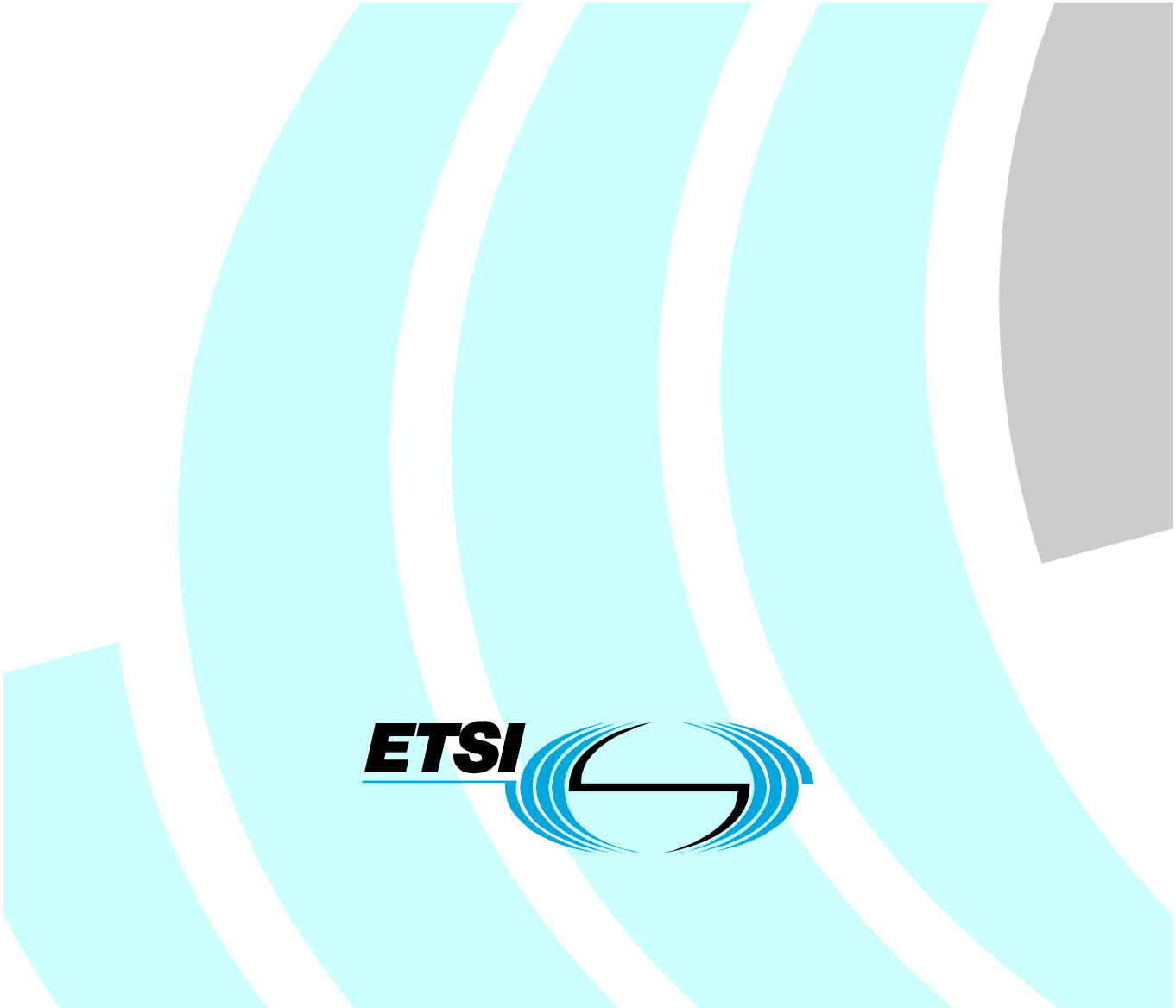
Final draft ETSI ES 201 377-2 V1.4.1 (2009-09)

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

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), and is now submitted for the ETSI standards Membership Approval Procedure.

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

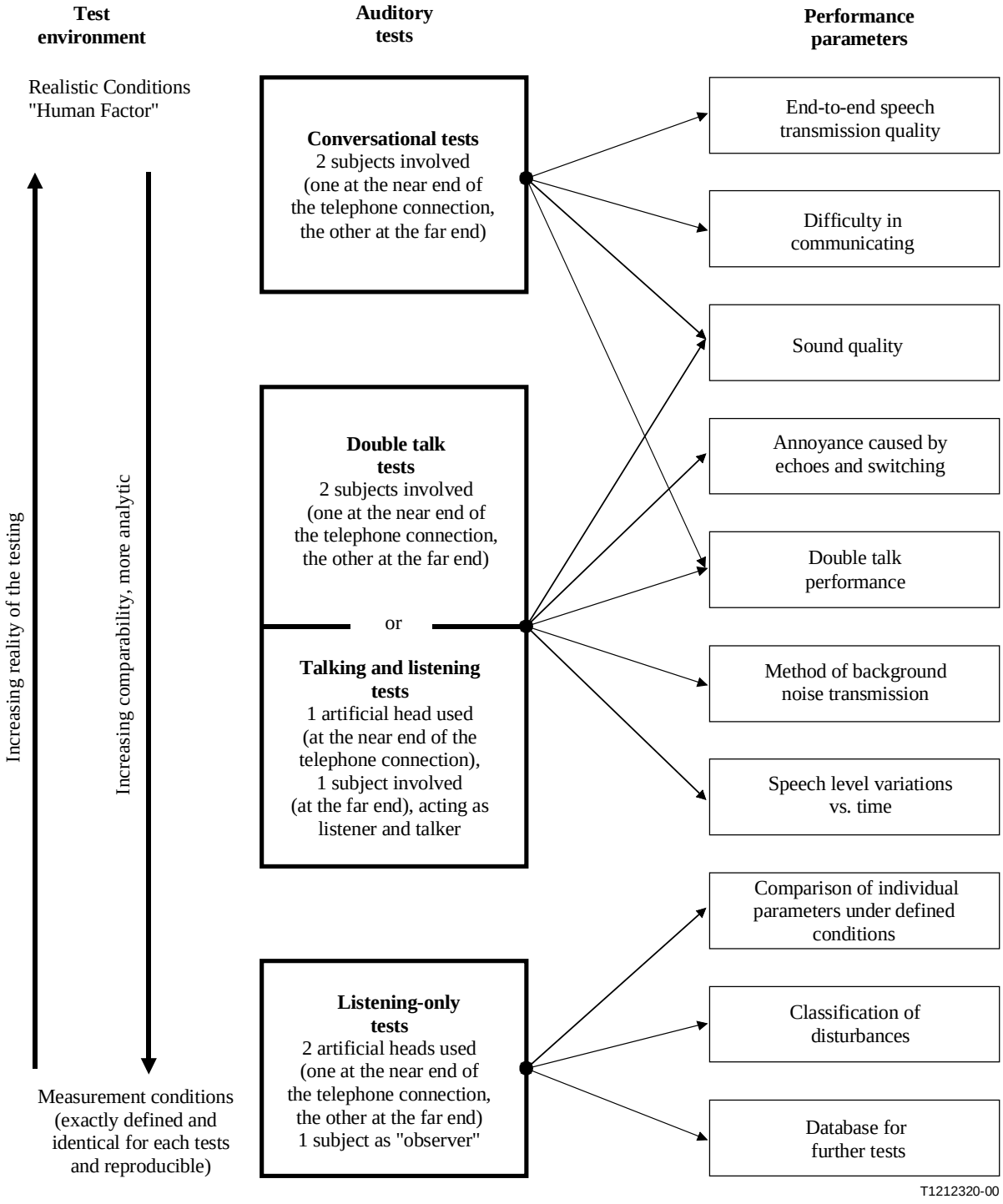
TCLw	Terminal Coupling Loss (weighted)
TELR	Talker Echo Loudness Rating
TOSQA	Telecommunication Objective Speech Quality Assessment
TRC	TRanscoder Controller
UTRAN	UMTS Terrestrial Radio Access Network
WEPL	Weighted Echo Path Loss

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.

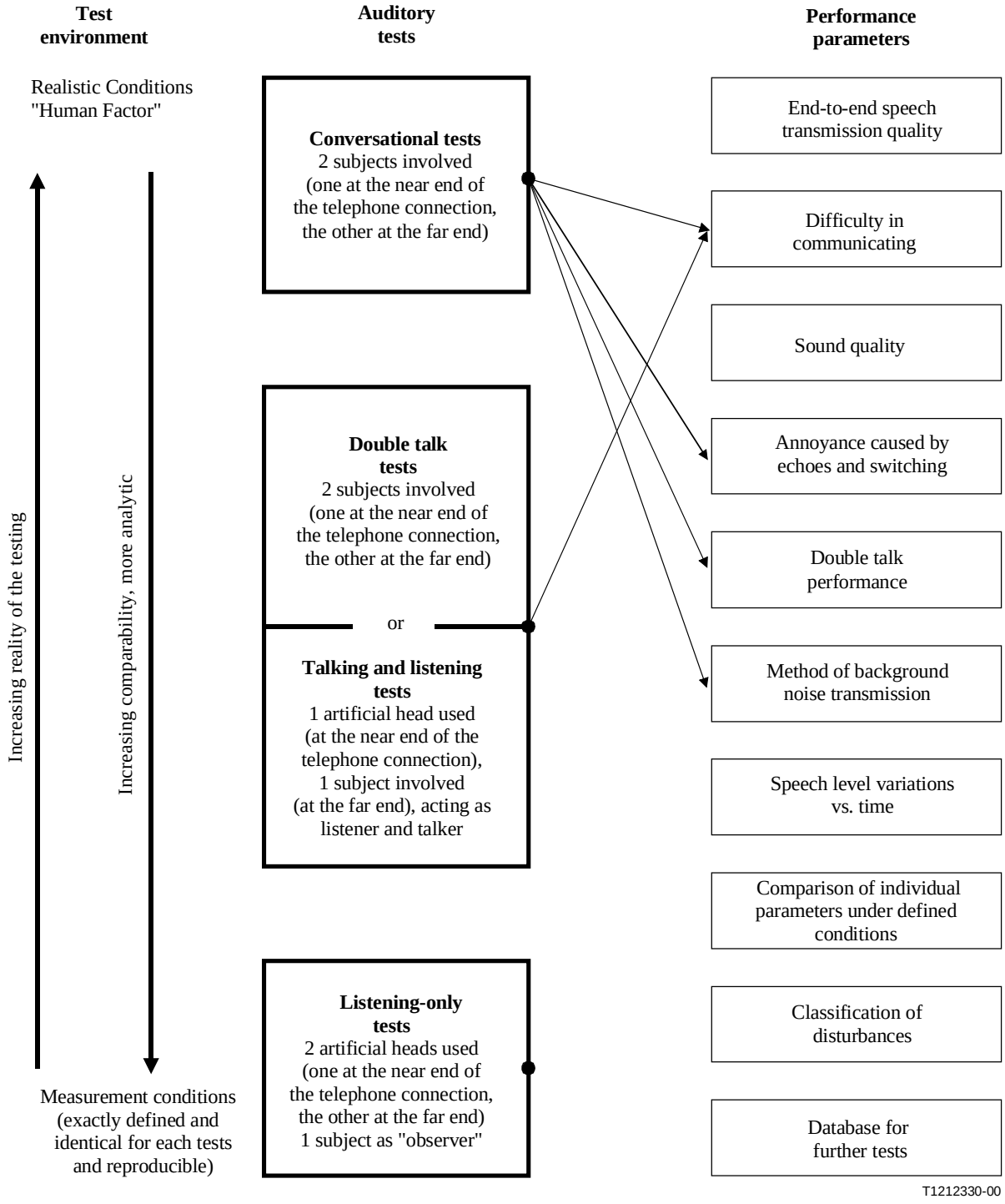
In order to find the determining factors a set of subjective test procedures have been developed allowing to extract the dominant quality aspects: Conversational test, talking and listening tests, double talk tests and listening only tests (as described in Speech Communication 20 (pp. 241 to 254) [3] 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.



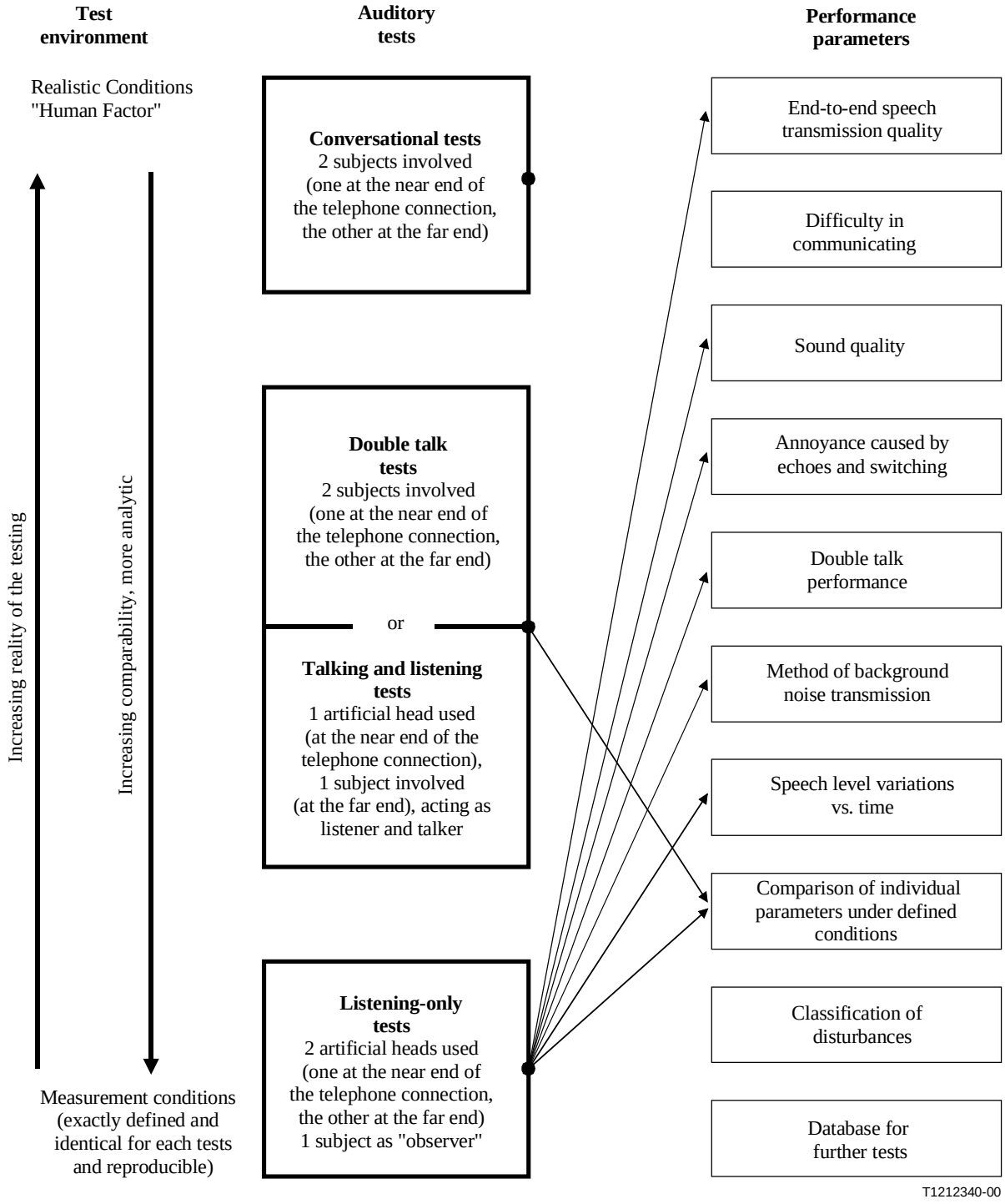
NOTE: The assignment of "near end" and "far end" is chosen according to the E-model (ITU-T Recommendation G.107 [7]).

