ETSI TS 103 189 V1.1.2 (2013-10)

Technical Specification

**IMS Network Testing (INT);
Specification of end-to-end QoS assessment for VoLTE and RCS Interop Events or Plugtests**

Reference
RTS/INT-00100

Keywords
assessment, QoS, video, voice

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 5
Foreword....5
Introduction 5
1 Scope 6
2 References 6
2.1 Normative references 6
2.2 Informative references....7
3 Abbreviations 7
4 The general structure of end-to-end QoS assessment for VoLTE and RCS Interop Events or Plugtests 9
4.1 Tests cases described in TS 186 011-2....10
4.1.1 User-initiated VoIP call setup and release (TS 186 011-2, clause 4.4.2)....10
4.1.2 Addition of media stream (TS 186 011-2, clause 4.4.11) 11
4.1.3 Ad-hoc Conferencing service (TS 186 011-2, clause 4.4.13) 11
4.1.4 IPTV service (TS 186 011-2, clause 4.4.14)....11
4.1.5 IMS-to-PSTN call (TS 186 011-2, clause 4.4.16.1)....12
4.1.6 PSTN-to-IMS call (TS 186 011-2, clause 4.4.16.2)....12
4.2 Tests cases described in TS 102 901 13
4.2.1 TD_IMS_SHARE_0001 (TS 102 901, clause 4.5.4.1.1)....13
4.2.2 TD_IMS_SHARE_0002 (TS 102 901, clause 4.5.4.1.2)....13
4.3 Tests and test sessions 13
4.4 Test Reports....14
5 Test overview 14
5.1 General Test Description....14
5.2 End-to-end Tests based on Instrumental Assessment of Speech Samples....15
5.3 End-to-end Tests based on Instrumental Assessment of Video Samples 15
5.4 Other end-to-end tests in the voice channel....16
5.5 End-to-end Tests based on functional parameters 16
5.5.1 Voice Service....16
5.5.1.1 Telephony Call Setup Time....16
5.5.2 Video Streaming Service 17
5.5.2.1 Streaming Service Access Time....17
5.6 Network Performance Parameters....17
5.7 Future extension of the tests....17
6 Detailed Test Description....18
6.1 [QoS_Voice_el_qualification]: End-to-end QoS assessment of voice service 18
6.1.1 End-to-end Speech Quality assessment 19
6.1.1.1 [QoS_Voice_ac_01]: Testing at Acoustical Interfaces 19
6.1.1.1.1 Measurement setup....20
6.1.1.1.2 Measurement methodology 21
6.1.1.1.3 Test signals....21
6.1.1.1.4 Instrumental assessment of listening quality 22
6.1.1.2 [QoS_Voice_el_01]: Testing at Electrical Interfaces....23
6.1.1.2.1 Measurement setup....23
6.1.1.2.2 Measurement methodology 23
6.1.1.2.3 Test signals....24
6.1.1.2.4 Instrumental assessment of listening quality 24
6.1.2 Other end-to-end tests in the voice channel....25
6.1.2.1 Testing at Acoustical Interfaces....25
6.1.2.1.1 Measurement setup....25
6.1.2.1.2 Measurement methodology 26
6.1.2.1.3 Test signals....29

6.1.2.2 Testing at Electrical Interfaces 30
6.1.2.2.1 Measurement setup 30
6.1.2.2.2 Measurement methodology 30
6.1.2.2.3 Test signals 31
6.1.3 [QoS_Voice_func_01]: Functional QoS parameters of the voice channel 31
6.1.4 [QoS_Voice_np_01]: Network performance parameters in the voice channel 31
6.2 End-to-end QoS assessment of video service 32
6.2.1 End-to-end Video Quality assessment 32
6.2.1.1 Testing at Optical Interfaces 32
6.2.1.2 [QoS_Video_filebased_01]: Testing with inserted and retrieved video files 32
6.2.1.2.1 Measurement setup 32
6.2.1.2.2 Measurement methodology 33
6.2.1.2.3 Test signals 33
6.2.1.2.4 Instrumental assessment of video quality 33
6.2.2 Other end-to-end tests in the video channel 34
6.2.3 [QoS_Video_func_01]: Functional QoS parameters of the video channel 34
6.2.4 [QoS_Video_np_01]: Network performance parameters in the video channel 34

7 Documentation and Interpretation of Test Results 35
7.1 Measurement configuration description 35
7.1.1 Mandatory documentation of test results 35
7.2 One-view visualization of performances 35
7.2.1 Creation of Quality Pies according to Recommendation ITU-T P.505 35
7.3 Visualization of performance variations 37
7.4 Result interpretations 39

Annex A (normative): Mandatory template for result reporting 40

History 41

Intellectual Property Rights

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

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

Foreword

This Technical Specification (TS) has been produced by ETSI Technical Committee IMS Network Testing (INT).

Introduction

The following assumptions are considered in the present document. QoS tests described below are to be integrated into functional tests. It is worth noting that a plugtest environment is typically noisy. During a plugtest event a functional test is carried out once only for each configuration. On the other hand, most of the QoS tests should be done a couple of times for each configuration in order to get statistically reliable results (to capture possible variations).

The following conclusions have been made in the context of the present document. Media related QoS tests can be carried out only after an IMS session has been established (LTE UE to LTE UE). Because of the noisy environment acoustical quiet rooms are used for voice QoS testing at the acoustical interface. On the basis of many discussions with crucial players in the terminal world, the experts concluded that electrical interfaces (headsets, etc.) are not recommended to be used in the plugtest, because of speech processing functions deployed in smart phones. This point is going to be solved by a first validation phase to be organized by ETSI CTI before the plugtest. Functional QoS parameters which are defined on a per call basis and which are true end-to-end functional parameters (e.g. call set-up time) are included in a group of tests recommended for the plugtest. On the other hand, functional QoS parameters which are defined on a statistical basis are excluded (e.g. call completion ratio). For video streaming QoS testing the insertion and retrieval of test video sequences is assumed to be possible.

The following observations apply to the tests described in the present document. While the IMS media session is established (LTE UE to LTE UE) all the media QoS testing has to be done. The LTE UEs are a part of the end-to-end QoS test architecture. In fact, a UE is considered a part of the transmission chain and for that reason it is also an integral part of the service. In principle only the media channel under test is active while testing, background traffic in this context is outside the scope of the plugtest. QoS results acquired during the plugtests may be better than those perceived by users in a real LTE environment. Analyzing or tracking down root causes of bad QoS results is outside the scope of the plugtest.

1 Scope

The present document specifies a set of end-to-end QoS test extensions to TS 186 011 [4], TS 102 901 [2] and TS 103 029 [3]. To some extent the end-to-end QoS tests are self-contained and are therefore provided as a self-standing document.

2 References

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

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

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

2.1 Normative references

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

[1] ETSI ES 201 168: "Speech Processing, Transmission and Quality Aspects (STQ); Transmission characteristics of digital Private Branch eXchanges (PBXs) for interconnection to private networks, to the public switched network or to IP gateways".

[2] ETSI TS 102 901 (V.5.1.1): "IMS Network Testing (INT); IMS NNI Interoperability Test Specifications; IMS NNI interoperability test descriptions for RCS".

[3] ETSI TS 103 029: "IMS Network Testing (INT); IMS & EPC Interoperability test descriptions".

[4] ETSI TS 186 011-1 (V5.1.1): "IMS Network Testing (INT); IMS NNI Interoperability Test Specifications; Part 1: Test Purposes for IMS NNI Interoperability".

[5] ETSI TS 186 011-2 (V5.1.1): "IMS Network Testing (INT); IMS NNI Interoperability Test Specifications; Part 2: Test Description for IMS NNI Interoperability".

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

[7] Recommendation ITU-T J.247: "Objective perceptual multimedia video quality measurement in the presence of a full reference".

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

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

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

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

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

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

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

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

[16] Recommendation ITU-T P.505: "One-view visualization of speech quality measurement results".

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

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

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

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

[21] Moorthy, A.K.; Lark Kwon Choi; Bovik, A.C.; de Veciana, G.: "Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies", In IEEE Journal of Selected Topics in Signal Processing, vol. 6, No. 6, October 2012, pp. 652-671, ISSN 1932-4553.

[22] Seshadrinathan, K.; Soundararajan, R.; Bovik, A.C.; Cormack, L.K.: "Study of Subjective and Objective Quality Assessment of Video, In IEEE Transactions on Image Processing", vol. 19, No. 6, June 2010, pp. 1427-1440, ISSN 1057-7149.

[23] Chikkerur, S.; Sundaram, V.; Reisslein, M.; Karam, L.J.: "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison", In IEEE Transactions on Broadcasting, vol.57, No.2, June 2011, pp. 165-182, ISSN 0018-9316.

[24] Hekstra, A.P.; Beerends, J.G.; Ledermann, D.; de Caluwe, F.E.; Kohler, S.; Koenen, R.H.; Rihs, S.; Ehrsam, M.; Schlauss, D.: "PVQM-A perceptual video quality measure, In Signal Processing: Image Communication". vol. 17, No.10, pp. 781-798, Nov. 2002, ISSN 0923-5965.

[25] Final report from the video quality experts group on the validation of objective models of multimedia quality assessment, Phase 1, Video Quality Experts Group (VQEG), 2008.

[26] Recommendation ITU-T Y.1541 (12-2011): "Network performance objectives for IP-based services".

2.2 Informative references

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

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

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

[i.3] ETSI TR 103 122: "Speech and multimedia Transmission Quality (STQ);QoS of connections from current technologies to LTE for delay sensitive applications".

3 Abbreviations

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

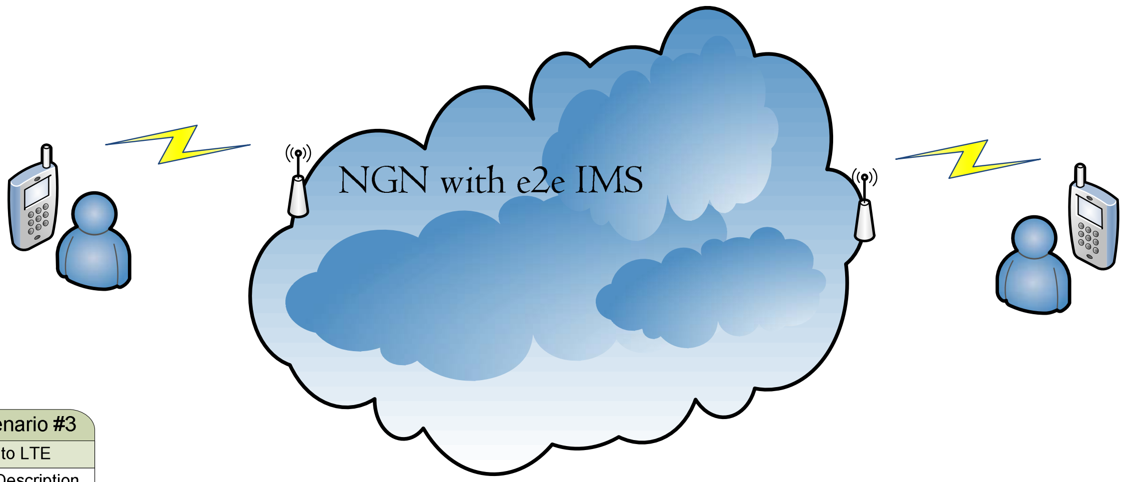
3GPP	3rd Generation Partnership Project
AGC	Automatic Gain Control
ASL	Active Speech Level
CIF	Common Intermediate Format
CS	Composite Source
CSS	Composite Source Signal
CTI	(ETSI) Center for Testing and Interoperability
DMOS	Difference Mean Opinion Score
DRP	ear Drum Reference Point
e2e	end-to-end
EPC	Evolved Packet Core

FFT	Fast Fourier Transform
HATS	Head And Torso Simulator
HW	Hardware
IMS	IP Multimedia Subsystem
INT	IMS Network Testing
IP	Internet Protocol
IPDV	IP Delay Variation
IPLR	IP Loss Ratio
IPTD	IP Transfer Delay
IPTV	IP Television
IRS	Intermediate Reference System
ITU-T	International Telecommunication Union, Telecommunication Standardization Sector
LTE	Long-Term Evolution
MOS	Mean Opinion Score
MOS-LQO	MOS Listening Quality Objective
MRP	Mouth Reference Point
NGN	Next Generation Networks
NNI	Network-Network-Interface
OLR	Overall Loudness Rating
PDD	Post Dialling Delay
PEVQ	Perceptual Evaluation of Video Quality
PLC	Packet Loss Concealment
POLQA®	Perceptual Objective Listening Quality Analysis
PSTN	Public Switched Telephone Network
PVQM	Perceptual Video Quality Measure
QCIF	Quarter Common Intermediate Format
QoS	Quality of Service
RCS	Rich Communication Services
RLR	Receive Loudness Rating
ROI	Region of Interest
RTP	Real-time Transport Protocol
SW	Software
SYN	(TCP) SYN(chronize)
TCP	Transmission Control Protocol
TELR	Talker Echo Loudness Rating
UE	User Equipment
UNI	User-Network-Interface
VAD	Voice Activity Detection
VGA	Video Graphics Array
VoIP	Voice over IP
VQEG	Video Quality Experts Group
WAP	Wireless Access Protocol
WSP	Web Service Provider

4 The general structure of end-to-end QoS assessment for VoLTE and RCS Interop Events or Plugtests

While the main purpose of interoperability test events or so called plugtests is the focus on end-to-end functional testing allowing to identify interoperability issues in protocols, base standards and implementations, it is highly desirable to retrieve at this early stage information about the end-to-end media quality. Of course, this is limited to such test cases where the establishment of an end-to-end call/session is part of the test.

