SLOVENSKI STANDARD
SIST ES 202 737 V1.8.1:2020

01-julij-2020

Kakovost prenosa govora in večpredstavnih vsebin (STQ) - Prenosne zahteve za ozkopasovne terminale VoIP (ročne in naglavne) glede na kakovost storitev (QoS), kot jo dojema uporabnik

Speech and multimedia Transmission Quality (STQ) - Transmission requirements for narrowband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user

Ta slovenski standard je istoveten z: ETSI ES 202 737 V1.8.1 (2020-05)

ICS:

33.050.01	Telekomunikacijska terminalska oprema na splošno	Telecommunication terminal equipment in general

SIST ES 202 737 V1.8.1:2020 en

ETSI ES 202 737 V1.8.1 (2020-05)

Speech and multimedia Transmission Quality (STQ); Transmission requirements for narrowband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user

Reference

RES/STQ-274

Keywords

narrowband, quality, speech, telephony, terminal, VoIP

ETSI

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

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

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

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

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

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Contents

Intellectual Property Rights

Essential patents

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

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

Trademarks

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

Foreword

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

Modal verbs terminology

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

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

Introduction

Traditionally, analogue and digital telephones were interfacing switched-circuit 64 kbit/s PCM networks. With the fast growth of IP networks, terminals 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, providing telephony service. Since the IP networks will be in many cases interworking with the traditional PSTN and private networks, many of the basic transmission requirements have to be harmonised with specifications for traditional digital terminals. However, due to the unique characteristics of the IP networks including packet loss, delay, etc. new performance specifications, as well as appropriate measurement methods, will have to be developed. Terminals are getting increasingly complex, advanced signal processing is used to address the IP specific issues. Also, the VoIP terminals may use other than 64 kbit/s PCM (Recommendation ITU-T G.711 [7]) speech coding algorithms.

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 narrowband VoIP handset and headset 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 3,4 kHz narrowband VoIP handset and headset terminals; it addresses all types of IP based terminals, including wireless and soft phones. DECT terminals are covered in ETSI EN 300 175-8 [i.6] and ETSI EN 300 176-2 [i.7].

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.

2 References

2.1 Normative references

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

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

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

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

[1] ETSI EN 300 726: "Digital cellular telecommunications system (Phase 2+) (GSM); Enhanced Full Rate (EFR) speech transcoding (GSM 06.60)".

[2] ETSI TS 126 171: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description (3GPP TS 26.171)".

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

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

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

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

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

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

[9] Recommendation ITU-T G.726: "40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)".

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

[11] 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".

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

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

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

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

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

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

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

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

[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 for hands-free and handset terminal testing".

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

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

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

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

[26] Recommendation ITU-T P.863: "Perceptual objective listening quality prediction".

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

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

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

[30] IETF RFC 5481: "Packet Delay Variation Applicability Statement".

2.2 Informative references

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

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

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

[i.1] ETSI EG 201 377-1: "Speech and multimedia Transmission Quality (STQ); Specification and measurement of speech transmission quality; Part 1: Introduction to objective comparison measurement methods for one-way speech quality across networks".

[i.2] ETSI EG 202 425: "Speech Processing, Transmission and Quality Aspects (STQ); Definition and implementation of VoIP reference point".

[i.3] 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.4] Netem™.

NOTE: Information available at https://wiki.linuxfoundation.org/networking/netem.

[i.5] IETF RFC 4737: "Packet Reordering Metrics".

[i.6] ETSI EN 300 175-8: "Digital Enhanced Cordless Telecommunications (DECT); Common Interface (CI); Part 8: Speech and audio coding and transmission".

[i.7] ETSI EN 300 176-2: "Digital Enhanced Cordless Telecommunications (DECT); Test specification; Part 2: Audio and speech".

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

3 Definition of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the following terms 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

Composite Source Signal (CSS): signal composed in time by various signal elements

diffuse field equalization: equalization of the HATS sound pick-up, equalization of the difference, in dB, between the spectrum level of the acoustic pressure at the ear Drum Reference Point (DRP) and the spectrum level of the acoustic pressure at the HATS Reference Point (HRP) in a diffuse sound field with the HATS absent using the reverse nominal curve given in table 3 of Recommendation ITU-T P.58 [14]

Ear Reference Point (ERP): virtual point for geometric reference located at the entrance to the listener's ear, traditionally used for calculating telephonometric loudness ratings

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

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.

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

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: when a receive volume control is provided, the setting which is closest to the nominal RLR of 2 dB

reordering: packet order changes during transfer over the network [i.5], packets arrive out of order at the receiver (i.e. RTP packets)

3.2 Symbols

Void.

