ETSI ES 201 377-2 V1.4.1 (2009-12)

ETSI Standard

Speech and multimedia Transmission Quality (STQ); Specification and measurement of speech transmission quality; Part 2: Mouth-to-ear speech transmission quality including terminals

Reference

RES/STQ-00105-2

Keywords

network, QoS, quality, speech, terminal, testing, transmission

ETSI

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

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

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

Important notice

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

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

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

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

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://webapp.etsi.org/IPR/home.asp).

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

The present document provides technical requirements for assessing the conversational speech quality performance parameters from mouth-to-ear independent of the technology used.

The present document is part 2 of a multi-part deliverable covering the specification and measurement of speech transmission quality, as identified below:

EG 202 377-1: "Introduction to objective comparison measurement methods for one-way speech quality across networks";

ES 202 377-2: "Mouth-to-ear speech transmission quality including terminals";

EG 202 377-3: "Non-intrusive objective measurement methods applicable to networks and links with classes of services".

Introduction

Various standards within ETSI, ITU, TIA and other standardization organizations describe performance requirements for different types of terminals, networks and network components. In each standard emphasis is given typically only to a part of the overall connection. The speech quality perceived by the user however is influenced by any component in the overall connection. In modern complex network and end-to-end (mouth-to-ear) configurations there is no guarantee for a sufficient overall performance if only the individual components conform to their relevant standards. Furthermore many of the existing testing specifications still assume a linear and time invariant behaviour of the components which due to complex signal processing in most of the modern communication devices can no longer be expected. Only a few standards exist which describe test procedures and requirements for the interaction of different network components with the different types of terminals.

The present document addresses the mouth-to-ear speech quality taking into account all conversational aspects. An overview about different network/terminal configurations and their specific impact on speech quality is given. The present document describes testing procedures and setups for different configurations.

1 Scope

The present document addresses mouth-to-ear (i.e. end-to-end speech quality for 3,1 kHz telephony). It both:

a) summarizes and gives guidance about the main factors that affect speech quality in end-to-end scenarios; and

b) specifies test methods for end-to-end speech quality testing.

The test methods can be used both for the complete transmission from mouth-to-ear and also for testing individual sections of a connection.

The end-end (mouth-to-ear) test methods specified in the present document are independent of the technology used in the network and the terminals. However when practical considerations make it necessary to test at electrical interfaces within or between equipments the present document explains how to handle the most common current technologies.

The present document is designed to be used by:

- terminal and terminal component (e.g. soundcard) developers who wish to evaluate the end-to-end performance of networks and their terminals (or components); or
- network designers who wish to evaluate the end-to-end performance of their networks with typical terminals.

And therefore it gives advice on how networks and representative terminals (respectively) can be selected or simulated for use in the end-to-end tests.

The test methods described allow the evaluation of all conversational situations such as single talk and double talk by means of objective procedures.

The present document takes account of:

a) all types of terminals, including handsets, headsets and dedicated hands-free arrangements such as are provided with some mobile terminals and PC based terminals;

b) both circuit switched and packet based networks, including IP and ATM.

The present document is not generally suitable for wideband telephony or other forms of wideband communication although the parametric approach and the measurement procedures for some of the parameters described in the present document are applicable for wideband communication as well.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific.

- For a specific reference, subsequent revisions do not apply.
- Non-specific reference may be made only to a complete document or a part thereof and only in the following cases:
  - if it is accepted that it will be possible to use all future changes of the referenced document for the purposes of the referring document;
  - for informative references.

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

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

2.1 Normative references

The following referenced documents are indispensable for the application of the present document. For dated references, only the edition cited applies. For non-specific references, the latest edition of the referenced document (including any amendments) applies.

[1] ITU-T Recommendation G.821: "Error performance of an international digital connection operating at a bit rate below the primary rate and forming part of an Integrated Services Digital Network".

[2] Gierlich, H.W.; Kettler, F., Diedrich, E.: "Speech Quality Evaluation of Hands-Free Telephones During Double Talk: New Evaluation Methodologies"; EUSIPCO 1998, Proceedings, Vol. II.

