommendation ITU-T P.501, annex C [19]. The preferred language is English since the objective method was validated with English language. The test signal level is +1,3 dBPa at the MRP.

Three signals are required for the tests:

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

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

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

N-MOS-LQOw, S-MOS LQOw and G-MOS LQOw are calculated as described in ETSI EG 202 396-3 [i.2].

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

8.5.5.5.1 Requirements

The test is carried out applying the Composite Source Signal in receive direction. During and after the end of Composite Source Signal bursts (representing the end of far end speech simulation) the signal level in send direction should not vary more than 10 dB (during transition to transmission of background noise without far end speech). The measurement is conducted for all types of background noise as defined in clause 7.1.

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.

8.5.5.5.2 Measurement Method

The test setup is described in clause 6.1.

The background noises are generated as described in clause 6.3.

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

In a second step the same measurement is conducted but with inserting the CS-signal at the far end. The exactly identical background noise signal is applied. The background noise signal 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 seconds in order to allow adaptation of the noise reduction algorithms. After at least 5 seconds a Composite Source Signal according to Recommendation ITU-T P.501, annex C [19] is applied in receive direction with a duration of ≥ 2 CSS periods. The test signal level is -16 dB_{m0} at the electrical reference point.

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

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

8.5.6 Quality of echo cancellation

8.5.6.1 Temporal echo effects

8.5.6.1.1 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.

8.5.6.1.2 Measurement Method

The test setup is described in clause 6.1.

The test signal consists of periodically repeated Composite Source Signal according to Recommendation ITU-T P.501, annex C [19] 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 CS signal. The integration time for the level analysis shall be 35 ms, the analysis is referred to the level analysis of the reference signal.

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

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.50 [12] to see time variant behaviour of EC. However for such tests the simple broadband attenuation based test principle as described above cannot be applied due to the time varying spectral content of the speech like signals.

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

8.5.6.2 Spectral Echo Attenuation

8.5.6.2.1 Requirements

The echo attenuation vs. frequency shall be below the tolerance mask given in table 19.

Table 19: 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.

8.5.6.2.2 Measurement Method

The test setup is described in clause 6.1.

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 [19].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 [19]. 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.

8.5.6.3 Occurrence of Artefacts

For further study.

8.5.7 Variant Impairments; Network dependant

8.5.7.1 Send and receive delay - Round trip delay

Requirement

Send and receive delays are tested separately but the requirement is defined for the combination of send and receive delays (round-trip delay).

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

The delay T_{rtd} in send direction T_s plus the delay in receive direction T_r shall be less than 50 ms.

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

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

NOTE 3: This delay does not take into account additional radio link if provided(e.g. Bluetooth link).

As the actual delay depends on the codec implementations, complementary information are given in annex B.

Measurement method

- **Send direction**

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

$$T_s + t_{System}$$

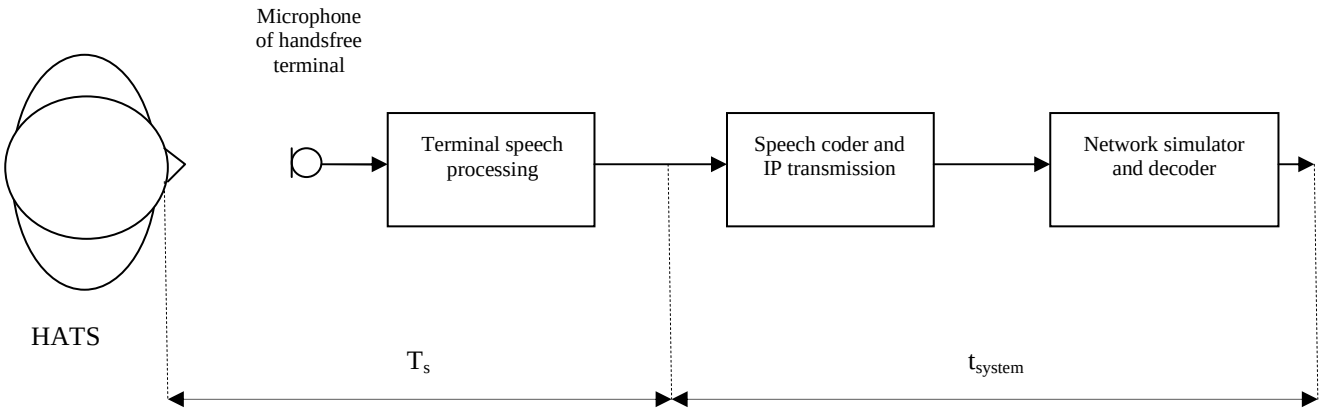

Figure 19: Different blocks contributing to the delay in send direction

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, annex C [19] is used. The pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal level is -4,7 dBPa at the MRP:

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

   - The setup of the handset/headset terminal is in correspondence to clause 5.2.

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

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

- **Receive direction**

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

$$T_r + t_{System}$$

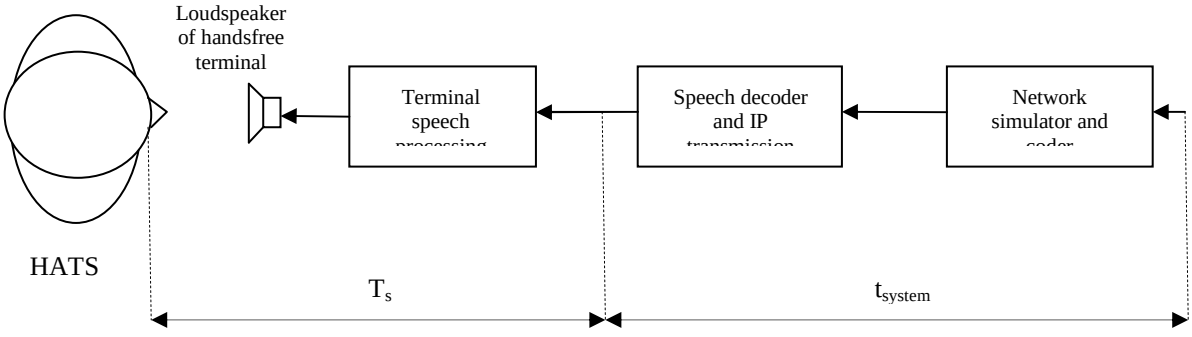

Figure 20: Different blocks contributing to the delay in receive direction

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 [19], annex C 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 5.2.

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

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

8.5.7.2 Delay versus Time Send

For further study.

8.5.7.3 Delay versus Time Receive

For further study.

8.5.7.4 Quality of Jitter buffer adjustment

The listening speech quality and the delay is measured, but with variant network impairments.

Requirements

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

Table 20: Requirements for variant network impairments

Codec	MOS-LQOS
G.722	> 3,6
G.722.1	> 3,8

The delay measured 20 seconds after ending of the IP delay variation should be max. 10 ms higher than the delay measured before the IP delay variation.

Test 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 [26], then again a CSS signal (20 seconds after the IP delay variation stops). The speech signal level is averaged over all used (original) sentences (10 sentences). This test is redone for all of the 10 sentences.

NOTE 1: The 10 used sentences consist of the 10 single sentences taken from the 5 sentence pairs used in clauses 6.3.3 and 6.3.4.

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 seconds 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 12 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-LQOS of the first sentence is measured using Recommendation ITU-T P.863 [27] in superwideband mode. The overall result is the average MOS-LQOS of the 12 measurements.

The second to the fifth sentence (every 5 seconds a sentence) are used to measure the speech quality during jitter buffer adaption (high to low). MOS-LQOS is measured using Recommendation ITU-T P.863 [27] in superwideband mode for each of these four sentences. The minimum MOS-LQOS of these four sentences is used for the averaging over all 12 measurements. The overall result for the speech quality during jitter buffer adaption (high to low) is the average of the minimum MOS-LQOS-value of the 12 measurements.