Figures 1 and 2 depict the e2e QoS test scenarios that can be distinguished.



LTE to LTE Scenario #3		
Voice Call LTE to LTE		
Symbol	Count	Description
	2	User
	2	Comm-link
	2	Smart phone
	2	LTE access point
	1	NGN with e2e IMS

Figure 1: e2e QoS scenario #1

As depicted in figure 1, the LTE terminal is connected to another LTE terminal. In such cases the IMS based QoS control will cover the entire end-to-end connection.

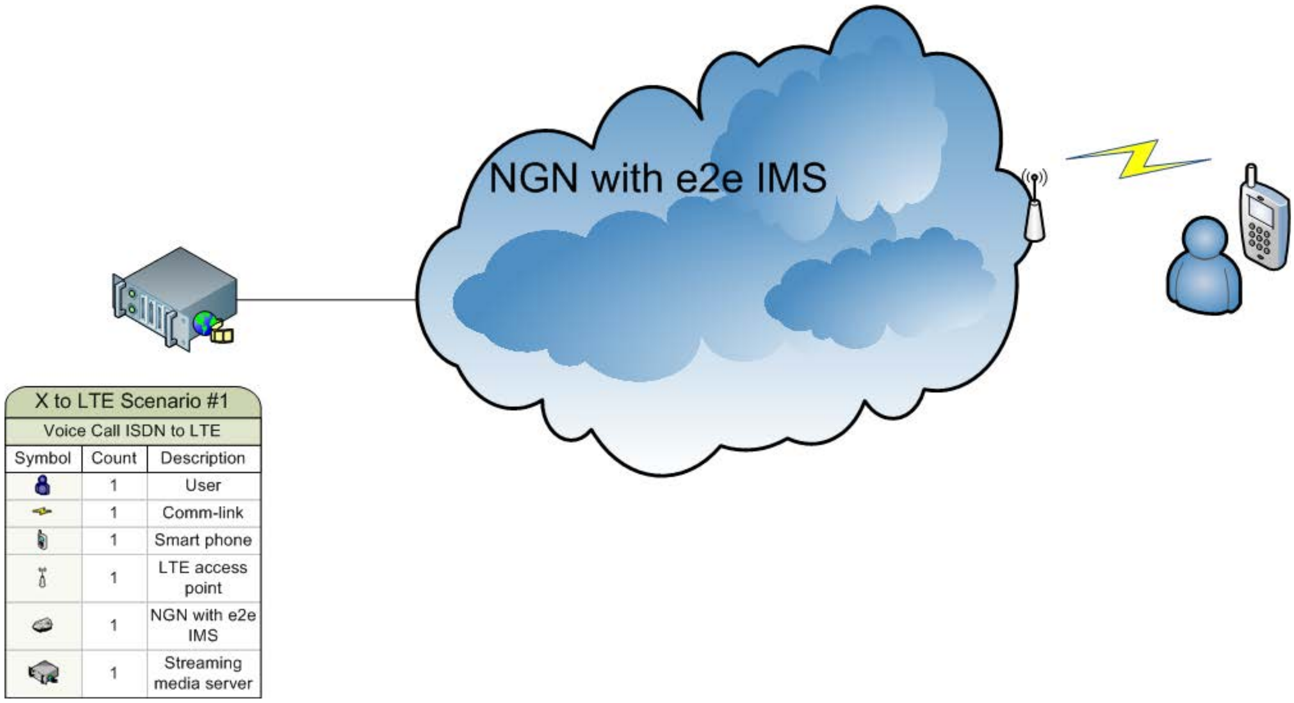


Figure 2: e2e QoS scenario #2

In figure 2, the video-capable LTE terminal is connected to a streaming media server via an NGN network. In this case the IMS based QoS control will cover the entire end-to-end connection.

4.1 Tests cases described in TS 186 011-2

4.1.1 User-initiated VoIP call setup and release (TS 186 011-2, clause 4.4.2)

After step 33 of the functional test UC_02_I (A calls B) and after step 48 of the functional test UC_02_R (A call B, B is roaming) the following QoS tests are carried out:

Table 1: Summary of QoS tests realized within the functional test UC_02_I and UC_02_R

QoS_Voice_ac_01
QoS_Voice_ac_02
QoS_Voice_ac_03
QoS_Voice_ac_04
QoS_Voice_ac_05
QoS_Voice_ac_06
QoS_Voice_ac_07
QoS_Voice_ac_08
QoS_Voice_ac_09
QoS_Voice_el_qualification
QoS_Voice_el_01
QoS_Voice_el_02
QoS_Voice_el_03
QoS_Voice_el_04
QoS_Voice_func_01
QoS_Voice_np_01

After the QoS tests are completed the functional test continues with step 34A or 34B (UC_02_I) and 49A or 49B (UC_02_R) respectively.

4.1.2 Addition of media stream (TS 186 011-2, clause 4.4.11)

After steps 63A and 63B of the functional test UC_13 the QoS tests defined in clause 4.1.1 (voice media stream) or clause 4.1.4 (video media stream) are carried out.

After the QoS tests are completed the functional test continues with step 64.

4.1.3 Ad-hoc Conferencing service (TS 186 011-2, clause 4.4.13)

After step 59 of the functional test UC_16 the following QoS tests are carried out:

Table 2: Summary of QoS tests realized within the functional test UC_16

QoS_Voice_ac_01
QoS_Voice_ac_02
QoS_Voice_ac_03
QoS_Voice_ac_04
QoS_Voice_ac_05
QoS_Voice_ac_06
QoS_Voice_ac_07
QoS_Voice_ac_08
QoS_Voice_ac_09
QoS_Voice_el_qualification
QoS_Voice_el_01
QoS_Voice_el_02
QoS_Voice_el_03
QoS_Voice_el_04
QoS_Voice_func_01
QoS_Voice_np_01

After the QoS tests are completed the functional test continues with step 60.

4.1.4 IPTV service (TS 186 011-2, clause 4.4.14)

After step 7 of the functional test UC_19 the following QoS tests are carried out:

Table 3: Summary of QoS tests realized within the functional test UC_19

QoS_Video_filebased_01
QoS_Video_func_01
QoS_Video_np_01

After the QoS tests are completed the functional test continues with step 9.

4.1.5 IMS-to-PSTN call (TS 186 011-2, clause 4.4.16.1)

After step 27 of the functional test UC_23 the following QoS tests are carried out:

Table 4: Summary of QoS tests realized within the functional test UC_23

QoS_Voice_ac_01
QoS_Voice_ac_02
QoS_Voice_ac_03
QoS_Voice_ac_04
QoS_Voice_ac_05
QoS_Voice_ac_06
QoS_Voice_ac_07
QoS_Voice_ac_08
QoS_Voice_ac_09
QoS_Voice_el_qualification
QoS_Voice_el_01
QoS_Voice_el_02
QoS_Voice_el_03
QoS_Voice_el_04
QoS_Voice_func_01
QoS_Voice_np_01

After the QoS tests are completed the functional test continues with step 28A or 28B.

4.1.6 PSTN-to-IMS call (TS 186 011-2, clause 4.4.16.2)

After step 28 of the functional test UC_24 the following QoS tests are carried out:

Table 5: Summary of QoS tests realized within the functional test UC_24

QoS_Voice_ac_01
QoS_Voice_ac_02
QoS_Voice_ac_03
QoS_Voice_ac_04
QoS_Voice_ac_05
QoS_Voice_ac_06
QoS_Voice_ac_07
QoS_Voice_ac_08
QoS_Voice_ac_09
QoS_Voice_el_qualification
QoS_Voice_el_01
QoS_Voice_el_02
QoS_Voice_el_03
QoS_Voice_el_04
QoS_Voice_func_01
QoS_Voice_np_01

After the QoS tests are completed the functional test continues with step 29A or 29B.

4.2 Tests cases described in TS 102 901

4.2.1 TD_IMS_SHARE_0001 (TS 102 901, clause 4.5.4.1.1)

After step 30 of the functional test TD_IMS_SHARE_0001 the following QoS tests are carried out:

Table 6: Summary of QoS tests realized within the functional test TD_IMS_SHARE_0001

QoS_Video_filebased_01
QoS_Video_func_01
QoS_Video_np_01

After the QoS tests are completed the functional test continues with step 32A or 32B.

4.2.2 TD_IMS_SHARE_0002 (TS 102 901, clause 4.5.4.1.2)

After step 45 of the functional test TD_IMS_SHARE_0002 the following QoS tests are carried out:

Table 7: Summary of QoS tests within the functional test TD_IMS_SHARE_0002

QoS_Video_filebased_01
QoS_Video_func_01
QoS_Video_np_01

After the QoS tests are completed the functional test continues with step 47A or 47B.

4.3 Tests and test sessions

The assessment of end-to-end QoS shall be an integral part of the interoperability tests. In order to allow an adaptation to the practical arrangements of a test event, the selection of QoS parameters from the present document is up to the test house and shall be documented together with the results.

Due to the global approach of interoperability test events and the resulting fact that equipment under test is distributed at various locations around the globe, it will be necessary to re-route both ends of the media stream to one central location, tag them with an unambiguous identifier and submit them to a post-processing database.

In the context of the present document, media refers to voice and video.

On the basis of many discussions with crucial players in the terminal world, the experts concluded that electrical measurements (e.g. via the headset interface) are not feasible in a modern mobile UE, since various speech processing algorithms inside the mobile UE would try to adapt to the device connected to this interface and interfere seriously with the measurements.

Recognizing that with a regular mobile UE electrical measurements are not feasible, it is mandatory to provide separate rooms at both send and receive end of the connection, which are sufficiently quiet.

However, for the purposes of a first validation event also tests between electrical interfaces have been added to the present document.

Therefore, interfaces and procedures have to be agreed upon between the test house and manufacturers, how voice and video sequences can "electrically" be injected on the sending side and how the received samples can be retrieved "electrically" at the receive side.

NOTE: For some parameters send and receive side may be inside the same terminal.

4.4 Test Reports

Test reports should contain as many details of the equipment under test, such as SW and HW versions, settings, type of codec and other signal enhancement procedures activated during the test.

The media sequences injected and those retrieved shall be stored in a database.

5 Test overview

This clause provides an overview of end-to-end QoS test suite realized as a part of VoLTE and RCS Interop Events or Plugtests. It starts with a general test description and carries on with basic descriptions of all tests associated with the end-to-end QoS test suite.

5.1 General Test Description

The tests are based on the ETSI INT specifications as described in the latest versions of TS 186 011-1 (Test purposes for IMS NNI Interoperability) [4], TS 186 011-2 (Test descriptions for IMS NNI Interoperability) [5], TS 102 901 (IMS NNI interoperability test descriptions for RCS) [2], TS 103 029 (IMS & EPC Interoperability) [3].

The test plan is subdivided into two different parts. The first part focuses on the instrumental assessment of speech and video quality by using real speech and video samples. In the second part, functional parameters of voice and video over LTE connections and other parameters in the end-to-end voice channel are measured. In addition a selection of network performance parameters are listed which can be monitored for comparison purposes.

For the purpose of evaluating the tests described in the present document, ETSI CTI is planning a small test event over a test LTE/IMS network before applying this test specification for a public event.

Therefore, the present document provides tests for the following scenarios:

- Electrical - Electrical Connection
- Acoustical - Acoustical Connection
- Network Performance UNI - UNI

Quality measurements are conducted using different kinds of input signals (e.g. from the ITU-T data base):

- Real speech samples (English) used to calculate MOS-LQO values (super wideband mode only) - samples coming from Recommendation ITU-T P.501 [15].
- Real video samples (different content).

The description of the signals is included in clauses 6.1.1.1.3, 6.1.1.2.3 and 6.2.1.2.3.

Speech quality measures according to Recommendation ITU-T P.863 [20] are performed for all voice scenarios.

Video quality assessment is done according to Recommendation ITU-T J.247 [7], namely PEVQ® model.

Assessment methods are described in clauses 6.1.1.1.4, 6.1.1.2.4 and 6.2.1.2.4.

