

**VoLTE QoS Assessment
Technology Evaluation Plugtests;
Sophia Antipolis, France;
18-20 November 2013**



ETSI

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

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

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

Important notice

Individual copies of the present document can be downloaded from:

<http://www.etsi.org>

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

Users of the present document should be aware that the document may be subject to revision or change of status.

Information on the current status of this and other ETSI documents is available at

<http://portal.etsi.org/tb/status/status.asp>

If you find errors in the present document, please send your comment to one of the following services:

http://portal.etsi.org/chaicor/ETSI_support.asp

Copyright Notification

No part may be reproduced except as authorized by written permission.
The copyright and the foregoing restriction extend to reproduction in all media.

© European Telecommunications Standards Institute 2013.
All rights reserved.

DECT™, **PLUGTESTS™**, **UMTS™**, **TIPHON™**, the TIPHON logo and the ETSI logo are Trade Marks of ETSI registered for the benefit of its Members.

3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners.

LTE™ is a Trade Mark of ETSI currently being registered for the benefit of its Members and of the 3GPP Organizational Partners.

Contents

1	Executive Summary	4
2	Introduction	5
3	Abbreviations	6
4	Participants	7
5	Technical and Project Management	8
5.1	Test Plan	8
5.1.1	Acoustic-to-acoustic	8
5.1.2	Acoustic-to-electric / Electric-to-acoustic	8
5.2	Test Schedule	9
5.3	Test Infrastructure	10
5.3.1	LTE / IMS Network	10
5.3.2	QoS Test Tools	13
5.3.3	Test Signals	13
5.3.4	Network Performance Test Tools	14
5.3.5	LTE devices	15
5.3.5.	User Equipments	15
5.3.5.2	Codecs	15
5.3.5.3	SIP Clients	15
5.4	Test Setups	15
5.4.1	Acoustic-to-acoustic	15
5.4.2	Acoustic-to-electric / Electric-to-acoustic	17
5.4.4	Electric-to-electric	18
5.4.4.2	VoIP Gateway and Headset interfaces	19
5.4.4.3	VoIP Gateway and Bluetooth interfaces	20
5.5	Support from STF 453	20
6	Achieved Results	21
6.1	Test Sessions	21
6.2	QoS tests	21
6.2.1	Acoustic-to-Acoustic Setup	21
6.2.2	Acoustic-to-Electric / Electric-to-Acoustic Setup	22
6.2.3	Electric-to-Electric Setup	24
6.2	Network performance	24
6.4	TS validation status	25
6.5	Findings	26
6.5.1	Findings in TS 103 189	26
6.5.2	Options for the electrical interface	27
Annex A Test Tool Report 28		
A.1.	Introduction	28
A.2.	Summary	28
A.2.1	Quality Pie Examples	29
A.3.	Description of Test Signals and Setup	29
A.3.1	Acoustic-to-Acoustic Setup	32
A.3.2	Acoustic-to-Electric/Electric-to-Acoustic Setup	33
A.3.3	Electric-to-Electric Setup	33
A.3.4	Headset Interface Level Adjustment	33
A.4.	Discussion of Results	34
A.4.1	Acoustic-to-Acoustic Setup	34
A.4.2	Acoustic-to-Electric / Electric-to-Acoustic Setup	37
A.4.3	Electric-to-Electric Setup	42
History	44

1 Executive Summary

ETSI's Centre for Testing and Interoperability organized the VoLTE QoS Assessment Technology Evaluation Plugtests from 18 – 20 November 2013 at ETSI premises in Sophia Antipolis, France. The event aimed at validating and evaluating the different methodologies described in ETSI TS 103 189 to assess the quality of speech of Voice over LTE calls.

The event was supported by several organisations, including:

- Com4Innov, who provided the access to a real size LTE / IMS test network, deployed in Sophia Antipolis, as well as a number of different LTE devices.
- HEAD acoustics, who provided the equipment required to run the end to end QoS assessment on both the acoustical and electrical interfaces
- AT4 wireless, who provided the test tools required to perform parallel network performance measures during the end to end QoS assessment sessions.

In addition, STF 453, from the ETSI Technical Committee for Core Networks and Interoperability Testing, provided QoS testing expertise and compiled the findings during the validation sessions that were used as input for a revision of ETSI TS 103 189.

The main outcome of the Plugtests was:

- The validation of 76% of the test cases in TS 103 189
- The identification of a number of gaps and potential improvements in TS 103 189, which include:
 - two new test setups, dealing with electrical-to-acoustic and acoustic-to-electrical configurations
 - the extension of existing test cases
 - the documentation of a procedure for electrical headset interface level adjustment
 - the documentation of the pertinence of testing in 2 directions in symmetrical configurations (acoustic-to-acoustic and electrical-to-electrical)
 - the need to report on the environmental conditions during the test sessions
 - some clarifications on the network performance parameters to be measured during the test sessions
 - some new tests addressing idle channel noise, delay vs. time and background noise

One of the conclusions of the Plugtests was the submission of a new Work Item to address those gaps and improvements in a revised version of TS 103 189.

2 Introduction

This Plugtests aimed to validate and evaluate the different methodologies described in ETSI TS 103 189 to assess the speech quality of speech of Voice over LTE calls.

The equipment used for the validation sessions included:

- A deployed LTE / IMS test network
- Several LTE devices from different manufacturers
 - 6 LTE capable smart phones
 - 2 LTE dongles
 - 1 LAN to LTE router
- test tools providing:
 - instrumental assessment of speech samples, as defined in ETSI TS 103 189
 - network performance measures, as defined in ETSI TS 103 189

4 different setups were used during the validation sessions:

- Acoustic-to-acoustic
- Electrical-to-acoustic / Acoustic-to-electrical
- Electrical-to-electrical

3 Abbreviations

Ac	Acoustic interface
AMR-WB	Adaptive Multi-Rate Wide Band
APN	Access Point Name
CSCF	Call Session Control Function
EI	Electric interface
EPC	Evolved Packet Core
IMS	IP Multimedia Subsystem
LTE	Long Term Evolution
MMTel	Multimedia Telephony
PCRF	Policy Charging Rules Function
QCI	QoS Class Identifier
QoE	Quality of Experience
QoS	Quality of Service
RTCP	Real-Time Transport Control Protocol
RTP	Real-time Transport Protocol
STF	Specialist Task Force
VoLTE	Voice over LTE

4 Participants

The organisations who contributed to the test event are listed in the table below by alphabetical order.

Company Name	Support provided
AT4 wireless	Network performance measurement tools
Com4Innov	LTE / IMS test network LTE devices
HEAD acoustics	QoS measurement tools
STF 453	QoS Testing expertise

5 Technical and Project Management

All the information presented in this chapter was prepared, reviewed and agreed during the conference calls held with participants to prepare the event.

5.1 Test Plan

The original test plan was based on TS 103 189, developed by TC INT (STF 453). During the regular conference calls which were held as part of the event preparation, organisations could propose additional tests, measures and setups. Eventually, the original test plan was extended with 2 new setups, and 22 additional tests cases and measures. Moreover, during the event preparation, different implementation options for each setup were investigated, agreed and prioritized.

The following clauses summarise the setups and test cases that were followed during the event.

5.1.1 Acoustic-to-acoustic

Test Name	Parameter	Origin
QoS_Voice_ac_01	MOS-LQO	TS 103 189 V1.1.2
	Delay variation over time	TS 103 189 V1.1.2
	MOS stability	TS 103 189 V1.1.2
QoS_Voice_ac_02	End-to-end frequency response	TS 103 189 V1.1.2
QoS_Voice_ac_03	Overall loudness rating	TS 103 189 V1.1.2
QoS_Voice_ac_04	End-to-end delay	TS 103 189 V1.1.2
QoS_Voice_ac_11	Idle channel noise	New
QoS_Voice_ac_12	Delay versus time	New
QoS_Voice_np_01	UNI-to-UNI delay over time	TS 103 189 V1.1.2
	UNI-to-UNI jitter	TS 103 189 V1.1.2
	UNI-to-UNI packet loss	TS 103 189 V1.1.2

5.1.2 Acoustic-to-electric / Electric-to-acoustic

Test Name	Parameter	Origin
QoS_Voice_ac-el/el-ac_01	MOS-LQO in sending/receiving direction	New
	Delay variation over time in sending/receiving direction	
	MOS stability in sending/receiving direction	
QoS_Voice_ac-el/el-ac_02	End-to-end frequency response in sending/receiving direction	New
QoS_Voice_ac-el/el-ac_03	End-to-end loudness rating in sending/receiving direction	New
QoS_Voice_ac-el/el-ac_04	End-to-end delay in sending/receiving direction	New
QoS_Voice_ac-e/el-ac_05	Overall echo attenuation	New
QoS_Voice_ac-el/el-ac_06	Echo Level vs. time	New

QoS_Voice_ac-el/el-ac_07	Spectral Echo Attenuation	New
QoS_Voice_ac-el/el-ac_08	Attenuation range in sending direction during double talk	New
QoS_Voice_ac-el/el-ac_09	Attenuation range in receiving direction during double talk	New
QoS_Voice_ac-el/el-ac_10	Detection of echo components during double talk	New
QoS_Voice_ac-el/el-ac_11	Idle channel noise in sending/receiving direction	New
QoS_Voice_ac-el/el-ac_12	Delay versus time in sending/receiving direction	New
QoS_Voice_ac-el/el-ac_13	Background noise transmission with far end speech (Car/Pub)	New

5.1.4 Electric-to-electric

Test Name	Parameter	Origin
QoS_Voice_el_01	MOS-LQO	TS 103 189 V1.1.2
	Delay variation over time	TS 103 189 V1.1.2
	MOS stability	TS 103 189 V1.1.2
QoS_Voice_el_02	End-to-end frequency response	TS 103 189 V1.1.2
QoS_Voice_el_03	Junction loudness rating	TS 103 189 V1.1.2
QoS_Voice_el_04	End-to-end delay	TS 103 189 V1.1.2
QoS_Voice_np_01	UNI-to-UNI delay over time	TS 103 189 V1.1.2
	UNI-to-UNI jitter	TS 103 189 V1.1.2
	UNI-to-UNI packet loss	TS 103 189 V1.1.2

5.2 Test Schedule

The preliminary test schedule was developed prior to the Plugtest and was circulated to all the participants in advance for comments. The initial test schedule allowed to test several variations of each test setup and to validate the test cases in scope with several LTE devices from different vendors. The day was organized in a morning test session from 8.30 to 12.30 and in an afternoon test session from 13.30 to 18.30. The very specific equipment required for running the test sessions made it not possible to run several test sessions in parallel.

During the test event the test schedule was constantly adapted according to the progress of the Plugtests setup, previous test sessions and findings. This was done during the daily wrap-up meetings at the end of each day and during regular informal updates with participants.

The figure below shows the last version of the test schedule as of Friday Wednesday the 20th of November.

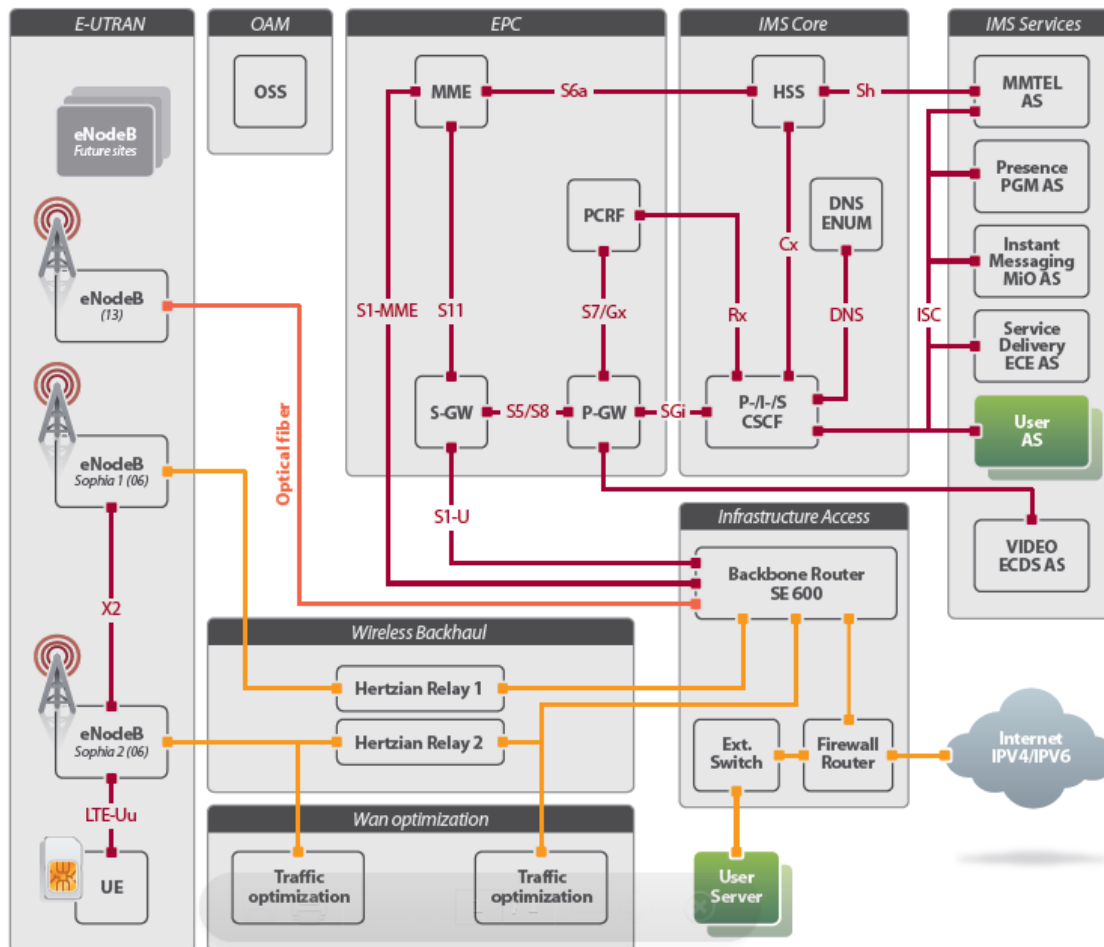
	Monday, 18/11/2013	Tuesday, 19/11/2013	Wednesday, 20/11/2013
8:30 -10:30	Equipment setup	Test Ac- Ac + Nw perf Ac-EI validation (dongle)	Test EI- Ac (router)
10:30 – 12:30	Ac-Ac validation UE installation	Test Ac- Ac + Nw perf EI-Ac validation (dongle)	Test Ac- EI (router)
13:30 – 15:30	Ac-EI /EI-ac validation (headset) Nw Perf validation	Test Ac- Ac + Nw perf EI-EI validation (dongle)	Test EI- EI (router)
15:30 – 17:30	EI-el setup validation (headset)	Test Ac- Ac + Nw perf EI-Ac validation (router)	Test EI- EI (router)
17:30 – 18:30	Wrap-up		

5.3 Test Infrastructure

5.3.1 LTE / IMS Network

The test network consisted on a 4G/LTE access network, an IMS Core and Application Servers supporting Voice over LTE. The test network was deployed in Sophia Antipolis and covered the Plugtests rooms, in ETSI premises over the air. The eNodeBs operated in the 2,6 Ghz frequency band (Band 7).

The architecture of the test network is depicted hereafter:



LTE/IMS Network architecture

In order to perform VoLTE calls, the network needs to support Policy management and QoS and be able to handle dedicated bearers, allow to limit congestion and to enhance the service quality.