3.3 Abbreviations

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

AGC	Automatic Gain Control
AM-FM	Amplitude Modulation-Frequency Modulation
AMR	Adaptative Multi-Rate
AMR-NB	Adaptive Multi-Rate NarrowBand
CS	Composite Source
CSS	Composite Source Signal
DRP	ear Drum Reference Point
DUT	Device Under Test
EC	Echo Canceller
ECRP	Earcap Reference Point
EFR	Enhanced Full Rate
EL	Echo Loss
ERP	Ear Reference Point
FFT	Fast Fourrier Transform
G-MOS-LQOn	Overall transmission quality narrowband
GSM	Global System for Mobile communications
HATS	Head And Torso Simulator
IEC	International Electrotechnical Commission
IP	Internet Protocol
IPDV	IP Packet Delay Variation
ITU-T	International Telecommunication Union - Telecommunication standardization sector
MOS	Mean Opinion Score
MOS-LQOy	Mean Opinion Score - Listening Quality Objective

NOTE: y being N for narrowband, M for mixed, S for super-wideband and F for fullband. See Recommendation ITU-T P.800.1 [23].

MRP	Mouth Reference Point
NLP	Non Linear Processor
N-MOS-LQOn	Transmission quality of the background noise narrowband
PBX	Private Branch eXchange
PC	Personal Computer
PCM	Pulse Code Modulation
PLC	Packet Loss Concealment
PN	Pseudo-random Noise
POI	Point Of Interconnect
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RLR	Receive Loudness Rating
RMS	Root Mean Square
RTP	Real-Time Transport Protocol
SLR	Send Loudness Rating
S-MOS-LQOn	Transmission quality of the speech narrowband
STD	STanDard position
STMR	SideTone Masking Rating
TCLw	Terminal Coupling Loss (weighted)
TDM	Time Division Multiplex
TOSQA	Telecommunication Objective Speech Quality Assessment
VAD	Voice Activity Detector

4 General considerations

4.1 Default coding algorithm

VoIP terminals shall support the coding algorithm according to Recommendation ITU-T G.711 [7] (both μ-law and A-law). VoIP terminals may support other coding algorithms.

NOTE: Associated Packet Loss Concealment (PLC) e.g. as defined in Recommendation ITU-T G.711 [7] appendix I 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 the general rules of transmission planning are carried out with the E-model of Recommendation ITU-T G.107 [3] taking into account that the E-model does not yet address headsets; this includes the a-priori determination of the desired category of speech transmission quality as defined in Recommendation ITU-T G.109 [5].

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 [4], 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 with its amendments [4] provides further guidance on this important issue.

The following optimum terminal parameters from a user's 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.

5 Test equipment

5.1 IP half channel measurement adaptor

The IP half channel measurement adaptor is described in ETSI EG 202 425 [i.2].

5.2 Environmental conditions for tests

The following conditions shall apply for the testing environment:

a) Ambient temperature: 15 °C to 35 °C;

b) Relative humidity: 5 % to 85 %;

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

5.3 Accuracy of measurements and test signal generation

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

Table 1: Measurement accuracy

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

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 14 000 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 a.c. the test shall be conducted within ±4 % of the rated frequency.

5.4 Network impairment simulation

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

An appropriate network simulator has to be used, for example Netem™ [i.4].

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

- Netem™ is part of the networking function of Linux™. With Netem™, there can be generated loss, duplication, dela, jitter and out of order packets (and the jitter distribution can be chosen during runtime). Netem™ can be run on a Linux™-PC running as a bridge or a router.
- 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.

NOTE: Netem™ and Linux™ are examples of suitable products available commercially. This information is given for the convenience of users of the present document and does not constitute an endorsement by ETSI of these products.

Requirements for the network impairment simulation can be found in annex D.

5.5 Acoustic environment

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

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

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

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

Standardized measurement methods for measurements with variable echo paths are for further study.

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 Requirements and associated measurement methodologies

6.1 Notes

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

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

6.2 Test setup

6.2.1 General

The preferred acoustical access to terminals is the most realistic simulation of the "average" subscriber. This can be made by using HATS (Head And Torso Simulator) with appropriate ear simulation and appropriate means to fix handset and headset terminals in a realistic and reproducible way to the HATS. HATS is described in Recommendation ITU-T P.58 [14], appropriate ears are described in Recommendation ITU-T P.57 [13] (type 3.3 and type 3.4 ear), a proper positioning of handsets under realistic conditions is to be found in Recommendation ITU-T P.64 [15].

