

ETSI TS 126 132 V16.2.0 (2020-11)



**Universal Mobile Telecommunications System (UMTS);
LTE;
5G;
Speech and video telephony terminal acoustic test
specification
(3GPP TS 26.132 version 16.2.0 Release 16)**



Reference

RTS/TSGS-0426132vg20

Keywords

5G,LTE,UMTS

ETSI

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

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

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

Important notice

The present document can be downloaded from:

<http://www.etsi.org/standards-search>

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

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

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

<https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx>

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

<https://portal.etsi.org/People/CommitteeSupportStaff.aspx>

Copyright Notification

No part may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm except as authorized by written permission of ETSI.

The content of the PDF version shall not be modified without the written authorization of ETSI.

The copyright and the foregoing restriction extend to reproduction in all media.

© ETSI 2020.

All rights reserved.

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

3GPP™ and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners.

oneM2M™ logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners.

GSM® and the GSM logo are trademarks registered and owned by the GSM Association.

Intellectual Property Rights

Essential patents

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

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

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Legal Notice

This Technical Specification (TS) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found under <http://webapp.etsi.org/key/queryform.asp>.

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the [ETSI Drafting Rules](#) (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

Intellectual Property Rights	2
Legal Notice	2
Modal verbs terminology.....	2
Foreword.....	9
Introduction	9
1 Scope	10
2 References	10
3 Definitions, symbols and abbreviations	12
3.1 Definitions	12
3.2 Abbreviations	12
4 Interfaces	13
5 Test configurations	14
5.1 Setup for terminals	14
5.1.1 Setup for handset terminals.....	15
5.1.2 Setup for headset terminals.....	16
5.1.3 Setup for hands-free terminals	16
5.1.3.1 Vehicle-mounted hands-free	16
5.1.3.2 Desktop hands-free.....	17
5.1.3.3 Hand-held hands-free	17
5.1.3.4 Softphone including speakers and microphone.....	18
5.1.3.5 Softphone with separate speakers	21
5.1.4 Position and calibration of HATS.....	25
5.1.5 Test setup for quality in the presence of ambient noise measurements	25
5.1.5.1 Handset	25
5.1.5.2 Hand-held hands-free	26
5.2 Setup of the electrical interfaces.....	29
5.2.1 Codec approach and specification.....	29
5.2.2 Direct digital processing approach.....	30
5.3 Accuracy of test equipment	30
5.4 Test signals.....	32
5.5 Void.....	32
5.5.1 Void	32
5.5.2 Void	32
6 Test conditions	32
6.1 Environmental conditions.....	32
6.1.1 Handset and headset terminals.....	32
6.1.2 Hands-free terminals.....	33
6.2 System simulator conditions.....	33
7 Narrowband telephony transmission performance test methods.....	34
7.1 Applicability.....	34
7.2 Overall loss/loudness ratings	34
7.2.1 General.....	34
7.2.2 Connections with handset UE.....	34
7.2.2.1 Sending loudness rating (SLR)	34
7.2.2.2 Receiving loudness rating (RLR).....	34
7.2.2.3 Receiving loudness rating (RLR) in the presence of background noise.....	35
7.2.3 Connections with desktop and vehicle-mounted hands-free UE.....	36
7.2.3.1 Sending loudness rating (SLR)	36
7.2.3.2 Receiving Loudness Rating (RLR)	36
7.2.4 Connections with hand-held hands-free UE	37
7.2.4.1 Sending loudness rating (SLR)	37

