Final draft ETSI ES 202 738 V1.5.1 (2016-10)

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

Reference

RES/STQ-243

Keywords

handsfree, loudspeaking, narrowband, quality, speech, 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 only prevailing document is the print of the Portable Document Format (PDF) version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
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

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 (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.

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**", "**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, terminals directly interfacing packet-switched networks (VoIP) are being rapidly introduced. Such IP network edge devices may include specifically designed IP phones, soft phones or other devices connected to the IP based networks and 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 specification, as well as appropriate measuring 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 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 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 will provide speech transmission performance requirements for narrowband 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 referenced 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 I-ETS 300 245-3: "Integrated Services Digital Network (ISDN); Technical characteristics of telephony terminals; Part 3: Pulse Code Modulation (PCM) A-law, loudspeaking and handsfree telephony".

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

[3] 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)".

[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 O.41: "Psophometer for use on telephone-type circuits".

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

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

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

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

[17] Recommendation ITU-T P.310: "Transmission characteristics for narrow-band digital handset and headset telephones".

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

[19] Recommendation ITU-T P.342: "Transmission characteristics for narrow-band digital loudspeaking and hands-free telephony terminals".

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

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

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

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

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

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

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

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 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] NIST Net.

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

[i.4] Netem.

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

[i.5] 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".

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): corresponds to a reference point "n" from Recommendation ITU-T P.58 [15]: "n" is one of the points numbered from 11 to 17 and defined in table 6a of Recommendation ITU-T P.58 [15] (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
CS	Composite Source
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
HFT	HandsFree Terminal
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

NOTE: y being N for narrow-band, W for wideband, M for mixed and S for superwideband. See Recommendation ITU-T P.800.1 [24].

MRP	Mouth Reference Point
NIST Net	Network Simulation Tool from National Institute of Standards and Technology
NIST	National Institute of Standards and Technology
NLP	Non Linear Processor
PC	Personal Computer
PCM	Pulse Code Modulation
PDA	Personal Digital Assistance
POLQA	Perceptual objective listening quality assessment
PLC	Packet Loss Concealment
PN	PseudoNoise
POI	Point Of Interconnect
PSTN	Public Switched Telephone Network

QoS	Quality of Service
RLR max	Receive Loudness Rating corresponding to the maximum setting of the volume control
RLR min	Receive Loudness Rating corresponding to the minimum setting of the volume control
RLR	Receive Loudness Rating
RMS	Root Mean Square
SLR	Send Loudness Rating
TCLw	Terminal Coupling Loss (weighted)
TCN	Trace Control for Netem
VoIP	Voice over Internet Protocol
xDSL	any Digital Subscriber Line technology

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: Packet Loss Concealment as defined in e.g. appendix I of Recommendation ITU-T G.711 [7] 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 [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 Standard is acceptable or representative for the specific configuration is the responsibility of the individual transmission planner.

Recommendation ITU-T G.108 [4] 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 300 245-3 [1]) parameters are applicable, if Recommendation ITU-T G.711 [7] is used.

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 N
Measured maximum frequency	20 kHz
NOTE: The measured maximum frequency is due to P.58 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 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 loss of performance of the packet network.

An appropriate network simulator has to be used, for example NIST Net [i.3] or Netem [i.4].

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

Here is a brief blurb about NIST Net:

- The NIST Net network emulator is a general-purpose tool for emulating performance dynamics in IP networks. The tool is designed to allow controlled, reproducible experiments with network performance sensitive/adaptive applications and control protocols in a simple laboratory setting. By operating at the IP level, NIST Net can emulate the critical end-to-end performance characteristics imposed by various wide area network situations (e.g. congestion loss) or by various underlying subnetwork technologies (e.g. asymmetric bandwidth situations of xDSL and cable modems).
- NIST Net is implemented as a kernel module extension to the Linux™ operating system and an X Window System-based user interface application. In use, the tool allows an inexpensive PC-based router to emulate numerous complex performance scenarios, including: tunable packet delay distributions, congestion and background loss, bandwidth limitation, and packet reordering/duplication. The X interface allows the user to select and monitor specific traffic streams passing through the router and to apply selected performance "effects" to the IP packets of the stream. In addition to the interactive interface, NIST Net can be driven by traces produced from measurements of actual network conditions. NIST Net also provides support for user defined packet handlers to be added to the system. Examples of the use of such packet handlers include: time stamping/data collection, interception and diversion of selected flows, generation of protocol responses from emulated clients.

The key points of Netem can be summarized as follows:

- Netem is nowadays part of most Linux™ distributions, it only has to be switched on, when compiling a kernel. With Netem, there are the same possibilities as with 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, Trace Control for Netem (TCN) 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 the exact same distortions with Netem.

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

5.5 Acoustic environment

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, the test laboratory, in the case where its test room does not conform to anechoic conditions as given in Recommendation ITU-T P.342 [19], has to present difference in results for measurements due to its test room.

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 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.

6.2 Test setup

6.2.1 General

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 [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 bite rate is used, it should be adopted, for testing terminal electroacoustical parameters, the bit rate recognized as giving the best characteristics is selected, e.g.:

- ETSI TS 126 171 [3]: 12,2 kbit/s.