[3] Gierlich, H.W. (December 1996): "The auditory perceived quality of hands-free telephones: Auditory judgements, instrumental measurements and their relationship", Speech Communication 20, pp. 241-254.

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

[5] IEC 61672 (all parts): "Electroacoustics - Sound level meters".

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

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

[8] ITU-T Recommendation G.111: "Loudness ratings (LRs) in an international connection".

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

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

[11] ITU-T Recommendation G.168: "Digital network echo cancellers".

[12] ITU-T Recommendation G.712: "Transmission performance characteristics of pulse code modulation channels".

[13] ITU-T Recommendation O.131: "Quantizing distortion measuring equipment using a pseudo-random noise test signal".

[14] ITU-T Recommendation O.132: "Quantizing distortion measuring equipment using a sinusoidal test signal".

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

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

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

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

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

[20] ITU-T Recommendation P.51: "Artificial mouth".

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

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

[23] ITU-T Recommendation P.581: "Use of Head and Torso Simulator (HATS) for hands-free terminal testing".

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

[25] ITU-T Recommendation P.79 and Corrigendum 2 (2001): "Calculation of loudness ratings for telephone sets".

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

[27] ITU-T Recommendation P.810: "Modulated noise reference unit (MNRU)".

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

[29] ITU-T Recommendation P.831: "Subjective performance evaluation of network echo cancellers".

[30] ITU-T Recommendation P.832: "Subjective performance evaluation of hands-free terminals".

[31] ITU-T Recommendation 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".

[32] ITU-T Recommendation Y.1541: "Network performance objectives for IP-based services".

[33] ITU-T COM12-42 (Federal Republic of Germany, January 1998): "Listening only test results for hands-free telephones and their dependence upon room surroundings".

[34] TIA/EIA 810-A: "Telecommunications - Telephone Terminal Equipment-Transmission Requirements for Narrowband".

[35] ITU-T Recommendation P.59: "Artificial conversational speech".

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

[37] ETSI TS 100 961: "Digital cellular telecommunications system (Phase 2+) (GSM); Full rate speech; Transcoding (GSM 06.10 Release 1998)".

[38] ETSI EN 300 969: "Digital cellular telecommunications system (Phase 2+) (GSM); Half rate speech; Half rate speech transcoding (GSM 06.20 version 8.0.1 Release 1999)".

[39] ETSI EN 300 726: "Digital cellular telecommunications system (Phase 2+) (GSM); Enhanced Full Rate (EFR) speech transcoding (GSM 06.60 version 8.0.1 Release 1999)".

[40] ETSI EN 300 903: "Digital cellular telecommunications system (Phase 2+) (GSM); Transmission planning aspects of the speech service in the GSM Public Land Mobile Network (PLMN) system (GSM 03.50 version 8.1.1 Release 1999)".

[41] ISO 9614 (all parts): "Acoustics - Determination of sound power levels of noise sources using sound intensity".

[42] Inter-Noise'96: "Evaluation of Acoustic-Quality Based on a Relative Approach", K. Genuit: 25th Anniversary Congress Liverpool, 30.07-02.08.1996, Conference Proceedings (Book 6 / ISBN: 1 873082 90 8), pp. 3233-3238, Liverpool, England.

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

[44] 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".

[45] ETSI ES 202 738: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for narrowband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user".

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

2.2 Informative references

The following referenced documents are not essential to the use of the present document but they assist the user with regard to a particular subject area. For non-specific references, the latest version of the referenced document (including any amendments) applies.

[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 TR 101 110: "Digital cellular telecommunications system (Phase 2+) (GSM); Characterisation, test methods and quality assessment for handsfree Mobile Stations (MSs) (GSM 03.58)".

[i.3] ETSI EG 201 050: "Speech Processing, Transmission and Quality Aspects (STQ); Overall Transmission Plan Aspects for Telephony in a Private Network".