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



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

Figure 1b: Overview about test methods used for subjective evaluation-parameter access via interviews



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

Figure 1c: Overview about test methods used for subjective evaluation - parameter access by including reference conditions

The subjectively relevant parameters determining the "speech transmission quality" are as follows.

The overall quality is determined by:

- Delay and echo.
- Sound quality.
- Quality of background noise transmission at idle, in single talk and double talk conditions.

- Speech level variations during single talk and double talk.
- Disturbances caused by switching during single talk and double talk (completeness of speech transmission).
- Disturbances caused by echoes during single talk and double talk.

Consequently the evaluation methods need to be divided into single talk measurements and double talk evaluations. In addition evaluations are required during periods of silence where only background noise is present.

Since the typical test setup should include all components involved in the mouth-to-ear transmission a test arrangement should include the terminals "attached" to a realistic substitution of a user and his typical environment. Figure 2 illustrates how a test setup from end-to-end may look like typically.

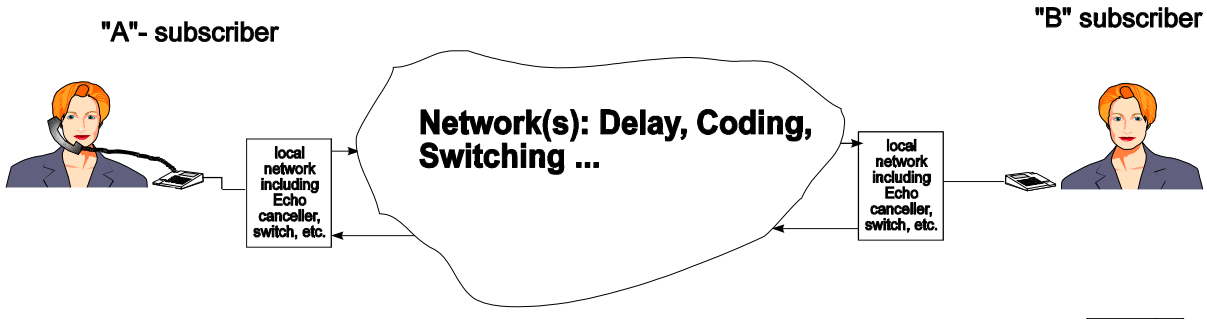


Figure 2: Typical test setup for determining the speech transmission quality from end-to-end (mouth-to-ear) by subjective evaluation of the speech quality relevant parameters (example for handset/hands-free communication)

Test setups as shown in figure 2 are used in auditory (subjective) tests to determine the quality aspects subjectively (see ITU-T Recommendations P.800 [26], P.810 [27], P. 830 [28], P.831 [29] and P.832 [30]). From the evaluations in ITU-T Recommendation P.800 [26], procedures have been derived which allow the objective testing of the relevant parameters of terminals (or even end-to-end scenarios).

5 Test configurations

This clause describes the test setups for terminals, networks and their various combinations. Since the present document describes the general aspects of end-to-end speech quality testing, specific test setups and configuration description are made only in general. In case any specific description of terminal or network setups is needed (e.g. buffer sizes, type of codecs, packet loss simulations) these descriptions need to be found in the relevant standards of such transmission systems.

5.1 Test setup for terminals

The general access to terminals is described in figure 3. The traditional way to test handset-terminals is the LRGP-position using Type 1 artificial ear and the artificial mouth according to ITU-T Recommendation P.51 [20]. positioned in LRGP (loudness rating guard ring position, ITU-T Recommendation P.64 [24]).The preferred acoustical access to terminal 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, headset or hands-free terminals in a realistic by reproducible way to the HATS. HATS is described in ITU-T Recommendation P.58 [22], appropriate ears are described in ITU-T Recommendation P.57 [21] (Type 3.3 and Type 3.4 ear), a proper positioning of handsets in realistic conditions is found in ITU-T Recommendation P.64 [24], the test setups for various types of hands-free terminals can be found in ITU-T Recommendation P.581 [23].

The preferred way of testing a terminal is either to connect it to a network simulator with exact defined settings and access points or, in case of end-to-end scenarios, to connect the terminal to the "typical" network it is used in. 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.

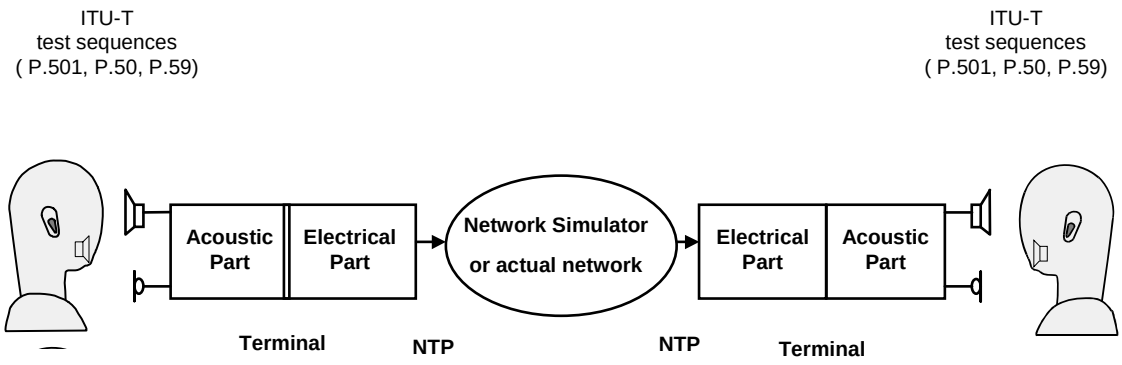


Figure 3: Test setup for terminals, acoustical access in end-to-end scenarios including a network or using a network simulator

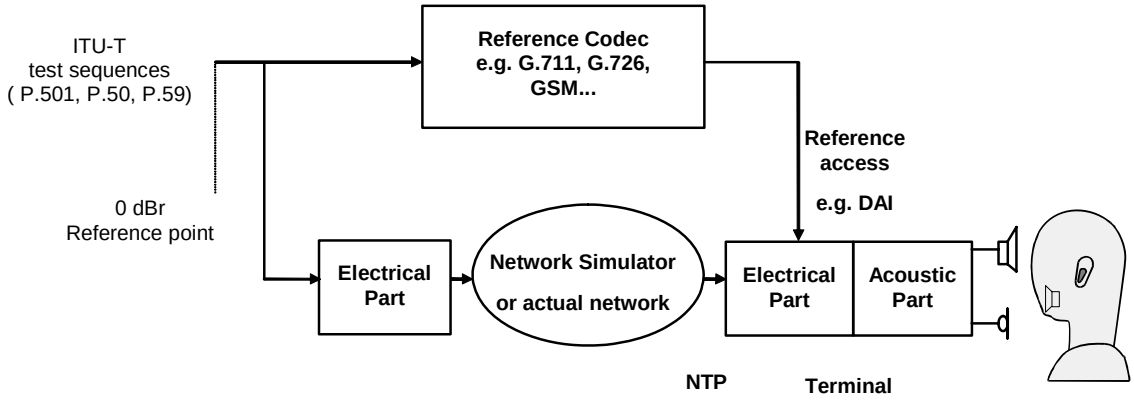


Figure 4: Test setup for terminals, electrical access using a "reference" access or a network simulator

NOTE: Instead of using HATS as shown in figures 3 and 4 the test may be conducted using the LRGP positioning for handsets. However it should be noted that this will lead to more simplified acoustical conditions and such may reduce the validity of the measurement, especially in noisy environments and for terminals with unknown acoustical properties.

5.1.1 Setup for handset terminals

When using a handset telephone the handset is placed in the HATS position as described in ITU-T Recommendation P.64 [24]. The artificial mouth shall conform with ITU-T Recommendation P.58 [22] when HATS is used. The artificial ear shall conform with ITU-T Recommendation P.57 [21], Type 3.3 or Type 3.4 ears shall be used. For (traditional) standard type handset terminals (the earpeace of which can be naturally sealed to the circular rim of a Type 1 or Type 3.2 coupler) LRGP positioning according to ITU-T Recommendation P.64 [24] may be used as well. The artificial mouth shall conform with ITU-T Recommendation P.51 [20]. The artificial ear shall conform with ITU-T Recommendation P.57 [21], Type 1 or Type 3.2 ears shall be used.

5.1.2 Setup for headset terminals

When using a headset the headset is placed on a HATS conforming to ITU-T Recommendation P.58 [22]. The artificial mouth shall conform with ITU-T Recommendation P.58 [22] when HATS is used. The artificial ear shall conform with ITU-T Recommendation P.57 [21], Type 3.3 or Type 3.4 ears shall be used.

The headset shall be placed in its recommended wearing position. Further information about setup and the use of HATS can be found in ITU-T Recommendation P.380 [16].

5.1.3 Setup for hands-free type terminals and loudspeaking terminals

General definition of hands-free terminals from ITU-T Recommendation P.340 [15] All types of terminals, which cannot be fit to the LRGP-position or the HATS-position -except headsets- need to be considered as hands-free type terminals. ITU-T Recommendation P.581 [23] describes the setup for hands-free terminals which are not covered by the setup description in ITU-T Recommendation P.340 [15].

5.1.4 Position and calibration of HATS

All the sending and receiving characteristics shall be tested with the HATS and it shall be indicated in the test report that HATS is used, it shall be indicated what type of ear was used at what application force. In case of hands-free measurements the HATSHFRP(s) shall be used for the calibration(s), the reference point chosen for the HATSHFRP shall be indicated.

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

The HATS shall be equipped with two Type 3.3 or 3.4 artificial ears. For hands-free measurements the HATS shall always be equipped with two artificial pinnas. The pinnas are specified in ITU-T Recommendation P.57 [21] for Types 3.3 and 3.4 artificial ears. The pinna shall be positioned on HATS according to ITU-T Recommendation P.58 [22].

The exact calibration and equalization procedures as well as the combination of the two ear signals for the purpose of measurements can be found in ITU-T Recommendation P.581 [23]. This Recommendation also describes the positioning for the various types of hands-free terminals.

5.2 Setup of the electrical interfaces

If electrical interfaces to terminals shall be simulated proper interface simulation is needed. The appropriate information typically can be found in the relevant terminal standards for all standardized interfaces.

For all interfaces the following points should be taken into account:

- General points:
  - Any nonlinearity or time variance measured should derive from the EUT and not from the test equipment or the test interface used.
  - The dynamic range of the test equipment has to exceed the dynamic range of the EUT by at least 10 dB, i.e. the noise floor of the test equipment has to be at least 10 dB below the noise floor of the EUT or at least 10 dB below the specified requirements.
  - All nonlinear distortions introduced by the test equipment must not influence any measurement result (e.g. distortion measurements).
  - Any delay introduced by the test equipment has to be taken into account for delay and echo measurements.
- Analogue interfaces:
  - Typically the impedance of the measuring device has to match the impedance of the interface of the EUT. Alternatively low impedance outputs are terminated by high impedance inputs.
  - The dynamic range of the EUT has to be matched by the test equipment, i.e. the reference levels for the electric or the acoustical interface have to be defined and matched.
- Digital interfaces:
  - The electrical specifications of the interface of the EUT (clock, frame, data, level, impedance, etc.) have to be matched by the test equipment.
  - Care must be take to avoid jitter problems e.g. by improper configuration of master/slave.

- It is advisable to use a common system clock for the measurement of complex configurations.
- Any speech codec used must fully meet the codec specification of the speech codec used in the EUT.

5.3 Test signals

Due to the extensive coding of the speech signals, standard test signals are not applicable for the tests, appropriate test signals (general description) are defined in ITU-T Recommendations P.50 [17] and P.501 [18]. More information can be found in the test procedures described below. Care should be taken, that the test signal used offers sufficient wideband signal energy in case a wideband system is tested.

For narrow band terminals the test signal used shall be bandfiltered between 200 Hz and 4 kHz with a bandpass filter providing a minimum of 24 dB/octet filter steepness, when feeding into the receiving direction.

The test signal levels are referred to the average level of the (band filtered in receiving direction) test signal, averaged over a period of 10 s. Unless specified otherwise, the averaging time for all measurements is 10 s.

5.4 Accuracy of test equipment

Unless specified otherwise, the accuracy of measurements is typically defined in the standard applicable. E.g. for narrowband test equipment the accuracy shall be better than:

Item	Accuracy
Electrical Signal Power	±0,2 dB for levels ≥ -50 dBm
Electrical Signal Power	±0,4 dB for levels < -50 dBm
Sound pressure	±0,7 dB
Time	±5 %
Frequency	±0,2 %

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

Quantity	Accuracy
Sound pressure level at MRP	±1 dB for 200 Hz to 4 kHz ±3 dB for 100 Hz to 200 Hz and 4 kHz to 8 kHz
Electrical excitation levels	±0,4 dB (see note 1)
Frequency generation	±2 % (see note 2)
NOTE 1: Across the whole frequency range. NOTE 2: When measuring sampled systems, it is advisable to avoid measuring at sub-multiples of the sampling frequency. There is a tolerance of ±2 % on the generated frequencies, which may be used to avoid this problem, except for 4 kHz where only the -2 % tolerance may be used.	

The measurements results shall be corrected for the measured deviations from the nominal level.

The sound level measurement equipment shall conform to IEC 61672 [5] Type 1.

These descriptions need to be found in the relevant standards of such transmission systems. Further information about the test of VoIP terminals can be found in [43], [44], [45] and [46].

6 Test conditions

6.1 Acoustic environment

In general two approaches need to be taken into consideration: Either the room and the background noise in the room contributes to the test parameter measured (e.g. performance measurements of acoustic echo cancellers) or the influence of the room and the background noise should be eliminated as much as possible (e.g. if loudness ratings of handsets are measured).

All measurements where no background noise or background noise simulation is needed should be conducted in quiet and "anechoic conditions". Depending on the distance of the transducers to 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. For hands-free terminals as well as for some headsets or small handset terminals an anechoic room is needed for these measurements. The potential error introduced by the environmental conditions should be less than 10 % of desired measurement accuracy.

In all conditions where the performance of acoustic echo cancellers is tested a realistic room representing the typical use condition for the terminal needs to be used.

In all conditions where the terminal performance in the presence of background noise is evaluated a representative background noise or different background noises are needed. Proper background noise simulation also may serve for this purpose. Care needs to be taken to realistically reproduce the background noise in level, frequency, temporal and spatial distribution (see TR 101 110 [i.2]).

6.2 Network conditions, general

The speech quality is dependent on a good transmission network. High Bit Error Rate (BER) packet loss without proper concealment and jitter contribute to degraded speech.

The speech levels are important for the user perception of the speech but also for the handling of the speech within the complete transmission system.

Low speech levels results in a low acoustical output volume. In analogue systems low speech levels may lead to insufficient signal to noise ratio, too high speech levels may lead to distorted and clipped speech signals. In digital systems speech coders and speech transcoders can not utilize the full dynamic range. Too high speech levels result in distorted speech. The level of distortion is dependent on the top and average speech level of the individual speaker, clipping can occur. Too low speech levels may lead to insufficient signal to noise ratio but to distorted speech signals as well due to the insufficient use of the dynamic range provided by the speech coders.

Unbalanced levels increase the risk for echo problems. The performance of the echo canceller is sensitive to a too large unbalance between sending and receiving.

The intention shall be to have the same line levels for the entire telephone network. Normal line levels have an average of -22 dBm0 to -16 dBm0 with a crest factor of about 16 dB. The individual differences can be large. For the average speaker this results in a peak level of -6 dBm0 to 0 dBm0. The clipping level is 3,14 dBm0 for PCM A-law coded speech and 3,17 dBm0 for PCM μ-law coded speech (ITU-T Recommendation G.711 [36]) and the clipping margin is then 3 dB, for the worst "normal" case. For wideband coding the relavant overload point depends on the codec chosen and is defined in the test standard applicable to the type of terminal under test.