An involved testlab or testlabs provide/s the test systems and the licensed implementations of the objective assessment algorithms. As already outlined above, acoustical and electrical end-to-end measurements are performed within this test. This may require different types of test equipment.

For video: Insertion and retrieval of the video sequence shall be possible with QoS judgement according to Recommendation ITU-T J.247 [7].

5.2 End-to-end Tests based on Instrumental Assessment of Speech Samples

Speech samples that are acquired during the tests are evaluated using an instrumental speech quality measurement called POLQA®. This analysis method leads to a one dimensional test result with a high correlation to auditory perceived speech quality for one-way transmissions (e.g. MOS-LQO values according to Recommendation ITU-T P.800.1 [19]). This model has been validated for VoIP and wireless transmission scenarios [20], and is applicable for the scenarios using electrical or acoustical interfaces.

The result is a one-dimensional score (MOS-LQO according to Recommendation ITU-T P.800.1 [19]). It is influenced by parameters like:

- the type of speech coder;
- the type of AGC, VAD and silence suppression at the sending side;
- comfort noise generation at the receiving side;
- the system reaction on packet loss and jitter in the network (e.g. the quality of PLC - packet loss concealment and jitter buffer design).

Further information about the speech quality assessment can be found in EG 201 377-1 [i.1].

Recommendation ITU-T P.863 [7] - Tests will give the following sub-parameters:

- MOS-LQO
- Delay variation over time
- Difference in loudness level
- MOS stability

5.3 End-to-end Tests based on Instrumental Assessment of Video Samples

Video samples that are acquired during the tests are evaluated using an instrumental video quality measurement called PEVQ® [7] (defined in Annex B of Recommendation ITU-T J.247 [7]). This analysis method leads to a one dimensional test result with a high correlation to video quality perceived by a human subject. PEVQ® is very robust model which is designed to predict the effects of transmission impairments on the video quality as perceived by a human subject. Its main targets are mobile and multimedia applications. This method has been validated for IP-based wireless transmission scenarios [7] and is therefore applicable for this test.

The result is a one-dimensional score. It is influenced by parameters like:

- the type of video coder;
- display size;
- the system reaction on packet loss and jitter in the network (e.g. the quality of PLC - packet loss concealment and jitter buffer design).

Further information about the video quality assessment can be found in [21], [22] and [23].

PEVQ® model will give the following sub-parameters:

- DMOS
- Delay variation over time
- Quality variation over time
- Jerkiness

- Blockiness
- Blur
- Chrominance
- Luminance
- Temporal Distortion
- Frozen Frames
- Skipped Frames

5.4 Other end-to-end tests in the voice channel

For video: Insertion and retrieval of the video sequence shall be possible with QoS judgement according to Recommendation ITU-T J.247 [7].

- End-to-end frequency response:
  Non-linearities in the overall end-to-end frequency response will be perceived as quality degradation. Although the main influence on this parameter comes from the terminals, there might be undesired influences caused by network elements, e.g. in case of transcoding gateways.
- Overall loudness rating:
  The optimum overall loudness rating is 10 dB, deviations will be perceived as quality degradation. Although the main influence on this parameter comes from the terminals, there might be undesired influences caused by network elements, e.g. in case of transcoding gateways.
- End-to-end delay:
  Delay is the most critical impairment, since additional delay increments are incurred by each element in the end-to-end connection chain and once incurred it cannot be compensated for. The perception of delay as an impairment depends on many factors, such as the context and the purpose of a conversation.
- Quality of echo cancellation:
  Echo cancellers are not just on or off, and there are configurations where more echo cancellers are active in a connection than there should be. This can lead to echoes being cancelled only part time or even worse situations. Low quality echo cancellation can be perceived as very annoying degradation.
- Double talk performance:
  This parameter looks into the situation when the talk spurts of both partners of a conversation overlap for a period of time. Degradations due to bad double talk performance can be perceived as very annoying because it has a potential to frequently interrupt the flow of the conversation.

5.5 End-to-end Tests based on functional parameters

5.5.1 Voice Service

5.5.1.1 Telephony Call Setup Time

Abstract Definition:

The telephony setup time describes the time period between sending the complete address information and receiving the call set-up notification.

Abstract Equation:

$$\boxed{\text{Telephony Setup Time}\,[s] = \left(t_{\text{connect established}} - t_{\text{user presses send button on UE}}\right)[s]}$$

where $t_{connectestablished}$ is a time when a corresponding connection is established and $t_{userpresssendbuttonon\ UE}$ is a time when user sends out a complete number of the called party (complete address information) and a call proceeding begins.

NOTE: This parameter is not calculated unless the telephony call setup attempt is successful. It is assumed that early traffic channel assignment is used.

5.5.2 Video Streaming Service

5.5.2.1 Streaming Service Access Time

Abstract Definition:

The parameter Streaming Service Access Time describes the duration of a service access from requesting the stream at the portal until the reception of the first stream data packet at the UE.

The first data packet refers to the RTP protocol.

Abstract Equation:

$$\boxed{\text{Streaming Service Access Time } [s] = \left(t_{\text{reception of first data packet}} - t_{\text{stream request}}\right)[s]}$$

where $t_{stream\ request}$ is the time when a stream is requested and $t_{reception\ of\ first\ data\ packet}$ is the time when a first data packet is received.

5.6 Network Performance Parameters

It is advised that it is beneficial to monitor the following parameters between UNI and UNI in both directions of transmission during the period when the end-to-end QoS parameters are being tested:

- UNI-to-UNI delay over time
- UNI-to-UNI jitter
- UNI-to-UNI packet loss

In case these parameters are measured in separately set-up connections, it is not possible to relate their values to the end-to-end QoS results.

The exact procedure of these monitoring tests shall be specified by the test house and it shall be documented together with the test results.

5.7 Future extension of the tests

In the part of testing the video quality it is foreseen that more detailed testing methods can be described in the future; this is captured in the structure of clause 6 by the notations "for further study".

6 Detailed Test Description

This clause deals with a detailed test description of end-to-end QoS assessment of voice service and video service, respectively.

Table 8 provides a link between the test names used in clause 4 and the clauses where the test is actually specified.

Table 8: Link between test names and relevant clauses

Test Name	Clause
QoS_Voice_ac_01	6.1.1.1
QoS_Voice_ac_02	6.1.2.1.2.1
QoS_Voice_ac_03	6.1.2.1.2.2
QoS_Voice_ac_04	6.1.2.1.2.3
QoS_Voice_ac_05	6.1.2.1.2.4.1
QoS_Voice_ac_06	6.1.2.1.2.4.2
QoS_Voice_ac_07	6.1.2.1.2.5.1
QoS_Voice_ac_08	6.1.2.1.2.5.2
QoS_Voice_ac_09	6.1.2.1.2.5.3
QoS_Voice_el_qualification	6.1
QoS_Voice_el_01	6.1.1.2
QoS_Voice_el_02	6.1.2.2.2.1
QoS_Voice_el_03	6.1.2.2.2.2
QoS_Voice_el_04	6.1.2.2.2.3
QoS_Voice_func_01	6.1.3
QoS_Voice_np_01	6.1.4
QoS_Video_filebased_01	6.2.1.2
QoS_Video_func_01	6.2.3
QoS_Video_np_01	6.2.4

6.1 [QoS_Voice_el_qualification]: End-to-end QoS assessment of voice service

The following clauses focus on all aspects of end-to-end QoS assessment of voice service, namely on end-to-end speech quality assessment, other end-to-end tests in the voice channel, functional QoS parameters of the voice channel and network performance parameters in the voice channel.

For any test between electrical interfaces of mobile terminals it is mandatory to confirm the qualification of the interfaces (e.g. headset interface, special test interface) by executing the following procedure. A telephony connection is established between the mobile terminal (with the electrical interface under consideration) and a LTE Reference Point according to EG 202 425 [i.2]. The impedance of each test interface shall be 600 ohms and be compliant with the return loss requirements for an M4 interface in ES 201 168 [1]. The Overall Loudness Rating (OLR) (see clause 6.1.2.2.2 of the present document) in both directions of transmission shall be 10 dB ± 6 dB. The MOS-LQO value according to Recommendation ITU-T P.863 [20] is measured according to the procedure described in clause 6.1.1.2.2 of the present document, but with the additional requirement that the measurement is executed simultaneously in both directions of transmission. All 16 results shall be MOS-LQO (Recommendation ITU-T P.863 [20]) > 4,0.