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 bit rate is used for testing terminal electro acoustical parameters, the bit rate recognized giving the best characteristics should be selected, e.g.:

- AMR-NB (ETSI TS 126 171 [2]): 12,2 kbit/s.
- Recommendation ITU-T G.729.1 [11]: 32 kbit/s.

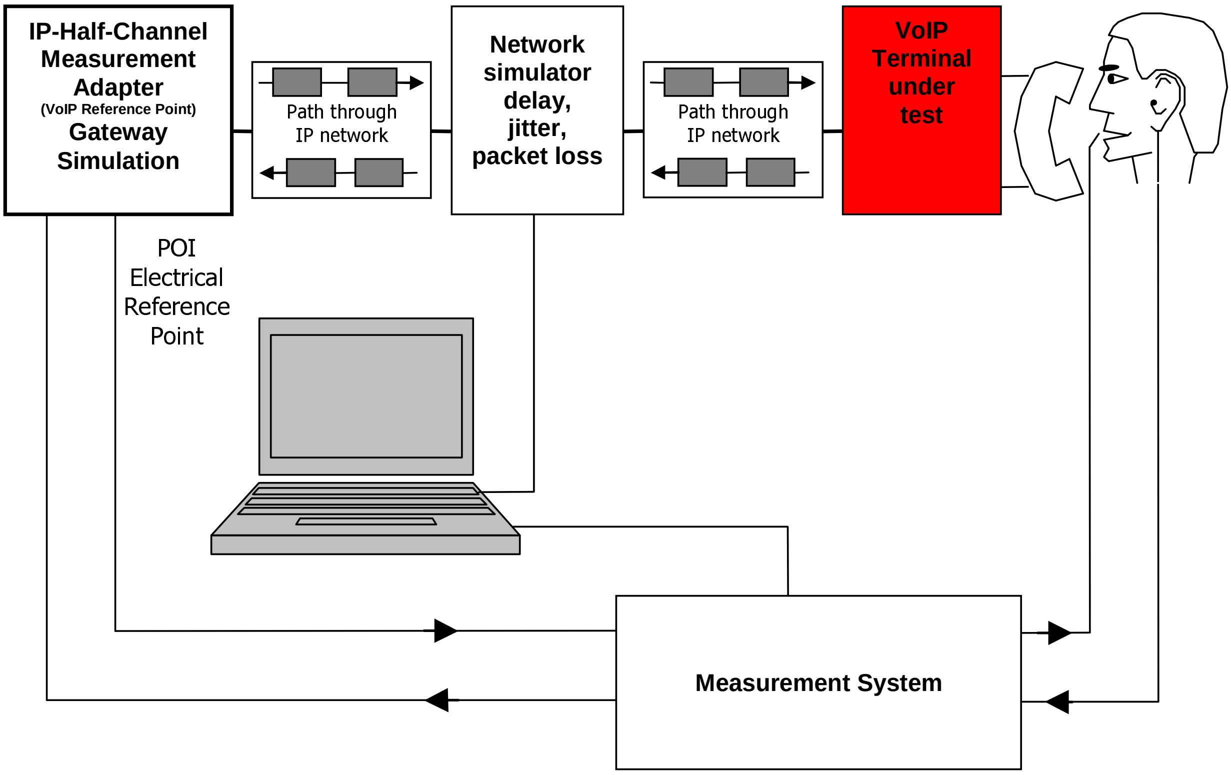

Figure 1: Half channel terminal measurement

6.2.2 Setup for handsets and headsets

When using a handset telephone the handset is placed in the HATS position as described in Recommendation ITU-T P.64 [15]. The artificial mouth shall be conforming to Recommendation ITU-T P.58 [14]. The artificial ear shall be conforming to Recommendation ITU-T P.57 [13], either type 3.3 or type 3.4 ears shall be used. In case of testing a flat handset (e.g. smartphone) with artificial ear of:

- Type 3.4, the *flat handset position* according to annex D.3 of Recommendation ITU-T P.64 [15] shall be used (A = 0°, B = 5° and C = 0°).
- Type 3.3, the *alternative handset position* according to annex E.2 of Recommendation ITU-T P.64 [15] shall be used with the definition A=0°, B=5° and C=0°.

This aligns measurements using artificial ears of types 3.3 and 3.4, where the flat handset position is explicitly specified (annex D.3 of Recommendation ITU-T P.64 [15]).

Recommendations for positioning headsets are given in Recommendation ITU-T P.380 [18]. If not stated otherwise headsets shall be placed in their recommended wearing position. Further information about setup and the use of HATS can be found in Recommendation ITU-T P.380 [18].

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

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

6.2.3 Position and calibration of HATS

All the send and receive characteristics shall be tested with the HATS, it shall be indicated what type of ear was used at what application force.

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