Echo effects are present in all types of networks. The echo is not normally noticed in the traditional PSTN because of the short round trip delay. As soon as the one way transmission time exceeds 15 ms however echo might be noticeable. Generally there are two sources of echoes: one is the hybrid echo the other is acoustic echo. The hybrid echoes are created in the PSTN network, by the hybrid used when changing from four to two wire circuits. The hybrid normally provides an Echo Return Loss (ERL) of > 15 dB, that means the echo is 15 dB lower than the unreflected signal. The acoustic echoes are created by acoustic coupling between speaker and microphone. The level of the acoustic echo is dependent on the TCL (terminal coupling loss) of the terminal.

If a satellite connection is used as part of the connection, the additional delay causes degradation in the conversational speech quality.

The test setup chosen should be representative for the type of network to be evaluated. Either a network simulation should be used or a real network (under controlled conditions) should be used. The network condition should be reproducible. If this can not be guaranteed a statistical approach should be chosen measuring the relevant parameters at various times in order to find average, maximum and minimum quality provided by the configuration.

6.2.1 Network conditions, PSTN

In general all parameters determining the speech quality in circuit switched networks are well controlled. PSTN type networks work isochronously. Typically loss of data is fairly unlikely. The delays in a connection are constant and within the control of the network operators. The network interfaces in such networks are well defined. In analogue networks the line characteristics (length and type of lines) has to be taken into account since it may have impact on loudness ratings, frequency responses, sidetone performance and trans-hybrid losses.

- Echo cancellation.

  As soon as a delay higher then 25 ms (mouth-to-ear) is introduced in a connection echo cancellation has to be provided. In PSTN networks this is typically concentrated at international exchanges since within national or continental connections the network delay is typically below 25 ms. The echo cancellation is intended to cancel echoes produced by the hybrid of analogue terminals. The maximum echo loss to be provided is generally determined by the maximum delay which may be inserted in a connection. However most of modern network echo cancellation should conform to ITU-T Recommendation G.168 [11] which assumes 250 ms one way transmission time and provides adequate performance limits.

- Speech levels.

  Different speech levels e.g. due to different national level settings, different attenuation in analogue systems can be expected and may degrade the performance of echocancellers and other signal processing the network and the terminals.

- Speech coding.

  Low bitrate speech coding is typically not found in PSTN networks although ADPCM type coding is used quite frequently used in cordless terminals.

When DCME are inserted in a PSTN network all considerations found in clause 6.2.2 apply.

6.2.2 Network conditions, packet based transmission

Packet based networks differ considerably from circuit switched networks. The input signal (speech signal) is segmented into packets typically of fixed length. Typical packet lengths used are in the range of 5 ms to 30 ms. By the packetization itself a minimum delay namely the packet length is introduced into the transmission.

- Delay, packet loss and jitter

  Furthermore the packets of different sources (e.g. terminals or subscribers) have to be sequenced for transmission. This process may add additional delay - as well as any buffering or routing in a connection. While in isochronous networks the delay of a connection is constant and predictable, the delay in IP-networks may vary between different connections depending on network load, routing, bandwidth, switching and others. In addition jitter may be introduced by the IP connection. Jitter occurs when the delay of each individual packet varies. Again, jitter is depending on network load, transmission bandwidth, switching and others. The delay and jitter is not predictable in IP-networks and therefore it is described by statistical parameters such as delay (average and distribution) and packet loss distribution.

  In IP-networks the receiver has to take care of collection of the packets and their correct ordering. Jitter buffers are used in order to collect and sequence the packets received in such a way that a mostly error free transmission is achieved.

While in isochronous TDM networks the packet loss is negligible and almost no jitter occurs, packet loss and jitter are important factors in IP-connection and may highly influence the speech quality. When assessing the speech quality of IP networks these parameters (packet loss and jitter) have to be taken into account and speech quality measurements have to be conducted for the various network conditions to be expected. It is advisable to make appropriate simulations in a lab-type environment in order to get an estimate of the speech quality range to be expected under real use conditions (see ITU-T Recommendation Y.1541 [32]).

- Echo cancellation

  As soon as a delay higher then 15 ms is introduced in a connection echo cancellation has to be provided, either in the terminal or in the network. In IP connections gateways have to be equipped with echo cancellers providing sufficient echo loss for all expected delays in a connection. The maximum echo loss to be provided is determined by the maximum delay which may be inserted in a connection. Since under worst case conditions the maximum one way transmission delay may be higher than 300 ms the minimum echo loss which has to be provided is 55 dB (see ITU-T Recommendation G.131 [10]). Consequently any IP-terminal connected to an IP network has to provide the same amount of echo loss (see TIA/EIA 810-A [34]).

- Speech levels

  Different speech levels e.g. due to computer terminals with insufficiently controlled LRs or equipment with codec which are not properly implemented may degrade the performance of echocancellers and other signal processing in an IP-network.

- Speech coding

  Speech coding especially in combination with packet loss and jitter may add further impairments to the transmitted speech signal. Speech codecs with integrated Packet Loss Concealment algorithms (PLC) add less impairments in presence of packet loss than speech codecs without PLC.

- Silence suppression

  Frequently silence suppression is used, voice activity is detected by voice activity detection, silent periods (periods where the signal level is below a defined threshold) are suppressed and not transmitted in order to save bandwidth. At the far end side silent periods typically are replaced by a low level noise signal (comfort noise). Especially in background noise situations this procedure may lead to background noise modulations in combination with comfort noise which may be quite annoying for the user. Furthermore voice activity detection may lead to clipping of the actual speech signal.

In IP-networks the classical differentiation between terminal and network no longer can be made, all sorts of impairments mentioned above may be introduced by IP-terminals as well.

6.2.3 Network conditions, GSM mobile and 3G mobile

The speech quality is dependent on a good transmission network. High Bit Error Rate (BER) on the A-bis and A-ter interface contributes to degraded speech. The PCM coded speech used between TRC and MSC is less sensitive to the effects of high BERs than the GSM coded speech transmitted between the BTS and the TRC.

The audible effects on speech quality due to increasing or decreasing BER, on the A-bis and A-ter interface, are unnoticeable providing that the transmission fulfils the performance objectives stated in ITU-T Recommendation G.821 [1].

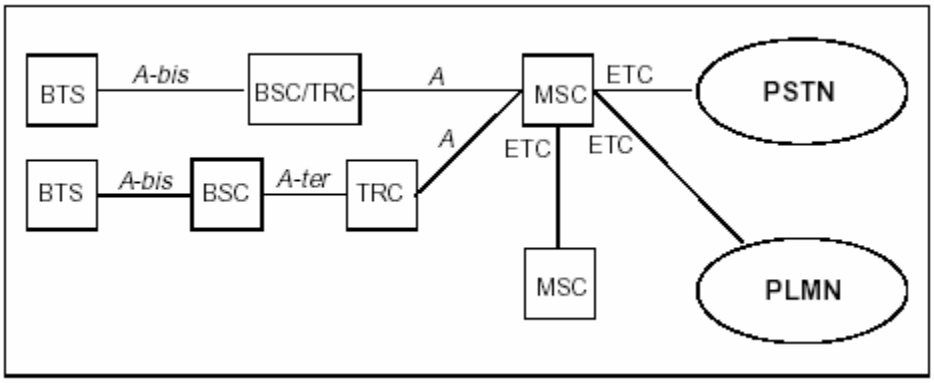


Figure 5: GSM Network configuration

Also for 3G networks (UTRAN), the speech quality is partly dependent on a good transmission network. High Packet Error Rate (PER) on the Iu, Iur and Iub interface contributes to degraded speech. The PCM coded speech used between TRC and MSC is less sensitive to the effects of high PERs than the GSM coded speech transmitted between the BTS and the TRC. The audible effects on speech quality due to increasing or decreasing PER, on the Iu and Iub interface, are unnoticeable providing that the transmission fulfils the performance objectives stated in ITU-T Recommendation G.821 [1].

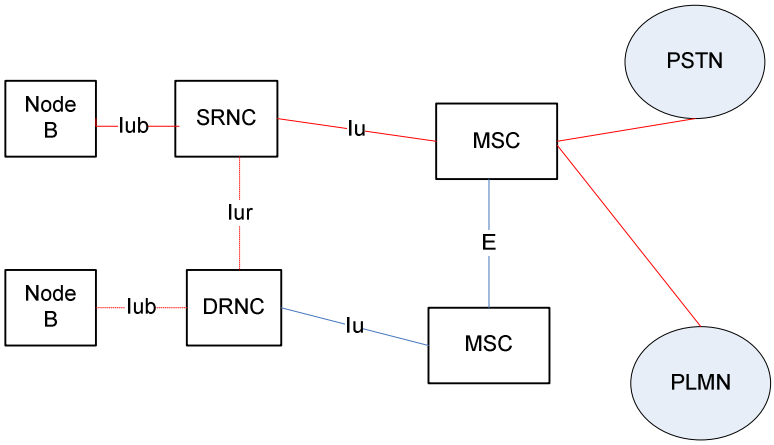


Figure 6: 3G CS voice network configuration

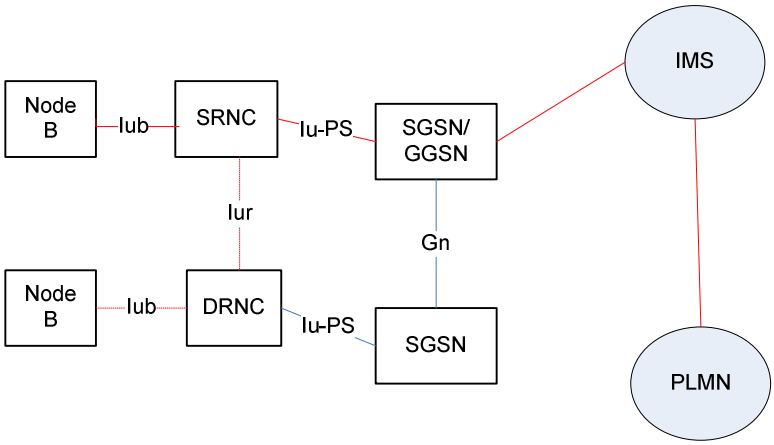


Figure 7: 3G PS voice network configuration

The upper limit of the speech quality is set by the codecs (e.g. TS 100 961 [37], Full rate speech transcoding, EN 300 969 [38], Half rate speech transcoding and EN 300 726 [39], Enhanced full rate speech transcoding). The speech codec quality is fixed and the focus of this clause is thus on the speech degradation factors. The following areas have been identified:

- Speech Levels:

  The speech levels affect speaker level, distortion and echo canceller performance.

- Echo control:

  Poor network echo cancelling results in disturbing echo for calls to the PSTN and PLMN with full duplex hands-free mobiles. Echo effect can also be caused by acoustic coupling in mobiles.

- Radio Network and Radio Network features:

  High bit error rate as a result of a low Carrier/Noise (C/N) and/or Carrier/Interference (C/I) ratio causes interruptions or degradation of the speech. The speech degradation because of radio environment factors can be minimized by using the radio network features Discontinuous Transmission (DTX), Power control and Frequency hopping.

  The use of the radio network feature DTX affect the speech quality by removing the true background. The handover procedure creates a short speech interruption.

  In UTRAN High PER as a result of a low Signal-to-Noise Ratio (SNR) and/or Signal-to-Interference Ratio (SIR) ratio causes interruptions or degradation of the speech. The speech degradation because of radio environment factors can be minimized by using a good uplink outer-loop power control for 3GPP Rel'99 CS speech operation. For CS over HSPA (High Speed Packet Access) introduced in 3GPP Rel'7 and for VoIP over HSPA intorduced in 3GPP Rel'6, a good scheduler and HARQ operation in the Node B is important to ensure that the required QoS can be met. Features such as DTX and DRX can be used to minimise the battery consumption for mobiles using voice.

- Transmission Network:

  High Bit Error Rate (BER) causes degradation of the speech. The speech quality is sensitive to BER on the A-bis and A-ter interface.

- User Equipment (UE):

  In UTRAN and GERAN the implementation of audio functions in the mobile stations affects the speech quality, i.e. speaker, microphone and speech processing. The mobile station implementation is not further discussed in this clause.

6.2.3.1 Speech levels

The speech levels are important for the user perception of the speech but also for the handling of the speech within the system.

In general the same considerations apply as for networks in general. Due to the specific size and shape of mobile terminals however specific signal processing especially echo cancellation is found in the terminals. Care should be taken to avoid unbalanced speech levels. The performance of the echo canceller is sensitive to a too large unbalance between the up- and downlink. The worst situation is when the uplink level is higher than the downlink level. High speech levels in the down link increase the risk for acoustic echo in the mobile.

The intention shall be to have the same line levels for the entire telephone network. This is dependent on the transmission, and adjustments of the line level might be used.

6.2.3.2 Echo control

The mean one-way delay of a GSM System is about 95 ms. The mean one-way delay of a UTRAN system is about 115 ms. This means that the echo returns to the mobile with a 190 ms respectively 230 ms delay (if satellite transmission are used the delay will be even longer). The echoes can be of two different types, hybrid and acoustic, see figure 6.

The echoes generated outside the own PLMN are handled by the echo cancellers in the GSM system. This is because the echo generated outside the own PLMN, can be heard in the GSM mobile.

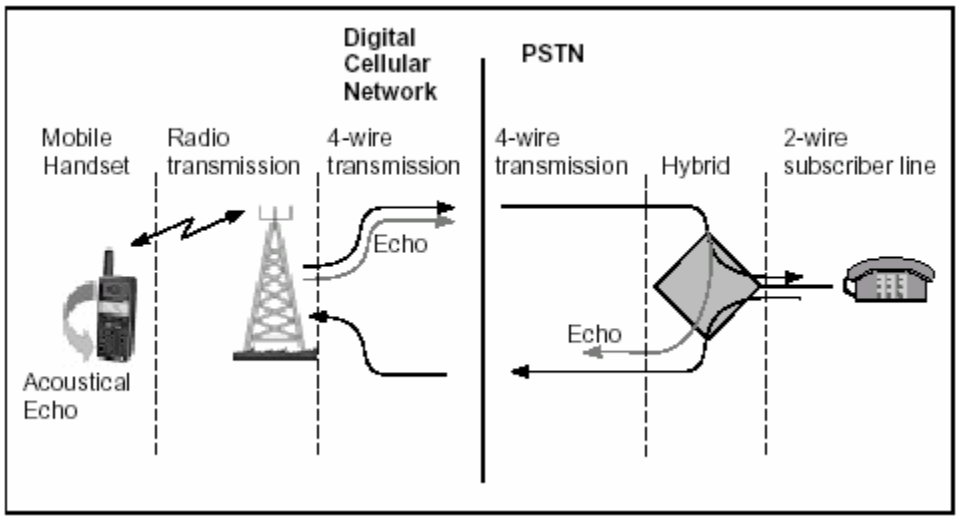


Figure 8: The echoes generated in PSTN and PLMN networks

Hybrid echo

The hybrid echoes are created in the PSTN network, by the hybrid used when changing from four to two wire circuits. The hybrid normally provides an Echo Return Loss (ERL) of > 15 dB, that means the echo is 15 dB lower than the unreflected signal. Thus echo cancellation needs to be provided by the PLMN in order to prevent echo signals being heard by the mobile user.

Acoustic echo

In the PLMN network the mobiles should handle the echo control. The mobiles should provide an EL of 46 dB (see EN 300 903 [40], Transmission Planning Aspects of the Speech Service in the GSM Public Land Mobile Network (PLMN) System) to eliminate the need for internal PLMN echo cancellers. This is hard to achieve with mobiles. Voice switching, echo cancellation or echo suppressors are commonly used to achieve the 46 dB EL requirement.