QoS management principles in LTE :

Default bearers are of type non-guaranteed bit rate (non-GBR). For services that do not require IMS, the QoS parameter stems from the HSS configuration. For example Internet access will use a priority with QCI 9 (see Table below) . For services based on IMS, there is a link between the Application Server and the PCRF (which manages the QoS and billing policy.) Through this link, and the configuration associated with the requested service, the LTE network can establish the necessary dedicated bearers with the appropriate QoS (QCI value). The Default Bearer is obtained during the attach process, and ensures continuous IP connectivity.

Dedicated bearers are obtained on demand, when it is required to carry GBR for certain delay-sensitive services, such as voice or video. The same dedicated bearer is used in case of multiple concurrent voice sessions (as in call waiting, conference supplementary services, etc...).

In a VoLTE session, three bearers are used:

- Bearer with QCI 9: Default bearer created when connecting to the network.
- Bearer with QCI5: Bearer created for IMS signaling (SIP messages) as soon as the connection to the IMS APN is established.
- Bearer with QCI1: This bearer is created once the voice session has been successfully initiated, to ensure audio transport (RTP messages)

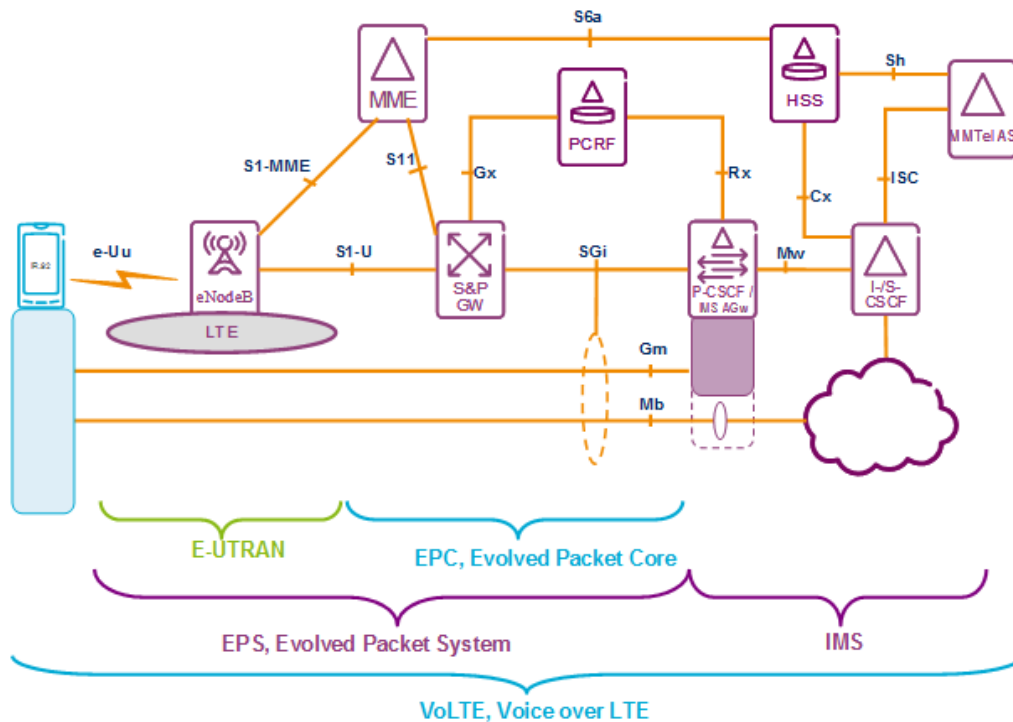
QCI	Resource Type	Priority	Packet Delay Budget	Packet Loss Rate	Example Services
1	GBR	2	100 ms	10^{-2}	Conversational Voice
2		4	150 ms	10^{-3}	Conversational Video (Live Streaming)
3		3	50 ms	10^{-3}	Real Time Gaming
4		5	300 ms	10^{-6}	Non-Conversational Video (Buffered Streaming)
5	Non-GBR	1	100 ms	10^{-6}	IMS Signalling
6		6	300 ms	10^{-6}	-Video (Buffered Streaming) -TCP-based (for example, www, e-mail, chat, FTP, P2P file sharing, progressive video, and so forth)
7		7	100 ms	4^{-3}	-Voice -Video (Live Streaming) -Interactive Gaming
8		8	300 ms	10^{-6}	-Video (Buffered Streaming)
9		9	300 ms	10^{-6}	-TCP-based (for example, www, e-mail, chat, FTP, P2P file sharing, progressive video, etc).

Standardized QCI Characteristics (abstract from 3GPP TS 23.203)

LTE/IMS – VoLTE infrastructure:

In addition to the LTE network, Voice over LTE requires an IMS infrastructure and its components, like the MMtel Application Server that provides all the functions related to Voice application (supplementary services) for mobile equipment.

IMS is a conceptual framework providing IP multimedia services to mobile subscribers such as Voice over LTE. The figure below depicts the various interfaces between components of the network that are involved in a VoLTE application.



IMS Services for Voice over LTE

5.3.2 QoS Test Tools

During the test event two artificial head measurement systems were used for the acoustic-to-acoustic end-to-end tests. These HATS according to ITU-T Recommendation P.58 were equipped with two ITU-T P.57 recommended artificial ears (type 3.3) and artificial mouth. A handset positioning device was used to mount the handsets to the artificial ear. All the tests were carried out using an 8 N application force between the different UEs and the artificial ear.

For the acoustic-to-electric/electric-to-acoustic tests one artificial head was substituted by an electric access point to the test network that, during the event, was implemented with a measurement frontend (reference IP gateway) together with a “LAN-to-LTE” router to access the LTE test network. In this test setup the sending direction was defined from the “UE point of view”, i.e. as the direction from the acoustic interface (microphone) to the electric interface. And the receiving direction was defined from the electric assessment point to the UE mounted on the artificial head (loudspeaker).

Tests between two electrical access points (electric-to-electric tests) could be carried out between the measurement frontend (reference IP gateway) in combination with the “LAN-to-LTE” router as described above on one side and one headset interface to an UE on the other side. Alternatively the Bluetooth interface for wideband speech transmission as described by the Bluetooth SIG Telephony Working Group in “Hands-free Profile (HFP) 1.6 specification” could have been used. Unfortunately, the UEs tested during the event did not provide wideband Bluetooth capability.

5.3.3 Test Signals

The following test signals and voice samples were used during the event.

- Composite Source Signal bursts (ITU-T Recommendation P.501) for short term delay measurements.
- Periodical repetition of Composite Source Signal bursts (ITU-T Recommendation P.501) with an overall duration of 120 s for delay analysis vs. time.
- Speech sequence consisting of eight sentences (British English, ITU-T Recommendation P.501) for average analyses like frequency response or loudness rating determination and MOS-LQO analysis.
- A subset of two sentences (one female voice and one male voice), the male voice providing the lowest pitch frequency the female voice showing the highest energy in the upper frequency range among the eight British English test sentences.

- A Voiced sound of the Composite Source Signal applied with decreasing and increasing test signal level to determine AGC characteristics or PLC implementations.
- A “compressed speech” signal from ITU-T Recommendation P.501 used for echo attenuation analysis.
- Further uncompressed speech samples taken from ITU-T P.501 for echo level vs. time analyses.
- A combination of uncorrelated Composite Source Signals for double talk analyses.
- Two background noise signals (car noise, pub noise, ETSI EG 202 396-1) played back via the artificial mouth at the near end coincident to the application of far end speech signals.

Additional details on the test signals can be found in Annex A.

5.3.4 Network Performance Test Tools

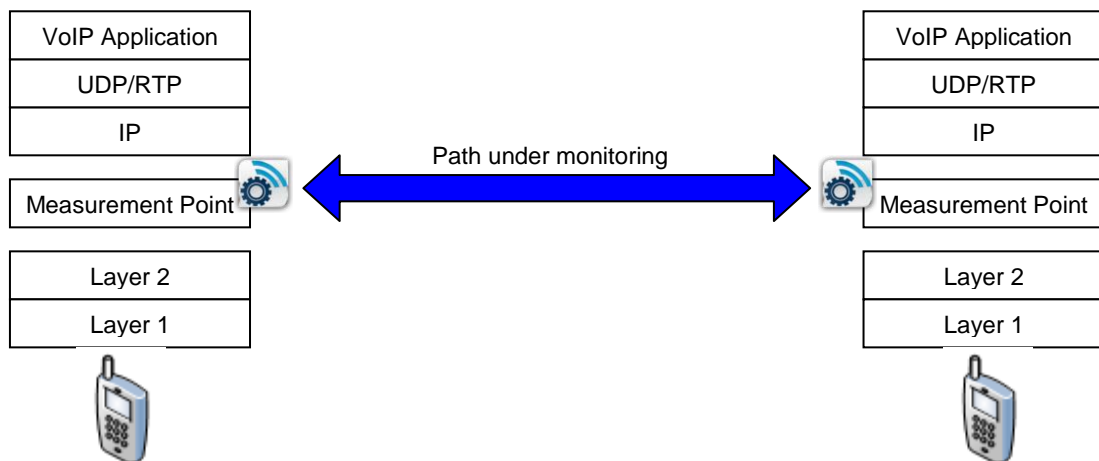
A dedicated tool was used for monitoring the performance parameters between UNI and UNI interfaces in both directions of transmission during the period while the end-to-end QoS parameters were being tested.

The performance tool was composed of two elements.

- **Controller:** The controller software was installed in a laptop, in the control room (B1B) and provided the GUI for the system. The controller performed the post-processing of the statistics captured by the agents, computed the KPIs and generated reports
- **Agent:** The software agents were installed in the UEs (smart phones) that were placed in the quiet rooms (B1A and B2). The agent captured the VoIP data from the UE and relayed it to the controller

The usage of the performance tools during this Plugtests is depicted in section 5.4.1.

In terms of OSI model the measurement points of the performance tool were located between layer 2 and layer 3 of the UEs. The next figure shows the points where the performance parameters were measured in the UEs protocol stacks.



5.3.5 LTE devices

5.3.5. User Equipments

The table below summarizes the different terminals that were used to validate TS 103 189:

Id	Vendor	MSISDN	IMPI	IMPU
UE4	V1	+336 38 06 00 10	33638060010@com4innov.com	sip:+33638060010@com4innov.com
UE4'	V1	+336 38 06 00 11	33638060011@com4innov.com	sip:+33638060011@com4innov.com
UE5	V2	+336 38 06 00 07	33638060007@com4innov.com	sip:+33638060007@com4innov.com
UE6	V3	+336 38 06 00 42	33638060042@com4innov.com	sip:+33638060042@com4innov.com
UE6'	V3	+336 38 06 00 47	33638060047@com4innov.com	sip:+33638060042@com4innov.com
UE8	V4	+336 38 06 00 46	33638060046@com4innov.com	sip:+33638060046@com4innov.com

In addition, an LTE VoIP Router was used to provide LTE access to the VoIP Reference Gateway used to test at the electric interface. .. The router allowed to access to the 4G network via Wifi or LAN interface. For the tests purposes, the reference VoIP Gateway was connected to the LAN interface of the router.

5.3.5.2 Codecs

The goal of the event was to validate TS 103 189 before native VoLTE user equipment are available for interoperability and end-to-end QoS testing, in order to have a stable and validated version of the TS published at the time when the native VoLTE UEs will be ready for testing. For that reason, native VoLTE UEs could not be used for the validation.

However, during the event preparation, and in order to obtain meaningful results, it was decided to target the codec that is expected to be used by native UEs: AMR-WB.

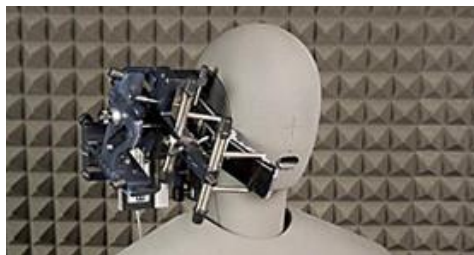
5.3.5.3 SIP Clients

Several SIP clients using AMR-WB codec were identified. A few of them were pretested in the test network and eventually one of them was selected and installed on all the UEs and used in all the validation sessions.

5.4 Test Setups

5.4.1 Acoustic-to-acoustic

This setup was implemented as described in TS 103 189. The acoustical access to terminals was a realistic simulation of the average subscriber. This was realized by using two HATS (Head And Torso Simulator) with appropriate ear simulation and artificial mouth and dedicated means to fix handset and headset terminals in a realistic and reproducible way.



The simulators and UEs were placed in 2 isolated and quiet rooms (B1A and B2) and controlled from a central room (B1B). All the participants could observe the results and listen to the recorded samples from this central room.

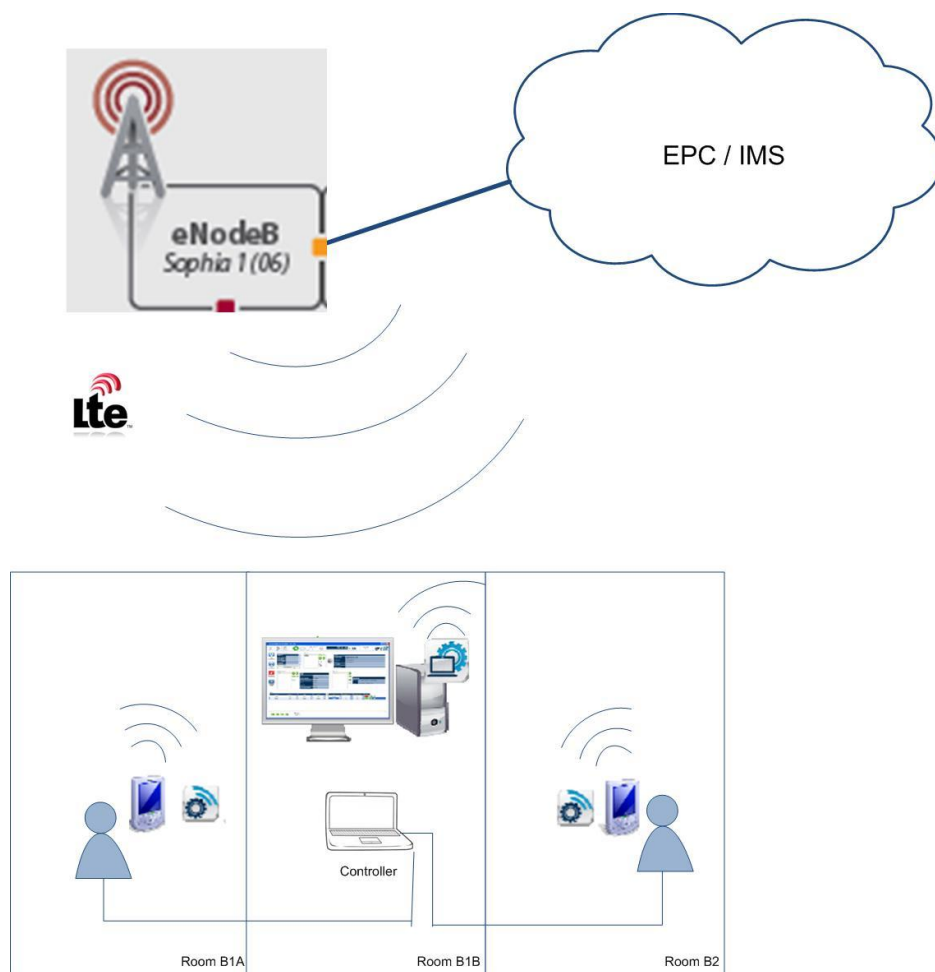
Note that these test rooms were not acoustically treated. However, considering the fact that the tests were carried out in handset mode with loudspeaker and microphone positioned close to the artificial ear respectively artificial mouth, the

room influence (reverberation) was minimized. The background noise level in both rooms could not be exactly controlled; however, air conditioning, ventilation and all other noise sources were disabled as far as possible. These environmental conditions were absolutely sufficient in order to validate the tests according to TS 103 189.