[i.4] ETSI TBR 008: "Integrated Services Digital Network (ISDN); Telephony 3,1 kHz teleservice; Attachment requirements for handset terminals".

[i.5] ETSI TR 102 251: "Speech Processing, Transmission and Quality Aspects (STQ); Anonymous Test Report from 2nd Speech Quality Test Event 2002".

[i.6] ETSI EG 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".

[i.7] ETSI EG 202 396-3: "Speech Processing, Transmission and Quality Aspects (STQ); Speech Quality performance in the presence of background noise Part 3: Background noise transmission - Objective test methods".

3 Definitions and abbreviations

3.1 Definitions

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

Acoustic Reference Level (ARL): acoustic level at MRP which results in a -10 dBm0 output at the digital interface

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

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 (see also ITU-T Recommendation P.58 [22]) using the reverse nominal curve given in table 3 of ITU-T Recommendation P.58 [22]

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

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

electric power and noise levels: the following electric power and noise level units are used in the present document:

dBm0: The absolute power level at a digital reference point of the same signal that would be measured as the absolute power level, in dBm, if the reference point was analogue. The absolute power in dBm is defined as 10 log (power in mW/1 mW). When the impedance is 600 ohm resistive, dBm can be referred to a voltage of 0,775 volts, which results in a reference active power of 1 mW. Note that 0 dBm0 is not the maximum digital code. For the L16-256 wideband codec adopted by TIA TR-41, 0 dBm0 is 3,14 dB below digital full scale.

end-to-end: endpoints of a (telephone) connection between two subscribers, either between the NTPs (e.g. for bearer services), or for speech communication between mouth and ear

G-MOS-LQOw: measure of the overall transmission quality in the presence of background noise (objective, wideband)

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

NOTE: HATS conforms to ITU-T Recommendation P.58 [22].

HATS position: correct handset position for measuring sensitivity and frequency response characteristics

NOTE: The HATS position has been shown to be essentially identical to the LRGP (loudness rating guard-ring position) position, except for the mouth simulator direction, which has been corrected with a 19 degrees downwards rotation to more closely match real talkers. For handsets with omnidirectional microphones, measurements on the two heads may differ slightly, typically less than 1 dB. For handsets with directional or noise-cancelling microphones, the differences will be larger, and the HATS position will give the more realistic results. See ITU-T Recommendation P.64 [24] (annexes D and E) and EUSIPCO 1998, Proceedings, Vol. II. [2].

Hands-Free Reference Point (HFRP): point located on the axis of the artificial mouth, at 50 cm from the outer plane of the lip ring, where the level calibration is made under free-field conditions

HATS Hands-Free Reference Point (HATS-HFRP): corresponds to a reference point "n" from ITU-T Recommendation P.58 [22]: "n" shall be one of the points numbered from 11 to 17 and defined in table 6a/P.58 (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 as close as possible to the axis lip-ring/HFT microphone under test. (see ITU-T Recommendation P.581 [23]).

mouth-to-ear: endpoints of a telephone connection between two subscribers between mouth and ear

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

N-MOS-LQOw: measure of the noise transmission quality in the presence of speech with background noise (objective, wideband)

pinna simulator: device which has the approximate shape and dimensions of a median adult human pinna

reference volume control setting: receive volume control setting which results in the Receive Loudness Rating (RLR) closest to the target value (centre of the RLR tolerance range)

NOTE: There may be separate settings for handset, headset and hands-free modes.

S-MOS-LQOw: measure of the speech transmission quality in the presence of background noise (objective, wideband)

sound pressure levels: value expressed as a ratio of the pressure of a sound to a reference pressure

NOTE 1: The following sound level units are used in the present document:

dBPa: The sound pressure level, in decibels, of a sound is 20 times the logarithm to the base 10 of the ratio of the pressure of this sound to the reference pressure of 1 Pascal (Pa).