7.2.4.2	Receiving loudness rating (RLR).....	37
7.2.5	Connections with headset UE.....	38
7.3	Idle channel noise (handset and headset UE).....	38
7.3.1	Sending.....	38
7.3.2	Receiving.....	38
7.4	Sensitivity/frequency characteristics.....	39
7.4.1	Handset and headset UE sending.....	39
7.4.2	Handset and headset UE receiving.....	39
7.4.3	Desktop and vehicle-mounted hands-free UE sending.....	39
7.4.4	Desktop and vehicle-mounted hands-free UE receiving.....	40
7.4.5	Hand-held hands-free UE sending.....	40
7.4.6	Hand-held hands-free UE receiving.....	40
7.5	Sidetone characteristics.....	40
7.5.1	Connections with handset UE.....	40
7.5.1.1	void.....	41
7.5.1.2	Connections with handset UE – HATS method.....	41
7.5.2	Headset UE.....	41
7.5.3	Hands-free UE (all categories).....	41
7.5.4	Sidetone delay for handset or headset.....	41
7.6	Stability loss.....	42
7.7	Acoustic echo control.....	43
7.7.1	General.....	43
7.7.2	Acoustic echo control in a hands-free UE.....	43
7.7.3	Acoustic echo control in handset UE.....	44
7.7.4	Acoustic echo control in a headset UE.....	44
7.8	Distortion.....	44
7.8.1	Sending distortion.....	44
7.8.2	Receiving.....	45
7.9	Void.....	46
7.10	Delay.....	47
7.10.0	UE Delay Measurement Methodologies.....	47
7.10.1	Delay in sending direction (Handset UE).....	47
7.10.1a	Delay in sending direction (headset UE).....	48
7.10.2	Delay in receiving direction (handset UE).....	48
7.10.2a	Delay in receiving direction (headset UE).....	49
7.10.3	Delay in sending + receiving direction using "echo" method (handset UE).....	50
7.10.3a	Delay in sending + receiving direction using "echo" method (headset UE).....	51
7.10.4	Delay and speech quality in conditions with packet arrival time variations and packet loss.....	51
7.10.4.1	Delay in sending direction.....	51
7.10.4.2	Delay in receiving direction.....	51
7.10.4.3	Speech quality loss in conditions with packet arrival time variations and packet loss.....	53
7.10.5	UE send clock accuracy.....	53
7.10.6	UE receiving with clock skew.....	53
7.11	Echo control characteristics.....	53
7.11.1	Test set-up and test signals.....	53
7.11.2	Test method.....	54
7.11.2.1	Signal alignment.....	55
7.11.2.2	Signal level computation and frame classification.....	55
7.11.2.3	Classification into categories.....	56
7.12	Quality (speech quality, noise intrusiveness) in the presence of ambient noise.....	56
7.12.1	Handset.....	56
7.12.2	Hand-held hands-free.....	58
7.13	Jitter buffer management behaviour.....	59
7.13.0	General.....	59
7.13.1	Delay histogram.....	59
7.13.2	Speech quality loss histogram.....	60
8	Wideband telephony transmission performance test methods.....	61
8.1	Applicability.....	61
8.2	Overall loss/loudness ratings.....	61
8.2.1	General.....	61
8.2.2	Connections with handset UE.....	61