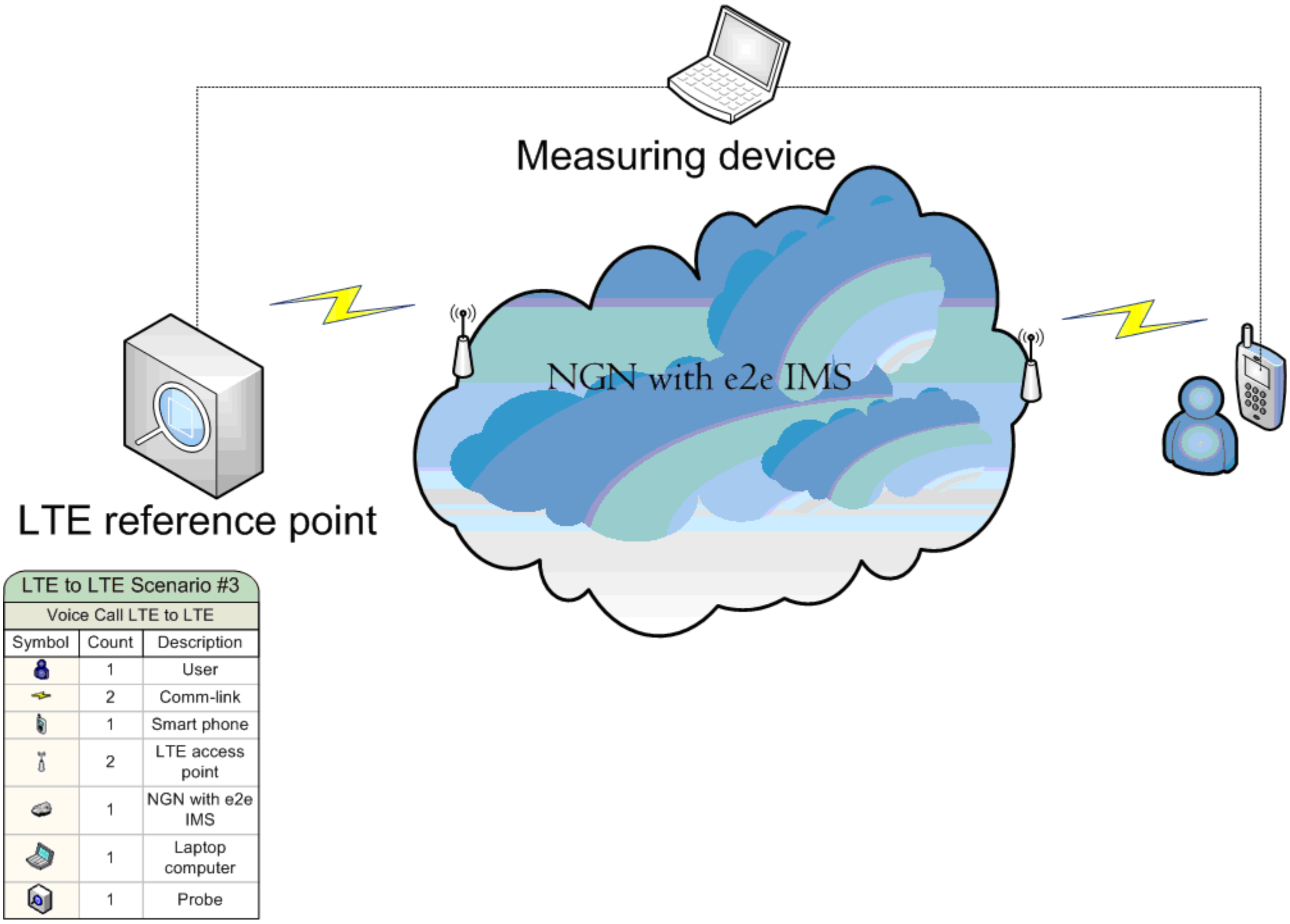


Figure 3: Measurement setup: LTE to LTE scenario including LTE reference point

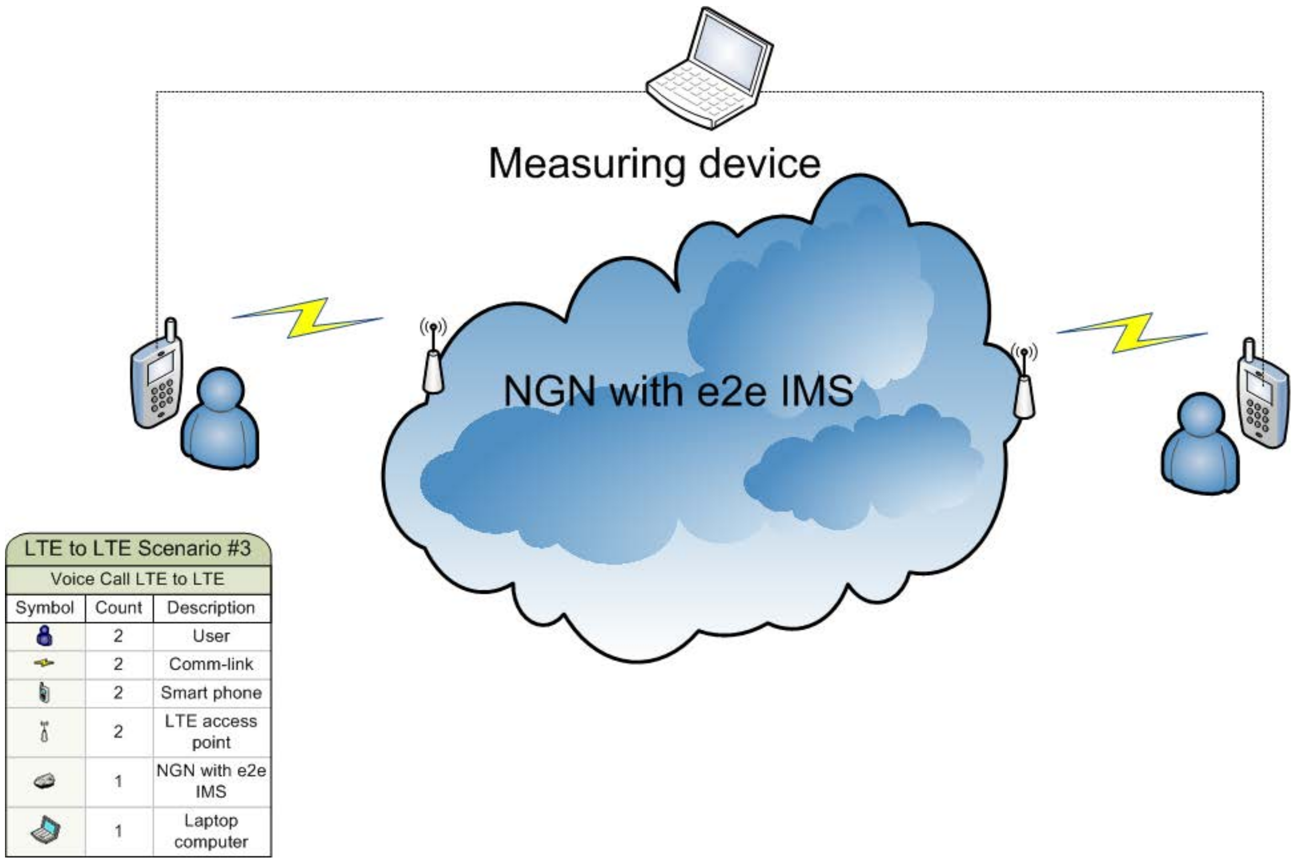
6.1.1 End-to-end Speech Quality assessment

In principle, the speech quality assessment can be done using two completely different ways, namely testing at acoustical interfaces and testing at electrical interfaces. Both methods are described below in more detail.

6.1.1.1 [QoS_Voice_ac_01]: Testing at Acoustical Interfaces

In the following clauses, the testing at acoustical interfaces is described from different perspectives, namely measurement setup, measurement methodology, test signals and instrumental assessment of listening quality.

6.1.1.1.1 Measurement setup



LTE to LTE Scenario #3		
Voice Call LTE to LTE		
Symbol	Count	Description
	2	User
	2	Comm-link
	2	Smart phone
	2	LTE access point
	1	NGN with e2e IMS
	1	Laptop computer

Figure 4: Measurement setup: LTE to LTE scenario

A measurement setup is depicted in figure 4. It should be noted here that one of the smart phones displayed in figure 4 can also be replaced by other types of terminals in order to also investigate an impact of other technologies interconnected with LTE on a quality perceived by the end user, see examples of such scenarios involving other transmission technologies in clause 4 of the present document. The input signals described in clause 6.1.1.1.3 are transmitted and recorded simultaneously, that means that the record process starts at the same time as the transmission process begins. Therefore exact delay assessment is possible. The measurement will be done at the acoustical interface.

The acoustical access to terminals is the most realistic simulation of the "average" subscriber. This can be made by using HATS (Head And Torso Simulator) with appropriate ear simulation and appropriate means to fix handset and headset terminals in a realistic and reproducible way to the HATS. HATS is described in Recommendation ITU-T P.58 [10], appropriate ears are described in Recommendation ITU-T P.57 [9] (type 3.3 and type 3.4 ear), a proper positioning of handsets under realistic conditions is to be found in Recommendation ITU-T P.64 [11].

When using a handset telephone the handset is placed in the HATS position as described in Recommendation ITU-T P.64 [11]. The artificial mouth shall be conform with Recommendation ITU-T P.58 [10]. The artificial ear shall be conform with Recommendation ITU-T P.57 [9], type 3.3 or type 3.4 ears shall be used and positioned on HATS according to Recommendation ITU-T P.58 [10]. Recommendations for positioning headsets are given in Recommendation ITU-T P.380 [14]. If not stated otherwise headsets shall be placed in their recommended wearing position. It shall be also indicated what type of ear was used at what application force. The horizontal positioning of the HATS reference plane shall be guaranteed within ±2º. The exact calibration and equalization of HATS can be found in Recommendation ITU-T P.581 [18]. If not stated otherwise, the HATS shall be diffuse-field equalized. The inverse nominal diffuse field curve as found in table 3 of Recommendation ITU-T P.58 [10] shall be used. Further information about setup and the use of HATS can be found in Recommendation ITU-T P.380 [14].

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

Unless stated otherwise, the application force of 8 N is used for handset testing.

The inverse average diffuse field response characteristics of HATS as found in Recommendation ITU-T P.58 [10] is used and not the specific one corresponding to the HATS used. Instead of using the individual diffuse field correction, the average correction function is used because, for handset and headset measurements, mostly the artificial ear, ear canal and ear impedance simulations are effective. The individual diffuse-field correction function of HATS includes all diffraction and reflection effects of the complete individual HATS which are not effective in the measurement and potentially would lead to bigger measurement uncertainties than using the average correction.

In general different acoustical environments have to be taken into account: Either room noise and background noise are an inherent part of the test environment or room noise and background noise shall be eliminated to such an extent that their influence on the test results can be neglected.

Unless stated otherwise, measurements shall be conducted under quiet conditions.

Depending on the distance of the transducers from mouth to 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.

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

6.1.1.1.2 Measurement methodology

In order to get statistically reliable results (to capture possible variations), a measurement should be done two times per test signal (4 different speech samples involved in the measurement) and transmission direction (uplink, downlink transmission). This approach provides eight quality values (MOS-LQO values) for each transmission direction, which will be averaged in post-processing and will form a single quality score characterizing the quality of the corresponding transmission direction, i.e. uplink quality score, downlink quality score. Finally, eight MOS-LQO values obtained for both directions will be averaged to form a global quality score describing an end-to-end media quality as perceived by the end user provided by a technology offered by particular vendor(s) in the investigated scenario.

In addition to the MOS-LQO values, the following parameters will be captured by the quality prediction model (described in more detail in clause 6.1.1.1.4) in parallel with the corresponding MOS-LQO value for each measurement:

- Delay variation over time
- Difference in loudness level
- MOS stability

6.1.1.1.3 Test signals

The transmitted voice signals coming from Recommendation ITU-T P.501 [15] are the basis for the instrumental evaluation of one-way speech transmission quality using the model defined in Recommendation ITU-T P.863 [20]. It is recommended to use an English subset of the database associated to Recommendation ITU-T P.501 [15].

Four speech files (8 s each) are generated and transmitted. Each of these files contains two different sentence pairs uttered by different male and female speakers. Each of these sentence pairs fulfils the requirements of Recommendation ITU-T P.800.1 [19] and shows a speech activity factor of about 50 %.

For testing the wideband telephony service provided by a terminal the test signal used shall be band limited between 100 Hz and 8 kHz with a bandpass filter providing a minimum of 24 dB/Oct. filter roll off, when feeding into the receive direction.

The test signal levels are referred to the average level of the (band limited in receive direction) test signal, averaged over the complete test sequence unless specified otherwise.

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

6.1.1.1.4 Instrumental assessment of listening quality

The latest psychoacoustic instrumental model according to Recommendation ITU-T P.863 [20] leads to a one dimensional "MOS-like" test result with a high correlation to quality scores gained by auditory listening only tests.

The basic structure of the POLQA® model is depicted in figure 5. In principle, the model is based on comparisons of the distorted signal with the undistorted reference input signal of the system. In addition, POLQA® uses psycho-acoustic models of the human speech perception. Procedures for instrumental speech quality estimation usually work in several steps. The first step eliminates signal differences that are irrelevant in the modelled auditory test (e.g. total delay and level differences). The next stage transforms both signals to an "internal representation" using psycho-acoustic models for the human sound perception. The spread (including multidimensional aspects) between both pre-processed signals is computed and will be used for estimating a quality value.

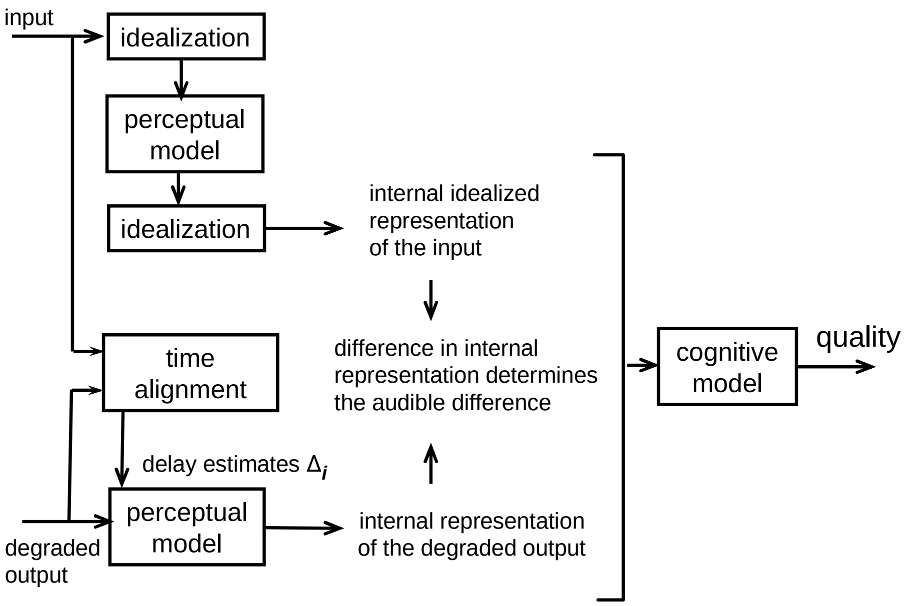


Figure 5: General structure of instrumental speech quality estimation approaches

The POLQA® model was validated for VoIP and wireless transmission scenarios and is therefore applicable for the scenarios to be tested during the event.