NOTE 2: 1 Pa = 1 N/m^2.

dBSPL: The sound pressure level, in decibels, of a sound is 20 times the logarithm to the base 10 of the ratio of the pressure of this sound to the reference pressure of 2×10^{-5} N/m^2 (0 dBPa corresponds to 94 dBSPL).

dB(A): The A-weighted sound level is the sound pressure level e.g. in dBSPL, weighted by use of metering characteristics and A-weighting specified in IEC 61672 [5].

3.2 Abbreviations

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

ARL	Acoustic Reference Level
BER	Bit Error Rate
BSC	Base Station Controller
BTS	Base Transceiver Station
C/A	adjacent channel interference
C/I	Carrier to Interference ratio
C/N	Carrier/Noise
CSS	Composite Source Signals
D	D-value of terminal
dBPa	decibel relative to one Pascal
dBSPL	decibel Sound Pressure Level
DCME	Digital Circuit Multiplication Equipment
DRP	ear Drum Reference Point
DTX	Discontinuous Transmission
EL	Echo Loss
ERL	Echo Return Loss
ERP	Ear Reference Point
FER	Frame Erasure Rate
GERAN	GSM/EDGE Radio Access Network
G-MOS-LQOn	Global mean opinion score (listening quality,objective, narrowband)
G-MOS-LQOw	Global mean opinion score (listening quality,objective, wideband)
HATS	Head And Torso Simulator
HFRP	Hands-Free Reference Point
HFT	Hands-Free Terminal
LRGP	Loudness Rating Guard-ring Position
LSTR	Listener SideTone Rating
LTI	Linear Time Invariant
MRP	Mouth Reference Point
MSC	Mobile service Switching Centre
Nc	circuit Noise referred to the 0 dBr-point
NLP	Non-Linear Processor
N-MOS-LQOn	Noise mean opinion score (listening quality,objective, narrowband)
N-MOS-LQOw	Noise mean opinion score (listening quality,objective, wideband))
OLR	Overall Loudness Rating
PCM	Pulse Code Modulation
PESQ	Perceptional Evaluation of Speech Quality
PLC	Packet Loss Concealment
PLMN	Public Land Mobile Network
PSTN	Public Switched Telephone Network
qdu	number of quantizing distortion units
RCV	Residual Capital Value
RLR	Receiving Loudness Rating
SLR	Sending Loudness Rating
S-MOS-LQOn	Speech mean opinion score (listening quality,objective, narrowband)
S-MOS-LQOw	Speech mean opinion score (listening quality,objective, wideband)
SND	Signal + Noise + Distortion
STMR	SideTone Masking Rating
TCL	Terminal Coupling Loss