8.2.2.1	Sending loudness rating (SLR)	61
8.2.2.2	Receiving loudness rating (RLR)	61
8.2.2.3	Receiving loudness rating (RLR) in the presence of background noise	62
8.2.3	Connections with desktop and vehicle-mounted hands-free UE	62
8.2.3.1	Sending loudness rating (SLR)	63
8.2.3.2	Receiving loudness rating (RLR)	63
8.2.4	Connections with hand-held hands-free UE	63
8.2.4.1	Sending loudness rating (SLR)	63
8.2.4.2	Receiving loudness rating (RLR)	64
8.2.5	Connections with headset UE	64
8.3	Idle channel noise (handset and headset UE)	64
8.3.1	Sending	64
8.3.2	Receiving	65
8.4	Sensitivity/frequency characteristics	65
8.4.1	Handset and headset UE sending	65
8.4.2	Handset and headset UE receiving	65
8.4.3	Desktop and vehicle-mounted hands-free UE sending	66
8.4.4	Desktop and vehicle-mounted hands-free UE receiving	66
8.4.5	Hand-held hands-free UE sending	66
8.4.6	Hand-held hands-free UE receiving	67
8.5	Sidetone characteristics	67
8.5.1	Connections with handset UE	67
8.5.2	Headset UE	67
8.5.3	Hands-free UE (all categories)	68
8.5.4	Sidetone delay for handset or headset	68
8.6	Stability loss	68
8.7	Acoustic echo control	70
8.7.1	General	70
8.7.2	Acoustic echo control in a hands-free UE	70
8.7.3	Acoustic echo control in a handset UE	70
8.7.4	Acoustic echo control in a headset UE	70
8.8	Distortion	70
8.8.1	Sending distortion	70
8.8.2	Receiving	72
8.9	Void	73
8.10	Delay	73
8.10.0	UE Delay Measurement Methodologies	73
8.10.1	Delay in sending direction (handset UE)	74
8.10.1a	Delay in sending direction (headset UE)	74
8.10.2	Delay in receiving direction (handset UE)	75
8.10.2a	Delay in receiving direction (headset UE)	76
8.10.3	Delay in sending + receiving direction using "echo" method (handset UE)	76
8.10.3a	Delay in sending + receiving direction using "echo" method (headset UE)	78
8.10.4	Delay and speech quality in conditions with packet arrival time variations and packet loss	78
8.10.4.1	Delay in sending direction	78
8.10.4.2	Delay in receiving direction	78
8.10.4.3	Speech quality loss in conditions with packet arrival time variations and packet loss	79
8.10.5	UE send clock accuracy	80
8.10.6	UE receiving with clock skew	80
8.11	Echo control characteristics	80
8.11.1	Test set-up and test signals	80
8.11.2	Test method	81
8.11.2.1	Signal alignment	82
8.11.2.2	Signal level computation and frame classification	82
8.11.2.3	Classification into categories	82
8.12	Quality (speech quality, noise intrusiveness) in the presence of ambient noise	83
8.12.1	Handset	83
8.12.2	Hand-held hands-free	84
8.13	Jitter buffer management behaviour	85
8.13.0	General	85
8.13.1	Delay histogram	85
8.13.2	Speech quality loss histogram	86

9	Super-wideband telephony transmission performance test methods.....	87
9.1	Applicability.....	87
9.2	Overall loss/loudness ratings.....	87
9.2.1	General.....	87
9.2.2	Connections with handset UE.....	87
9.2.2.1	Sending loudness rating (SLR).....	87
9.2.2.2	Receiving loudness rating (RLR).....	87
9.2.2.3	Receiving loudness rating (RLR) in the presence of background noise.....	87
9.2.3	Connections with desktop and vehicle-mounted hands-free UE.....	88
9.2.3.1	Sending loudness rating (SLR).....	88
9.2.3.2	Receiving loudness rating (RLR).....	88
9.2.4	Connections with hand-held hands-free UE.....	88
9.2.4.1	Sending loudness rating (SLR).....	88
9.2.4.2	Receiving loudness rating (RLR).....	88
9.2.5	Connections with headset UE.....	88
9.3	Idle channel noise (handset and headset UE).....	88
9.3.1	Sending.....	88
9.3.2	Receiving.....	89
9.4	Sensitivity/frequency characteristics.....	89
9.4.0	General.....	89
9.4.1	Handset and headset UE sending.....	89
9.4.1.1	Handset UE sending.....	89
9.4.1.2	Headset UE sending.....	90
9.4.2	Handset and headset UE receiving.....	90
9.4.2.1	Handset UE receiving.....	90
9.4.2.2	Headset UE receiving.....	90
9.4.3	Desktop and vehicle-mounted hands-free UE sending.....	90
9.4.4	Desktop and vehicle-mounted hands-free UE receiving.....	91
9.4.5	Hand-held hands-free UE sending.....	91
9.4.6	Hand-held hands-free UE receiving.....	91
9.5	Sidetone characteristics.....	92
9.5.1	Connections with handset UE.....	92
9.5.2	Headset UE.....	92
9.5.3	Hands-free UE (all categories).....	92
9.5.4	Sidetone delay for handset or headset.....	92
9.6	Stability loss.....	92
9.7	Acoustic echo control.....	93
9.7.1	General.....	93
9.7.2	Acoustic echo control in a hands-free UE.....	93
9.7.3	Acoustic echo control in a handset UE.....	94
9.7.4	Acoustic echo control in a headset UE.....	94
9.8	Distortion.....	94
9.8.1	Sending distortion.....	94
9.8.2	Receiving.....	94
9.9	Void.....	94
9.10	Delay.....	94
9.10.0	UE Delay Measurement Methodologies.....	94
9.10.1	Delay in sending direction (handset UE).....	94
9.10.1a	Delay in sending direction (headset UE).....	95
9.10.2	Delay in receiving direction (handset UE).....	95
9.10.2a	Delay in receiving direction (headset UE).....	95
9.10.3	Delay in sending + receiving direction using "echo" method (handset UE).....	95
9.10.3a	Delay in sending + receiving direction using "echo" method (headset UE).....	95
9.10.4	Delay and speech quality in conditions with packet arrival time variations and packet loss.....	95
9.10.4.1	Delay in sending direction.....	95
9.10.4.2	Delay in receiving direction.....	95
9.10.4.3	Speech quality loss in conditions with packet arrival time variations and packet loss.....	96
9.10.5	UE send clock accuracy.....	96
9.10.6	UE receiving with clock skew.....	97
9.11	Echo control characteristics.....	97
9.11.1	Test set-up and test signals.....	97
9.11.2	Test method.....	97