Tests using this acoustic to acoustic setup were only carried out in one direction, i.e. from room B1A (sending side) to room B2 (receiving side). Conversational aspects like double talk or echo tests were carried out in an acoustic-to-electric/electric-to-acoustic setups. If a different UE was used on both sides and both directions should be tested, the UE itself was exchanged between the two rooms.

The voice call was established between 2 LTE user equipments connected to the LTE / IMS network. During the audio session, 3 EPS dedicated bearers were allocated for each UE (as described in 3GPP TS 23.203 for standardized QCI characteristics) :

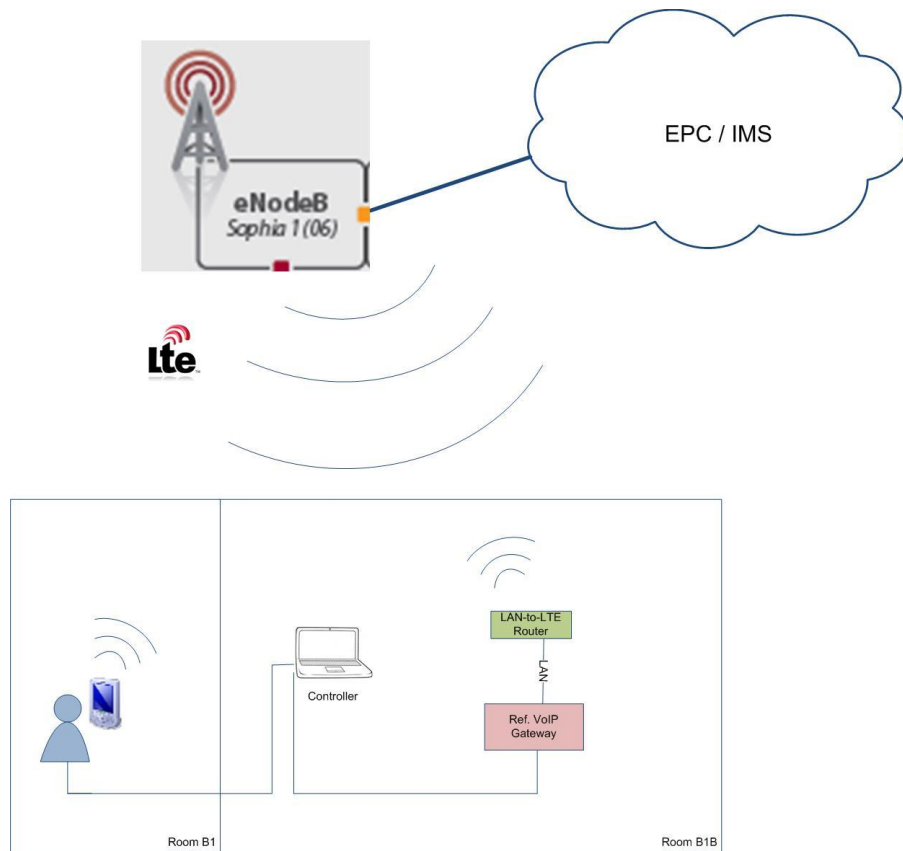
- QCI 1 : Audio flow (RTP)
- QCI 5 : IMS signalling (SIP)
- QCI 9 : default bearer - created as soon as a connection is done to the network and the PDN-GW give an IP address to the UE



A dedicated Android application had been previously installed on the LTE terminals. The applications were controlled remotely from the control room (B1B) and allowed to take performance measures during the instrumental speech assessment.

5.4.2 Acoustic-to-electric / Electric-to-acoustic

These setups were discussed during the preparation of the event and will be submitted as a potential improvement in TS 103 189. The acoustical access to the terminal in the receiver side is implemented via one HATS using its artificial ear and artificial mouth.



On the far end, the second HATS and UE were replaced by a Reference VoIP Gateway connected to the LTE network via a LAN-to-LTE router.

The voice call was established between the Reference VoIP Gateway and the user equipment placed in the isolated room. 3 dedicated bearers were allocated as described in previous sections.

In this setup different kinds of tests can be performed:

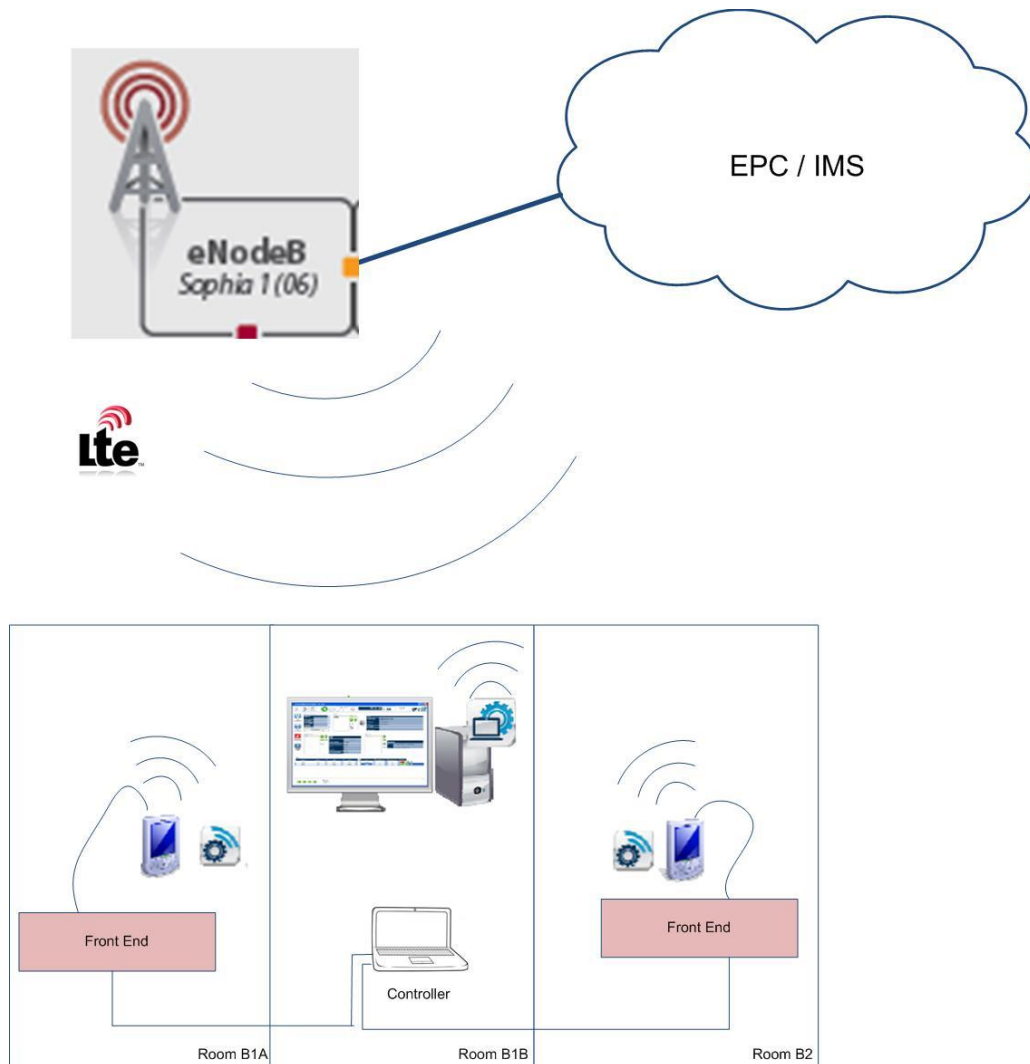
- Tests in sending direction: Signals played back via the artificial mouth of the HATS, captured and transmitted by the microphone path of the UE and measured electrically at the far end side using the frontend.
- Tests in receiving direction: Signals played back via the frontend on the far end side, transmitted to the UE, played back at the loudspeaker of the UE and measured using the artificial ear of the HATS.
- Echo tests: Signals played back via the frontend on the far end side, transmitted to the UE, played back at the loudspeaker of the UE and coupled into the microphone path of the UE. This coupled signal is processed by the echo cancellation algorithm implemented in the UE. The residual echo components transmitted through the network to the far end side are again measured using the frontend.
- Double talk tests: Signals are simultaneously transmitted in sending and receiving direction and either the transmitted send signal or the transmitted receive signal can be measured in order to perform analyses under double talk conditions.

5.4.4 Electric-to-electric

This setup is described in TS 103 189. During the event preparation several possible implementations of the access to the electrical interface were identified. Finally the 2 following ones were tested.

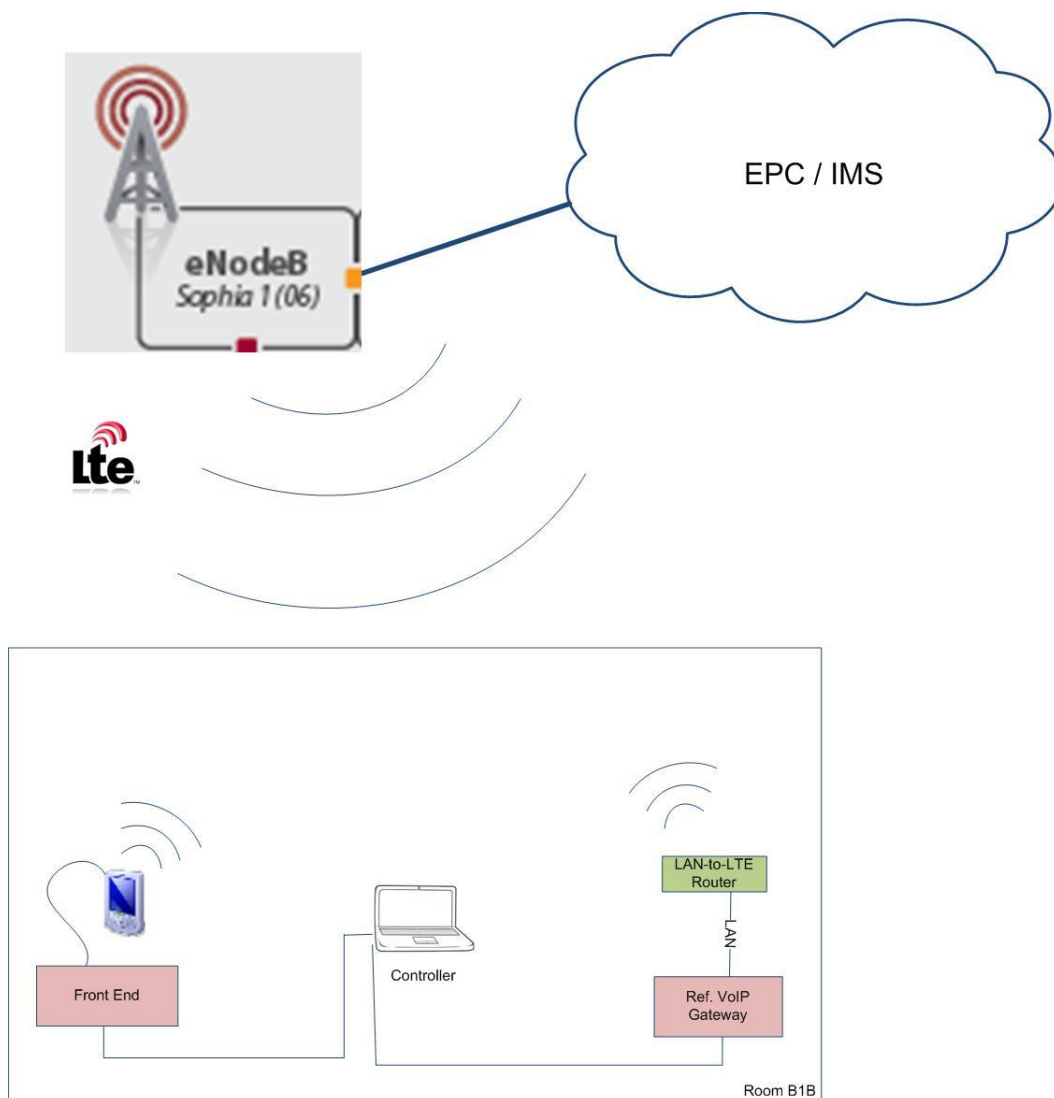
5.4.4.1 Headset interfaces

This is a symmetric setup based on the acoustic-to acoustic one, where the electric access to the terminals is implemented via headset interfaces on both sides.



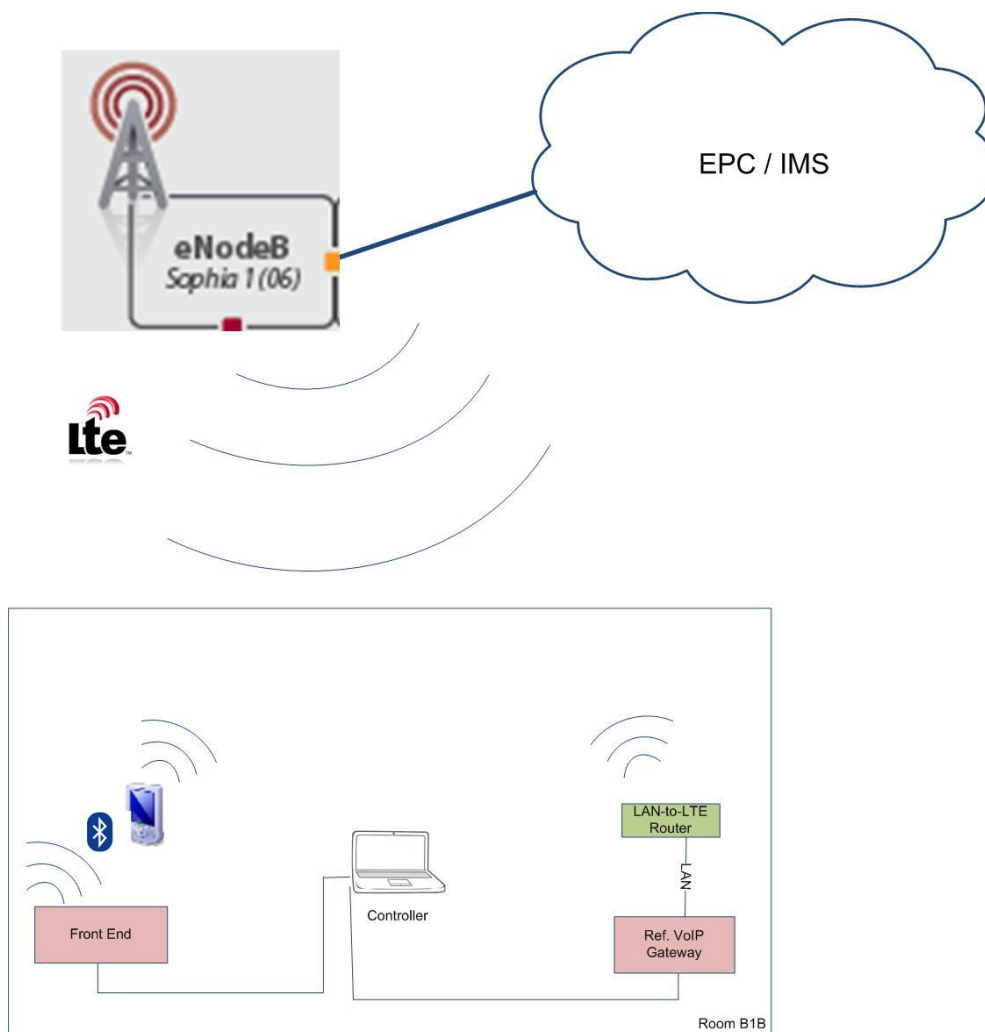
5.4.4.2 VoIP Gateway and Headset interfaces

This is an asymmetric setup based on the electrical-to acoustic setup providing a defined POI at the far end side. The electric access to the terminal is implemented via headset interface cable on one side, while on the other side a reference VoIP Gateway is connected to the network via a LAN-to-LTE router.