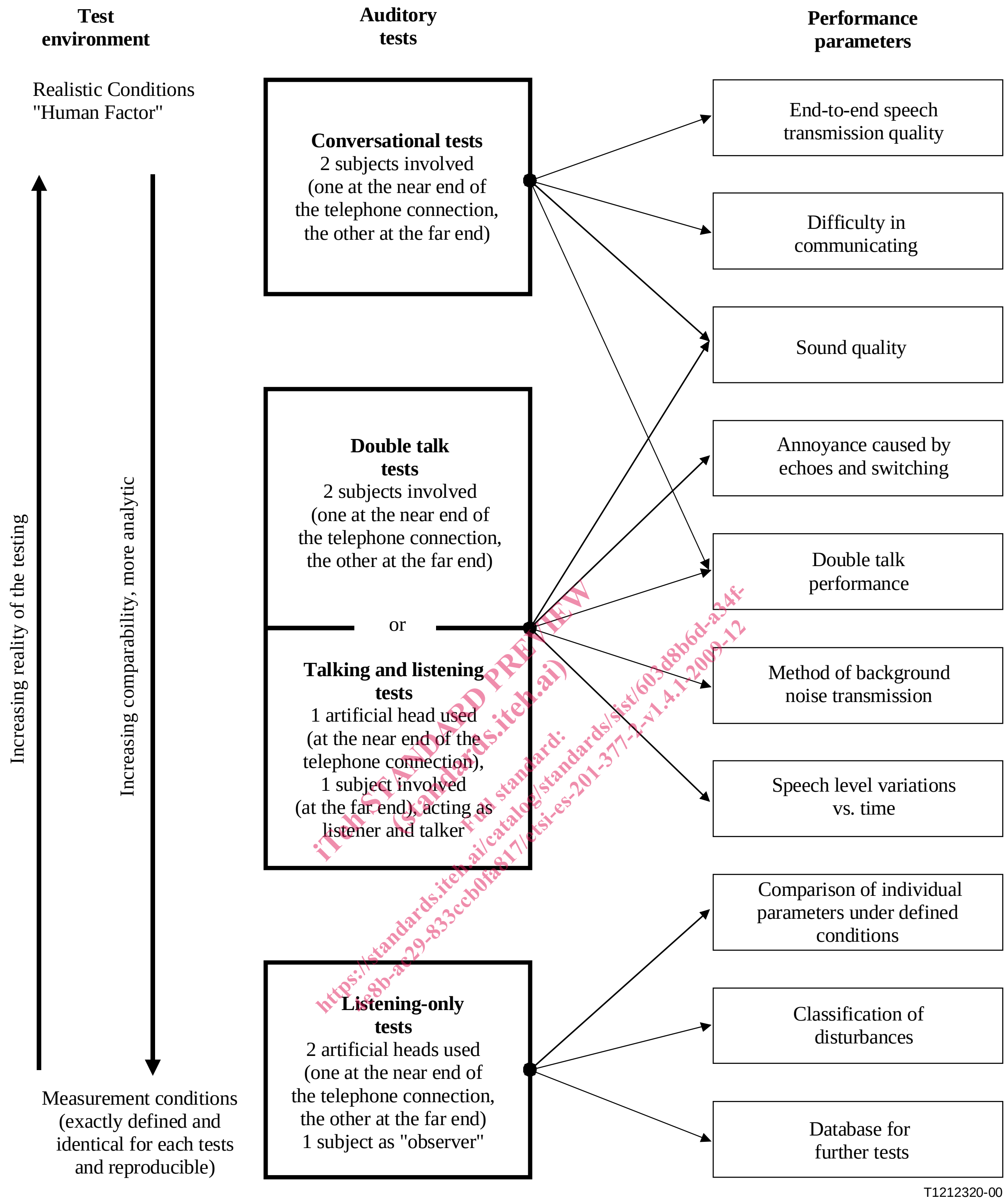
TCLw	Terminal Coupling Loss (weighted)
TELR	Talker Echo Loudness Rating
TOSQA	Telecommunication Objective Speech Quality Assessment
TRC	TRanscoder Controller
UTRAN	UMTS Terrestrial Radio Access Network
WEPL	Weighted Echo Path Loss

4 General considerations for end-to-end speech quality evaluations

When evaluation the overall speech transmission quality, networks and terminals may influence quite significantly the speech quality of a connection: Coding, delay and processing techniques like speech echo cancellers packetizing or DCME are mainly introduced by the network(s) but similar signal processing can be found in terminals as well. The transfer functions and loudness ratings of a connection are mainly determined by the terminals, the background noise and the background noise transmission are highly influenced by the terminal and the acoustical environment the terminal is exposed to. The conversational properties which are the most important ones in a conversation are determined by the terminal in combination with the network: double talk capability, switching characteristics, echo and delay are dominant impairments often introduced.

In order to find the determining factors a set of subjective test procedures have been developed allowing to extract the dominant quality aspects: Conversational test, talking and listening tests, double talk tests and listening only tests (as described in Speech Communication 20 (pp. 241 to 254) [3] and ITU-T Recommendations P.800 [26], P.810 [27], P.830 [28], P.831 [29] and P.832 [30]) are the basis of the parameter extraction procedure.

An overview of the methodologies is given in figures 1a to 1c.



NOTE: The assignment of "near end" and "far end" is chosen according to the E-model (ITU-T Recommendation G.107 [7]).

Figure 1a: Overview of test methods used for subjective evaluation - direct parameter access