The characteristics of the echo originated in the MS are very different from those of the PSTN echo. The acoustics echo from the MS is non-linear (due to speech coding and decoding in the echo path) and the end path delay is very long (of the order of 200 ms). These are the reasons why one cannot remove the MS echo with the same type of echo canceller that is developed for the PSTN echo, the echo generated by the mobile needs to be cancelled by the mobile itself before the speech codec is involved.

6.2.3.3 Radio network and radio network features

Provided that the rest of the system(s) is functioning correctly, i.e. do not contribute to degradation of speech quality, poor performance of the radio network (air interface) is the major contributor to degraded speech quality of a call. In GERAN the radio network environment is affected by the factors:

- co-channel interference (C/I);
- noise limitations (C/N);
- mobile speed (fading frequency);
- time dispersion;
- adjacent channel interference (C/A).

In UTRAN the radio network environment is affected by the factors:

- outer-loop power control (which sets the target for inner-loop power control);
- intra-frequency interference (intra-cell and inter-cell) (I_0);
- thermal noise (N_0);
- multipath fading and shadowing;
- adjacent channel interference (C/A).

During a call these factors together contribute to:

- Bit Error Rate (BER): the average amount of bit errors in a speech frame;
- Frame Erasure Rate (FER): the percentage of erased frames.

When the BER and the FER increases the speech decoder will get less and less information about the coded speech and thus the speech quality will degrade, see figure 9. This is why a good cellplanning is needed to avoid these kinds of interferences as much as possible. The quality of the radio environment can further be improved by using the radio network features DTX, Power control and Frequency hopping. Another degrading factors are the handovers. While moving from cell to cell they will also introduce a speech quality disturbance.

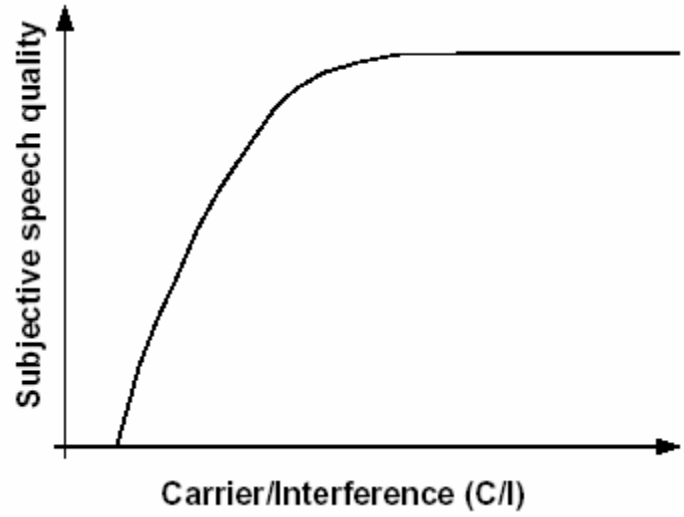


Figure 9: Speech quality vs C/I

Discontinuous Transmission (DTX)

The radio network feature Discontinuous Transmission reduces the total interference in the network, it only transmits when speech is detected. In normal conversation this will lead to a decreased transmitting time of about 50 %. The general speech quality for the network can therefore be increased. The use of DTX affects the speech quality, both positively and negatively. Positively by reducing the total interference level in the network. Negatively by introducing speech clipping and by injecting comfort noise instead of the true background. This could be disturbing for connections with high background noise levels where the background is very different from that of the comfort noise generated. It should be noted that the DTX and comfort noise generation is part of the speech codec and can therefore differ from one codec to another.

Frequency hopping

Frequency hopping can reduce the effect on the speech quality caused by multipath fading and interference. The signal strength variations, or the interference, are broken up into pieces of duration short enough for the interleaving and speech coding process to correct the errors. The average speech quality is thus increased compared to a non frequency hopping network.

Power control

In GERAN dynamic power control used together with frequency hopping and DTX results in a further increase in the protection against fading dips and interference. This is achieved by the fact that the output power is controlled with respect to received signal strength as well as with respect to speech quality (BER). The improvement in interference protection can be utilized directly as an improvement in general speech quality. It can also be utilized in replanning the radio network for high capacity, while leaving the general speech quality constant.

Quality based power control contributes to better speech quality by:

- enhanced signal strength based part that gives increased C/I gains in the radio network compared with the existing power control;
- a quality based part that raises the power in the presence of interference, which gives generally better speech quality.

In UTRAN a good outer-loop power control operation means that the resource needed for the call is optimal such that it meets the Block Error Rate (BLER) required for the call. This means that the resource is neither too little that the BLER is too high, nor that the resource is too high, which would mean that the UE and/or network is needlessly transmitting at a higher power than required, potentially causing capacity degradation and higher UE battery consumption in the uplink case.

Handover

At handover there will be a short speech interruption in both downlink and uplink. There will also be signalling going on in conjunction with the handover. This signalling will steal speech frames and can by that introduce some degradation. One message (the Physical Information message) will be repetitively sent in the new cell until the mobile and the system has established a connection in this cell. Since there is a delay in the communication between the system and the mobile, the system should not repeat the message to fast so as to avoid unnecessary repetitive signalling.

In UTRAN both, soft handover and hard handover are supported. Soft handover is supported in uplink and downlink for dedicated channels. However for voice over HSPA in downlink, it is not supported, but for voice over HSPA in uplink it is supported by the 3GPP standard. The trade-off for the use of soft-handover is between speed of packet delivery vs. amount of resource needed from a single base-station to maintain the connection. It was observed that it would be useful to maintain this in the 3GPP HSPA architecture for uplink only. Hard handover is supported for HSPA in both uplink and downlink. This causes some interruption to service, however it is very much dependent on the delay of the handover, and this is dependent on the signalling delay, network and UE processing delay. In 3GPP release 8, work has been done to speed up the handover mechanism for HSDPA, particularly for voice and other real-time services.

7 Measurement of "standard" parameters

The standard parameter and the according measurements to be included here are the following:

- Frequency Response in Sending and Receiving Direction.
- Overall Frequency Responses.
- SLR Sending Loudness Rating.
- RLR Receiving Loudness Rating.
- OLR Overall Loudness Rating.
- STMR Sidetone Masking Rating.
- LSTR Listener Sidetone Rating.
- D D-Value of Terminal.
- TCLw Terminal Coupling Loss (weighted).
- WEPL Weighted Echo Path Loss.
- TELR Talker Echo Loudness Rating.
- qdu number of quantizing distortion units.
- Nc circuit Noise referred to the 0 dBr-point.
- Distortion in Sending and Receiving Direction.
- Out-of-Band Signals in Sending and Receiving Direction.

In general the measurement principles are the same as used in other standards e.g. TBR 008 [i.4] but special consideration needs to be given to the following points:

- The appropriate measurement signal (especially with respect to the codec) needs to be chosen.

  It is recommended to use a speech-like test signal. Speech like test stimuli can be found in ITU-T Recommendations P.50 [17] and P.501 [18].

- The averaging times used to determine the transfer characteristics need to be adapted to the measurement signal chosen.
- Instead of level measurements using sine wave signal excitation, typically Fourier transformation is used for calculation/estimation of the output spectra.
- The measured output spectrum is always referred to the signal spectrum in order to determine transfer functions, loudness ratings etc.
- New procedures need to be established in order to determine distortion, especially to determine the parameter "speech sound quality" in combination with the acoustical interface. The basis for such procedures may be found in EG 201 377-1 [i.1].

NOTE: When using the E-model [7] for predicting the speech quality in terms of R-values the following, additional parameter need to be determined:

- Ie Equipment Impairment Factor (low bit-rate Codecs).
- Nfor Noise Floor at the Receive-side.
- Ps Room Noise at the Send-side.
- Pr Room Noise at the Receive-side.

For the evaluation of Ie values for low bit-rate codecs, some objective measurement methods have been developed or are under development (see ITU-T Recommendation P.862 [31] and EG 201 377-1 [i.1]).

7.1 Sending frequency response

a) The test signal to be used for the measurements shall be the artificial voice according to ITU-T Recommendation P.50 [17] or a speech like test signal as described in ITU-T Recommendation P.501 [18]. The type of test signal used shall be stated in the test report. 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.

b) The handset terminal is setup as described in clause 5. The handset is mounted either in the LRGP or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report.

   The hands-free terminal setup is described in clause 5. In addition to the MRP calibration, the broadband signal level is adjusted to -28,7 dBPa at the HFRP and the spectrum is not altered.

   Measurements shall be made at one twelfth-octave intervals as given by the R.40 series of preferred numbers in ISO 3 [6] for frequencies from 100 Hz to 4 kHz (100 Hz to 8 kHz for wideband systems) inclusive. 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.

c) The sensitivity is expressed in terms of dBV/Pa.

7.2 Receiving frequency response

a) The test signal to be used for the measurements shall be the artificial voice according to ITU-T Recommendation P.50 [17] or a speech like test signal as described in ITU-T Recommendation P.501 [18]. The type of test signal used shall be stated in the test report. The test signal level shall be -16 dBm0, measured at the digital reference point or the equivalent analogue point. The test signal level is averaged over the complete test signal sequence.

b) The handset terminal is setup as described in clause 5. The handset is mounted either in the LRGP position or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report.

   The hands-free terminal is setup as described in clause 5. The HATS is diffuse field equalized as described in ITU-T Recommendation P.581 [23]. The equalized output signal of each artificial ear is power-averaged on the total time of analysis; the "right" and "left" signals are voltage-summed for each 1/3 octave band frequency band; these 1/3 octave band data are considered as the input signal to be used for calculations or measurements. In symmetrical setups alternatively the output signal of just one ear ear can be chosen for analysis.

   Measurements shall be made at one twelfth-octave intervals as given by the R.40 series of preferred numbers in ISO 3 [6] for frequencies from 100 Hz to 4 kHz (100 Hz to 8 kHz for wideband systems) 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.

c) The sensitivity is expressed in terms of dBPa/V.

7.3 Overall frequency response

a) The test signal to be used for the measurements shall be the artificial voice according to ITU-T Recommendation P.50 [17] or a speech like test signal as described in ITU-T Recommendation P.501 [18]. The type of test signal used shall be stated in the test report. The spectrum of the 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.

b) Handset terminals are setup as described in clause 5. The handsets are mounted either in the LRGP or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report.

Hands-free terminals are setup as described in clause 5. In addition to the MRP calibration, the broadband signal level is adjusted to -28,7 dBPa at the HFRP and the spectrum is not altered. The HATS is diffuse field equalized as described in ITU-T Recommendation P.581 [23]. The equalized output signal of each artificial ear is power-averaged for the total time of analysis; the "right" and "left" signals are voltage-summed for each 1/3 octave band frequency band; these 1/3 octave band data are considered as the input signal to be used for calculations or measurements. In symmetrical stetups aternatively the output signal of just one ear ear can be chosen for analysis.

Measurements shall be made at one twelfth-octave intervals as given by the R.40 series of preferred numbers in ISO 3 [6] for frequencies from 100 Hz to 4 kHz (100 Hz to 8 kHz for wideband systems) 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.

c) The sensitivity is expressed in terms of dBPa/Pa.

7.4 Sending (and connection) loudness rating

a) The test signal to be used for the measurements shall be the artificial voice according to ITU-T Recommendation P.50 [17] or a speech like test signal as described in ITU-T Recommendation P.501 [18]. The type of test signal used shall be stated in the test report. 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.

b) The handset terminal is setup as described in clause 5. The handset is mounted either in the LRGP or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report.

The hands-free terminal setup is described in clause 5. In addition to the MRP calibration, the broadband signal level is adjusted to -28,7 dBPa at the HFRP and the spectrum is not altered.

The sending sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [25], bands 4 to 17 (1-20 for wideband systems). 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.

c) The sensitivity is expressed in terms of dBV/Pa. For narrowband systems the SLR shall be calculated according to ITU-T Recommendation P.79 [25], formula 5-1, over bands 4 to 17, using m = 0,175 and the sending weighting factors from ITU-T Recommendation P.79 [25], table 1.

For wideband systems the SLR shall be calculated according to ITU-T Recommendation P.79 [25], annex A.

7.5 Receiving (and connection) loudness rating

a) The test signal to be used for the measurements shall be the artificial voice according to ITU-T Recommendation P.50 [17] or a speech like test signal as described in ITU-T Recommendation P.501 [18]. The type of test signal used shall be stated in the test report. The test signal level shall be -16 dBm0, measured at the digital reference point or the equivalent analogue point. The test signal level is averaged over the complete test signal sequence.

b) The handset terminal is setup as described in clause 5. The handset is mounted either in the LRGP or in the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report. The DRP-ERP correction as described in ITU-T Recommedation P.57 [21] is used.

The hands-free terminal is setup as described in clause 5. The HATS is freefield equalized as described in ITU-T Recommendation P.581 [23]. The equalized output signal of each artificial ear is power-averaged on the total time of analysis; the "right" and "left" signals are voltage-summed for each 1/3 octave band frequency band; these 1/3 octave band data are considered as the input signal to be used for calculations or measurements. In symmetrical stetups alternatively the output signal of just one ear ear can be chosen for analysis.

For narrowband systems the receiving sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [25], bands 4 to 17. For the calculation the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

For wideband systems the receiving sensitivity shall be calculated from each band of the 20 frequencies given in table 1 of ITU-T Recommendation P.79 [25], 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.

c) For narrowband systems the sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [25], formula 5-1, over bands 4 to 17, using m = 0,175 and the receiving weighting factors from table 1 of ITU-T Recommendation P.79 [25].

   For wideband systems the RLR shall be calculated according to ITU-T Recommendation P.79 [25], annex A. No leakage correction shall be applied for the measurement.

   For hands-free terminals the calculated result shall be corrected by subtracting 8 dB.

d) No leakage correction shall be applied when using HATS or Type 3.2 artificial ear for the measurement.

7.6 Overall loudness rating

a) The test signal to be used for the measurements shall be the artificial voice according to ITU-T Recommendation P.50 [17] or a speech like test signal as described in ITU-T Recommendation P.501 [18]. The type of test signal used shall be stated in the test report. The spectrum of the 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.

b) Handset terminals are setup as described in clause 5. The handsets are mounted either in the LRGP or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report. . The DRP-ERP correction as described in ITU-T Recommendation P.57 [21] is used.

   Hands-free terminals are setup as described in clause 5. In addition to the MRP calibration, the broadband signal level is adjusted to -28,7 dBPa at the HFRP and the spectrum is not altered. The HATS is freefield equalized as described in ITU-T Recommendation P.581 [23]. The equalized output signal of each artificial ear is power-averaged for the total time of analysis; the "right" and "left" signals are voltage-summed for each 1/3 octave band frequency band; these 1/3 octave band data are considered as the input signal to be used for calculations or measurements. In symmetrical stetups alternatively the output signal of just one ear ear can be chosen for analysis.

   Measurements shall be made at one twelfth-octave intervals as given by the R.40 series of preferred numbers in ISO 3 [6] for frequencies from 100 Hz to 4 kHz (100 kHz to 8 kHz for widenband systems) 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.

c) For narrowband sytems the sensitivity is expressed in terms of dBPa/Pa and the OLR shall be calculated according to ITU-T Recommendation P.79 [25], formula 5-1 over bands 4 to 17, using m = 0,175 and the overall weighting factors from ITU-T Recommendation P.79 [25], table D.1.

   For wideband systems the OLR shall be calculated according to ITU-T Recommendation P.79 [25], annex A. No leakage correction shall be applied for the measurement.

   For hands-free terminals the calculated result shall be corrected by subtracting 8 dB.