5.4.4.3 VoIP Gateway and Bluetooth interfaces

This is an asymmetric setup based on the electrical-to acoustic setup providing a defined POI at the far end side. The electric access to the terminal is implemented via a Bluetooth connection to a Bluetooth wideband capable and transparent UE on one side using a reference Bluetooth access point, while on the other side a reference VoIP Gateway is connected to the network via a LAN-to-LTE router. The Bluetooth wideband transparency of the UE should be verified separately before the actual QoS tests. This solution was not verified due to lack of Bluetooth Wide Band capable User Equipemnts.



5.5 Support from STF 453

ETSI STF 453 provided QoS expertise and supported ETSI Centre for Testing and Interoperability (CTI) during the different phases of the Plugtests namely planning, preparation and execution phase. QoS experts also provided the link between the two technical committees that were at the origin of this event (TC INT and TC STQ), which allowed CTI to find the required tool support for the validation of the TS.

During the event they provided guidance on the test execution and compiled the feedback and findings from the different participants that is intended to be used as an input for TS 103 189 revision.

6 Achieved Results

6.1 Test Sessions

The results described in the following sections were obtained through a number of different combinations of user equipments, setups and methods. The table below summarizes the ones that were used to validate TS 103 189

Setup	Method	Nw perf	Send	Receive
Acoustic-to-acoustic	-	No	UE8	UE4
			UE4	UE8
			UE6	UE6'
		Yes	UE8	UE6
			UE4	UE4'
			UE8	UE5
Acoustic-to-electrical	Headset	No	UE8	UE4
	Gateway		UE4	-
Electrical-to-Acoustic	Headset	No	UE8	UE4
	Gateway		-	UE4
Electrical-to-electrical	Headset (x2)	No	UE8	UE4
			UE4	UE8
	Gateway + Headset	No	-	UE4
	Headset + Gateway		UE4	-

6.2 QoS tests

6.2.1 Acoustic-to-Acoustic Setup

The MOS-LQO results and other calculated one-dimensional parameters for the end-to-end acoustical setup are listed exemplarily for two symmetrical test setups in the two tables below. The complete test results of the VoLTE QoS Assessment Plugtest can be found in Annex A.

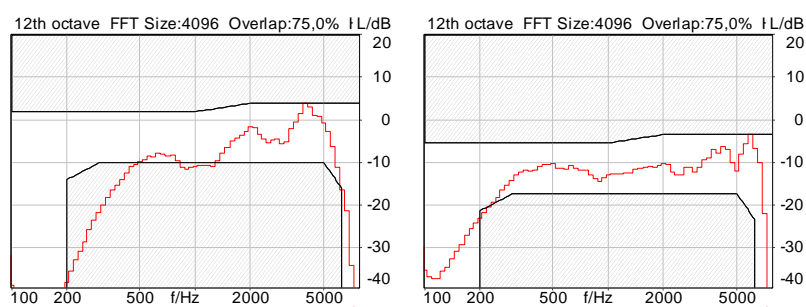
	UE4/UE4'	UE8/UE8'
MOS-LQO (sentence pairs)	3.0	3.4

	UE4/UE4'	UE8/UE8'
Setting x/y	2 / 6	3 / 6
OLR (8 sent.)	9.4 dB	10.2 dB

Δ MOS-LQO (sentence pairs)	2.8 – 3.4	3.1 – 3.9	OLR (2 sent.)	9.4 dB	10.3 dB
Delay [P.863]	426.1 – 427.8 ms	632.7 – 638.1 ms	Idle noise level	-56.6 dB _{Pa} (A)	-56.8 dB _{Pa} (A)
			Delay	421.0 ms	654.4 ms

The MOS-LQO scores differ significantly between the two UE. Furthermore they also differ significantly between the single sentences respectively the different sentence pairs for one UE. The loudness and the idle noise level are in a reasonable range for both setups. The end-to-end delay is high for both setups but significantly higher for the UE8 compared to the UE4 setup.

The generally high delay is most likely due to the fact that not an integrated VoLTE client but a VoIP client installed on the application layer of the phone was used for the evaluation session.

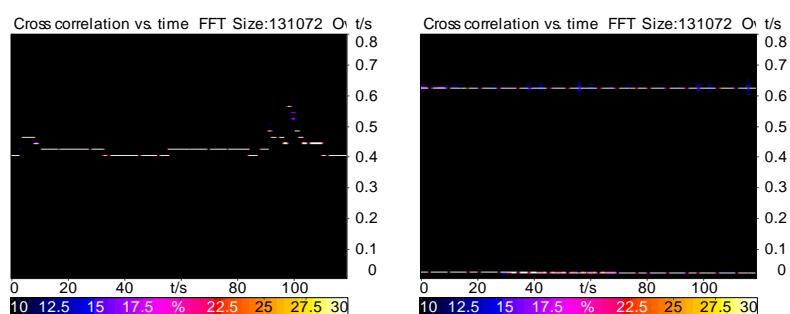


The corresponding overall frequency responses are shown in the left hand diagrams. The curve for the UE4 setup shows much stronger limitation than for the UE8 setup. This is the main reason for the lower MOS-LQO.

UE4 to UE4'

UE8 to UE8' (UE5)

The end-to-end delay was further determined vs. time in order to verify the stability and temporal changes of the delay parameter.



UE4 to UE4'

UE8 to UE8' (UE5)

The two connections lead to very different results. An average delay of approx. 420 ms could be determined for UE4. However, relatively strong delay variations can be observed. The delay increases up to approx. 600 ms at maximum.

Vice versa the use of UE8 on both sides leads to a very high delay of more than 600 ms. This delay is stable vs. time. A potential reason could be the implemented jitter buffer control in this UE that covers all varying delay from the network at the price of a very high absolute delay.

6.2.2 Acoustic-to-Electric / Electric-to-Acoustic Setup

In the following analyses carried out on the UE4 in the acoustic-to-electric / electric-to-acoustic setup are presented.

Receiving	UE4
MOS-LQO sending (sentence pairs)	3.6
MOS-LQO receiving (sentence pairs)	3.2
MOS-LQO "mouth-to-ear" (sentence pairs)	3.0

The MOS-LQO in sending direction was determined to 3.6. A lower result of 3.2 MOS-LQO was measured for the receiving direction. For comparison the MOS-LQO result from the acoustic-to-acoustic setup (mouth-to-ear) is also given. The three results indicate that the receiving direction of a UE can be seen as the limiting factor for listening speech quality.

The tables below show further one-dimensional results measured in both directions.

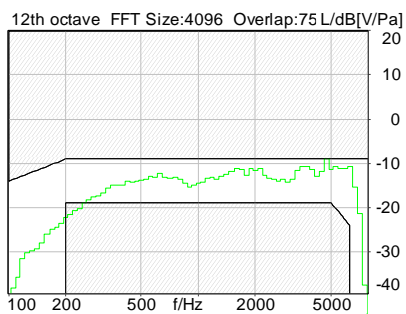
Sending	UE4
(SLR+JLR) (2 sent.)	8.6 dB
Delay sending	401.1 ms

Receiving	UE4
(JLR+RLR) (2 sent.)	1.0 dB (3/6)
Idle noise level	-62.3 dB _{Pa} (A)
Delay receiving	288 ms

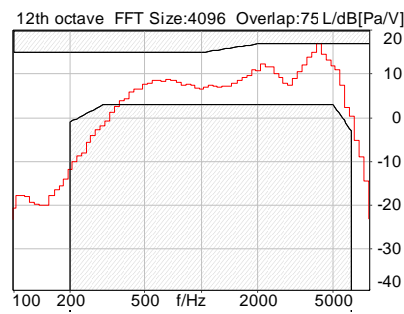
The measured frequency responses in the acoustic to electric setup and the electric-to-acoustic setup respectively are given in the following diagrams. For comparison the mouth-to-ear frequency response derived in the acoustic-to-acoustic setup is indicated below.

The comparison of the three curves indicate that the mouth-to-ear frequency response (acoustic-to-acoustic) can be estimated as a linear combination of the sending and receiving frequency responses measured in the acoustic-to-electric and electric-to-acoustic setup. Furthermore it can be seen that the main spectral limitation is due to the receiving side of UE4. The sending side is rather transparent only introducing some additional attenuation in the low frequency range (high pass characteristic).

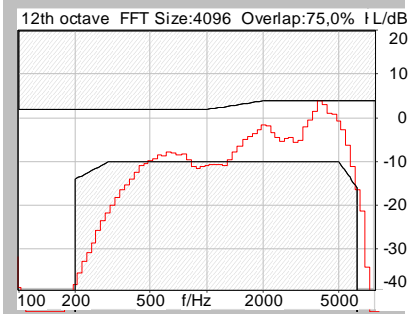
**Acoustic-to-electric “sending”
frequency response**



**Electric-to-acoustic “receiving”
frequ. response**



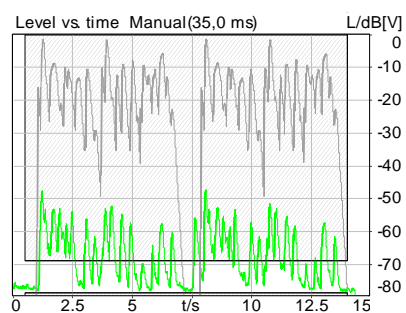
**Acoustic-to-acoustic “mouth-to-ear”
frequ. resp.**



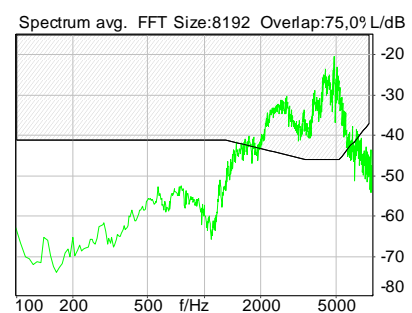
The echo performance of the device under test is described by three analysis results: the overall echo attenuation, the echo level vs. time and the spectral echo attenuation.

One-dimensional results	
Overall echo attenuation	36.7 dB

Echo level vs. time



Spectral echo attenuation



As demonstrated by all three analysis results significant residual echoes are detected for the UE4. This is most likely due to the fact that the used VoIP application does not have access to the echo control implemented in the DSP of the phone but only provides in-built software for echo cancellation. It is to be expected that real VoLTE phones will provide more reliable echo performance.

6.2.3 Electric-to-Electric Setup

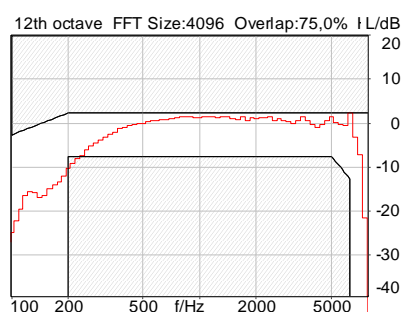
The results of the tests described in the TS 103 189 specification are summarized in the table and the diagrams below.

	UE4 to VoIP Ref. Interface	UE4 to UE8 (Headset Interface)	UE8 to UE4 (Headset Interface)
MOS-LQO (sentence pairs)	3.9	3.7	3.0
Delay [P.863]	298 – 368 ms	703 – 705 ms	404 – 460 ms
Junction Loudness Rating	JLR = 0.0 dB	---	---

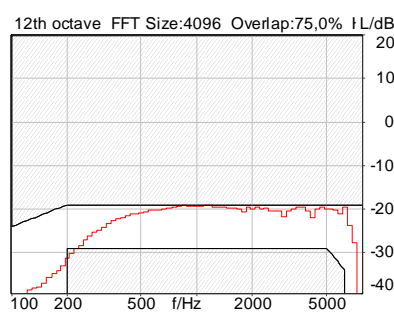
The quality scores calculated by POLQA are high if the first test setup is used. This could be expected since the network should not significantly degrade the signal transmission.

Using another UE at the far end side leads to degradations of the quality scores depending on the receiving characteristics of the UE. This is additionally confirmed by the frequency response representations below.

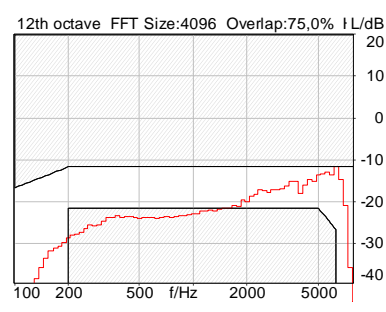
UE4 to MFEV8.1



UE4/UE8 (Headset Int.)



UE8/UE4 (Headset Int.)



6.2 Network performance

Network performance measures and end-to-end QoS instrumental assessment were performed by two independent tools, which required the implementation of a manual synchronization process to enable the correlation of the results collected from both tools.

The average values of the network performance parameters monitored during the test sessions are listed in the table below.

Send	Receive	One-way Delay	Delay variation	Packet Loss Average
UE8	UE6	28.237 ms	5.983 ms	0.056 %
UE4	UE4'	22.629 ms	6.313 ms	0,000 %
UE8	UE5	23.557 ms	5.662 ms	0,016 %

6.4 TS validation status

Overall, more than 75% of the test cases described in TS 103 189 were successfully validated during the event, see details in the validation status for each of them in the table below. In total, 9 test cases could not been tested. Five of them have not been tested due to test equipment not being available and the rest due to the non-availability of the test house. They are planned to be tested later. All the validated test cases can be applied during VoLTE and RCS Test Events.