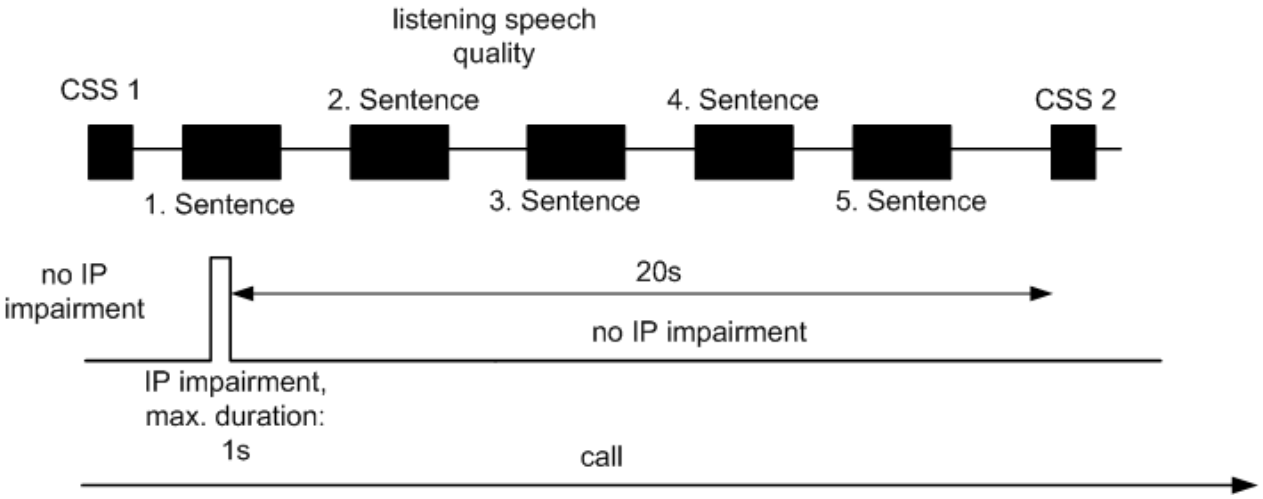

Figure 21: Test Sequence to measure quality of Jitter buffer adjustment (with 1 of 12 sentences)

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 C of ETSI ES 202 718 [28].

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

Annex A (informative): Processing delays in VoIP terminals

This annex gives some elements about delays generated in VoIP terminals. At first, only wired terminals are considered. These terminals could be schematized as shown in figure A.1.

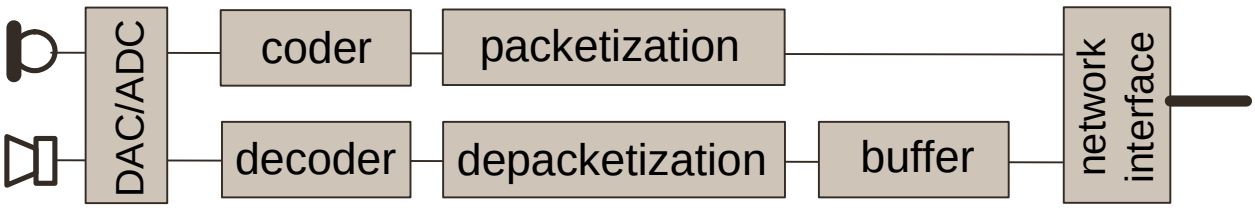

Figure A.1: Synoptic of the different functions implemented in a VoIP terminal

The implemented functions in the send part of the terminal are:

- The analogue-digital conversion.
- The encoding.
- The packetization.
- The interfacing with the network.

The implemented functions in the receive part of the terminal are:

- The interfacing with the network.
- The depacketization.
- The buffering.
- The decoding.
- The digital-analogue conversion.

Let us examine each function's contribution to the processing delay characterizing VoIP terminals.

On the send part of the terminal, the **network interface** operates the transfer of digital data from IP stack to IP network. At the reception, the network interface operates the transfer of digital data from IP network to IP stack. The network interface has a low contribution to the delay. The contribution is estimated at less than 2 ms per transmission way (send and receive direction).

The **packetization** represents the transfer of the audio frames through the IP stack, from the telephony applicative part of the terminal to the transmission network. The packetization consists in adding specific headers (associated to different protocols) to audio frames. The delay associated to the packetization is considered as no significant and included into encoding time.

Encoding corresponds to the compression of the speech signal. The delay associated to the encoding process depends on the implemented codec and the payload's length (number of audio frames) inserted into each IP packet. On the send part of the terminal, encoding is the main contribution to the processing delay. The delay can strongly change according to the codec and the payload's length.