7.7 Sidetone masking rating

The measurement of STMR is applicable under the following conditions:

- The terminal only respectively connected through a typical line is measured.
- The Type 1 artificial ear is used.
- The measurement is only applicable to handset terminals which can be sealed to the Type 1 artificial ear.

The measurement is not applicable for end-to-end measurements including terminals.

Test procedure:

a) The test signal to be used for the measurements shall be the artificial voice according to ITU-T Recommendation P.50 [17] or a speech like test signal as described in ITU-T Recommendation P.501 [18]. The type of test signal used shall be stated in the test report. The spectrum of the 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.

b) Handset terminals are setup as described in clause 5. The handset is mounted in the LRGP position (see ITU-T Recommendation P.64 [24]) and the earpiece is sealed to the knife-edge of the artificial ear.

c) Where a user controlled volume control is provided, the measurements shall be carried out at a setting which is as close as possible to the nominal value of the RLR (RLR = 3 dB).

   Measurements shall be made at one twelfth-octave intervals as given by the R.40 series of preferred numbers in ISO 3 [6] for frequencies from 100 Hz to 8 kHz inclusive. For the calculation the averaged measured level at each frequency band (ITU-T Recommendation P.79 [25], table 4, bands 1 to 20) is referred to the averaged test signal level measured in each frequency band.

d) The Sidetone path loss (LmeST), as expressed in dB, and the SideTone Masking Rate (STMR) (in dB) shall be calculated from the formula 5-1 of ITU-T Recommendation P.79 [25], using m = 0,225 and the weighting factors of in table 3 of ITU-T Recommendation P.79 [25].

NOTE 1: STMR is needed in the E-Model [7].

NOTE 2: Further investigations are required when applying Type 3.x artificial ears for the measurement. For the time being values measured using Type 3.x artificial ears should not be used directly in the E-model since values required here depend on the electrical sidetone measured. Measurements with 3.x artificial ears always include the acoustical sidetone which is included due to the acoustical leakage present in those artificial ears.

7.8 Listener sidetone

NOTE 1: With 3.x types of ears it is not possible to conduct LSTR measurements. Instead the D-value measurement should be applied.

The measurement of LSTR is applicable under the following conditions:

- The terminal only respectively connected through a typical line is measured.
- The Type 1 artificial ear is used.
- The measurement is only applicable to handset terminals which can be sealed to the Type 1 artificial ear.

The measurement is not applicable for end-to-end measurements including terminals.

Test procedure:

a) Sound field calibration: The diffuse sound field is calibrated in the absence of any local obstacles. The averaged field shall be uniform to within ±3 dB within a radius of 0,15 m of the MRP, when measured in one-third octave bands according to IEC 61260 [4] from 100 Hz to 8 kHz (bands 1 to 20).

NOTE 2: The pressure intensity index, as defined in ISO 9614 [41], may prove to be a suitable method for assessing the diffuse field.

NOTE 3: Where more than one loudspeaker is used to produce the desired sound field, the loudspeakers may require to be fed with non-coherent electrical signals to eliminate standing waves and other interference effects.

b) Where adaptive techniques or voice switching circuits are not used (need to be declared by the supplier) the spectrum shall be band limited (50 Hz to 10 kHz) "pink noise" (see ITU-T Recommendation P.64 [24], annex B) to within ±3 dB and the level shall be adjusted to 70 dB(A) (-24 dBPa(A)). The tolerance for this level is ±1 dB.

In other cases the level shall be adjusted to 50 dB(A) (-44 dBPa(A)). The tolerance for this level is ±1 dB.

c) Handset terminals are setup as described in clause 5. The handset is mounted in the LRGP position (see ITU-T Recommendation P.64 [24]) and the earpiece is sealed to the knife-edge of the artificial ear.

d) Measurements are made on one-third octave bands according to IEC 61260 [4] for the 20 bands centred at 100 Hz to 8 kHz (bands 1 to 20). For each band the sound pressure in the artificial ear shall be measured by connecting a suitable measuring set to the artificial ear.

NOTE 4: There may be problems with the signal to noise ratio. If it is less than 10 dB in any band, the microphone noise level and the noise level of any out-of-band signals need to be subtracted from the measured sidetone level (power subtraction).

e) The listener sidetone path loss is expressed in dB and the LSTR shall be calculated from ITU-T Recommendation P.79 [25], formula 5-1, using m = 0,225 and the weighting factors in table 3 of ITU-T Recommendation P.79 [25].

NOTE 5: LSTR is needed for the E-Model [7]. LSTR can be calculated from D-value measurements.

NOTE 6: Further investigations are required when applying Type 3.x artificial ears for the measurement. For the time being values measured using Type 3.x artificial ears should not be used directly in the E-model since values required here depend on the electrical sidetone measured. Measurements with 3.x artificial ears always include the acoustical sidetone which is included due to the acoustical leakage present in those artificial ears.

7.9 Measurement and calculation of the value of the D-factor (DelSM)

In general the D-factor can be measured for terminals in combination with a suitable network simulation but as well from end-to-end using the far end terminal for assessing the signal transmitted in sending.

NOTE 1: Wideband calculation is for further study, provisionally the measurement is based on narrowband.

a) Sound field calibration: The diffuse sound field is calibrated in the absence of any local obstacles. The averaged field shall be uniform to within ±3 dB within a radius of 0,15 m of the MRP, when measured in one-third octave bands according to IEC 61260 [4] from 100 Hz to 8 kHz (bands 1 to 20).

NOTE 2: The pressure intensity index, as defined in ISO 9614 [41], may prove to be a suitable method for assessing the diffuse field.

NOTE 3: Where more than one loudspeaker is used to produce the desired sound field, the loudspeakers may require to be fed with non-coherent electrical signals to eliminate standing waves and other interference effects.

Besides the measurement in a diffuse pink or hoth noise the measurement with realistic background noise is highly recommended. The setup and equalization as well as a database with different types of background noises is described in EG 202 396-1 [i.6].

b) Where adaptive techniques or voice switching circuits are not used (need to be declared by the supplier) the spectrum shall be band limited (50 Hz to 10 kHz) "pink noise" (see ITU-T Recommendation P.64 [24], annex B) to within ±3 dB and the level shall be adjusted to 70 dB(A) (-24 dBPa(A)). The tolerance for this level is ±1 dB. For other types of terminals a realistic background noise and the background noise setup as described in EG 202 396-1 shall be used. The backround noise level depends on the type of background noise chosen and can be found in [i.6].

c) Handset or headset terminals are mounted as described in clause 5. Measurements are made on one-third octave bands according to IEC 61260 [4] for the 14 bands centred at 200 Hz to 4 kHz (bands 4 to 17). For each band the diffuse sound sensitivity S_{si}(diff) is measured. The sensitivity shall be expressed in terms of dBV/Pa.

The D-factor may be used for hands-free terminals as well. The requirements however have to be adapted. Hands-free terminals are setup as described in clause 5.

NOTE 4: Additional types of background noise should be used to conduct tests with background noise simulating the real use conditions of the terminal as close as possible. Therefore level and spectrum as well as distribution of sound sources may be different as compared to the requirements when simulating a diffuse sound field with pink noise.

d) The direct sound sensitivity shall be measured using the test set-up specified in clause 5.1 and a speech like test signal as defined in ITU-T Recommendation P.50 [17] or P.501 [18]. The type of test signal used shall be stated in the test report. The direct sound sensitivity is measured in one-third octave bands according to IEC 61260 [4] for the 14 bands centred at 200 Hz to 4 kHz (bands 4 to 17). For each band the direct sound sensitivity S_{si}(direct) is measured. The sensitivity shall be expressed in terms of dBV/Pa.

e) The value of the D-factor shall be calculated according to ITU-T Recommendation P.79 [25], annex E, formulas E2 and E3, over the bands from 4 to 17, using the coefficients K_i from table E1 of ITU-T Recommendation P.79 [25].

7.10 Delay

7.10.1 Delay in sending direction

a) The test signal to be used for the measurements shall be a Composite Source Signal (CSS) as described in ITU-T Recommendation P.501 [18]. 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.

b) The handset terminal is setup as described in clause 5. The handset is mounted either in the LRGP or in the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear shall be stated in the test report.

The hands-free terminal setup is described in clause 5. In addition to the MRP calibration, the broadband signal level is adjusted to -28,7 dBPa at the HFRP and the spectrum is not altered.

The delay is calculated using the cross correlation function between the signal at the electrical test point and the signal at the MRP. The measurement is corrected by the delay introduced by the test equipment.

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

NOTE: 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 (ITU-T Recommendation P.862 [31]) may be used.

7.10.2 Delay in receiving direction

a) The test signal to be used for the measurements shall be a Composite Source Signal (CSS) as described in ITU-T Recommendation P.501 [18]. 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.

b) The handset terminal is setup as described in clause 5. The handset is mounted either in the LRGP or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear shall be stated in the test report.

The hands-free terminal is setup as described in clause 5. The HATS is diffuse field equalized as described in ITU-T Recommendation P.581 [23]. The equalized output signal of one artificial ear is used for the delay calculation.

The delay is calculated using the cross correlation function between the signal at the electrical test point and the signal at the DRP. The measurement is corrected by the delay introduced by the test equipment.

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

NOTE: 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 (ITU-T Recommendation P.862 [31]) may be used.

7.10.3 Overall delay

a) The test signal to be used for the measurements shall be a Composite Source Signal (CSS) as described in ITU-T Recommendation P.501 [18]. 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.

b) The handset terminals are setup as described in clause 5. The handsets are mounted either in the LRGP or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear shall be stated in the test report.

Hands-free terminals are setup as described in clause 5. In addition to the MRP calibration, the broadband signal level is adjusted to -28,7 dBPa at the HFRP and the spectrum is not altered. The HATS is diffusefield equalized as described in ITU-T Recommendation P.581 [23]. The equalized output signal of one artificial ear is used for the delay calculation.

The delay is calculated using the cross correlation function between the signal at the MRP on the one side of the connection and the signal at the DRP at the other side of the connection. The measurement is corrected by the delay introduced by the test equipment.

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

NOTE: 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 (ITU-T Recommendation P.862 [31]) may be used.

7.11 Terminal coupling loss

The measurement is only applicable for terminals respectively terminals connected through a typical line.

The measurement is not applicable for end-to-end measurements including terminals at both ends.

a) For conducting the tests the typical environments where the terminal is typically used shall be used. E.g. office rooms for all office types of telephones, a car cabin for hands-free telephones in vehicles. The terminal is setup as described in clause 5. The ambient noise level shall be less than -64 dBPa(A) for handset and headset terminals, less than -70 dBPa(A) for hands-free type terminals. The attenuation from electrical reference point input to electrical reference point output shall be measured using a speech like test signal.

b) Before the actual test a training sequence consisting of 10 s artificial voice male and 10 s artificial voice female according to ITU-T Recommendation P.50 [17] is altered. The training sequence level shall be -16 dBm0 in order not to overload the codec.

c) The test signal is a PN-sequence complying with ITU-T Recommendation P.501 [18] 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. The test signal level is -3 dBm0. The low-crest factor is achieved by random-alternation of the phase between -180° and 180°.

d) The TCLw is calculated according to ITU-T Recommendation G.122 [9], clause B.4 (trapezoidal rule). For wideband terminals the trapezoidal rule is used as well but the frequency range is extended to 300 to 6 700 Hz 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 length of the test signal shall be at least one second (1,0 s). For the measurement a time window has to be applied adapted to the duration of the actual test signal (200 ms).

NOTE: In addition requirements on spectral and temporal echo loss should be considered.

7.12 Talker echo loudness rating

NOTE 1: Wideband calculation is for further study, provisionally the measurement is based on narrowband.

In general the talker echo loudness rating can be expressed by:

$$TELR = SLR + RLR + Le$$

where Le is the echo loss provided by the far end. Le includes the complete echo loss provided by the far end terminal and all echo cancellers active in the connection. The measurement of SLR and RLR is described in clauses 7.4 and 7.5.

In scenarios where the near end terminal is not connected, the measurement of Le can be made by assessing the connection at the 0 dBr point and calculating Le according to ITU-T Recommendation G.111 [8]. If no echo cancellation is involved in the network and no loss is inserted, the Le value is identical to TCL W. In these scenarios the test setup is as follows:

a) For conducting the tests the typical environments where the far end terminal is typically used shall be used. E.g. office rooms for all office types of telephones, a car cabin for hands-free telephones in vehicles, etc. The terminal is setup as described in clause 5. The ambient noise level shall be less than -64 dBPa(A) for handset and headset terminals, less than -70 dBPa(A) for hands-free type terminals. The attenuation from electrical reference point input to electrical reference point output shall be measured using a speech like test signal.

b) Before the actual test a training sequence consisting of 10 s artificial voice male and 10 s artificial voice female according to ITU-T Recommendation P.50 [17] is altered. The training sequence level shall be -16 dBm0 in order not to overload the codec.

c) The test signal is a PN-sequence complying with ITU-T Recommendation P.501 [18] with a length of 4 096 points b (for the 48 kHz sampling rate) and a crest factor of 6 dB. The duration of the test signal is 250 ms. The test signal level is -3 dBm0. The low-crest factor is achieved by random-alternation of the phase between -180° and 180°.

d) Le is calculated according to ITU-T Recommendation G.111 [8], annex A, formula A 4-7 (m = 1). 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 length of the test signal shall be at least one second (1,0 s). For the measurement a time window has to be applied adapted to the duration of the actual test signal (200 ms).

NOTE 2: It should be noted that Le may be highly time variant depending on the type of terminal and the performance of the echo cancellers involved in the connections. Therefore in ITU-T Recommendation G.168 [11] various tests are described which allow the testing of the echo canceller performance parameters in various conditions. Instead of the echo path simulations described in ITU-T Recommendation G.168 [11] the real echo path (terminal) used in the connection should be tested.

NOTE 3: Time varying echo loss as well as echo loss variations e.g. in double talk conditions may be measured but are not taken into account in the E-model yet.

In end-to-end scenarios the TELR typically can not be assessed directly. If however the sidetone-path of the near end terminal and the acoustical coupling of the test setup can be separated from the echo signal the measurement is possible. In this case it is required that the delay between acoustical coupling and the echo signal is long enough to apply a time window to the echo signal. In addition the frequency responses and SLR and RLR of the near end terminal must be known. The procedure now is as follows:

a) For conducting the tests the typical environments where the far end terminal is typically used shall be used. E.g. office rooms for all office types of telephones, a car cabin for hands-free telephones in vehicles. The room where the near end terminal is setup should be ideally anechoic. The terminal is setup as described in clause 5. The ambient noise level shall be less than -64 dBPa(A) for handset and headset terminals, less than -70 dBPa(A) for hands-free type terminals in both rooms. The attenuation from mouth-to-ear shall be measured using a speech like test signal.

b) Before the actual test a training sequence consisting of 10 s artificial voice male and 10 s artificial voice female according to ITU-T Recommendation P.50 [17] is altered. The training sequence level shall be -4,7 dBPa.

c) The test signal is a PN-sequence complying with ITU-T Recommendation P.501 [18] with a repetition period > than the expected delay of the echo signal. and a crest factor of 6 dB. The duration of the test signal is adapted to the expected echo signal. The test signal level is +5 dBPa. The low-crest factor is achieved by random-alternation of the phase between -180° and 180°.

d) The impulse response of the measured signal is calculated by inverse Fourier transformation. The echo impulse response is cut off in the time domain and used for the further calculations.

e) From the impulse response the echo spectrum is generated by Fourier transformation. This spectrum is corrected by the sending and receiving frequency responses of the near end terminal. The corrected spectrum is used for Le calculation.

f) Le is calculated according to ITU-T Recommendation G.111 [8], annex A, formula A 4-7 (m = 1). 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 length of the test signal shall be at least one second (1,0 s). For the measurement a time window has to be applied adapted to the duration of the actual test signal.