6.1.1.2 [QoS_Voice_el_01]: Testing at Electrical Interfaces

6.1.1.2.1 Measurement setup

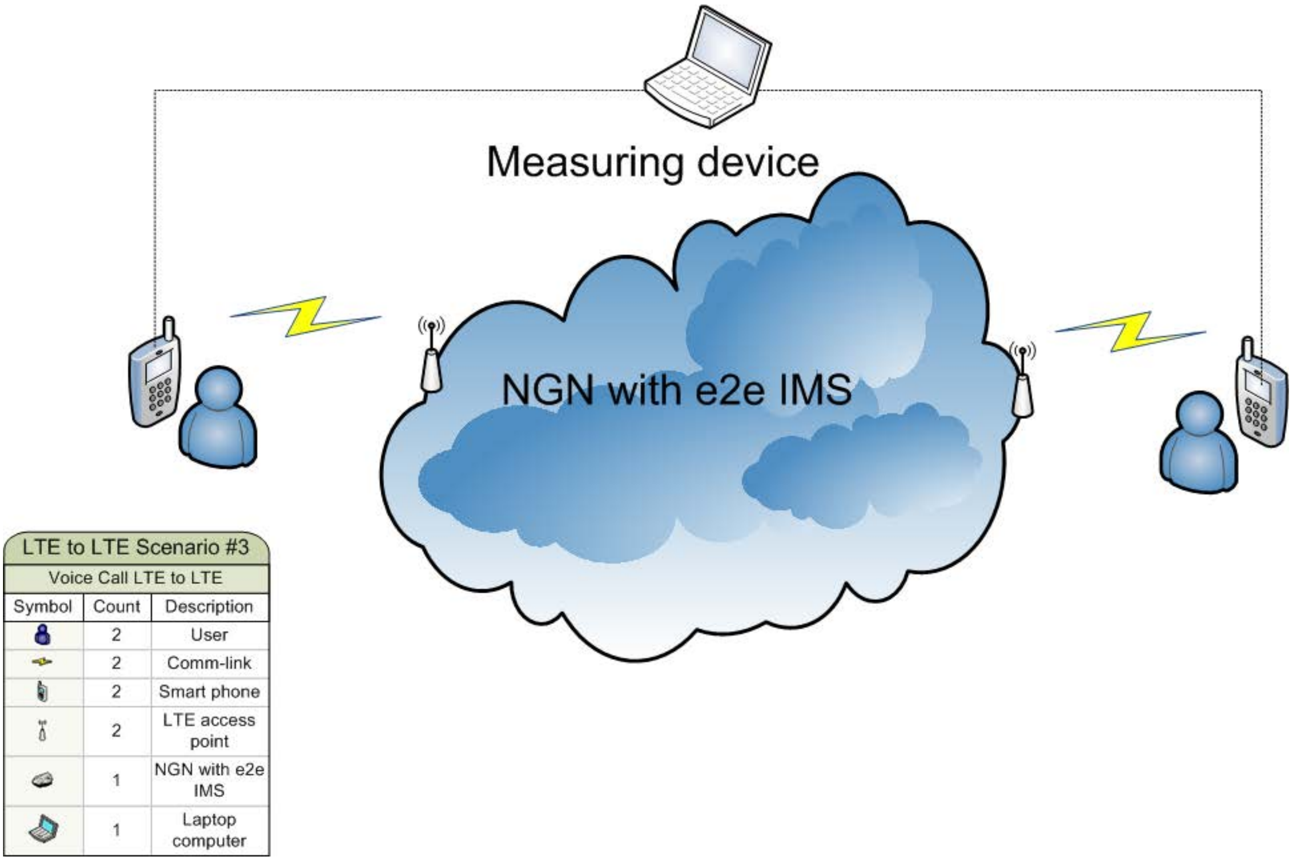


Figure 6: Measurement setup: LTE to LTE scenario

A measurement setup is depicted in figure 6. It should be noted here that one of the smart phones displayed in figure 6 can be also replaced by other types of terminals in order to also investigate an impact of other technologies interconnected with LTE on a quality perceived by the end user, see examples of such scenarios involving other transmission technologies in TR 103 122 [i.3]. The input signals described in clause 6.1.1.2.3 are transmitted and recorded simultaneously, that means that the record process starts at the same time as the transmission process begins. Therefore exact delay assessment is possible. In comparison to the test setup presented in clause 6.1.1.1, a measurement is performed at the electrical interface.

The tests described hereafter assume that a 4-wire analogue interface is provided by the terminal. This may either be a specially provided test interface or a regular headset interface. In either case care should be taken, that input and output levels are in a nominal range and that additional speech processing functions (which have the potential to corrupt the test results) are disabled (if possible).

6.1.1.2.2 Measurement methodology

In order to get statistically reliable results (to capture possible variations), a measurement should be done two times per test signal (four different speech samples involved in the measurement) and transmission direction (uplink, downlink transmission). This approach provides eight quality values (MOS-LQO values) for each transmission direction, which will be averaged in post-processing and will form a single quality score characterizing a quality of corresponding transmission direction, i.e. uplink quality score, downlink quality score. Finally, eight MOS-LQO values obtained for both directions will be averaged to form a global quality score describing an end-to-end media quality as perceived by the end user provided by a technology offered by particular vendor(s) in the investigated scenario.

In addition to the MOS-LQO values, the following parameters will be also captured by the quality prediction model (described in more detail in clause 6.1.1.2.4) in parallel with the corresponding MOS-LQO value for each measurement:

- Delay variation over time
- Difference in loudness level
- MOS stability

6.1.1.2.3 Test signals

The transmitted voice signals coming from Recommendation ITU-T P.501 [15] are the basis for the instrumental evaluation of one-way speech transmission quality using the model defined in Recommendation ITU-T P.863 [20]. It is recommended to use an English subset of the database associated to Recommendation ITU-T P.501 [15].

Four speech files (8 s each) are generated and transmitted. Each of these files contains two different sentence pairs uttered by different male and female speakers. Each of these sentence pairs fulfils the requirements of Recommendation ITU-T P.800.1 [19] and shows a speech activity factor of about 50 %.

All speech samples which are used as electrical input signals should be pre-filtered with a modified IRS(send) filter [15]. The active speech level (ASL, [14]) at the sending side is adjusted in these conditions to -16 dBm0 at 600 Ohm.

6.1.1.2.4 Instrumental assessment of listening quality

The latest psychoacoustic instrumental model according to Recommendation ITU-T P.863 [20] leads to a one dimensional "MOS-like" test result with a high correlation to quality scores gained by auditory listening only tests.

The basic structure of the POLQA® model is depicted in figure 7. In principle, the model is based on comparisons of the distorted signal with the undistorted reference input signal of the system. In addition, POLQA® uses psycho-acoustic models of the human speech perception. Procedures for instrumental speech quality estimation usually work in several steps. The first step eliminates signal differences that are irrelevant in the modelled auditory test (e.g. total delay and level differences). The next stage transforms both signals to an "internal representation" using psycho-acoustic models for the human sound perception. The spread (including multidimensional aspects) between both pre-processed signals is computed and will be used for estimating a quality value.

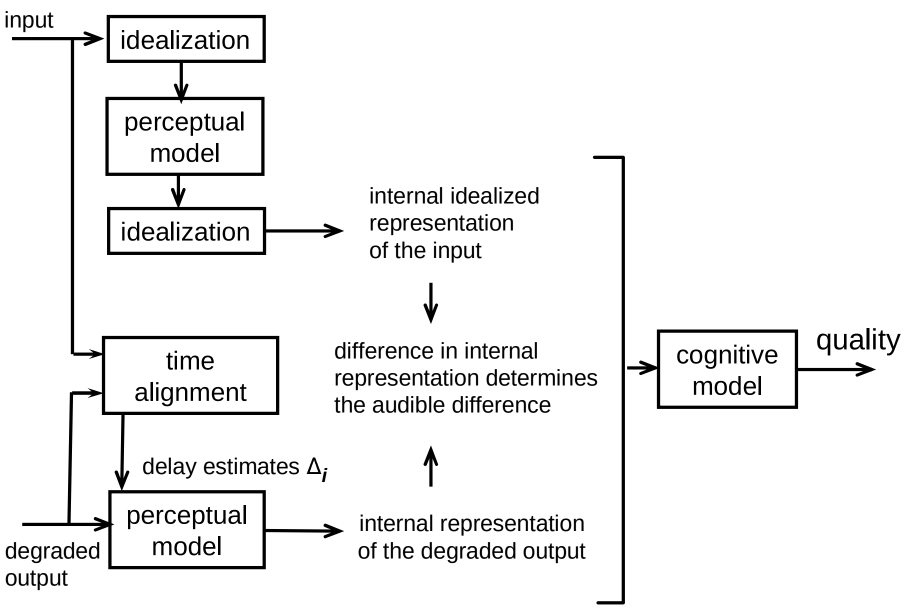


Figure 7: General structure of instrumental speech quality estimation approaches

The POLQA® model was validated for VoIP and wireless transmission scenarios and therefore is applicable for the scenarios to be tested during the event.

6.1.2 Other end-to-end tests in the voice channel

Like the speech quality assessment, the end-to-end tests in the voice channel described in this clause can be done using two completely different ways, namely testing at acoustical interfaces and testing at electrical interfaces. All tests in the voice channel presented in this clause are described below for acoustical and electrical interfaces separately.

6.1.2.1 Testing at Acoustical Interfaces

In the following clauses, the testing at acoustical interfaces is described from different perspectives, namely measurement setup, measurement methodology and test signals.

6.1.2.1.1 Measurement setup

The measurement setup depicted in fgure 4 (clause 6.1.1.1.1) also applies to all tests in the voice channel described in this clause.

The acoustical access to terminals is the most realistic simulation of the "average" subscriber. This can be made by using HATS (Head And Torso Simulator) with appropriate ear simulation and appropriate means to fix handset and headset terminals in a realistic and reproducible way to the HATS. HATS is described in Recommendation ITU-T P.58 [10], appropriate ears are described in Recommendation ITU-T P.57 [9] (type 3.3 and type 3.4 ear), a proper positioning of handsets under realistic conditions is to be found in Recommendation ITU-T P.64 [11].

When using a handset telephone the handset is placed in the HATS position as described in Recommendation ITU-T P.64 [11]. The artificial mouth shall be conform with Recommendation ITU-T P.58 [10]. The artificial ear shall be conform with Recommendation ITU-T P.57 [9], type 3.3 or type 3.4 ears shall be used and positioned on HATS according to Recommendation ITU-T P.58 [10]. Recommendations for positioning headsets are given in Recommendation ITU-T P.380 [14]. If not stated otherwise headsets shall be placed in their recommended wearing position. It shall be also indicated what type of ear was used at what application force. The horizontal positioning of the HATS reference plane shall be guaranteed within ±2º. The exact calibration and equalization of HATS can be found in Recommendation ITU-T P.581 [18]. If not stated otherwise, the HATS shall be diffuse-field equalized. The inverse nominal diffuse field curve as found in table 3 of Recommendation ITU-T P.58 [10] shall be used. Further information about setup and the use of HATS can be found in Recommendation ITU-T P.380 [14].

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

Unless stated otherwise, the application force of 8 N is used for handset testing.

The inverse average diffuse field response characteristics of HATS as found in Recommendation ITU-T P.58 [10] is used and not the specific one corresponding to the HATS used. Instead of using the individual diffuse field correction, the average correction function is used because, for handset and headset measurements, mostly the artificial ear, ear canal and ear impedance simulations are effective. The individual diffuse-field correction function of HATS includes all diffraction and reflection effects of the complete individual HATS which are not effective in the measurement and potentially would lead to bigger measurement uncertainties than using the average correction.

In general different acoustical environments have to be taken into account: either room noise and background noise are an inherent part of the test environment or room noise and background noise shall be eliminated to such an extent that their influence on the test results can be neglected.

Unless stated otherwise, measurements shall be conducted under quiet conditions.

Depending on the distance of the transducers from mouth to 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.

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

6.1.2.1.2 Measurement methodology

6.1.2.1.2.1 [QoS_Voice_ac_02]: End-to-end frequency response

The test signal to be used for the measurements shall be the British-English speech sequences from Recommendation ITU-T P.501 [15]. 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.

The handsets are mounted at the HATS position (see Recommendation ITU-T P.64 [11]). The application force used to apply the handset against the artificial ear shall be within the range specified in Recommendation ITU-T P.64 [11].

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 (Recommendation ITU-T P.79 [12], table 3, bands 1 to 20) is referred to the averaged test signal level measured in each frequency band.

6.1.2.1.2.2 [QoS_Voice_ac_03]: Overall loudness rating (OLR)

The test signal to be used for the measurements shall be the British-English speech sequences from Recommendation ITU-T P.501 [15]. 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.

The handsets are mounted at the HATS position (see Recommendation ITU-T P.64 [11]). The application force used to apply the handset against the artificial ear shall be within the range specified in Recommendation ITU-T P.64 [11].

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 (Recommendation ITU-T P.79 [12], table 3, bands 1 to 20) is referred to the averaged test signal level measured in each frequency band.

The overall sensitivity shall be calculated from each band of the 20 frequencies given in table 1 of Recommendation ITU-T P.79 [12], bands 1 to 20. For the calculation the averaged measured level at the receive side for each frequency band is referred to the averaged test signal level measured in each frequency band at the send side.

The OLR shall be calculated according to Recommendation ITU-T P.79 [12], formula (A - 23b), over bands 1 to 20, using m = 0,175 and the weighting factors from Recommendation ITU-T P.79 [12], annex A, table A.2.

6.1.2.1.2.3 [QoS_Voice_ac_04]: End-to-end delay

The end-to-end delay is measured from the MRP to the Drum Reference Point (DRP).

1) For the measurements a Composite Source Signal (CSS) according to Recommendation ITU-T P.501 [15] is used. The pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal level is -4,7 dBPa at the MRP.

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

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

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

6.1.2.1.2.4 Quality of echo cancellation

6.1.2.1.2.4.1 [QoS_Voice_ac_05]: Temporal echo effects

The test signal consists of periodically repeated Composite Source Signal according to Recommendation ITU-T P.501 [15] with an average level of -5 dBm0 as well as an average level of -25 dBm0. The echo signal is analysed during a period of at least 2,8 s which represents 8 periods of the CS signal. The integration time for the level analysis shall be 35 ms, the analysis is referred to the level analysis of the reference signal.

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

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

6.1.2.1.2.4.2 [QoS_Voice_ac_06]: Spectral Echo Attenuation

During the measurement it should be ensured that the measured signal is really the echo signal and not the Comfort Noise which possibly may be inserted in send direction in order to mask the echo signal.