Analogue to digital conversion consists in transforming speech signal from analogue to digital format. The processing delay associated to the conversion is considered as no significant.

Digital to analogue conversion consists in transforming speech signal from digital to analogue format. As analogue to digital conversion, the processing delay associated to digital to analogue conversion is considered as not significant.

The **depacketization** represents the transfer of the audio frames through the IP stack, from transmission network to the telephony applicative part of the terminal. The depacketization consists in tacking off the headers associated to protocols to get back audio frames after transmission. The delay associated to the depacketization is considered as not significant and included into the decoding processing time.

The first role of the **jitter buffer** is to ensure synchronization between send and receive terminals. This synchronization is carried out by buffering the audio frames received from the IP stack before send them to the decoder. The second role of the jitter buffer is to smooth a possible variation of the transmission time. If synchronization of send and receive terminals requires a minimum size of buffer, smoothing transmission delay variation requires a buffer size depending on jitter produced by the network. High variations of transmission time involve an important size of the buffer to smooth jitter. Jitter buffers can be implemented either as buffer with static size(s) (several sizes are possible) or as dynamic buffer. In the last case, size management is carried out according to QoS present on the network interface. Jitter buffer is the main contribution to the processing time on the reception part of VoIP terminal.

Decoding corresponds to the rebuilding of speech signal from receive audio frames. The delay associated to decoding depends on the codec implemented. Decoding contributes in a significant way to the processing time on the reception part of VoIP terminal.

Table A.1 presents the processing times of VoIP terminals for different codecs and IP packet payload's lengths.

In this table, x1, x2, x3, x4, y5, x6 and x7 represent the encoding delays according to selected codec. In the same way, y1, y2, y3, y4, y5, y6 and y7 represent the decoding delays according to selected codec.

According to selected codec and payload's length, columns 5 and 6 show overall encoding and decoding delays respectively. Overall encoding time takes into account algorithm, encoding and packetization delays. Overall decoding time takes into account algorithm, decoding and depacketization delays.

Column 7 shows for each codec and payload's length the real time condition. It stands for the maximum duration to encode and decode at the same time. IP terminals have to meet this requirement.

Column 10 shows the minimum delay induced by the jitter buffer. To ensure a correct running of the VoIP terminal, the minimal size of jitter buffer has to correspond to the IP packet payload's length. Furthermore, a double buffering operation induces 10 additional ms in the overall jitter buffer processing.

Column 12 shows the minimum End to End delay induced by two terminals connected to a "perfect" network (i.e. with no jitter, no packet loss and with a null transmission delay), with real time condition at the lower limit (i.e. no significant encoding and decoding times).

Column 13 shows the minimum End to End delay induced by two terminals connected to a "perfect" network (i.e. with no jitter, no packet loss and with a null transmission delay), with real time condition at the upper limit (i.e. encoding + decoding times very close to the payload size).