NOTE 4: It should be noted that for low level echo signals the method is not applicable due to insufficient signal to noise ratio.

7.13 Weighted echo path loss

NOTE 1: Wideband calculation is for further study, provisionally the measurement is based on narrowband.

WEPL is the frequency weighted sum of all losses and gains in inserted in a network. The weighting is defined in ITU-T Recommendation G.122 [9]. It includes TCLw and any transhybrid loss and such cannot be measured directly. Further information can be found in EG 201 050 [i.3].

NOTE 2: It should be noted that WEPL may be highly time variant depending on the type of terminal and the performance of the echo cancellers involved in the connections.

NOTE 3: Time varying echo loss as well as echo loss variations e.g. in double talk conditions may be measured but are not taken into account in the E-model yet.

7.14 Distortion

7.14.1 Distortion in sending

a) The handset terminal is setup as described in clause 5. The handset is mounted either in the LRGP or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report.

The hands-free terminal setup is described in clause 5. In addition to the MRP calibration, the broadband signal level is adjusted to -28,7 dBPa at the HFRP and the spectrum is not altered.

b) The type of test and test signals depend on the type of signal processing used and introduced in the system to be tested. For systems introducing highly time variant and non linear signal processing (e.g. all types of low bit rate codecs based on CELP type codecs) the tests described below are not applicable or should be applied with some caution.

- When testing non linear time invariant transmission systems a pure tone measurement can be used. An instrument capable of measuring the harmonic distortion of signals with fundamental frequencies in the range of 315 Hz to 1 kHz shall be connected to the artificial ear. A pure-tone signal in the range of -45 dBPa to -4,7 dBPa shall be applied at the MRP at frequencies of 315 Hz, 502 Hz and 1 004 kHz. For wideband systems 2008 Hz is used in addition. Alternatively the level is adjusted until the output of the terminal is -10 dBm0. The level of the signal at the MRP is then the ARL. The level of this signal is in the range of -35 dB to +10 dB rel. to the ARL. The ratio of the signal to total distortion power at the electrical interface shall be measured with the psophometric noise weighting (see ITU-T Recommendations G.712 [12] and O.132 [14]).

- Alternatively a band-limited noise signal corresponding to ITU-T Recommendation O.131 [13] can be used. The test signal shall be applied at the MRP. The level of this signal is in the range of -45 dBPa to -4,7 dBPa. Alternatively the level is adjusted until the output of the terminal is -10 dBm0. The level of the signal at the MRP is then the ARL. The level of this signal is in the range of -45 dB to +7 dB rel. to the ARL. The ratio of signal to total distortion power at the electrical interface shall be measured (see ITU-T Recommendations G.712 [12], annex A and O.131 [13]).

- For systems requiring speech like test stimuli a test-signal which is more "speech-like" e.g. an AM-FM modulated sinewave composed signal having a fundamental frequency similar to speech and the typical speech like harmonics can be used. Limit the spectrum to 1 kHz and measure the total distortion in the frequency range from 1 kHz to 3,4 kHz. The test signal is defined as:

$$s(t) = \sum_i [[A + \mu_{AM} \cos(2\pi t n \times f_{AM})] \times \cos[(2\pi t \times f_{0i}) + \mu_{FM} \times \sin(2\pi t \times f_{FM})]]$$

with A = 0,5

f_{AM} = 3 Hz, μ_{AM} = 0,5

f_{FM} = 5 Hz, μ_{FM} = 1

$F_{0\,i} = i \times 240$ Hz ;i = 1..4

The spectrum should be shaped using a shaping filter as described in table 3 of ITU-T Recommendation P.501 [18] which provides a slope of 5 dB/octet.

The test signal level is adjusted to -4,7 dBPa at the MRP.

The total energy of the distortion components is measured in the frequency range from 1 kHz to 4 kHz at the electrical interface.

NOTE: In order to ensure a reliable activation of the configuration measured, an activation signal may be generated before the actual measurement starts. The activation signal may consist of a sequence of 4 Composite Source Signals (CSS) according to ITU-T Recommendation P.501 [18]. Alternatively artificial voice according to ITU-T Recommendation P.50 [17] can be used. The level of the activation signal is -4,7 dBPa, measured at the MRP. The level of the activation signal is averaged over the complete activation sequence signal.

7.14.2 Distortion in receiving

a) The handset terminal is setup as described in clause 5. The handset is mounted either in the LRGP or in the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report.

The hands-free terminal is setup as described in clause 5. The HATS is diffuse field equalized as described in ITU-T Recommendation P.581 [23].

b) The type of test and test signals depend on the type of signal processing used and introduced in the system to be tested. For systems introducing highly time variant and non linear signal processing (e.g. all types of low bit rate codecs based on CELP type codecs) the tests described below are not applicable or should be applied with some caution:

- When testing non linear time invariant transmission systems a pure tone measurement can be used. An instrument capable of measuring the harmonic distortion of signals with fundamental frequencies in the range of 315 Hz to 1 kHz shall be connected to the electrical interface. A pure-tone signal in the range of -45 dBm0 to 0 dBm0 shall be applied at the electrical interface at frequencies of 315 Hz, 502 Hz and 1 004 kHz. For wideband systems 2008 Hz is used in addition.The ratio of the signal to total distortion power at the artificial ear shall be measured with the psophometric noise weighting (see ITU-T Recommendations G.712 [12] and O.132 [14]).

- Alternatively a band-limited noise signal corresponding to ITU-T Recommendation O.131 [13] can be used. The test signal shall be applied at the MRP. The level of this signal is in the range of -45 dBPa to -4,7 dBPa. The ratio of the signal-to-total distortion power shall be measured with the psophometric noise weighting in the artificial ear (see ITU-T Recommendations G.712 [12] and O.132 [14]).

- For systems requiring speech like test stimuli a test-signal which is more "speech-like" e.g. an AM-FM modulated sinewave composed signal having a fundamental frequency similar to speech and the typical speech like harmonics, limit the spectrum to 1 kHz and measure the total distortion in the frequency range from 1 kHz to 3,4 kHz. The test signal is defined as:

$$s(t) = \sum_i [[A + \mu_{AM} \cos(2\pi t n \times f_{AM})] \times \cos[(2\pi t \times f_{0i}) + \mu_{FM} \times \sin(2\pi t \times f_{FM})]]$$

with A = 0,5

f_{AM} = 3 Hz, μ_{AM} = 0,5

f_{FM} = 5 Hz, μ_{FM} = 1

$F_{0\,i} = i \times 240$ Hz ;i = 1..4

The spectrum should be shaped using a shaping filter as described in table 3 of ITU-T Recommendation P.501 [18] which provides a slope of 5 dB/octet.

The test signal level is adjusted to -16 dBm0 at the electrical interface.

The total energy of the distortion components is measured in the frequency range from 1 kHz to 4 kHz in the artificial ear.

NOTE: In order to ensure a reliable activation of the configuration measured, an activation signal may be generated before the actual measurement starts. The activation signal consists of a sequence of 4 Composite Source Signals (CSS) according to ITU-T Recommendation P.501 [18]. Alternatively artificial voice according to ITU-T Recommendation P.50 [17] can be used. The level of the activation level is -16 dBm0, measured at the electrical interface. The level of the activation signal is averaged over the complete activation sequence signal.

7.14.3 Overall distortion

a) Handset terminals are setup as described in clause 5. The handsets are mounted either in the LRGP or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report.

Hands-free terminals are setup as described in clause 5. In addition to the MRP calibration, the broadband signal level is adjusted to -28,7 dBPa at the HFRP and the spectrum is not altered. The HATS is diffuse field equalized as described in ITU-T Recommendation P.581 [23]. The equalized output signal of each artificial ear is power-averaged for the total time of analysis; either the "right" or the "left" signals are used measurements and calculations.

b) The type of test and test signals depend on the type of signal processing used and introduced in the system to be tested. For systems introducing highly time variant and non linear signal processing (e.g. all types of low bit rate codecs based on CELP type codecs) the tests described below are not applicable or should be applied with some caution.

- When testing non linear time invariant transmission systems a pure tone measurement can be used. A pure-tone signal in the range of -45 dBPa to -4,7 dBPa shall be applied at the MRP at frequencies of 315 Hz, 502 Hz and 1 004 kHz. For wideband systems 2008 Hz is used in addition. The ratio of the signal to total distortion power at the artificial ear shall be measured with the psophometric noise weighting (see ITU-T Recommendations G.712 [12] and O.132 [14]).

- Alternatively a band-limited noise signal corresponding to ITU-T Recommendation O.131 [13] can be used. The test signal shall be applied at the MRP. The level of this signal is in the range of -55 dBm0 to -3 dBm0. The ratio of the signal-to-total distortion power shall be measured with the psophometric noise weighting in the artificial ear (see ITU-T Recommendations G.712 [12] and O.132 [14]).

- For systems requiring speech like test stimuli a test-signal which is more "speech-like" e.g. an AM-FM modulated sinewave composed signal having a fundamental frequency similar to speech and the typical speech like harmonics, limit the spectrum to 1 kHz and measure the total distortion in the frequency range from 1 kHz to 3,4 kHz. The test signal is defined as:

$$s(t) = \sum_i [[A + \mu_{AM} \cos(2\pi t n \times f_{AM})] \times \cos[(2\pi t \times f_{0i}) + \mu_{FM} \times \sin(2\pi t \times f_{FM})]]$$

with A = 0,5

f_{AM} = 3 Hz, μ_{AM} = 0,5

f_{FM} = 5 Hz, μ_{FM} = 1

$F_{0\,i}$ = i × 240 Hz ;i = 1..4

The spectrum should be shaped using a shaping filter as described in table 3 of ITU-T Recommendation P.501 [18] which provides a slope of 5 dB/octet.

The test signal level is adjusted to -4,7 dBPa at the mouth reference point.

The total energy of the distortion components is measured in the frequency range from 1 kHz to 4 kHz in the artificial ear.

NOTE: In order to ensure a reliable activation of the configuration measured, an activation signal may be generated before the actual measurement starts. The activation signal consists of a sequence of 4 Composite Source Signals (CSS) according to ITU-T Recommendation P.501 [18]. Alternatively artificial voice according to ITU-T Recommendation P.50 [17] can be used. The level of the activation level is -4,7 dBPa, measured at the MRP. The level of the activation signal is averaged over the complete activation sequence signal.

7.15 Sensitivity against out-of-band signals in sending

NOTE 1: Typically this measurement is applied only in narrowband system. If out of band measurements for wideband systems are required, the principle is kept but no frequencies higher than 10 kHz are produced by the artifial mouth since this would be out of the mouth specification.

a) The handset terminal is setup as described in clause 5. The handset is mounted either in the LRGP or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report.

The hands-free terminal setup is described in clause 5. In addition to the MRP calibration, the broadband signal level is adjusted to -28,7 dBPa at the HFRP and the spectrum is not altered.

b) The type of test and test signals depend on the type of signal processing used and introduced in the system to be tested. For systems introducing highly time variant and non linear signal processing (e.g. all types of low bit rate codecs based on CELP type codecs) the tests described below are not applicable or should be applied with some caution.

- When testing non linear time invariant transmission systems a pure tone measurement can be used. For input signals at frequencies of 4,65 kHz, 5 kHz, 6 kHz, 6,5 kHz, 7 kHz and 7,5 kHz at the level specified of -4,7 dBPa, the level of any image frequencies at the electrical interface shall be measured.

- Alternatively a white Gaussian noise band-limited to the frequency range between 4,6 kHz and 8 kHz with a level of -4,7 dBPa at the MRP can be used. The total level - measured in a frequency range from 300 Hz to 3,4 kHz is measured at the electrical reference point and shall be less than 40 dB referred to the reference level. The reference level is determined using artificial voice according to ITU-T Recommendation P.50 [17], band-limited to the frequency range between 300 Hz and 3,4 kHz with a level of -4,7 dBPa at the MRP. For this signal the in-band level averaged over the complete reference signal length is determined at the electrical reference point.

NOTE 2: In order to ensure a reliable activation of the terminal an activation signal may be generated before the actual measurement starts. The activation signal consists of a sequence of 4 Composite Source Signals (CSS) according to ITU-T Recommendation P.501 [18]. The level of the activation level is -4,7 dBPa, measured at the MRP. The level of the activation signal is averaged over the complete activation sequence signal.

NOTE 3: Appropriate limits can be found e.g. in TBR 008 [i.4].

7.16 Spurious out-of-band signals in receiving

NOTE 1: Typically this measurement is applied only in narrowband system. If out of band measurements for wideband systems are required, the principle is kept but no frequencies higher than 12 kHz are produced by the artifial ear since this would be out of the HATS specification.

a) The handset terminal is setup as described in clause 5. The handset is mounted either in the LRGP or in the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report.

The hands-free terminal is setup as described in clause 5. The HATS is diffuse field equalized as described in ITU-T Recommendation P.581 [23].

b) The type of test and test signals depend on the type of signal processing used and introduced in the system to be tested. For systems introducing highly time variant and non linear signal processing (e.g. all types of low bit rate codecs based on CELP type codecs) the tests described below are not applicable or should be applied with some caution.

- For input signals at the frequencies 500 Hz, 1 000 Hz, 2 000 Hz and 3 150 Hz applied at the level of -10 dBm0, the level of spurious out-of-band image signals at frequencies of up to 8 kHz shall be measured selectively in the artificial ear.

- Alternatively the test signal used is artificial voice according to ITU-T Recommendation P.50 [17], band-limited in a frequency range between 300 Hz and 3,4 kHz with a level of -12 dBm0 in receiving direction. The level of the out-of-band signal is measured in a frequency range between 4,6 kHz and 8 kHz at the DRP and shall be at least 45 dB below the level of the reference signal. The level of the reference signal is determined by measuring the acoustical level of the in-band signal at the artificial ear. The in-band signal is the artificial voice signal band-limited between 300 Hz and 3,4 kHz applied with a level of -12 dBm0.

NOTE 2: In order to ensure a reliable activation of the terminal an activation signal may be generated before the actual measurement starts. The activation signal consists of a sequence of 4 Composite Source Signals (CSS) according to ITU-T Recommendation P.501 [18]. The level of the activation level is -16 dBm0, measured at the electrical interface. The level of the activation signal is averaged over the complete activation sequence signal.

NOTE 3: Appropriate limits can be found e.g. in TBR 008 [i.4].

8 Advanced measurement procedures, taking into account the conversational situation

All measurements and parameters listed in clause 7assume that the systems under test are broadly linear and time invariant (LTI-systems). For LTI-systems, the single figure measures of clause 7 can be used effectively to define speech quality performance. In cases where the signal processing in the components cannot be assumed to be linear and time invariant (except the codec), signal processing may influence the speech transmission quite substantially, especially in the conversational situation. The signal processing procedures to be expected are voice activated switching and amplification, echo cancellation (acoustic and electric), noise reduction, etc. The importance of double talk performance and background noise transmission was derived by conversational tests and investigated more in detail by using specific double talk tests and listening only tests.

The signal processing components expected in non-LTI (non-linear time variant) systems are found e.g. in small (mobile) terminals, hands-free terminals, echocancellers, packetizing equipment.

Table 1 gives an overview of the relevant subjective and objective parameters, which in addition to the parameters defined in clause 7, influence the speech transmission quality.