Before the actual measurement a training sequence is fed in consisting of 10 s CS signal according to Recommendation ITU-T P.501 [15]. The level of the training sequence is -16 dBm0.

The test signal consists of a periodically repeated Composite Source Signal. The measurement is carried out under steady-state conditions. The average test signal level is -16 dBm0, averaged over the complete test signal. 4 CS signals including the pauses are used for the measurement which results in a test sequence length of 1,4 s. The power density spectrum of the measured echo signal is referred to the power density spectrum of the original test signal. The analysis is conducted using FFT analysis with 8 k points (48 kHz sampling rate, Hanning window).

The spectral echo attenuation is analysed in the frequency domain in dB.

6.1.2.1.2.5 Double talk performance

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

In order to guarantee sufficient quality under double talk conditions the Talker Echo Loudness Rating (TELR) should be high and the attenuation inserted should be as low as possible. Connections which do not allow double talk in any case should provide a good echo attenuation which is realized by a high attenuation range in this case.

The most important parameters determining the speech quality during double talk are (see Recommendations ITU-T P.340 [13] and P.502 [17]):

- Attenuation range in send direction during double talk $A_{H,S,dt}$.
- Attenuation range in receive direction during double talk $A_{H,R,dt}$.
- Echo attenuation during double talk.

The categorization of a connection is based on the three categories defined in the following clauses and this categorization is given by the "lowest" of the three parameters, e.g. if $A_{H,S,dt}$ provides 2a, $A_{H,R,dt}$ 2b and echo loss 1, the categorization of the terminal is 2b.

6.1.2.1.2.5.1 [QoS_Voice_ac_07]: Attenuation Range in Send Direction during Double Talk

Based on the level variation in the uplink direction during double talk $A_{H,S,dt}$ the behaviour of the connection can be classified according to table 9.

The category of the connection according to table 9 shall be noted in the test report.

Table 9: Categories of the terminal defined in Recommendation ITU-T P.340 [13] for send direction

Category	**1**	**2a**	**2b**	**2c**	**3**
	Full Duplex Capability	Partial Duplex Capability			No Duplex Capability
$A_{H,S,dt}$ [dB]**	≤ 3	≤ 6	≤ 9	≤ 12	> 12

In general this table provides a quality classification of connections regarding double talk performance. However, this does not mean that a connection which is category 1 based on the double talk performance is of high quality concerning the overall quality as well.

Measurement Method

The test signal to determine the attenuation range during double talk is shown in figure 8. A sequence of uncorrelated CS signals is used which is inserted in parallel in send and receive direction.

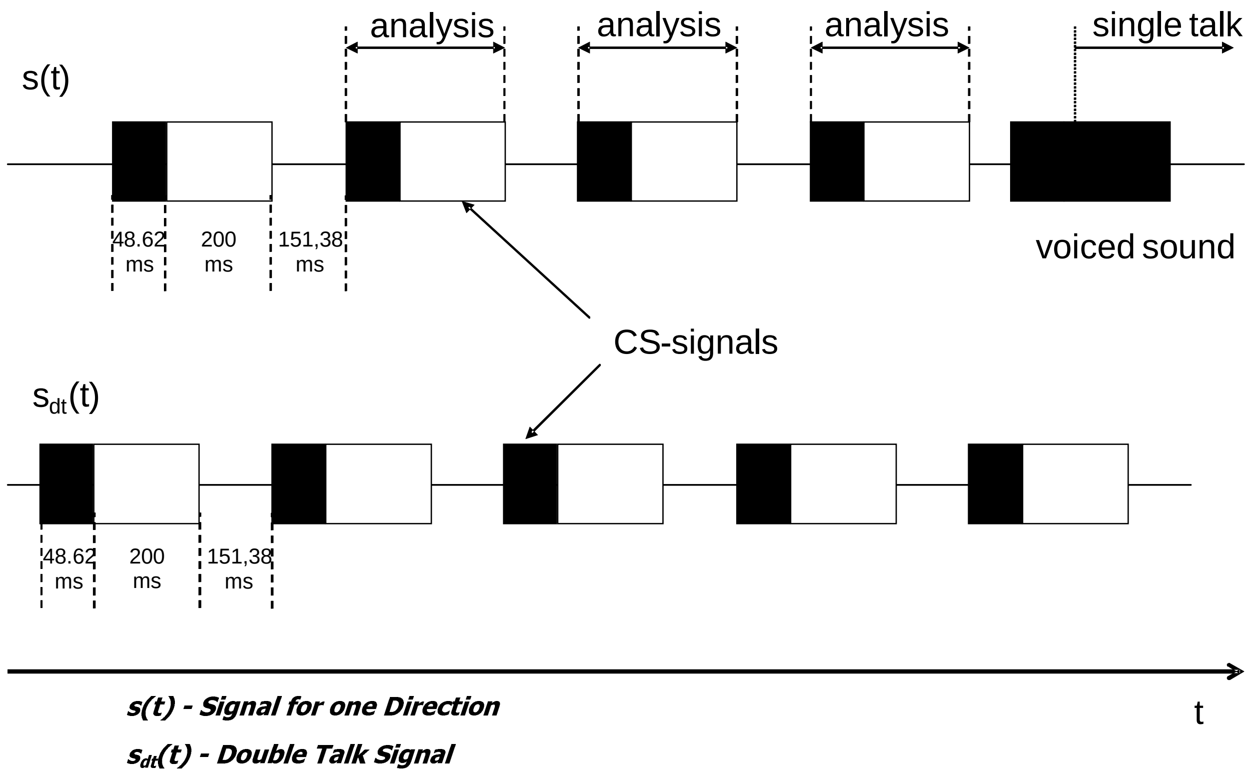


Figure 8: Double Talk Test Sequence with overlapping CS signals in send and receive direction

Figure 8 indicates that the sequences overlap partially. The beginning of the CS sequence (voiced sound, black) is overlapped by the end of the pn-sequence (white) of the opposite direction. During the active signal parts of one signal the analysis can be conducted in send and receive direction. The analysis times are shown in figure 8 as well. The test signals are synchronized in time at one acoustical interface. The delay of the test arrangement should be constant during the measurement.

NOTE: The length of voiced sound of the double talk signal is achieved by repeating one period of the voiced sound for double talk according to Recommendation ITU-T P.501 [15] 10 times and cutting off the initial 3,3 ms of the period of the first voiced sound.

The settings for the test signals are as follows:

Table 10: CS Signal Parameters in send direction

	Send Direction (sdt(t))	Receive Direction (s(t))
Pause Length between two Signal Bursts	151,38 ms	151,38 ms
Average Signal Level (Assuming an Original Pause length of 101,38 ms)	-4,7 dBPa	-4,7 dBPa
Active Signal Parts	-3 dBPa	-3 dBPa

When determining the attenuation range in the send direction the signal measured at the artificial ear is referred to the test signal inserted.

The level is determined as level vs. time from the time domain. The integration time of the level analysis is 5 ms. The attenuation is determined from the level difference measured at the beginning of the double talk always with the beginning of the CS-signal in send direction until its complete activation (during the pause in the receive channel). The analysis is performed over the complete signal starting with the second CS-signal. The first CS-signal is not used for the analysis.

6.1.2.1.2.5.2 [QoS_Voice_ac_08]: Attenuation Range in Receive Direction during Double Talk

Based on the level variation in the downlink direction during double talk $A_{H,R,dt}$ the behaviour of the connection can be classified according to table 11.

The category of the terminal according to table 11 shall be noted in the test report.

Table 11: Categories of the terminal defined in Recommendation ITU-T P.340 [13] in receive direction

Category	1	2a	2b	2c	3
	Full Duplex Capability	Partial Duplex Capability			No Duplex Capability
$A_{H,R,dt}$ [**dB]	≤ 3	≤ 5	≤ 8	≤ 10	> 10

In general this table provides a quality classification of connections regarding double talk performance. However, this does not mean that a connection which is category 1 based on the double talk performance is of high quality concerning the overall quality as well.

The test signal to determine the attenuation range during double talk is shown in figure 8. A sequence of uncorrelated CS signals is used which is inserted in parallel in send and receive direction. The test signals are synchronized in time at the acoustical interface. The delay of the test arrangement should be constant during the measurement.

The settings for the test signals are as follows:

Table 12: CS Signal Parameters in receive direction

	Receive Direction (s(t))	Send Direction (sdt(t))
Pause Length between two Signal Bursts	151,38 ms	151,38 ms
Average Signal Level (Assuming an Original pause Length of 101,38 ms)	-4,7 dBPa	-4,7 dBPa
Active Signal Parts	-3 dBPa	-3 dBPa

When determining the attenuation range in the receive direction the signal measured at the artificial ear is referred to the test signal inserted.

The level is determined as level vs. time from the time domain. The integration time of the level analysis is 5 ms. The attenuation is determined from the level difference measured at the beginning of the double talk always with the beginning of the CS-signal in receive direction until its complete activation (during the pause in the send channel). The analysis is performed over the complete signal starting with the second CS-signal. The first CS-signal is not used for the analysis.

6.1.2.1.2.5.3 [QoS_Voice_ac_09]: Detection of echo components during double Talk

Echo Loss during double talk is the echo suppression provided by the connection during double talk.

The measurement method for echo during double talk between two acoustical interfaces is for further study.

6.1.2.1.3 Test signals

A description of the test signals recommended for each of the tests defined in the voice channel are available above, as a part of the measurement methodology. For that reason, this clause only outlines crucial facts related to the test signals which have to be carefully considered when performing the tests described above.

Due to the coding of the speech signals, care should be taken when using single frequency for wireless terminals/networks (e.g. LTE) acoustic tests. Appropriate test signals (general description) are defined in Recommendations ITU-T P.50 [8] and P.501 [15].

For testing the wideband telephony service provided by a terminal the test signal used shall be band limited between 100 Hz and 8 kHz with a bandpass filter providing a minimum of 24 dB/Oct. filter roll off, when feeding into the receive direction.

The test signal levels are referred to the average level of the (band limited in receive direction) test signal, averaged over the complete test sequence unless specified otherwise.

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

6.1.2.2 Testing at Electrical Interfaces

In the following clauses, the testing at electrical interfaces is described from different perspectives, namely measurement setup, measurement methodology and test signals.

6.1.2.2.1 Measurement setup

The measurement setup depicted in figure 6 (clause 6.1.1.2.1) also applies to all tests in the voice channel described in this clause.

The tests described hereafter assume that a 4-wire analogue interface is provided by the terminal. This may either be a specially provided test interface or a regular headset interface. In either case care should be taken, that input and output levels are in a nominal range and that additional speech processing functions (which have the potential to corrupt the test results) are disabled (if possible).

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

6.1.2.2.2 Measurement methodology

6.1.2.2.2.1 [QoS_Voice_el_02]: End-to-end frequency response

The test signal to be used for the measurements shall be the British-English speech sequences from Recommendation ITU-T P.501 [15]. The spectrum of acoustic signal produced by the artificial mouth is calibrated under free field conditions at the MRP. The test signal level is averaged over the complete test signal sequence.

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 (Recommendation ITU-T P.79 [12], table 3, bands 1 to 20) is referred to the averaged test signal level measured in each frequency band.

6.1.2.2.2.2 [QoS_Voice_el_03]: Overall loudness rating (OLR)

The test signal to be used for the measurements shall be British-English speech sequences from Recommendation ITU-T P.501 [15]. The spectrum of acoustic signal produced by the artificial mouth is calibrated under free field conditions at the MRP. The test signal level is averaged over the complete test signal sequence.

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 (Recommendation ITU-T P.79 [12], table 3, bands 1 to 20) is referred to the averaged test signal level measured in each frequency band.

The overall sensitivity shall be calculated from each band of the 20 frequencies given in table 1 of Recommendation ITU-T P.79 [12], bands 1 to 20. For the calculation the averaged measured level at the receive side for each frequency band is referred to the averaged test signal level measured in each frequency band at the send side.

The OLR shall be calculated according to Recommendation ITU-T P.79 [12], formula (A - 23b), over bands 1 to 20, using m = 0,175 and the weighting factors from Recommendation ITU-T P.79 [12], annex A, table A.2.

6.1.2.2.2.3 [QoS_Voice_el_04]: End-to-end delay

The end-to-end delay is measured from the ingress of the UE at the send side to the egress of the UE at the receive side.

1) For the measurements a Composite Source Signal (CSS) according to Recommendation ITU-T P.501 [15] is used. The pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The reference signal is the original signal (test signal).

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

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

6.1.2.2.3 Test signals

A description of the test signals recommended for each of the tests defined in the voice channel are available above, as a part of the measurement methodology. For that reason, this clause only outlines crucial facts related to the test signals which have to be carefully considered when performing the tests described above.

Due to the coding of the speech signals, care should be taken when using single frequency for wireless terminals/networks (e.g. LTE) acoustic tests. Appropriate test signals (general description) are defined in Recommendation ITU-T P.50 [8] and Recommendation ITU-T P.501 [15].