9.11.2.1	Signal alignment	97
9.11.2.2	Signal level computation and frame classification	97
9.11.2.3	Classification into categories	97
9.12	Quality (speech quality, noise intrusiveness) in the presence of ambient noise	97
9.12.1	Handset	97
9.12.2	Hand-held hands-free	99
9.13	Jitter buffer management behaviour	100
9.13.0	General	100
9.13.1	Delay histogram	100
9.13.2	Speech quality loss histogram	101
10	Fullband telephony transmission performance test methods	102
10.1	Applicability	102
10.2	Overall loss/loudness ratings	102
10.2.1	General	102
10.2.2	Connections with handset UE	102
10.2.2.1	Sending loudness rating (SLR)	102
10.2.2.2	Receiving loudness rating (RLR)	102
10.2.2.3	Receiving loudness rating (RLR) in the presence of background noise	102
10.2.3	Connections with desktop and vehicle-mounted hands-free UE	102
10.2.3.1	Sending loudness rating (SLR)	102
10.2.3.2	Receiving loudness rating (RLR)	103
10.2.4	Connections with hand-held hands-free UE	103
10.2.4.1	Sending loudness rating (SLR)	103
10.2.4.2	Receiving loudness rating (RLR)	103
10.2.5	Connections with headset UE	103
10.3	Idle channel noise (handset and headset UE)	103
10.3.1	Sending	103
10.3.2	Receiving	103
10.4	Sensitivity/frequency characteristics	104
10.4.0	General	104
10.4.1	Handset and headset UE sending	104
10.4.2	Handset and headset UE receiving	104
10.4.3	Desktop and vehicle-mounted hands-free UE sending	104
10.4.4	Desktop and vehicle-mounted hands-free UE receiving	104
10.4.5	Hand-held hands-free UE sending	104
10.4.6	Hand-held hands-free UE receiving	104
10.5	Sidetone characteristics	104
10.5.1	Connections with handset UE	104
10.5.2	Headset UE	104
10.5.3	Hands-free UE (all categories)	104
10.5.4	Sidetone delay for handset or headset	104
10.6	Stability loss	105
10.7	Acoustic echo control	106
10.7.1	General	106
10.7.2	Acoustic echo control in a hands-free UE	106
10.7.3	Acoustic echo control in a handset UE	106
10.7.4	Acoustic echo control in a headset UE	107
10.8	Distortion	107
10.8.1	Sending distortion	107
10.8.2	Receiving	107
10.9	Void	107
10.10	Delay	107
10.10.0	UE Delay Measurement Methodologies	107
10.10.1	Delay in sending direction (handset UE)	107
10.10.1a	Delay in sending direction (headset UE)	107
10.10.2	Delay in receiving direction (handset UE)	107
10.10.2a	Delay in receiving direction (headset UE)	107
10.10.3	Delay in sending + receiving direction using "echo" method (handset UE)	107
10.10.3a	Delay in sending + receiving direction using "echo" method (headset UE)	108
10.10.4	Delay and speech quality in conditions with packet arrival time variations and packet loss	108
10.10.4.1	Delay in sending direction	108