Table 1: Subjectively relevant parameters and their correlating objective parameters

Subjectively relevant parameter	Description	Correlating objective parameter
Delay and echo loss	One way transmission time caused by signal processing, packetizing and transmission	- delay - (time variant) echo loss - TCL - Switching characteristics
Quality of background noise transmission	Transmission effect experienced in the send direction during - idle mode - only far end speech active - only near end speech active	- attenuation range - attenuation in send direction - switching characteristics - minimum activation level in send direction - frequency response - EC design of NLP or centre clippers - design of noise reduction systems - sensitivity of background noise detection (activation level, absolute level, level fluctuations)
Double talk performance	Typically in send and receive direction: - a loudness variation between single and double talk periods - loudness variation during double talk - echo disturbances - occurrence of speech gaps	- attenuation range - attenuation in send and receive direction during double talk - switching characteristics - minimum activation level to switch over from receive to send direction and from send to receive direction - echo attenuation - spectral and time dependent echo characteristics - design of NLP or centre clippers in conjunction with ECs
Echo disturbances under single talk conditions	Measured between receive and send direction	- echo level - echo level fluctuation vs. time - spectral echo attenuation
Speech sound quality	In send and receive direction	- frequency responses - distortions
Loudness	In send and receive direction	- loudness ratings in send and receive
Noise	In send and receive direction	- noise level - level fluctuations - spectral characteristics

NOTE 1: The behaviour of users during subjective tests clearly demonstrates that the individual speech levels on both sides of the connection highly influence the transmission performance of a system under test. Consequently the measurement levels should be adapted in an appropriate way to represent the possible level variations at the "receive" and "send" inputs.

NOTE 2: Room characteristics highly influence the perceived transmission quality (see ITU-T COM12-42 [33]). This demands that terminals be tested in an appropriate environment.

NOTE 3: Recent subjective testing has found that echo levels during double talk may influence the speech quality more than previously expected. Echo loss requirements during double talk can be found in the appendix to ITU-T Recommendations G.131 [10] and P.340 [15].

8.1 Measurement setup for objective tests

The measurement setup should be chosen as described in clause 5. In addition some changes are suggested in order to improve the accuracy of the measurements:

- A possible delay in sending or receiving of terminals and associated network components and the acoustical propagation delay between artificial mouth and the HFT microphone must be taken into account. This is especially important for analyses, where the measured signals are referred to the original test signals. These analyses require an exact synchronization of both signals. Therefore the delay has to be compensated.

- When analyzing hands-free type telephones, in addition a microphone should be positioned very close to the HFT loudspeaker in order to record the RCV signal. This recorded signal can easier be distinguished from the signal introduced by the artificial mouth under double talk measurement conditions.

8.2 Practical realization of test signals

The signals to be used for the advanced measurements are specifically adapted to the measurement and can be found in ITU-T Recommendations P.501 [18], P.50 [17] and P.59 [35]. The tests described in the following are suggested to determine additional parameters as given in table 1 and are described in ITU-T Recommendation P.502 [19]. In cases where background noise simulation is required, care should be taken when choosing the type of background noise. Time and frequency characteristics should be chosen to simulate the typical environment the system is supposed to operate, e.g. car type noise for HFTs in car type environments. In case of specific acoustic preprocessing provided by the terminal, the spatial characteristics of the background noise may be important. In order to simplify the description, long term power density spectrum and long term level density during the measurement should be noted.

8.3 Quality of background noise transmission

In general the simulation of background noise can be a continuous noise signal (with shaped spectrum), however a more sophisticated signal is recommended to represent realistic conditions (e.g. office voice babble). In such cases the background noise should be characterized by its long term power density spectrum and its average level applied during the measurement. Suitable background noise signals can be found in [i.6].

For the following tests the background noise signal is regarded as the measurement signal and not as a disturbing component. Consequently analyses should be applied to the noise signal. The transmission quality of background noise (from the near end in SND direction) can be evaluated:

- at idle mode;

- with far end speech;

- with near end speech.

In all these cases important parameters are:

- the sensitivity of background noise detection in terms of activation level;

- the absolute level of the transmitted signal;

- level fluctuations.

8.3.1 Test setup for background noise transmission tests

Besides the test setups for stationary background noise as found in the individual standards (e.g. in [i.4]) a setup for simulating realistic background noises in a lab-type is described in [i.6].

The general procedure for setup a background noise simulation arrangement is described in [i.6]. The present document 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 10.

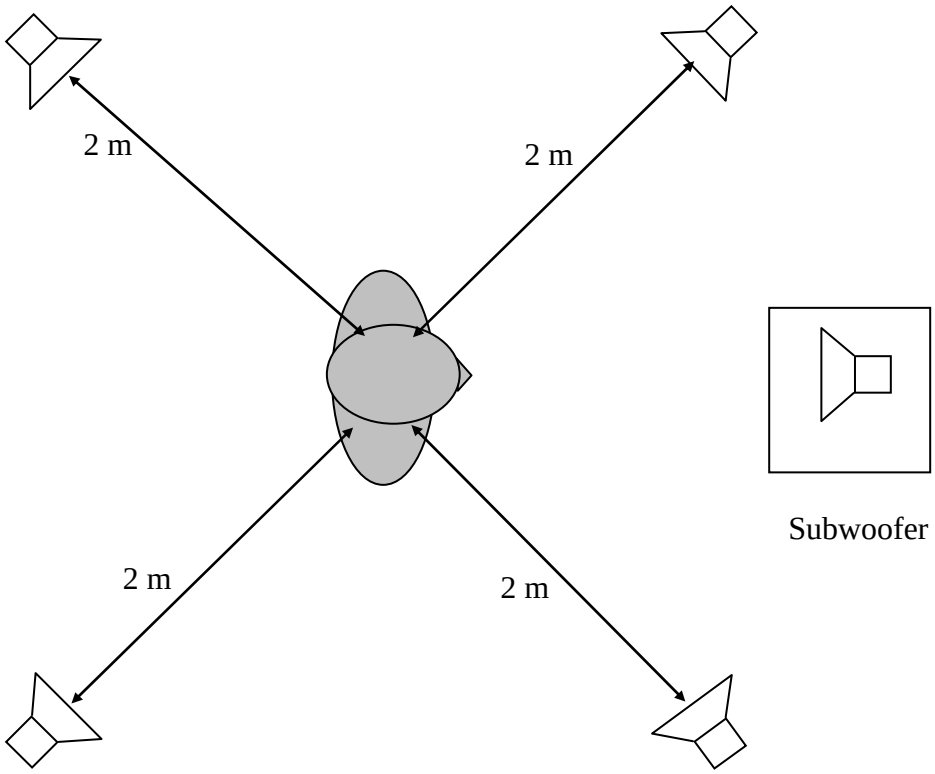


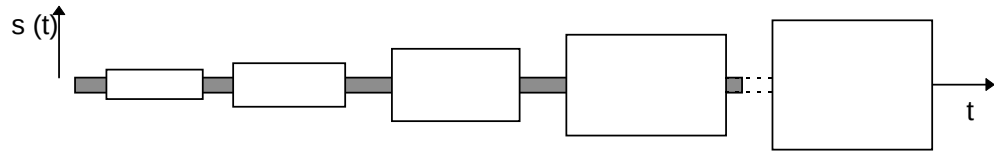
Figure 10: Structure of test signal for attenuation range measurement

The equalization and calibration procedure for the setup is described in detail in [i.6].

Whenever a performance evaluation in realistic background noise scenarios is required this setup in combination with the different analysis methods as described below can be used.

8.3.2 Background noise transmission with far end speech

The following signal structure can be used to evaluate the quality of background noise transmission in SND direction coincident **with far end speech**. Exemplary the following figure 10a represents a signal with a continuous noise signal applied at the near end (SND direction, grey colour) and a simulation of far end speech in RCV direction (white colour, bursts of CSS can be used). The measurement is carried out in SND direction. In figure 10a the level of the CSS bursts vary and the simulation of background noise is applied with a constant level.



NOTE: The dotted line indicates the repetition or elongation of the test signal to achieve the suitable length for the measurement.

Figure 10a: Example of a test signal structure to evaluate the quality of background noise transmission in SND direction (with far end speech simulation)

The test is carried out applying the Composite Source Signal in receiving direction. After the end of the last Composite Source Signal burst (representing the end of far end speech simulation) the signal level in sending direction is measured (and typically should not vary by more than 10 dB (during transition to transmission of background noise without far end speech)).

a) The handset terminal is setup as described in clause 5. The handset is mounted either in the LRGP or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report.

   The hands-free terminal setup is described in clause 5.

b) According to the specification of the manufacturer/test lab the background noise is played back. The test should be carried out using a mostly constant background noise (e.g. constant driving condition in a car). The background noise should be applied for at least 5 seconds in order to adapt noise reduction algorithms.

c) A Composite Source Signal as shown in figure 10a is applied as test signal in receiving direction with a duration of ≥ 2 CSS periods. The test signal level is -16 dBm0 at the electrical reference point. If the dynamic behaviour should be investigated a signal with rising level as indicated in figure 10a can be used.

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

e) The signal level variation in sending direction is determined immediately after the end of the last CSS burst in receiving direction until the transmitted background noise signal in sending direction reached the maximum level. The resulting level difference corresponds to the signal level variation in sending direction.

In addition to the level analysis as described above also a spectral analysis or a hearing model based signal analysis can be used. One possibility is the so called "Relative Approach" [42]. The "Relative Approach" is a single ended method which does not need a reference signal and is purely based on the analysis of the transmitted background noise signal. The algorithm takes into account the sensitivity of the human ear on signal fluctuation in the time domain as well as on significant spectral structures. It is recognized that slow variations of a signal in time and/or frequency are typically not disturbing. The algorithm is based on foreword estimation using the signal history. The new signal values are predicted and compared to the actual acquired signal. The basis for the new procedure is the hearing adequate spectral representation of the time and frequency domain. Therefore a time frequency analysis is required. Like in ordinary hearing models the nonlinear relationship between sound pressure level and loudness perceived subjectively is taken into account by time-frequency warping using a Bark filter bank and proper integration of the individual outputs. The filter bank is realized in the time domain. The output signals of the filter bank are rectified and integrated, thus the envelope is generated. This three-dimensional output of the hearing model is the basis of the "Relative Approach".

After spectral analysis and nonlinear transformation the foreword estimation of new signal value is made and compared to the actual signal. The difference can be colour coded and represented as new type of spectrography: bright colours represent big differences between estimation and actual signal, dark colours indicate low differences.

8.3.3 Background noise transmission with near end speech

A similar signal structure can be used to determine the quality of background noise transmission coincident **with near end speech**. In this case the background noise and the speech signal simulation (again CSS can be used) are applied and measured on the same direction (in opposite to the upper figure 10a), e.g. the SND direction.

The test is carried out applying a Composite Source Signal in sending direction. After the end of the last Composite Source Signal burst (representing the end of near end speech simulation) the signal level in sending direction is measured (and typically should not vary more than 10 dB (during transition to transmission of background noise without near end speech).

a) The handset terminal is setup as described in clause 5. The handset is mounted either in the LRGP or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report.

   The hands-free terminal setup is described in clause 5. In addition to the MRP calibration, the broadband signal level is adjusted to -28,7 dBPa at the HFRP and the spectrum is not altered.

b) According to the specification of the manufacturer/test lab the background noise is played back. The test should be carried out using a mostly constant background noise (e.g. constant driving condition in a car). The background noise should be applied for at least 5 s in order to adapt noise reduction algorithms.

c) The near end speech is simulated using the Composite Source Signal according to ITU-T Recommendation P.501 [18] with a duration of ≥ 2 CSS periods (see figure 10a). The test signal level is -4,7 dBPa at the MRP.

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

e) The signal level variation in sending direction is determined immediately after the end of the last CSS burst in receiving direction until the transmitted background noise signal in sending direction reached the maximum level. The resulting level difference corresponds to the signal level variation in sending direction.

In addition to the level analysis as described above also a spectral analysis or a hearing model based signal analysis (e.g. "Relative Approach") can be used.

8.3.4 Speech transmission quality with near end background noise

Speech Quality for wideband systems can be tested based on EG 202 396-3 [i.7]. The test method is applicable for narrowband (100 Hz to 4 kHz) and wideband (100 Hz to 8 kHz) transmission systems. The test method described leads to three MOS-LQO quality numbers:

N-MOS-LQO: Transmission quality of the background noise

S-MOS-LQO: Transmission quality of the speech

G-MOS-LQO: Overall transmission quality

The requirements for the individual numbers should be set in the standards specifying the terminal performance characteristics in detail.

The test is carried out applying a speech signal consisting of 8 sentences spoken by two male and two female speakers in sending direction. The preferred language is French. Appropriate test sentences can be found in ITU-T Recommendation P.501 [18]. Background noise sequences are available in EG 202 396-1 [i.6]. The background noises chosen for the tests should be defined in the standards specifying the terminal performance characteristics in detail.

a) The handset terminal is setup as described in clause 5. The background noise setup is defined in clause 8.3.1. The handset is mounted either in the LRGP or the HATS position (see ITU-T Recommendation P.64 [24]). The application force used to apply the handset against the artificial ear is noted in the test report.

   The hands-free terminal setup is described in clause 5. In addition to the MRP calibration, the broadband signal level is adjusted to -28,7 dBPa at the HFRP and the spectrum is not altered. The background noise setup is defined in clause 8.3.1.

b) According to the specification of the manufacturer/test lab the background noise is played back. The background noise should be applied for at least 5 s in order to adapt noise reduction algorithms.

c) 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 ITU-T Recommendation P.501 [18]. The preferred language is French since the objective method was validated with French language. The test signal level is -4,7 dBPa at the MRP.

d) Three signals are required for the tests:

1) The clean speech signal is used as the undisturbed reference (see [i.7]).

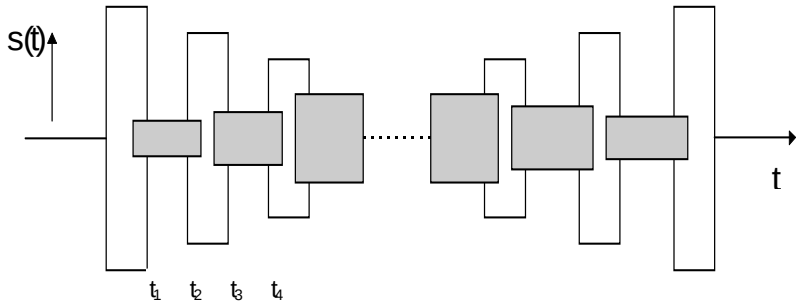
2) The speech plus undisturbed background noise signal is recorded at the terminal's microphone position using an omni directional measurement microphone with a linear frequency response between 50 Hz and 12 kHz.

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

e) N-MOS-LQO, S-MOS LQO and G-MOS LQO are calculated as described in EG 202 396-3 [i.7].

8.4 Double talk performance

Figure 11 demonstrates the structure of a test signal, based on periodically repeated Composite Source Signals in SND and RCV direction, which can be used to determine important parameters during periods of double talk. The signal and examples for applications are described in ITU-T Recommendation P.502 [19].



NOTE: The dotted line indicates the repetition or elongation of the test signal to achieve the suitable length for the measurement.

Figure 11: Structure of test signal for double talk evaluation

The signal represents an ongoing double talk period. Periods of the CSS are fed to the RCV input port (grey colour) and via the artificial mouth to the HFT microphone (white colour). The measurement level of the RCV and SND signals varies for 0,5 dB between two CSS periods. In addition the level distribution between the SND and RCV direction spreads over a wide range. Typical signal parameters can be chosen as follows.

Table 2

	active duration / pause duration	**highest signal level (active part)**	**lowest signal level (active part)**	**level of the first signal period**
CSS in **SND direction**	248,62 ms / 151,38 ms	-3 dBPa	-23 dBPa	-3 dBPa
CSS in **RCV direction**	248,62 ms / 151,38 ms	-10 dBm0	-29,5 dBm0	-29,5 dBm0
NOTE: All levels measured in dBV are referred to the 0 dBm0 reference point.				