Table A.1

Codec	Frame	Lookahead	Payload	Sending procesing delay = Algorithm delay + coding and packetization delay	Receiving procesing delay = Algorithm delay + coding and packetization delay	Real time condition	Network interface and ADC delay	Network interface and DAC delay	Minimum delay of the jitter buffer	Maximum delay of the jitter buffer	Minimum End to End delay with the lower jitter buffer processing time when real time condition is minimum (x+y=0)	Minimum End to End delay with the lower jitter buffer processing time when real time condition is maximum (x+y=upper limit)	Maximum End to End delay with the higher jitter buffer processing time when real time condition is minimum (x+y=0)	Minimum End to End delay with the higher jitter buffer processing time when real time condition is maximum (x+y=upper limit)
G.711	1	0	10	10+x1	y1	x1+y1<10 ms	2	2	20	400	34	44	414	424
	1	0	20	2*(10+x1)	2*y1	2*(x1+y1)<20 ms	2	2	30	400	54	74	424	444
	1	0	30	3*(10+x1)	3*y1	3*(x1+y1)<30 ms	2	2	40	400	74	104	434	464
	1	0	40	4*(10+x1)	4*y1	4*(x1+y1)<40 ms	2	2	50	400	94	134	444	484
	1	0	50	5*(10+x1)	5*y1	5*(x1+y1)<50 ms	2	2	60	400	114	164	454	504
	1	0	60	6*(10+x1)	6*y1	6*(x1+y1)<60 ms	2	2	70	400	134	194	464	524
G.729	10	5	10	(10+x2)+5	y2	x2+y2<10 ms	2	2	20	400	39	49	419	429
	10	5	20	(2*(10+x2))+5	2*y2	2*(x2+y2)<20 ms	2	2	30	400	59	79	429	449
	10	5	30	(3*(10+x2))+5	3*y2	3*(x2+y2)<30 ms	2	2	40	400	79	109	439	469
	10	5	40	(4*(10+x2))+5	4*y2	4*(x2+y2)<40 ms	2	2	50	400	99	139	449	489
	10	5	50	(5*(10+x2))+5	5*y2	5*(x2+y2)<50 ms	2	2	60	400	119	169	459	509
	10	5	60	(6*(10+x2))+5	6*y2	6*(x2+y2)<60 ms	2	2	70	400	139	199	469	529
G.723.1	30	7,5	30	(30+x3)+7,5	y3	x3+y3<30 ms	2	2	40	400	81,5	111,5	441,5	471,5
	30	7,5	60	(2*(30+x3))+7,5	2*y3	2*(x3+y3)<60 ms	2	2	70	400	141,5	201,5	471,5	531,5
NB-AMR	20	5	20	(20+x4)+5	y4	x4+y4<20 ms	2	2	30	400	59	79	429	449
	20	5	40	(2*(20+x4))+5	2*y4	2*(x4+y4)<40 ms	2	2	50	400	99	139	449	489
	20	5	60	(3*(20+x4))+5	3*y4	3*(x4+y4)<60 ms	2	2	70	400	139	199	469	529
G.722	10	1,5	10	(10+x5)+1,5	y5	x5+y5<10 ms	2	2	20	400	35,5	45,5	415,5	425,5
	10	1,5	20	(2*(10+x5))+1,5	2*y5	2*(x5+y5)<20 ms	2	2	30	400	55,5	75,5	425,5	445,5
	10	1,5	30	(3*(10+x5))+1,5	3*y5	3*(x5+y5)<30 ms	2	2	40	400	75,5	105,5	435,5	465,5
	10	1,5	40	(4*(10+x5))+1,5	4*y5	4*(x5+y5)<40 ms	2	2	50	400	95,5	135,5	445,5	485,5
	10	1,5	50	(5*(10+x5))+1,5	5*y5	5*(x5+y5)<50 ms	2	2	60	400	115,5	165,5	455,5	505,5
	10	1,5	60	(6*(10+x5))+1,5	6*y5	6*(x5+y5)<60 ms	2	2	70	400	135,5	195,5	465,5	525,5
WB-AMR	20	5	20	(20+x6)+5	y6+0,94	x6+y6<20 ms	2	2	30	400	59,94	79,94	429,94	449,94
	20	5	40	(2*(20+x6))+5	2*y6+0,94	2*(x6+y6)<40 ms	2	2	50	400	99,94	139,94	449,94	489,94
	20	5	60	(3*(20+x6))+5	3*y6+0,94	3*(x6+y6)<60 ms	2	2	70	400	139,94	199,94	469,94	529,94
G.729.1	20	25	20	(20+x7)+25+1,97	y7+1,97	x7+y7<20 ms	2	2	30	400	82,94	102,94	452,94	472,94
	20	25	40	(2*(20+x7))+25+1,97	2*y7+1,97	2*(x7+y7)<40 ms	2	2	50	400	122,94	162,94	472,94	512,94
	20	25	60	(3*(20+x7))+25+1,97	3*y7+1,97	3*(x7+y7)<60 ms	2	2	70	400	162,94	222,94	492,94	552,94

Annex B (informative): Bibliography

Recommendation ITU-T G.107: "The E-model, a computational model for use in transmission planning".

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

Recommendation ITU-T G.131: "Talker echo and its control".

History

Document history		
V1.2.1	October 2007	Publication
V1.3.1	September 2009	Publication
V1.3.2	September 2010	Publication
V1.4.1	January 2015	Membership Approval Procedure MV 20150316: 2015-01-15 to 2015-03-16