Test case #	Clause	Parameter	Status, see Note
QoS_Voice_ac_01	6.1.1.1	POLQA®	1
QoS_Voice_ac_02	6.1.2.1.2.1	e2e frequency response	1
QoS_Voice_ac_03	6.1.2.1.2.2	Overall loudness rating	1
QoS_Voice_ac_04	6.1.2.1.2.3	End-to-end delay	1
QoS_Voice_ac_05	6.1.2.1.2.4.1	Overall Echo Attenuation	2
QoS_Voice_ac_06	6.1.2.1.2.4.2	Echo Level versus Time	2
QoS_Voice_ac_07	6.1.2.1.2.4.3	Spectral Echo Attenuation	2
QoS_Voice_ac_08	6.1.2.1.2.5.1	Attenuation Range in Send Direction during Double Talk	2
QoS_Voice_ac_09	6.1.2.1.2.5.2	Attenuation Range in Receive Direction during Double Talk	2
QoS_Voice_ac_10	6.1.2.1.2.5.3	Detection of echo components during double Talk	2
QoS_Voice_ac_11	6.1.2.1.2.6	Idle channel noise	1
QoS_Voice_ac_12	6.1.2.1.2.7	Delay versus time	1
QoS_Voice_ac_el_01	6.1.1.2	POLQA®	1
QoS_Voice_ac_el_02	6.1.2.2.2.1	e2e send frequency response	1
QoS_Voice_ac_el_03	6.1.2.2.2.2	e2e send loudness rating	1
QoS_Voice_ac_el_04	6.1.2.2.2.3	End-to-end delay	1
QoS_Voice_ac_el_05	6.1.2.2.2.4.1	Overall Echo Attenuation	1
QoS_Voice_ac_el_06	6.1.2.2.2.4.2	Echo Level versus Time	1
QoS_Voice_ac_el_07	6.1.2.2.2.4.3	Spectral Echo Attenuation	1
QoS_Voice_ac_el_08	6.1.2.2.2.5.1	Attenuation Range in Send Direction during Double Talk	1
QoS_Voice_ac_el_09	6.1.2.2.2.5.2	Attenuation Range in Receive Direction during Double Talk	1
QoS_Voice_ac_el_10	6.1.2.2.2.5.3	Detection of echo components during double Talk	1
QoS_Voice_ac_el_11	6.1.2.2.2.6	Idle channel noise	1
QoS_Voice_ac_el_12	6.1.2.2.2.7	Delay versus time	1
QoS_Voice_ac_el_13	6.1.2.2.2.8	Background Noise Test	1
QoS_Voice_el_ac_01	6.1.1.3	POLQA®	1
QoS_Voice_el_ac_02	6.1.2.3.2.1	e2e receive frequency response	1
QoS_Voice_el_ac_03	6.1.2.3.2.2	e2e receive loudness rating	1
QoS_Voice_el_ac_04	6.1.2.3.2.3	End-to-end delay	1
QoS_Voice_el_ac_05	6.1.2.3.2.4.1	Overall Echo Attenuation	1
QoS_Voice_el_ac_06	6.1.2.3.2.4.2	Echo Level versus Time	1
QoS_Voice_el_ac_07	6.1.2.3.2.4.3	Spectral Echo Attenuation	1
QoS_Voice_el_ac_08	6.1.2.3.2.5.1	Attenuation Range in Send Direction during Double Talk	1
QoS_Voice_el_ac_09	6.1.2.3.2.5.2	Attenuation Range in Receive Direction during Double Talk	1
QoS_Voice_el_ac_10	6.1.2.3.2.5.3	Detection of echo components during double Talk	1
QoS_Voice_el_ac_11	6.1.2.3.2.6	Idle channel noise	1
QoS_Voice_el_ac_12	6.1.2.3.2.7	Delay versus time	1
QoS_Voice_el_ac_13	6.1.2.3.2.8	Background Noise Test	1
QoS_Voice_el_qualification	6.1	Qualification of electrical interfaces	1
QoS_Voice_el_01	6.1.1.4	POLQA®	1
QoS_Voice_el_02	6.1.2.4.2.1	Frequency response	1
QoS_Voice_el_03	6.1.2.4.2.2	Junction loudness rating	1

QoS_Voice_el_04	6.1.2.4.2.3	End-to-end delay	1
QoS_Voice_func_01	6.1.3	Functional QoS parameters of the voice channel	4
QoS_Voice_np_01	6.1.4	Network performance parameters in the voice channel	1
QoS_Video_filebased_01	6.2.1.2	PEVQ	4
QoS_Video_func_01	6.2.3	Functional QoS parameters of the video channel	4
QoS_Video_np_01	6.2.4	Network performance parameters in the video channel	4

NOTE: 1 ==> Test cases have been validated and can be applied during VoLTE and RCS Test Events
2 ==> Test cases have not been tested due to non-availability of test equipment
3 ==> Test cases have not been tested due missing functionality in the network / access
4 ==> Test cases have not been tested due to non-availability of test house
5 ==> Test cases have not been validated and cannot be applied during VoLTE and RCS Test Events

6.5 Findings

6.5.1 Findings in TS 103 189

[F1] A new setup, acoustic-to-electrical was identified and discussed. It was agreed that this intermediate setup should be tested in between acoustic-to-acoustic and electrical-to-electrical setups, and that level adjustment at the electrical interface would allow to make the testing more meaningful (comparable to the acoustic-to-acoustic testing). This new setup will be included in the TS revision

[F2] A procedure for level adjustment at the electrical headset interface was discussed, tested and validated. This procedure will be submitted to be documented in an annex of the TS 103 189 revision

[F3] It was demonstrated that in asymmetrical setups (different terminals or different interfaces), it is required to test in both directions, and that 2 sets of results should be recorded. This requirement will be documented in the TS 103 189 revision.

[F4] For tests providing average performance parameters like frequency responses or loudness ratings it may be recommended to calculate the average over 8 test sentences. However, the experience shows that in many cases 2 sentences are sufficient, if the sentences are properly selected, e.g. male voice with lowest pitch frequency, female voice with highest energy in the high frequency range. This was confirmed by the results obtained during the event. [F5] "Double sentences" instead of "single sentences" should be used as a basis for assessing the quality perceived by the user, when using the POLQA prediction model.

[F6] A new test measuring idle noise level of the phones should be documented and will be contributed for the TS revision.

[F7] In case of "asymmetric" connections, i.e. UE manufacturer X vs UE manufacturer Z, it is recommended to test in both directions, as each direction may lead to very different results. Symmetric connections UE manufacturer X vs UE manufacturer X' one way/ direction should be sufficient

[F8] Delay analysis versus time was found to be complementary to the "single value delay", delay profiles vs time (e.g. over 2 min) provide a lot of information about different equipment under test and their behaviour. This was confirmed by the results observed during the event and will be submitted to be documented in the TS revision.

[F9] The integration of test under defined PL and jitter conditions should be mentioned, perhaps not mandatory, but they are typically very important to distinguish between different implementations. Can be done via stochastic impairment models or by using real measured network profiles and apply them during testing.

[F10] The description of the network parameters to be monitored at the UNI to UNI interfaces during the test sessions should be clarified and it would be beneficial to add some additional meaningful measures to be recorded. A revised text for this section will be submitted for TS 103 189 revision.

[F11] Variation of loudness as calculated by POLQA is not meaningful (for a test event), a revised text will be submitted.

[F12] Additional explanation on RLR / SLR measurements via network should be added to the TS, a proposed text will be provided.

[F13] The acoustic-to-electric / electric-to-acoustic setups using a reference VoIP gateway connected via LAN to a LAN-to-LTE router were validated and a text proposal for the test cases description will be submitted for the TS revision.

6.5.2 Options for the electrical interface.

During the event preparation, several options to implement the access to the electrical interface were identified, discussed and prioritized. The table below summarizes the studied options:

Priority	Method	Description	Comments
1	Router	An LTE router is used to enable the electrical access to the LTE network. The LTE router is connected to the reference VoIP interface via LAN.	The SIP client running on the reference VoIP interface needs to be validated in the test network. This method was validated during the event.
1	Dongle	An LTE dongle is used to enable the electrical access to the LTE network. The LTE dongle is directly connected to the reference VoIP interface via USB.	The LTE dongle compatibility with the reference VoIP interface needs to be validated. The SIP client running on the reference VoIP interface needs to be validated in the test network. This method was not validated during the event.
2	Bluetooth	An LTE User Equipment providing wide band transparent Bluetooth access is used to enable the electrical access to the LTE network. The UE is connected to the test system via WB Bluetooth	The wide band transparent Bluetooth needs to be validated with a specific tool. This method was not validated during the event.
3	Headset	A regular LTE User Equipment is used to enable the electrical access to the LTE network. The UE is connected to the test system via the headset interface	The headset interface introduces an additional element in the end-to-end chain. This method was validated during the event.

Annex A Test Tool Report

The following report was produced by HEAD acoustics as part of their contribution to the test event.

A.1. Introduction

Speech quality tests according to the preliminary version of TS 103 189 V0.2.3 (2013-07), “IMS Networking Testing (INT); Specification of End-to-End QoS Assessment for LTE and RCS Interop Events or Plugtests” were implemented in the HEAD acoustics test system ACQUA. These measurements were conducted during the VoLTE test event for validation using different user equipment.

The focus of these tests was the verification of the implemented tests itself, suggestions of test adaptations and the feedback of the most important findings to STF 453 after this validation test. These findings are summarized in chapter 1 of this report. The setup and the test signals are briefly described in chapter 3. Chapter 4 discusses the results separated for the acoustic-to-acoustic setup, the acoustic-to-electric / electric-to-acoustic setup (tested in both directions) and the electric-to-electric setup.

The tests were carried out between November 18 and 20 at ETSI premises in Sophia Antipolis, France.

A.2. Summary

The tests according to the current version of TS 103 189 were implemented in the HEAD acoustics test system and carried out during the event. The tests were conducted and verified with several devices (user equipment, UE). Additional tests were prepared and verified for suitability in discussion during the event.

The following suggestions were derived by HEAD acoustics during the test preparation and from the TS 103 189 validation tests, respectively, and were partly already discussed during the test event:

- For tests of average performance parameters like frequency responses or loudness ratings it may be recommended to calculate the average over 8 test sentences as it is currently described in TS 103 189. However, experience shows that in many cases 2 sentences are sufficient, if these sentences are properly selected, e.g. male voice with lowest pitch frequency, female voice with highest energy in the high frequency range. This was also verified during the TS validation event. Comparison results are given within this report.
- ITU-T P.863 POLQA should be analyzed based on “sentences pairs” basis instead of single sentences. For information purpose both tests were conducted for comparison during the event and the results are discussed.
- The “variation of loudness” tests described in the current TS version are unclear and deleted after discussion.
- The idle noise level of UE should be measured and is suggested as new test case. It is easy to measure and may be helpful for fault detection.
- In case of “unsymmetrical” connections, i.e. UE model x vs. UE model y, tests in both directions are needed. This may lead to very different results as shown during the test event. For symmetrical connections (UE model x vs. UE model x) one way analyses are typically sufficient.
- Delay analyses vs. time are very helpful in addition to the “single value delay”. Delay profiles vs. time (e.g. over 2 min) provide a lot of information about the performance of different UEs. Interesting results have been analyzed during the event even the setup represents a test environment and not a real commercially used VoLTE network.
- The integration of tests under defined PL and jitter conditions should be mentioned in the TS, perhaps not as mandatory, but as recommended. They are typically very important to distinguish between different implementations (PL implementation, jitter buffer implementation). This can be realized by stochastic impairment models or by using real measured network profiles and apply them during testing [See also Tdoc S4 (13)1160 presented during the SA4#76 meeting 4th-8th November 2013].

- ### A.2.1 Quality Pie Examples

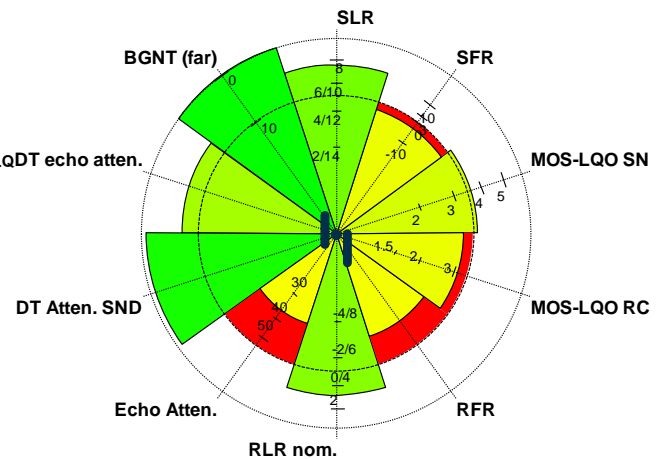


Fig. 2.2: Quality Pie example Acoustic-to-Electric / Electric-to-Acoustic, UE4

A.3. Description of Test Signals and Setup

For short term delay measurements a Composite Source Signal as given in **figure 3.1** is used. The noise part consists of a random, white noise sequence (duration 1 s) which provides the possibility for exact and robust cross correlation analysis between delayed signal and original test signal.

In principal the applied cross correlation analysis between transmitted signal and received signal can also be applied using real speech as test signal.

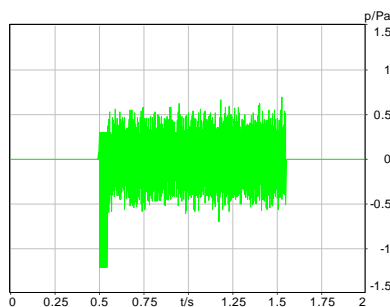


Fig. 3.1: CS Signal for delay tests

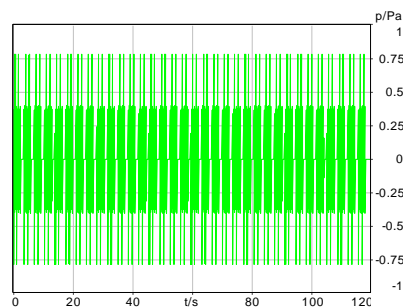


Fig. 3.2: Periodical repetition of CS bursts (120 s)

A periodical repetition of Composite Source Signals described in ITU-T Recommendation P.501 with an overall duration of 120 s is used for delay analysis vs. time (see **figure 3.2**).

For average analyses like frequency response or loudness rating determination, a sequence consisting of eight sentences (British English) as described in ITU-T Recommendation P.501 is recommended in TS 103 189. The test sequence is shown in **figure 3.3**. The eight sentences represent two male and two female voices (two sentences each). However, in many cases the use of two sentences (one female voice and one male voice) may be sufficient if these sentences are properly selected. For this purpose a male voice (“The last switch cannot be turned off”) and a female voice (“The hogs were fed chopped corn and garbage”) were selected from the test sentences recommended in ITU-T P.501 (see **figure 3.4**).

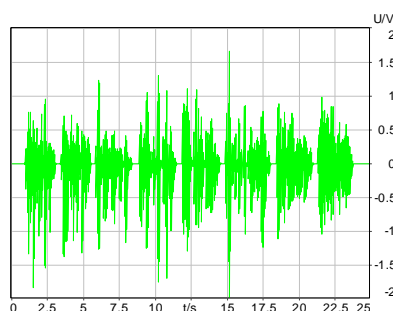


Fig. 3.3: British English test sentences (8 sentences)

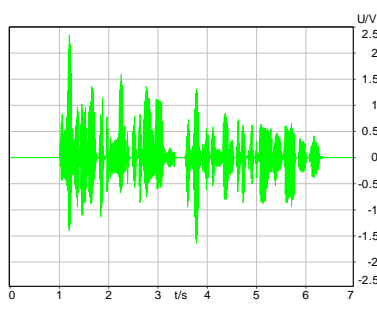


Fig. 3.4: 2 sentences subset

The male voice provides the lowest pitch frequency whereas the female voice shows the highest energy in the upper frequency range among the eight test sentences. Comparison tests using the 8 sentence version and the 2 sentences were carried out during the TS validation event.

MOS-LQO analyses using the ITU-T recommended P.863 POLQA analysis method are based on eight sentences (two male and two female voices, two sentences each).

Figure 3.5 shows this speech sequence used during the test event. In contrary to the sequence shown in **figure 3.3** (identical sentences) the signal activity is lowered to approximately 50 % over the entire sequence length by inserting appropriate pauses.

In general all languages as recommended in ITU-T P.501 can be used for MOS-LQO analysis. In order to limit the test duration the British English test sequence shown in **figure 3.5** was used during the TS validation event.