For testing the wideband telephony service provided by a terminal the test signal used shall be band limited between 100 Hz and 8 kHz with a bandpass filter providing a minimum of 24 dB/Oct. filter roll off, when feeding into the receive direction.

The test signal levels are referred to the average level of the (band limited in receive direction) test signal, averaged over the complete test sequence unless specified otherwise.

Unless specified otherwise, the test signal level shall be -16 dBm0 at the sending side.

6.1.3 [QoS_Voice_func_01]: Functional QoS parameters of the voice channel

As can be seen in clause 5.5.1, only one functional QoS parameter of the voice channel has been recommended in this case, namely telephony call setup time. Its definition and abstract equation are available in clause 5.5.1.

The exact derivation of the trigger points from the signalling protocol under test in the plugtest has to be determined by the test house and shall be stated in the test report.

6.1.4 [QoS_Voice_np_01]: Network performance parameters in the voice channel

It is advised that it is beneficial to monitor the following parameters between UNI and UNI in both directions of transmission during the period when the end-to-end QoS parameters are being tested:

- UNI-to-UNI delay over time
- UNI-to-UNI jitter
- UNI-to-UNI packet loss

In case these parameters are measured in separately set-up connections, it is not possible to relate their values to the end-to-end QoS results.

As outlined in the general notes to table 1 of Recommendation ITU-T Y.1541 [26], an evaluation interval of 1 minute is suggested for IPTD, IPDV, and IPLR and, in all cases, the interval shall be recorded with the observed value.

6.2 End-to-end QoS assessment of video service

The following clauses focus on all aspects of end-to-end QoS assessment of video service, namely on end-to-end video quality assessment, other end-to-end tests in the video channel, functional QoS parameters of the video channel and network performance parameters in the video channel.

6.2.1 End-to-end Video Quality assessment

In principle, the video quality assessment can be done using two completely different ways, namely testing on optical interface or testing with inserted and retrieved video files. As the first one is currently under development, it is left for father study. A detailed description of the well-established method (testing with inserted and retrieved video files) is given below.

6.2.1.1 Testing at Optical Interfaces

This is for further study.

6.2.1.2 [QoS_Video_filebased_01]: Testing with inserted and retrieved video files

In the following clauses, the testing with inserted and retrieved video files is described from different perspectives, namely measurement setup, measurement methodology, test signals and instrumental assessment of video quality.

6.2.1.2.1 Measurement setup

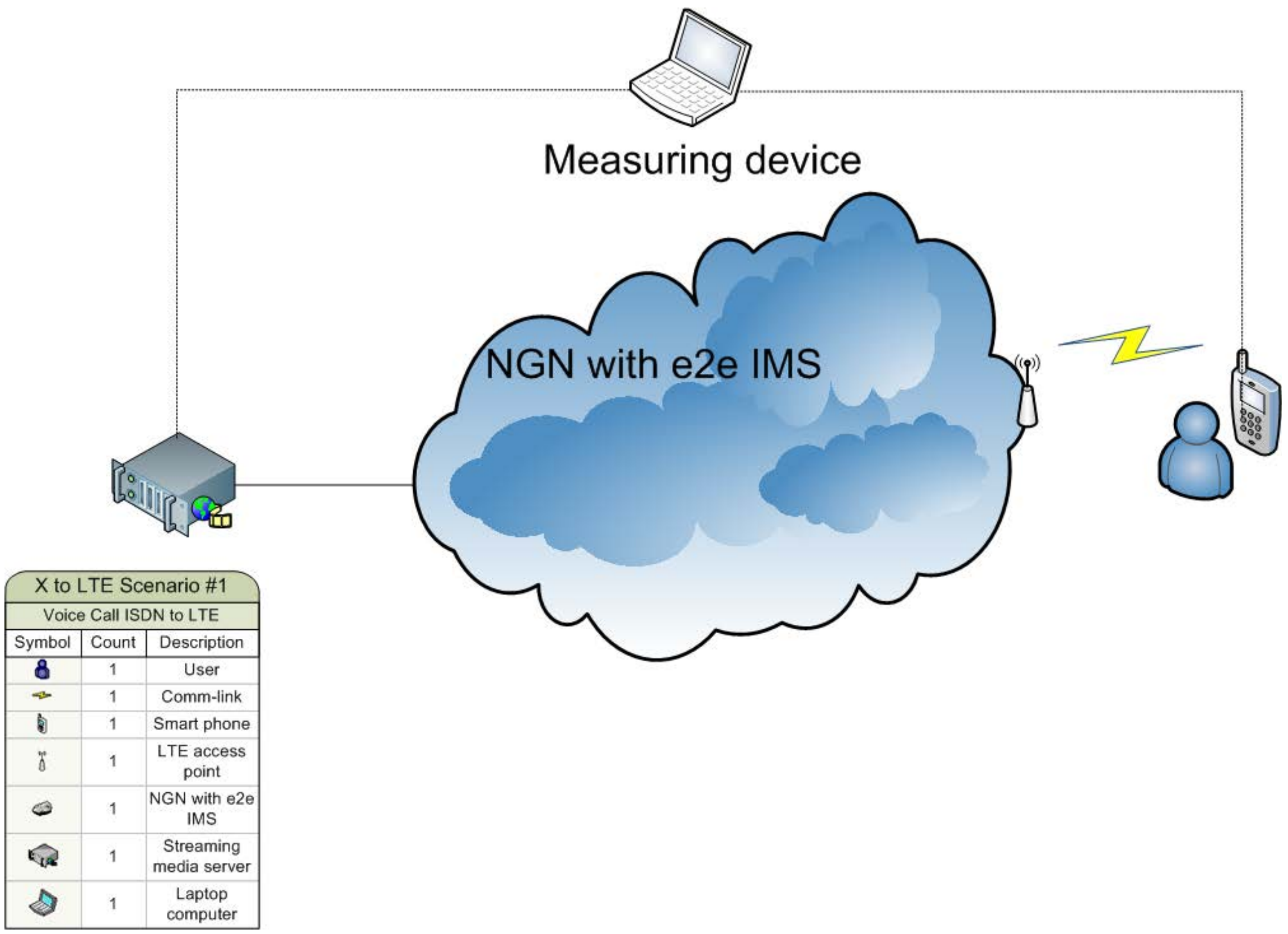


Figure 9: Measurement setup

A measurement setup is depicted in figure 9. The input signals described in clause 6.2.1.2.3 are transmitted and recorded simultaneously, that means that the retrieval process starts at the same time as the transmission process begins. Therefore exact delay assessment is possible.

6.2.1.2.2 Measurement methodology

In order to get statistically reliable results (to capture possible variations), a measurement should be done two times per each test signal (at least three different video sequences from a content perspective involved in the measurement). This approach provides at least six quality values characterized by DMOS scores (depends on a number of content types involved in the test) for each test condition (transmission from A to B involving a technology provided by corresponding vendor(s)), which will be averaged in post-processing and will form a global quality score describing an end-to-end media quality as perceived by the end user provided by a technology offered by particular vendor.

In addition to the DMOS values, the following parameters will be also captured by the quality prediction model (described in more detail in clause 6.2.1.2.4) in parallel with the corresponding DMOS value for each measurement:

- Delay variation over time
- Quality variation over time
- Jerkiness
- Blockiness
- Blur
- Chrominance
- Luminance
- Temporal Distortion
- Frozen Frames
- Skipped Frames

6.2.1.2.3 Test signals

Video test sequences with a length of at least ten seconds are recommended for this test. The selected video sequences should cover different types of content. In particular, it is proposed to use at least three different types of content ranging from mostly stable scenes to fast moving scenes. Naturally, a type of the content chosen for a test should be carefully described in a test report.

6.2.1.2.4 Instrumental assessment of video quality

A proprietary metric called Perceptual Evaluation of Video Quality (PEVQ) based on the PVQM model [24] is the most accurate full-reference video quality prediction model currently available.

The quality evaluation consists of five main stages. In the first stage, both the original and the distorted video signals are pre-processed by extracting the region of interest (ROI). The ROI is derived by cropping the actual frame, with a cropping size defined by the video format. These ROI-derived frames are used in subsequent stages. In stage No.2, the pre-processed video signals are spatially and temporally aligned. In stages No.3 and No.4, four spatial distortion measures, namely (edginess in luminance, edginess in chrominance, and two temporal variability indicators), as well as a temporal distortion measure are computed. In particular, a gradient filter is applied on both the luminance and chrominance part of the video signals to obtain the edge information. From the edge information for each frame, the normalized change in edginess for the distorted video signal with respect to the original video signal is computed and averaged over all frames to obtain the edginess in luminance and chrominance. The temporal variability of a frame is defined as the difference of :

- the absolute difference between the current and the previous frame of the original signal;
- the absolute difference between the current and the previous frame of the distorted signal.

The negative part of the temporal variability measures the new spatial information introduced in the signal, and the positive part of the temporal variability measures the effect of spatial information lost in the signal. The temporal distortion is computed from the amount of frame freezing as well as frame delay or loss information. A sigmoid approach to map the distortions to the DMOS video quality measure, with the mappings defined based on the input video format (QCIF, CIF, or VGA) is used in stage No.5. PEVQ was one of the two best performing methods in the VQEG Multimedia Quality Assessment, Phase I [25], and included as normative model in Recommendation ITU-T J.247 [7]. It should be noted here that PEVQ was validated for IP-based wireless transmission scenarios [7] and is therefore applicable for this test.

6.2.2 Other end-to-end tests in the video channel

This is for further study.

6.2.3 [QoS_Video_func_01]: Functional QoS parameters of the video channel

As can be seen in clause 5.5.2, only one functional QoS parameter of the video channel has been recommended in this case, namely streaming service access time. Its definition and abstract equation are available in clause 5.5.2. Table 13 shows trigger points, technical description and protocol parts for this parameter.

Table 13: Trigger Points for streaming service access time

Event from abstract equation	Trigger point from user's point of view	Technical description/protocol part
$t_{\text{stream request}}$: Time when stream is requested	Start: Stream request.	Start: • WAP 1.x, WAP 2.x: WSP Disconnect; • WAP 2.x: TCP SYN towards streaming platform.
$t_{\text{reception of first data packet}}$: Time when first data packet is received	Start: "Buffering" message.	Stop: Reception of first data packet.

The exact derivation of the trigger points from the signalling protocol under test in the plugtest has to be determined by the test house and shall be stated in the test report.

6.2.4 [QoS_Video_np_01]: Network performance parameters in the video channel

It is advised that it is beneficial to monitor the following parameters between UNI and UNI in both directions of transmission during the period when the end-to-end QoS parameters are being tested:

- UNI-to-UNI delay over time
- UNI-to-UNI jitter
- UNI-to-UNI packet loss

In case these parameters are measured in separately set-up connections, it is not possible to relate their values to the end-to-end QoS results.

As outlined in the general notes to table 1 of Recommendation ITU-T Y.1541 [26], an evaluation interval of 1 minute is suggested for IPTD, IPDV, and IPLR and, in all cases, the interval shall be recorded with the observed value.

7 Documentation and Interpretation of Test Results

The presentation of the end-to-end QoS results shall be done in a way which is also appropriate for non-experts, such as technical management staff.

Detailed interpretation of test results should be unambiguous and follow state-of-the-art statistical analysis and techniques. All necessary details shall either be contained in the final report or in related data bases.

Results should be presented in an anonymous report reproducing all obtained results during the measurement campaign or during the Plugtests.

7.1 Measurement configuration description

In the report, it is necessary to describe the analysis chain including network architecture and speech quality analyser implemented to carry out the measurements.

All information concerning the measurement chain (type of equipment, manufacturer and model of these equipments, firmware version and configuration) should be present in these reports. In the same manner, information concerning device tested (model and firmware ID, configuration during analysis) will be presented.

To respect the anonymity of manufacturers, all the results of the analysis, the information identifying the tested devices will be removed.

It is also necessary to introduce the methodology: Evaluated indicators, what they represent and the method to determine them. The means implemented to realize analyses (analyzers) as well as the measurement methods should be described.

7.1.1 Mandatory documentation of test results

In order to provide an easy means of comparison of tests carried out by different test houses, it is mandatory to provide all results in the format of the template given in Annex A of the present document.

7.2 One-view visualization of performances

In order to give a quick overview of all quality parameters, a specific representation (overview visualization defined in Recommendation ITU-T P.505 [16]) of the metric value will be used. This representation reveals at one glance the strengths and weaknesses of devices under test with respect to target values.

This specific presentation (Pie Diagram) provides an "aggregate" results view. This presentation is particularly adapted to show all results (all tested equipment performances), allowing an easy performance comparison of the tested equipment. The Pie Diagram makes also easier the comparison of the results when several configurations of the device under test are analyzed.

7.2.1 Creation of Quality Pies according to Recommendation ITU-T P.505

For each test case a Quality Pie according to Recommendation ITU-T P.505 [16] shall be created.