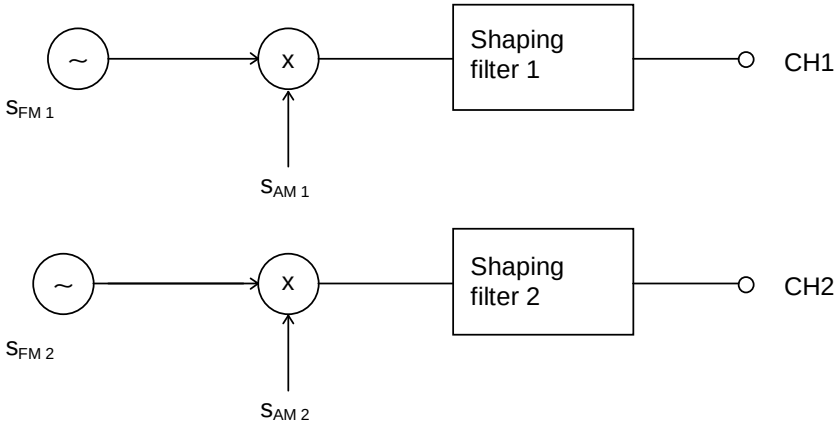
The SND and RCV signal periods overlap for 48,62 ms, which corresponds to the duration of the voiced sound of the CSS. The signals for SND and RCV direction should be uncorrelated. The complete duration amounts to 32 s.

The following parameters can be determined during double talk:

- attenuation range;
- attenuation in SND/RCV direction;
- switching characteristics, e.g. switching times;
- minimum activation level to switch over from RCV to SND direction and from SND to RCV direction;

- frequency responses in SND and RCV direction;
- loudness ratings in SND and RCV direction;
- echo attenuation;
- design of NLP or centre clippers in conjunction with ECs.

The evaluation of **echo during double talk** (including speech activity in SND direction from the near end speaker) requires the separation of echo signal and near end signal. A possible construction of a test signal is shown in figure 10.



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

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

Figure 12: Two channel test signal generation for double talk evaluations based on AM-FM signals

Typical settings are given in table 3.

Table 3

	f_{AM}	f_{FM}	F_0	shaping filter
Channel 1 (CH 1)	f_{AM1} = 3 Hz	f_{FM1} = 5 Hz	F_{01} = 270 Hz	LP, 5 dB/octet
Channel 2 (CH 2)	f_{AM2} = 3 Hz	f_{FM2} = 5 Hz	F_{02} = 290 Hz	LP, 5 dB/octet

Average measurement levels can be chosen for the signals, i.e. -4,7 dBPa (SND) and -20 dB_V (RCV). The test signal may be embedded in speech or speech like sequences. The measured signal in SND direction is filtered in order to eliminate the near end signal component.

The parameter, which can be determined, is:

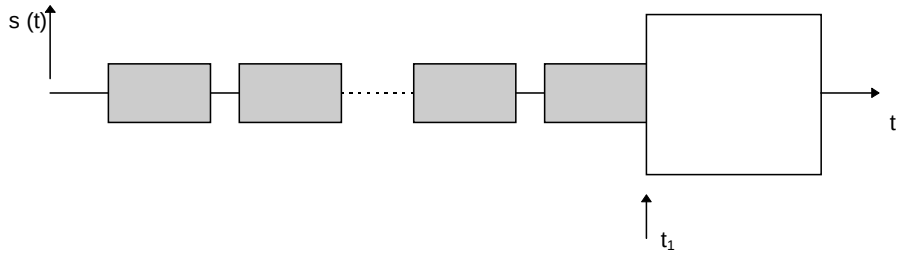
- the echo level during double talk.

8.5 Switching characteristics

For these subjective and the correlating objective parameters measurement signals according to ITU-T Recommendation P.501 [18] can be used.

NOTE: The switching characteristics influence the transmission quality in various ways as given by table 1 in ITU-T Recommendation P 501 [18] in detail (see column: correlating objective parameters).

One of the most important parameters, especially for implementations with level switching devices is the **attenuation range**. This parameter can be determined using a test signal structure shown in figure 13.



NOTE: The dotted line indicates the repetition or elongation of the test signal to achieve the suitable length for the measurement.

Figure 13: Structure of test signal for attenuation range measurement

A periodical repetition of CSS bursts as a simulation of speech is used to activate one transmission path (grey colour). At the end of one burst, indicated by t_1 on the time scale, the measurement signal is applied in the opposite path (white colour). This signal part consists of a periodical repetition of a voiced sound.

Typical settings are as follows:

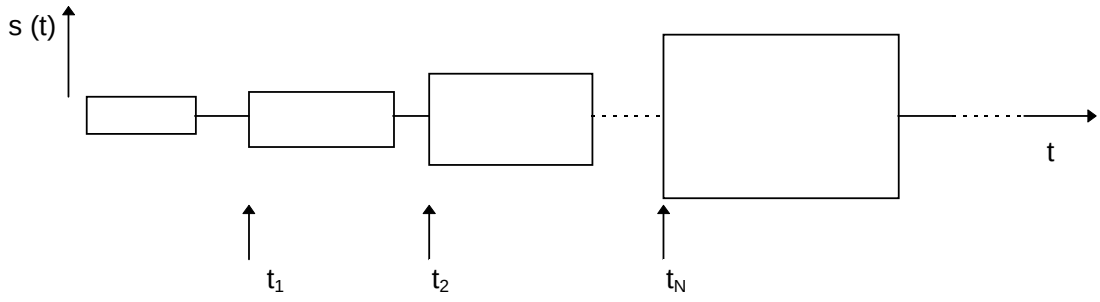
Table 4

	measurement signal	**measurement signal level**	**activation signal (in opposite direction)**	**level of the activation signal (in opposite direction)**
Switching **RCV → SND**	voiced sound in SND direction, period. repetition	-3 dBPa	CSS in RCV	-21,7 dBm0 (including pauses)
Switching **SND → RCV**	voiced sound in RCV direction, period repetition	-20 dBm0	CSS in SND	-4,7 dBPa (including pauses)

The following parameters can be measured:

- attenuation range;
- switching characteristics (for speech like signals), e.g. switching times.

The signal structure shown in figure 14 represents signal parts with increasing levels. The **minimum activation level** to switch on the RCV or SND direction from idle mode can be determined using these sequences. Periods of the CSS (as a simulation of speech) with increasing levels are suited for this signal.



NOTE: The dotted line indicates the repetition or elongation of the test signal to achieve the suitable length for the measurement.

Figure 14: Structure of test signal to determine the minimum activation level

Typical settings can be chosen as follows:

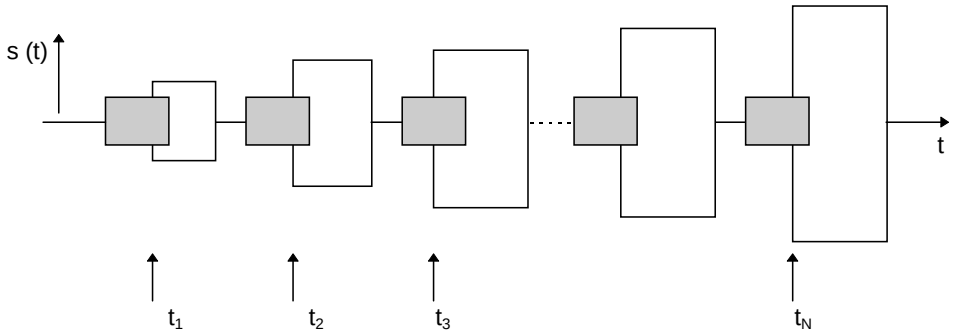
Table 5

	active duration / pause duration	level of the first period	level difference between two periods
CSS for switching in **SND direction**	248,62 ms / 451,38 ms	-23 dBPa (see note)	1 dB
CSS for switching in **RCV direction**	248,62 ms / 451,38 ms	-40 dBm0 (see note)	1 dB
NOTE: These levels should be sufficiently low, to ensure that a wide level range is measured.			

If the transmitted signals are measured and referred to the original measurement signal, the minimum activation level can be determined. The activation can be analysed at the beginning of each signal burst (t_1, t_2, ..., t_N). The parameters which can be determined using this signal are:

- the minimum activation level (for speech like signals);
- the switching times (for speech like signals, level dependent).

If the **minimum activation levels to switch over** from RCV to SND direction (or vice versa, i.e. from SND to RCV direction) shall be measured, the given test signals can be used with slight modifications. As shown in figure 15 an additional signal is needed in the opposite transmission direction (grey colour). The level of the measurement signal (white colour) increases again periodically. Periods of the CSS are suited for both signals in figure 15, if the switching characteristics shall be determined applying speech like signals. Again the signals should be chosen to be uncorrelated.



NOTE: The dotted line indicates the repetition or elongation of the test signal to achieve the suitable length for the measurement.

Figure 15: Structure of test signal to determine the minimum activation level to switch over

Suitable settings are given in table 6.

Table 6

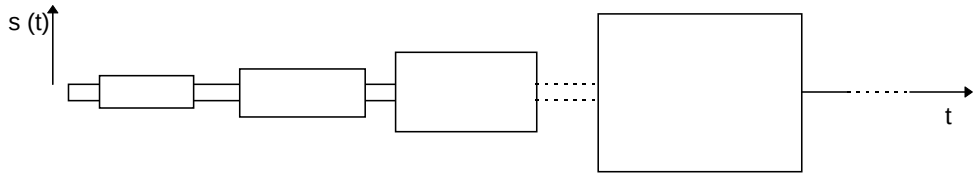
	active duration / pause duration	level of the first period	level difference between two periods	level (active part) in opposite transmission direction
CSS to switch over to **SND direction**	248,62 ms / 451,38 ms	-13 dBPa	1 dB	-20 dBm0 (RCV)
CSS in to switch over to **RCV direction**	248,62 ms / 451,38 ms	-30 dBm0	1 dB	-3 dBPa (SND)

Again the activation can be analyzed at the beginning of the signal bursts (t_1, t_2, ..., t_N).

In addition the same tests can be performed with a simulation of background noise applied at the opposite transmission path. Assessable parameters are:

- the minimum activation level (for speech like signals) to switch over;
- the switching times (for speech like signals).

The signal structure given in figure 16 can be used to determine the **switching characteristics in the presence of background noise**. In this case a speech like signal (CSS) and a background noise simulation are applied simultaneously on the same channel. The parameters for the CSS can be taken from the tables above.



NOTE: The dotted line indicates the repetition or elongation of the test signal to achieve the suitable length for the measurement.

Figure 16: Structure of test signal to determine the minimum activation level in the presence of background noise

The parameters to be determined are:

- the minimum activation level in the presence of background noise;
- the switching characteristics in the presence of background noise.

8.6 Level adjustments by companding or AGC

The following figures 17 and 18 represent test signals, generated by a periodical repetition of voiced sounds. These signals can be used to measure level adjustments for HFTs, which react comparable on the periodical repetition of a voiced sound and real speech. Additionally artificial voice can be used.

The signal given through figure 17 represents an input signal with a constantly changing level, whereas the signal levels are adjusted in certain steps for the signal in figure 18.

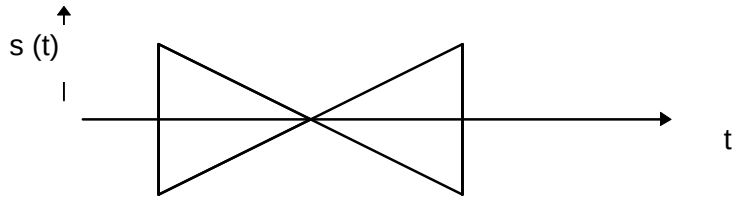


Figure 17: Structure of test signal to determine level adjustments (constantly changing input level)

Suggested parameters for the signal in figure 17 are as follows:

Table 7

	signal generation	highest level	lowest level	level variation
SND direction	voiced sound, period repeated	-3,0 dBPa	infinite	linear
RCV direction	voiced sound, period repeated	-20 dBm0	infinite	linear

The complete signal duration can be chosen to 10 s. The signal is suited to determine the range of level adjustments as a function of the input signal level.

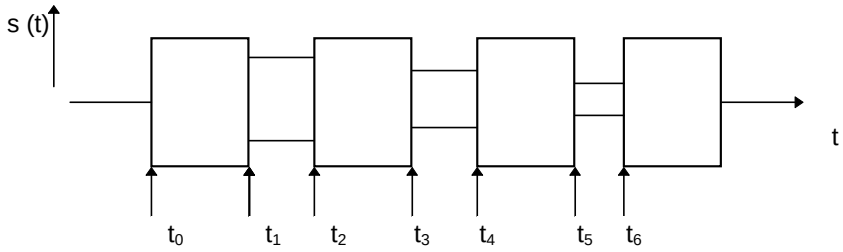


Figure 18: Structure of test signal to determine level adjustments

Suggested parameters for the signal in figure 18 are as follows:

Table 8

	signal generation	**signal level during (t_1 - t_0)**	**signal level during (t_2 - t_1)**	**signal level during (t_4 - t_3)**	**signal level during (t_6 - t_5)**
SND direction	voiced sound, periodically repeated	-3,0 dBPa	-8,0 dBPa	-13,0 dBPa	-18,0 dBPa
RCV direction	voiced sound, periodically repeated	-20 dBm0	-25 dBm0	-30 dBm0	-35 dBm0

The signal duration of the single periods can be chosen to 2,5 s each. The signal is suited to determine especially the time duration for level adjustments.

8.7 Additional echo disturbances

Echo impairments may be generated at various places:

- acoustical echo due to insufficient decoupling between loudspeaker and microphone of a terminal;
- electrical echo due to echoes generated in the network typically hybrid echoes.

The annoyance caused by echoes is independent of the source of echo and therefore echo loss requirements and the measurement setups are typically independent of the type of echo.

The general measurements for TCL and TELR are described in clauses 7.11 and 7.12. Echo measurements in double talk condition are described in clause 8.4.

In addition to these "static" echo measurements the following parameters should be measured in systems with non linear and/or time variant signal processing elements like echocancellers, speech activated/speech controlled attenuation/switching, comfort noise injection:

- Parameters described in ITU-T Recommendation P.340 [15], chapter 10.3 in case acoustic echo cancellation is involved.
- Parameters described in ITU-T Recommendation G.168 [11], chapter 3.4 in case only network echo cancellers are involved.

8.8 Speech sound quality

Speech sound quality is influenced by many parameters as they have been described before:

Frequency response, Loudness Rating, distortion, switching and others.

Since speech sound quality is mostly important during one way transmission, the one way transmission is discussed in this clause.

In the presence of non linear and time variant systems, e.g. when speech coding is used, packet loss occurs and switching is introduced, the "traditional" parameters as mentioned above may no longer give sufficient results. When evaluating only between electrical access points PESQ as described in ITU-T Recommendation P.862 [31] or the methods described in EG 201 377-1 [i.1] may give sufficient information about the impact of the various impairments on speech transmission quality.

When evaluating the complete connection including terminals, ITU-T Recommendation P.862 [31] in its present form cannot be applied. TOSQA 2001 (see EG 201 377-1 [i.1]) as used in the 1st and 2nd Speech Quality Test Event 2002 [i.5] is one method, which may be used in end-to-end scenarios. This method may be replaced by a new ITU-T approved method as soon as the standardization process is finalized.

Annex A (informative): Bibliography

Diedrich and H. W. Gierlich: "Mouth-to-ear Speech Quality: Some General Considerations", ETSI EP TIPHON 10 WG5, TD 78. http://docbox.etsi.org/zArchive/TIPHON/TIPHON/ARCHIVES/1998/05-9810-Tel_Aviv/

ETSI ES 202 740: "Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user".

History

Document history		
V1.1.1	April 2004	Publication as EG 201 377-2
V1.2.1	March 2005	Publication as EG 201 377-2
V1.3.1	November 2005	Publication as EG 201 377-2
V1.4.1	September 2009	Membership Approval Procedure MV 20091127: 2009-09-29 to 2009-11-27