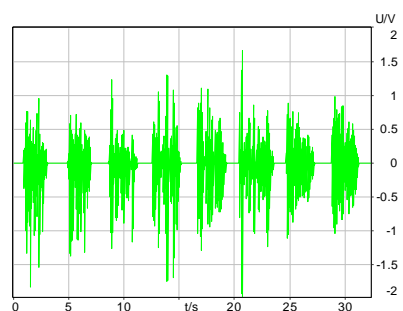


Fig. 3.5: British English used for MOS-LQO analysis

The voiced sound of the composite source signal provides deterministic characteristics. The periodical repetition of this voiced sound with decreasing and increasing test signal level is shown in **figure 3.6**. The enlarged sequence in **figure 3.7** shows the periodicity of this voiced sound. This test signal is used for detecting the presence of an automatic gain control or can be applied in receiving direction of terminals in order to evaluate packet loss concealment algorithms, the phase accuracy and audible disturbances (see **figure 3.8**).

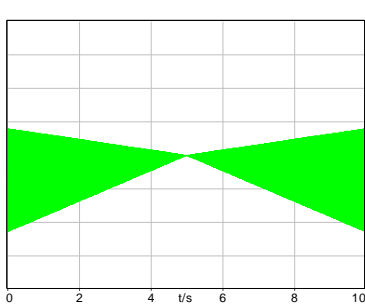


Fig. 3.6: AGC test signal

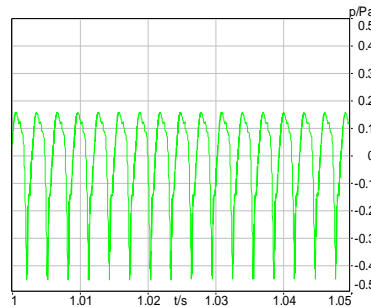


Fig. 3.7: Periodical repetition of voiced sound

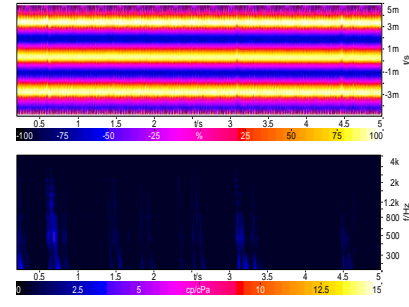


Fig. 3.8: PLC Analysis, example [5th ETSI SQTE]

The echo attenuation is analyzed using the “compressed speech” signal from ITU-T P.501 as shown in **figure 3.9**. This signal provides a low CREST factor and is therefore suitable to be applied with a high test signal level without saturation or overmodulation. It therefore provides a high analysis resolution. This is necessary especially for echo analyses where the requirements are set to ≥ 55 dB. The time signal of the “compressed speech” is shown in **figure 3.9**.

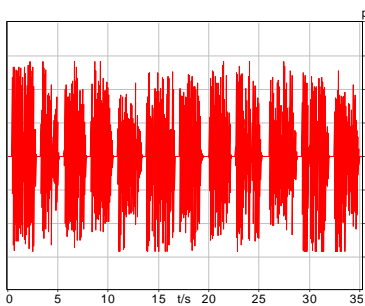


Fig. 3.9: Compressed speech

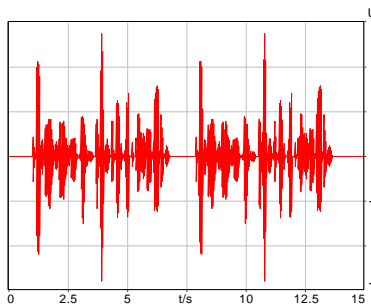


Fig. 3.10: 2 sentences subset

Further echo analyses like echo level vs. time use a real, uncompressed speech sample taken from ITU-T P.501 (**figure 3.7**). The sequence consists of a male and a female voice (“The birch canoe slid on the smooth planks”; “The hogs were fed chopped corn and garbage”).

The same test sequence is used for further aurally motivated echo analyses as described in [Scalable Perceptual based echo assessment method for aurally adequate evaluation of residual single talk echoes, M.Lepage et. al.; IWAENC 2013]. This analysis is available as EQUEST algorithm and provides MOS scores representing the expected echo disturbance on a 5-point Degradation Category Rating scale (DCR scale according to ITU-T P.800).

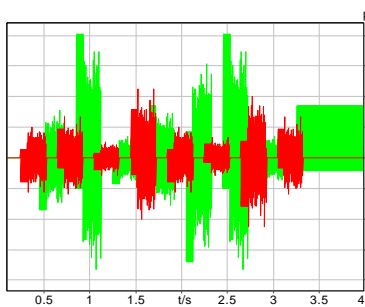


Fig. 3.11: Double talk test signal

Double talk analyses determining the attenuation in sending or receiving direction are carried out using the combination of uncorrelated Composite Source Signals as shown in **figure 3.11**. The test signals are applied with varying level vs. time (level variation of ± 3 dB in both directions). For electrical signal feeding the test signal level is adjusted to -16 dB_{m0} in average (red colored time signal in **figure 3.11**). In case this test signal is applied via the artificial mouth at the acoustical interface the average test signal level is adjusted to -4.7 dB_{Pa} (green colored time signal in **figure 3.11**).

For information purpose two background noise signals were played back via the artificial mouth at the near end. In general a background noise playback system as given in ETSI ES 202 396-3 is recommended. However, this setup was not used during the TS validation phase in order to limit the testing effort during the three days test event.

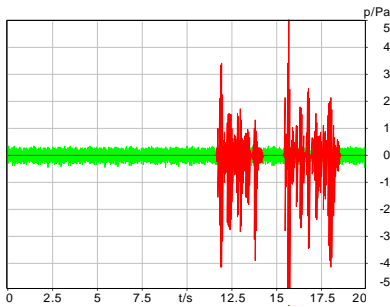


Fig. 3.12: BGNT with far end speech (Car), test signal

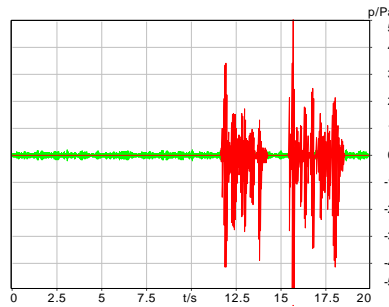


Fig. 3.13: BGNT with far end speech (Pub), test signal

The two test signals used during the event are shown in **figure 3.12** and **3.13**. The near end signal played back via the artificial mouth is given in green (stationary car noise, background noise level of 69 dB(A) at the MRP in **figure 3.12**, pub noise in **figure 3.13**, background noise level of 73 dB(A) at the MRP).

A.3.1 Acoustic-to-Acoustic Setup

During the test event two artificial head measurement systems (HMS II.3, HEAD acoustics) were used. These HATS according to ITU-T Recommendation P.58 are equipped with two ITU-T P.57 recommended artificial ears (type 3.3) and artificial mouth. A handset positioning device was used to mount the handsets to the artificial ear. If not stated otherwise all tests were carried out using an 8 N application force between the different UEs and the artificial ear.

Figure 3.14 shows the artificial head in room B1A.



Fig. 3.14: HMS II.3 in room B1A



The HATS in room B2 is shown in **figure 3.15**.



Fig. 3.15: HMS II.3 in room B2



Note that these test rooms are not acoustically treated. However, considering the fact that the tests were carried out in handset mode with loudspeaker and microphone positioned close to the artificial ear respectively artificial mouth, the room influence (reverberation) is minimized. The background noise level in both rooms could not be exactly controlled; however, air conditioning, ventilation and all other noise sources were disabled as far as possible. These environmental conditions were absolutely sufficient in order to validate the tests according to TS 103 189.

Tests using this acoustic to acoustic setup were only carried out in one direction, i.e. from room B1A (sending side) to room B2 (receiving side). Conversational aspects like double talk or echo tests were carried out in an acoustic-to-electric/electric-to-acoustic setup. If a different UE was used on both sides and both directions should be tested, the UE itself was exchanged between the two rooms.

A.3.2 Acoustic-to-Electric/Electric-to-Acoustic Setup

For the electric-to-acoustic tests described in TS 103 189 one artificial head was substituted by an electric access point to the test network. During the Test Event it was possible to configure the HEAD acoustics measurement frontend MFE VIII.1 together with a “LAN-to-LTE” router to access the LTE test network. In this test setup the sending direction is defined from the “UE point of view”, i.e. as the direction from the acoustic interface (microphone) to the electric interface. Vice versa the receiving direction describes the transmission direction from the electric assessment point to the UE mounted to the artificial head (loudspeaker).

Echo measurements are carried out by feeding signals via the electric interface and recording it in sending direction at the same interface.

For double talk testing signals are fed in both directions: using the artificial mouth on the sending side and electrically using the electric access point on the receiving side.

A.3.3 Electric-to-Electric Setup

This test could be carried out between the MFE VIII.1 in combination with the “LAN-to-LTE” router as described above and one headset interface on one UE. Alternatively the Bluetooth interface for wideband speech transmission using the Hands-free Profile according to the specification version 1.6 can be used. In this case it needs to be ensured beforehand that the UE is Bluetooth wideband capable, switches off all signal processing and provides a transparent signal transmission.

It should be considered that the use of the headset interface accessed via specific cables to feed signals in sending direction of the headset interface or analyze them at the receiving direction of this interface, does not ensure that the internal signal processing in the user interface is disabled. Therefore this setup can only be seen as a backup solution. However, it may be useful for specific tests like one way transmission quality analysis (MOS-LQO) if acoustic interfaces are not appropriate for this test. This setup may therefore be valid when testing a specific UE but explicitly excluding acoustic interface like loudspeaker and microphone.

A.3.4 Headset Interface Level Adjustment

When using the headset interface for testing it needs to be considered that these interfaces may be differently configured on each user equipment (UE). The headset interface cable replaces headset microphone and headset loudspeaker. However, the headset interface on each user equipment may still provide different characteristics like equalizers or different sensitivities in sending and receiving direction in order to optimize the performance for specific headsets delivered as bundle together with the user equipment.

In order to adapt the sensitivities and adjust the appropriate test signal levels during tests using a wired connection to the headset interface the following procedure is suggested and was successively verified during the test event.

- In an acoustic to acoustic setup using two mobile phones (user equipment) on the sending side (UE A) and receiving side (UE B) the Overall Loudness Rating (OLR) is determined as described in test [QoS_Voice_ac_03] in the Test Specification. The volume setting at the receiving side (UE B) is set so that an OLR closest to 10 dB is achieved. The Active Speech Level (ASL) is determined for this volume setting.
- The headset interface cable is then connected to the UE A on the sending side. The - loudness rating test is repeated and the ASL of transmitted speech at the artificial ear on the receiving side (UE B) is again calculated. If both calculated ASL values are identical the sending sensitivity of the headset cable connected to UE A meets the sensitivity of the UE A when operated in handset mode. In case the ASL differs from the test result in acoustic to acoustic setup, the test signal level at the sending side is adjusted (amplified or attenuated) in order to meet this target level measured at the artificial ear on the receiving side.
- The default level used for the sending direction as input level of the headset interface is -60 dB_V. The necessary signal level adjustment as described above should be documented. It represents a value to characterize the headset interface sensitivity of UE A.

- It should be noted that this adaptation was done using the comparison of ASL according to ITU-T Recommendation P.56 but can in principle also be made on basis of loudness ratings. The results may slightly differ but lead to similar results for the necessary gain adjustment at the headset interface cable.
- In a next step the headset interface cable can be connected to the UE B on the receiving side. Again the loudness rating measurement (JLR instead of RLR) is now carried out. A JLR of 0 dB corresponds to a RLR + SLR = OLR in the acoustic setup.
- The necessary adjustment in order to meet a JLR of 0 dB can be made via the playback volume on user interface B for the receiving direction.

A.4. Discussion of Results

A.4.1 Acoustic-to-Acoustic Setup

Several End-to-End connections including different UEs were tested during the 3-days event. The listening speech quality tests as described in TS 103 189, carried out on the ITU-T recommended P.863 basis are summarized in **table 4.1**. The connections were tested with identical UE at both sides and – for verification purpose – with UE of different manufacturers. This is especially interesting in order to verify the “symmetry” of a connection in terms of speech quality. The MOS-LQO results are calculated on a single sentence basis together with the Δ MOS-LQO scores among these eight single values in **table 4.1**. Additionally the MOS-LQO results are analyzed on a sentence-pair basis. Furthermore the Δ MOS-LQO between the lowest and maximum score for this analysis method is given in **table 4.1** as well as the average delay range.

	UE4 to UE4'	UE8 to UE8' (UE5)	UE6 to UE6'	UE4 to UE8	UE8 to UE4	UE8 to UE6
MOS-LQO (single sentence)	3.1	3.6	3.0	3.2	3.1	3.8
Δ MOS-LQO (single sentence)	2.6 – 3.7	3.2 – 4.0	2.0 – 4.0	2.7 – 3.8	2.7 – 3.6	3.4 – 4.1
MOS-LQO (sentence pairs)	3.0	3.4	2.7	3.3	3.0	3.7
Δ MOS-LQO (sentence pairs)	2.8 – 3.4	3.1 – 3.9	2.0 – 3.7	2.8 – 3.6	2.6 – 3.3	3.3 – 4.0
Delay [P.863]	426.1 – 427.8 ms	632.7 – 638.1 ms	370.2 – 439.0 ms	729.3 – 735.8 ms	372.3 – 374.7 ms	384.6 – 405.9 ms

Table 4.1: ITU-T P.863 test result

The MOS-LQO scores differ between 3.0 up to 3.8 (single sentence based) for the different tested connections. It is worth to mention that the highest score is achieved with different UE on the sending and receiving side (connection between UE8 on the sending side and UE6 on the receiving side). Surprisingly the lowest MOS-LQO was measured for the “symmetrical” setup between UE6 and UE6' used on both ends.

The Δ MOS-LQO results differ significantly between the single sentences, the widest range could be determined between 2.0 up to 4.0 MOS-LQO for the connection between two UE6 devices.

In principal the same tendency can be derived from the analysis using the sentence pairs for MOS-LQO calculation. The results also differ significantly between 2.7 MOS-LQO up to 3.7 MOS-LQO. However, the Δ MOS-LQO range does not show such strong variations as it could be expected when averaging over sentence pairs.

The end-to-end delay from “mouth to ear” also differs significantly between approximately 370 ms up to 735 ms for the different equipment. It needs to be clearly stated that this does not represent the end-to-end delay in a VoLTE connection because the speech transmission is realized over the BRIA App running on each UE. This definitely

increases the delay. The results here shall therefore only be analyzed on an informative basis. The aim of the Test Event was the validation of the tests itself and not the analysis of the absolute results.

The overall frequency responses are summarized in **table 4.2** for the three different symmetrical connections. This analysis was carried out by averaging the curves over eight sentences. The main difference between these measured curves can be seen in the low frequency range. The connection between the two UE4 devices on both sides shows the strongest high pass characteristic. Furthermore a slight limitation in the upper frequency range can also be detected compared to the other two curves. If both UE8 devices are used the end-to-end transmission is significantly wider. The widest range of transmitted frequencies can be analyzed for the connection between the UE6 and UE6' devices.