The numerous complex parameters that determine the quality of telecommunication equipment as well as end-to-end quality can be interpreted by technical experts only.

The tool specified in Recommendation ITU-T P.505 [16] provides a novel quality representation methodology which is easy to use and also easy to understand for non-experts and which can serve as a basis for commercial decisions on a management or marketing level with:

- Quick and easy recognition of expected speech quality problems for selected parameters (limit value violation).
- Assessment of strengths and weaknesses of signal processing implemented in a terminal or other telecommunication equipment, including end-to-end considerations (quality statement).

- Easy comparison of different equipment or connections based on the corresponding representations.
- Easy extension of the representation by new parameters relevant to quality in the future.

The one-view visualization methodology is based on the allocation of individual circle segments to the selected parameters - the so-called "quality pie"; a maximum number of 16 different segments is considered here for practical reasons.

The total number of parameters represented determines the size of the individual segments in the quality pie. The axes are shown with a common origin. The individual circle segments have the same size (spanned angle 360° divided by number of selected quality parameters).

The representation of individual segment sizes is not interdependent, thus guaranteeing the independence of the different quality parameters from each other, which leads to the following advantages:

- Independent representation of individual quality parameters.
- Segment sizes are determined by the number of selected parameters and are identical.
- Segment size (radius) is a measure for the quality regarding this parameter.
- A concentric circle around the origin is defined ($1/\sqrt{2}$) which represents a minimum quality measure; falling below this segment size (radius) indicates a non-compliance with this limit value.
- By means of a suitable colour selection results lying within the tolerance or transgressing the limit values can be easily visualized.

This online application of Recommendation ITU-T P.505 [16] available at ITU-T website (http://www.itu.int/net/itu-t/qualitypie/workplace/index.php) can help to produce high quality graphs for individual sets of parameters. It is intended to support the use of this methodology in the field, e.g. for recurring reporting tasks, but also for benchmarking or for test events.

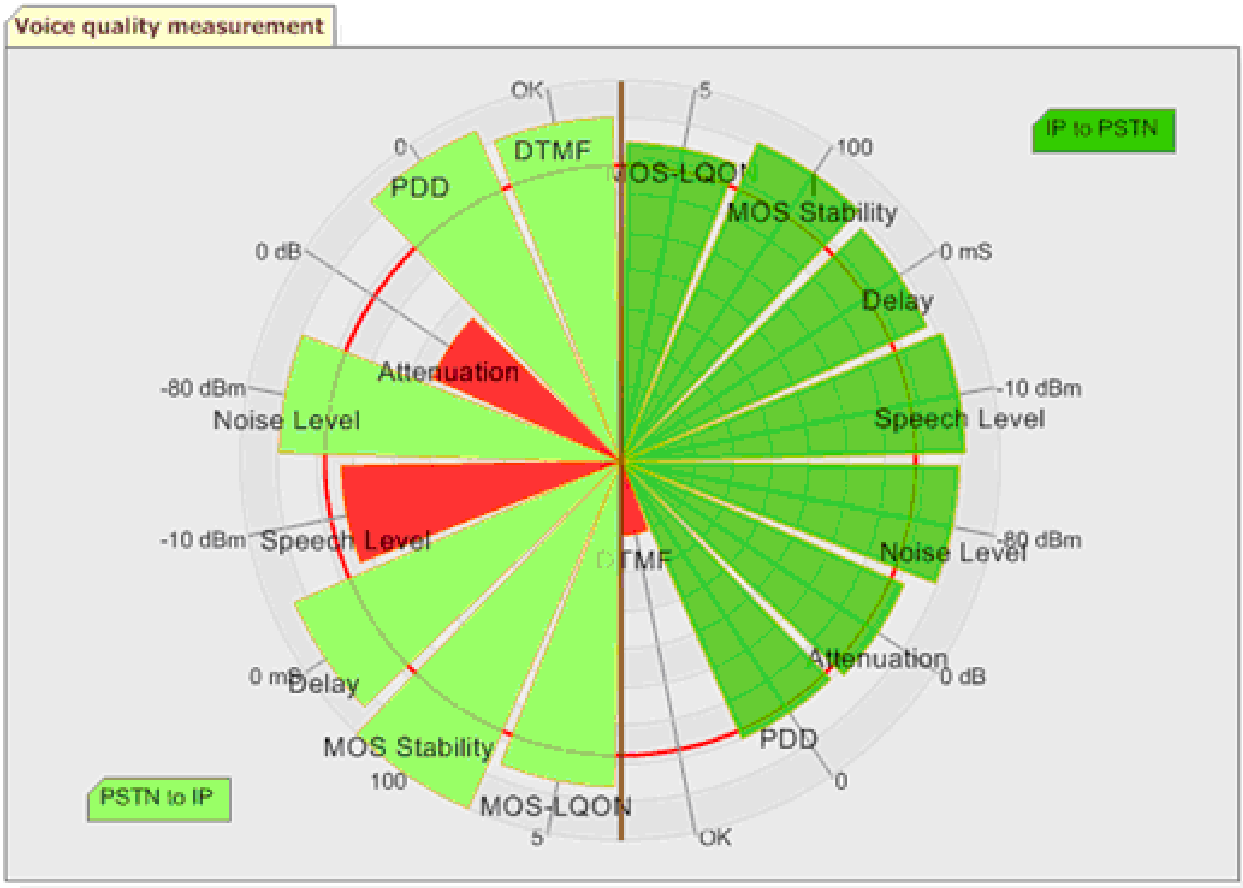


Figure 10: Example of results presented with the Pie Diagram presentation

In figure 10, results are presented for the two transmission paths (IP to PSTN and PSTN to IP) and for every indicator the measured value is compared to an acceptance threshold.

7.3 Visualization of performance variations

It is also important to present the variation of indicators. For the speech quality, other parameters captured in parallel with the corresponding MOS value for each measurement, like delay, etc. and network performance parameters, the indicators show the time variation within the same communication. In addition, average values of the monitored parameters obtained for each individual measurement should be also displayed in one graph in order to show the differences between successive analyses/measurements. The presentation of these indicator variations can be shown through graphs as those below.

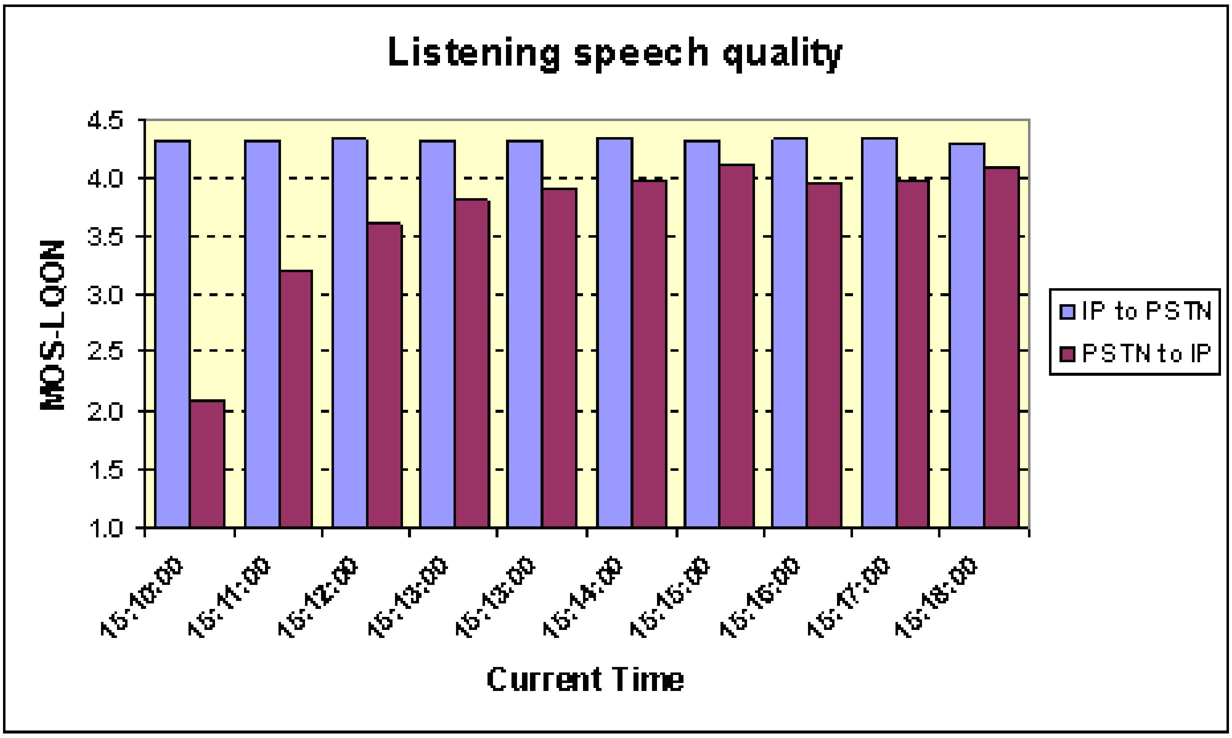


Figure 11: Example of graph for speech quality indicator

In figure 11, MOS scores (determined successively in the same call) are presented for the two transmission paths (IP to PSTN and PSTN to IP). For speech quality, the graphic presentation of MOS score values versus time completes the information given by the **speech quality stability indicator**.

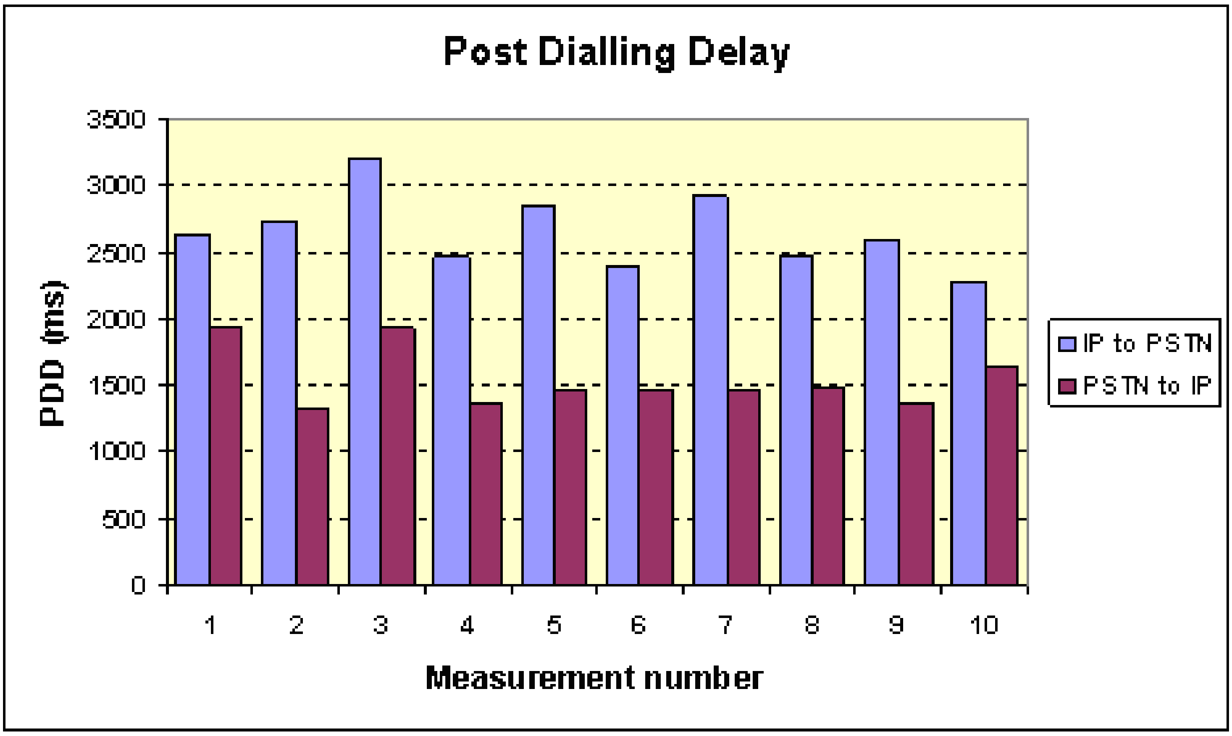


Figure 12: Example of graph for PDD

In figure 12, indicator values (determined by successive call attempts) are presented for the two call paths (IP to PSTN and PSTN to IP).

7.4 Result interpretations

An interpretation of the results would be presented in the report of the analysis. For each metric, it is necessary to explain the results obtained during the tests. The analysis of the results has to show the comparison of the performance levels (maximum, medium and minimum) of the tested devices.

Annex A (normative): Mandatory template for result reporting

This annex has an electronic attachment mandatory-result-template.xls contained in archive ts_103189v010102p0.zip which accompanies the present document.

History

Document history		
V1.1.1	October 2013	Publication
V1.1.2	October 2013	Publication