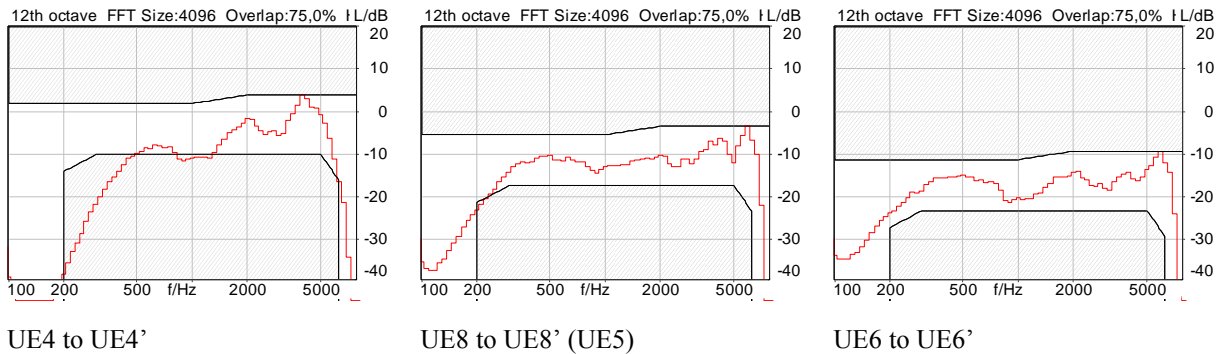


Table 4.2: Overall frequency responses, symmetrical connection

Similar conclusions can be drawn for the unsymmetric connection by comparing the overall frequency responses in **table 4.3**. Comparing these curves with the ones analyzed before for the symmetric connection it can be seen that the frequency characteristics are very similar in case the same device has been used on the receiving side. This is an indication that the frequency characteristic of the receiving terminal is in general the dominating factor for this type of analyses.

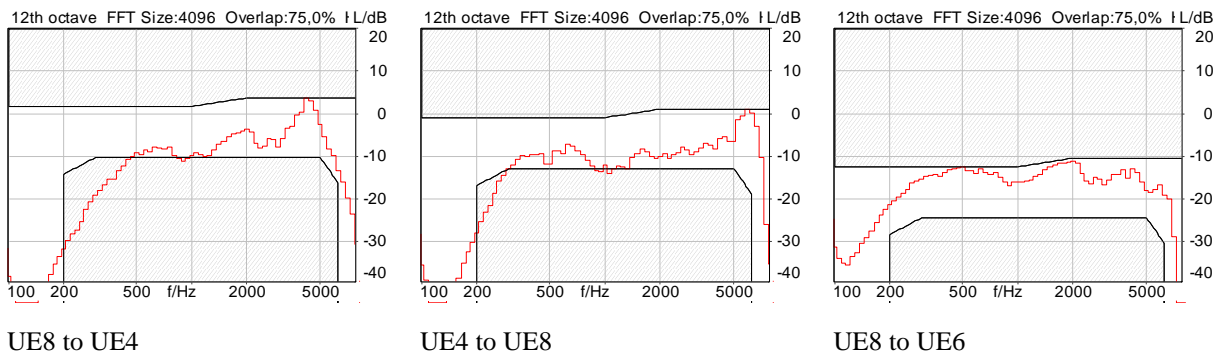


Table 4.3: Overall frequency responses, unsymmetrical connection

The frequency response comparison using the eight sentence average analysis and the two sentence approach leads to very similar results. The frequency responses are compared for the three symmetrical setups using the same UE on both sides (UE4 to UE4', UE8 to UE8', UE6 to UE6') in **table 4.4**. This comparison leads to the assumption that the overall frequency response can be analyzed on a two sentence basis using the selected sentence pair as described above. However, care should be taken by the test lab that the results are not influenced by long term AGC. In case the analysis during the event gives hints for such an implemented AGC, the results should be repeated several times when using the two sentence approach.

In principle the same precautions need to be considered for the eight sentence averaging process. In this case the test should also be repeated in order to get results under steady state conditions.

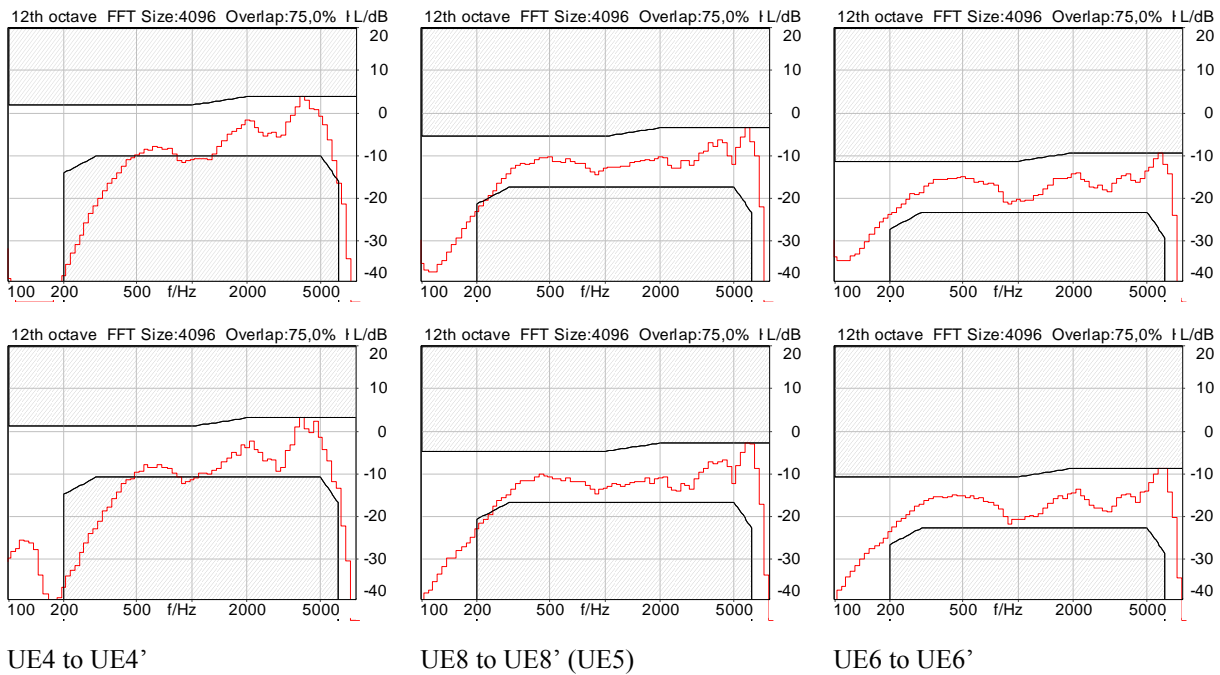


Table 4.4: Overall frequency responses, long sequence (8 sentences) vs. short sequence (selected sentence pair)

In the end-to-end scenario (acoustic-to-acoustic, “mouth to ear”) the Overall Loudness Rating (OLR) was determined - again using the eight sentence approach and comparing the results to the analysis based on two sentences. The measured results together with the adjusted volume settings on the receiving UE are given in **table 4.5**. The OLR was adjusted as close as possible to 10 dB. This suggested requirement is derived from an expected “nominal” SLR on the sending side of 8 dB in combination with a Receiving Loudness Rating (RLR) at nominal playback volume of 2 dB. The Junction Loudness Rating (JLR) of the connected network is assumed to be 0 dB. The UE6 only allowed a volume adjustment in steps of 10 dB(!). The OLR could therefore only be adjusted to approximately 15 dB instead of 10 dB as shown in **table 4.5**.

The idle noise level was additionally determined even up to now not described in the TS 103 189 version. The results did not differ significantly between the mobile phones. However, it also needs to be taken into account that the test environment was not completely quiet. In general it is recommended to add the idle noise level analysis in the TS test description.

	UE4 to UE4'	UE8 to UE8' (UE5)	UE6 to UE6'	UE4 to UE8	UE8 to UE4	UE8 to UE6
Setting x/y	2 / 6	3 / 6	5 / 6	3 / 6	2 / 6	5 / 6
OLR (8 sent.)	9.4 dB	10.2 dB	15.0 dB	9.3 dB	9.0 dB	13.0 dB
OLR (2 sent.)	9.4 dB	10.3 dB	14.8 dB	9.5 dB	9.0 dB	12.8 dB
Idle noise level	-56.6 dB _{Pa} (A)	-56.8 dB _{Pa} (A)	-55.9 dB _{Pa} (A)	-56.5 dB _{Pa} (A)	-57.2 dB _{Pa} (A)	-56.7 dB _{Pa} (A)
Delay	421.0 ms	654.4 ms	352.8 ms	721.4 ms	384.6 ms	396.0 ms

Table 4.5: One-dimensional parameters

The delay calculated from a cross correlation analysis using Composite Source Signal bursts was determined between approximately 350 ms up to 720 ms. Again the same statements as above need to be considered, the speech transmission includes a specific App on the mobile phones. Consequently, the measured delay does not represent the VoLTE connection as expected in real life conditions. However, the measured delays are easy to determine in an acoustic-to-acoustic setup and lead to very accurate results.

A further analysis was carried out determining the end-to-end delay vs. time. This is especially important in an end-to-end scenario in order to verify the stability and temporal changes of the delay parameter. This analysis therefore carries even more information than the short term delay analysis given in **table 4.5** or the average delay analysis on a sentence basis or based on sentence pairs as given in **table 4.1**.

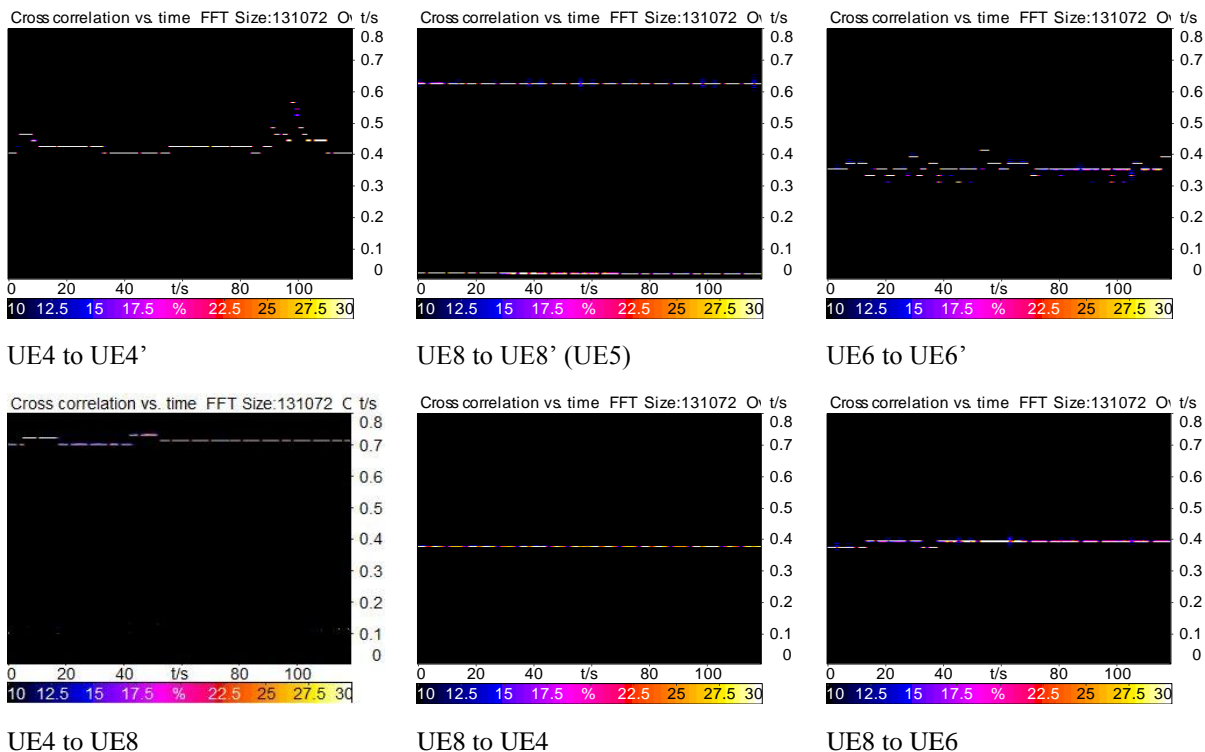


Table 4.6: End-to-end delay vs. time

The three connections between identical UE on both sides lead to very different results. An average delay of approximately 420 ms could be determined for the UE4. However, relatively strong delay variations could be observed as shown in the corresponding analysis in **table 4.6**. The delay increases up to approximately 600 ms at maximum.

Vice versa UE8 used on both sides leads to a very high delay of more than 600 ms. This delay is stable vs. time. A potential reason could be the implemented jitter buffer control in this UE that covers all varying delay from the network on the price of a very high absolute delay. A significantly lower delay but with strong delay variations vs. time could be measured for the UE6. These delay analyses clearly distinguish the different implementations in the three different UEs and therefore provide very informative results.

Asymmetric connections between UE4, UE8 and UE6 respectively are analyzed and shown in the lower diagrams in **table 4.6**. Again the delay is extremely high when the UE8 is used on the receiving side (see left hand diagram). This can be seen as another clear indication that the delay is mainly determined by the device which is used on the receiving side.

A.4.2 Acoustic-to-Electric / Electric-to-Acoustic Setup

In the following analyses the Δ MOS-LQO is not additionally analyzed. This analysis was carried out already in the acoustic to acoustic setup as indicated above in **table 4.1**. The following MOS-LQO results are derived on a sentence pair analysis.

These tests were carried out between one UE exemplarily and the HEAD acoustics measurement frontend MFE VIII.1 which could be connected to the com4Innov test network via a “LAN to VoLTE” router.

The MOS-LQO in sending direction was determined to 3.6. A lower result of 3.2 MOS-LQO was measured for the receiving direction. For comparison the MOS-LQO result from the acoustic to acoustic setup is also given in **table 4.7** (3.0 MOS-LQO). The comparison of these three results indicates that the receiving direction of a UE can be seen as the

limiting factor for listening speech quality. The sending direction is typically more transparent and provides better MOS-LQO results than the receiving direction.

Receiving	UE4
MOS-LQO sending (sentence pairs)	3.6
MOS-LQO receiving (sentence pairs)	3.2
MOS-LQO “mouth-to-ear” (sentence pairs)	3.0

Table 4.7: MOS-LQO results

Table 4.8 shows further one-dimensional results measured in receiving direction. The volume setting 3/6 leads to a RLR of 1.0 dB. It needs to be considered that this result does not only represent the RLR value because the real network is also part of the transmission chain. This parameter should be renamed to (JLR+RLR) in order to clearly distinguish from the RLR influence of the mobile phone in isolation.

The idle noise level of -62.3 dB_{Pa}(A) is significantly lower than the result indicated in **table 4.5** for the same UE. This is an indication that the sending direction also contributes to the perceived noise level in the ear when two UE are connected.

The measured delay in receiving direction could be determined to approximately 290 ms for UE4 in comparison with the measured delay of 421 ms in the end-to-end scenario as given in **table 4.5**, a delay of approximately 140 ms can be estimated for the UE4 in sending direction without the connection to the VoLTE test network.

Receiving	UE4
Setting x/y	3 / 6
(JLR+RLR) (8 sent.)	---
(JLR+RLR) (2 sent.)	1.0 dB
Idle noise level	-62.3 dB _{Pa} (A)
Delay receiving	288 ms

Table 4.8: One-dimensional parameters, receiving direction

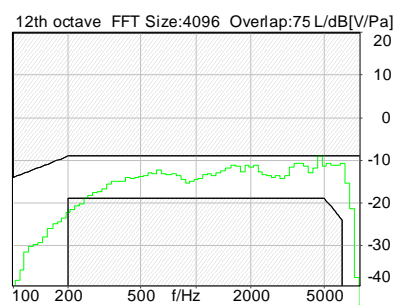
Corresponding to the receiving direction the sending direction was tested between the acoustic interface (artificial mouth) and the VoLTE network access through the MFE VIII.1 with the connected gateway. Instead of the Sending Loudness Rating (SLR) of the phone in isolation the measured sensitivity in sending direction is expressed by the combination of SLR and JLR provided by the connected VoLTE network. The corresponding parameter (SLR+JLR) was determined to 8.6 dB which meets the recommended SLR of a phone in isolation (SLR 8 ± 3 dB according to ETSI TS 103 739). The measured delay in sending direction of 401 ms was unexpectedly high. However, it needs to be considered that this value again covers the sending delay of the UE plus the network delay.

Sending	UE4
(SLR+JLR) (8 sent.)	---
(SLR+JLR) (2 sent.)	8.6 dB

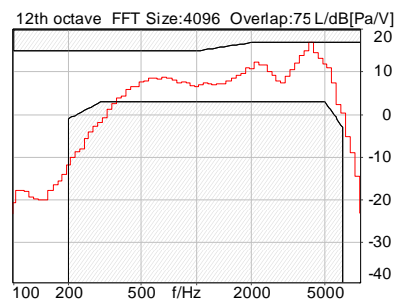
Delay sending	401.1 ms
---------------	----------

The measured frequency responses in the acoustic to electric setup and the electric to acoustic setup respectively are given in **table 4.10**. For comparison the mouth to ear frequency response derived in the acoustic to acoustic setup is indicated below.

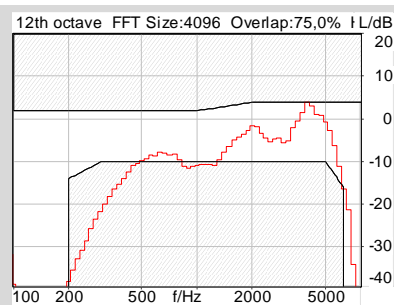
Acoustic-to-electric
“sending” frequency
response



Electric-to-acoustic
“receiving” frequency
response



Acoustic-to-acoustic
“mouth-to-ear” frequency
response

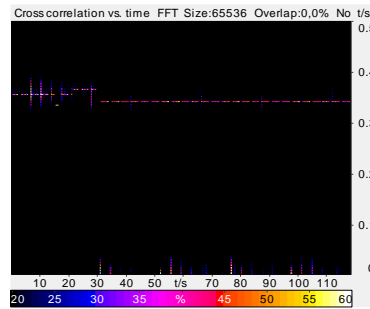


UE4 to UE4'

The comparison of the three measured phones indicate that the mouth to ear frequency response (acoustic to acoustic) can be seen as a linear combination of the sending and receiving frequency responses measured in the acoustic to electric and electric to acoustic setup. Furthermore it can be seen that the main spectral limitation is due to the receiving side of the UE4. The sending side is rather transparent only introducing some additional attenuation in the low frequency range (high pass characteristic).

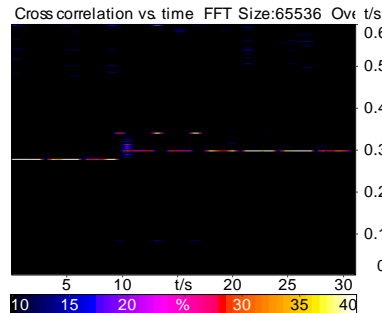
Table 4.10: Frequency response curves

Acoustic-to-electric “sending”
delay vs. time



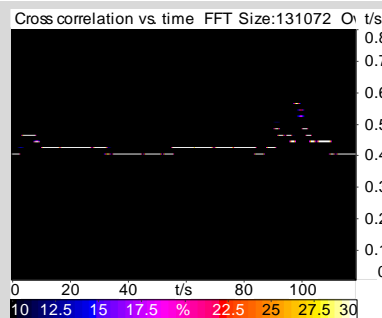
UE4 to MFEVIII.1

Electric-to-acoustic
“receiving” delay vs. time



UE4 to MFEVIII.1

Acoustic-to-acoustic “mouth-
to-ear” delay vs. time



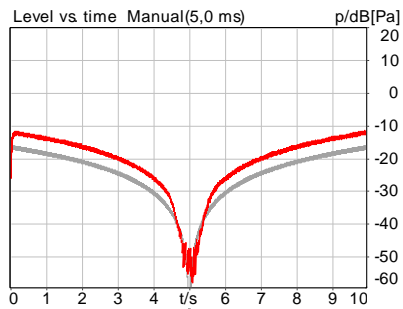
UE4 to UE4'

As for the acoustic-to-acoustic setup the delay vs. time analysis is also performed for the acoustic-to-electric and electric-to-acoustic scenario respectively. The resulting cross-correlation vs. time diagrams are shown in **table 4.11**. Comparing with the acoustic-acoustic result the acoustic-electric/electric-acoustic results show a rather constant behaviour vs. time. However, it needs to be considered that the acoustic-acoustic measurement was performed one day earlier thus the network condition might have been very different.

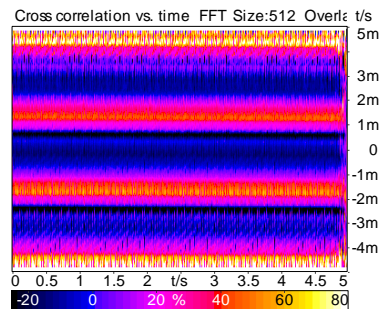
Table 4.11: Delay vs. time acoustic-to-electric and electric-to-acoustic

Two additional informative test results performed during the evaluation of the tests specified in TS 103 189 are presented in **table 4.12**. The AGC test uses the signal described in **figure 3.6** to compare the level of the transmitted signal with the level of the source signal over time. If significant level adjustments are detected this is an indication for an active gain control mechanism in the UE. The test result shown below for UE4 does not show any important level adjustment in receiving direction.

AGC Test of UE4



Packet Loss Concealment (PLC) test of UE4



Additionally the first five seconds of this test signal can be used for testing the performance of an implemented packet loss concealment. Therefore the transmitted signal is analysed using a cross-correlation vs. time in order to detect missing signal parts or phase shifts in the acoustically recorded signal in receiving direction.

Table 4.12: AGC receive test and PLC analysis

The result shown here for UE4 does not indicate any lost packets or phase shifts due to jitter.

Replacing the far end terminal by an electrical reference LTE gateway further allows detailed echo and double talk tests.

The echo performance of the device under test is described by three analysis results: the overall echo attenuation, the echo level vs. time and the spectral echo performance.

As demonstrated by all three analysis results significant residual echoes are detected for the UE4. This is most likely due to the fact that the BRIA application does not have access to the echo control implemented in the DSP of the phone but only provides in-built software for echo cancellation. It is to be expected that real VoLTE phones will provide more reliable echo performance.

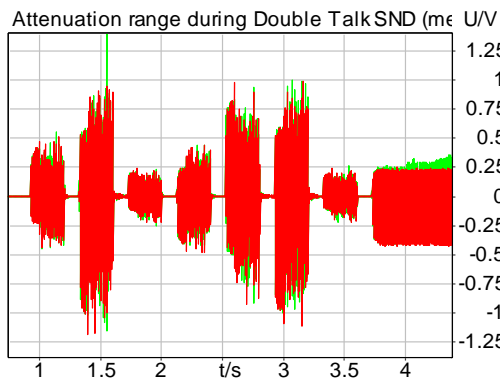
Informatively the perceptual echo performance using the EQUEST algorithm is also given. The result of this analysis is an objective MOS on a DCR scale from 1 (echo is very annoying) to 5 (echo is not audible). The rather high result can be explained by the fact that the used source signal level in receive during this test was rather low and thus echo conditions were quite uncritical for the UE4.

One-dimensional results		Echo level vs. time		Spectral echo attenuation	
Overall echo attenuation	36.7 dB				
Perceptual echo performance result (EQUEST)	4.0 MOS				

Table 4.13: Echo performance test results

The double talk tests performed on the UE4 determine the attenuation introduced in sending direction and the residual echo under double talk conditions. The attenuation introduced in uplink amounts to less than 1 dB and no significant echo components are detected. The double talk attenuation in receiving direction can in principle be analysed using the same method used for the sending direction. This test was not performed during the evaluation session since the receiving direction is usually transparent for mobile phones.

Attenuation during double talk in sending direction



Attenuation in SND: 0.6 dB

Echo during double talk

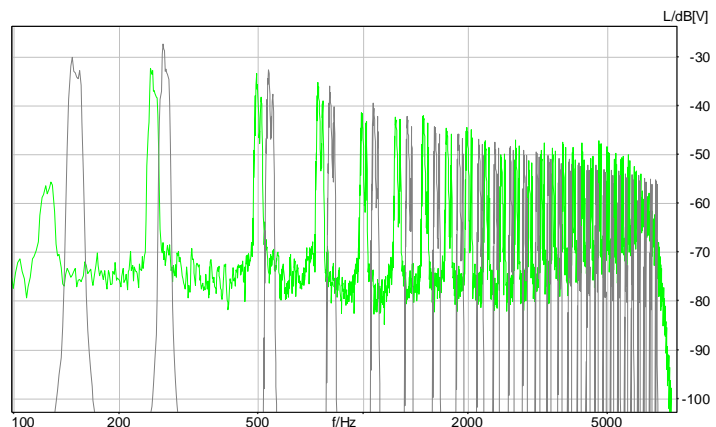
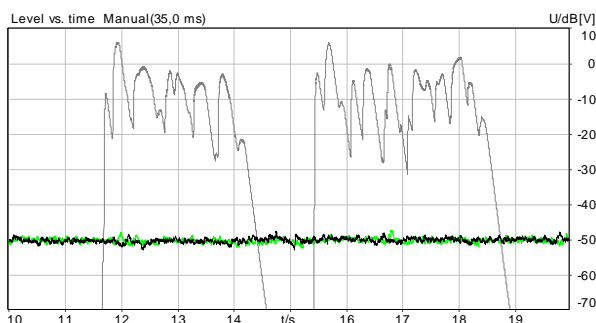


Table 4.14: Double talk test results

Considering the “mobility” of VoLTE phones the use in an environment with background noise is a very typical use case that should be considered when testing VoLTE UE. Ideally a background noise simulation as described in chapter 2 is used for this purpose. The setup of such a background noise simulation would have consumed valuable time during the test specification evaluation session therefore a simplified approach using the artificial mouth for playing back realistic background noise was used instead. This is absolutely sufficient for testing the feasibility of the background noise transmission tests with application of far end speech analysed below in **table 4.15**.

BGNT with far end speech (Car)



BGNT with far end speech (Pub)

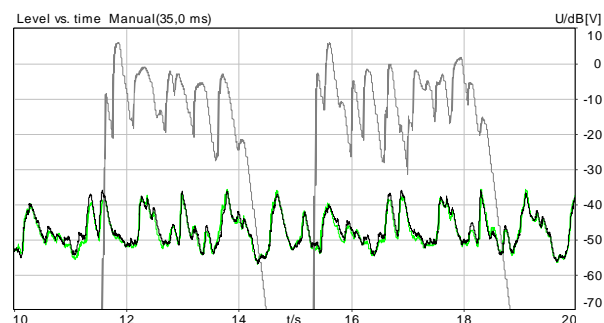


Table 4.15: Background noise transmission with far end speech, UE4

This test was performed two times using two different realistic background noise scenarios: a car interior noise (130 km/h) and a more non-stationary pub noise.

In both diagrams the green curve represents the level vs. time of the background noise transmitted in sending direction of the UE4 coincident to the application of a far end speech signal. The far end source signal is indicated by the grey curve. Additionally a black curve, representing the same background noise transmission in sending direction without application of the far end speech sequence is shown as a reference. When a far end signal is applied this activates the echo cancellation and suppression mechanisms implemented in the device under test in order to cancel or attenuate echo in sending direction. If strong attenuations are introduced this can also directly affect the transmission of the background noise and result in strong level modulations audible at the far end side. The analysis results for the UE4 do not show any attenuation since the green curve and the black reference curve are practically identical.

A.4.3 Electric-to-Electric Setup

The electric-to-electric test described in TS 103 189 have been verified using two different setups:

- Headset interface cable connected to the UE on the sending side and VoLTE reference gateway (MFE VIII.1) on the receiving side

- Headset interface cable connected to the UE on the sending side and a second UE with another headset interface cable on the receiving side

The first setup is to be preferred since it provides a defined point of interconnection (POI) at the far end side.

The results of the tests described in the TS 103 189 specification are summarized in **table 4.16** and in the diagrams of **table 4.17**.

	UE4 to MFEVIII.1	UE4 to UE8 (Headset Interface)	UE8 to UE4 (Headset Interface)
MOS-LQO (single sentence)	4.0	3.8	3.3
MOS-LQO (sentence pairs)	3.9	3.7	3.0
Delay [P.863]	298 – 368 ms	703 – 705 ms	404 – 460 ms
Junction Loudness Rating	JLR = 0.0 dB	---	---

Table 4.16: One-dimensional test results electric-to-electric

The quality scores calculated by POLQA are high if the first test setup is used. This could be expected since the network should not significantly degrade the signal transmission. A slight limitation by the sending side of UE4 – most likely its high pass implementation – was to be expected. The measured junction loudness rating of exactly 0 dB indicates that the headset level adjustment procedure applied to the UE headset interface as described in chapter 3.4 works correctly assuming that the com4innov LTE network provides a JLR of 0 dB.

Using another UE at the far end side (second test setup) leads to very slight degradations of the quality scores. This can also be expected considering the receiving characteristic of UE8 and comparing the electric-to-electric frequency response in **table 4.17**. Furthermore the delay significantly increases in this scenario as already detected for previous tests.

The quality degradations are even higher if UE4 is used on the receiving side. This again confirms the limitation in the receiving path of UE4 already detected in previous tests (e.g. acoustic-to-acoustic). This is additionally confirmed by the right hand frequency response representation in **table 4.17** below.

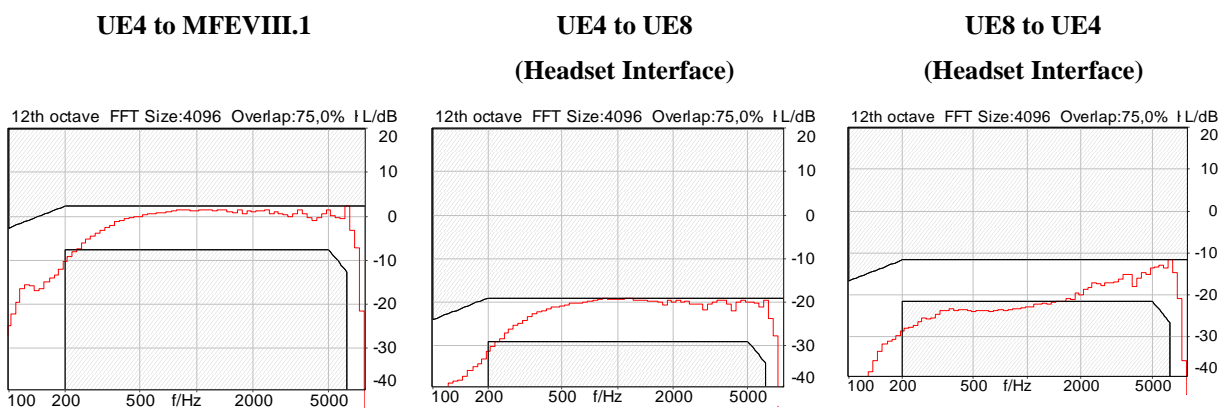


Table 4.17: Electric-to-electric frequency response representations

History

Document history		
V0.0.1	November 2013	Initial draft
V0.0.2	December 2013	+ initial input from AT4 Wireless, HEAD acoustics, and STF 453.
V0.0.3	December 2013	+ additional input from Com4Innov and HEAD ac + review from HEAD ac and STF 453
V0.1.0	December 2013	Final version