10.10.4.2	Delay in receiving direction	108
10.10.4.3	Speech quality loss in conditions with packet arrival time variations and packet loss.....	108
10.10.5	UE send clock accuracy	108
10.10.6	UE receiving with clock skew	108
10.11	Echo control characteristics.....	108
10.11.1	Test set-up and test signals	108
10.11.2	Test method	108
10.11.2.1	Signal alignment	108
10.11.2.2	Signal level computation and frame classification	108
10.11.2.3	Classification into categories	108
10.12	Quality (speech quality, noise intrusiveness) in the presence of ambient noise	109
10.12.1	Handset	109
10.12.2	Hand-held hands-free.....	109
10.13	Jitter buffer management behaviour	109
Annex A (informative):	Void	110
Annex B (informative):	Reference algorithm for echo control characteristics evaluation.....	111
B.1	General	111
B.2	Test script	112
B.3	Reference algorithm	113
B.3.1	Main algorithm	113
B.3.2	Delay compensation	115
B.3.3	Signal level computation and frame classification	115
B.3.4	Level vs time computation	118
B.3.5	Categorization	119
B.3.6	Auxiliary functions for reporting data	121
B.3.7	Other helper functions	122
Annex C (informative):	Measurement method for determining the one way radio delays of LTE radio network simulators	124
C.1	Measurement setup.....	124
Annex D (normative):	Clock skew measurement.....	125
D.1	Test procedure	125
D.2	Clock skew estimation algorithm	125
Annex E (normative):	Packet delay and loss profiles for UE delay testing of MTSI-based speech with LTE access	127
E.1	General	127
E.2	Simulation model for generating packed delay and loss profiles.....	127
E.3	Packet delay and loss profiles for simulated stationary operating conditions with DRX 20 ms and DRX 40 ms.....	130
Annex F (normative):	Packet delay and loss profiles for jitter buffer management behaviour	131
F.1	General	131
F.2	Packet delay and loss profiles	131
Annex G (informative):	Change history	133
History		136

Foreword

This Technical Specification has been produced by the 3GPP.

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of this TS, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

- x the first digit:
 - 1 presented to TSG for information;
 - 2 presented to TSG for approval;
 - 3 or greater indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the specification.

Introduction

The present document specifies test methods to allow the minimum performance requirements for the acoustic characteristics of GSM, 3G, LTE, NR and WLAN terminals when used to provide narrowband, wideband, super-wideband or fullband telephony to be assessed.

The objective for narrowband services is to reach a quality as close as possible to ITU-T standards for PSTN circuits. However, due to technical and economic factors, there cannot be full compliance with the general characteristics of international telephone connections and circuits recommended by the ITU-T.

The performance requirements are specified in TS 26.131; the test methods and considerations are specified in the main body of the text.

1 Scope

The present document is applicable to any terminal capable of supporting narrowband, wideband, super-wideband or fullband telephony, either as a stand-alone service or as the telephony component of a multimedia service. The present document specifies test methods to allow the minimum performance requirements for the acoustic characteristics of GSM, 3G, LTE, NR and WLAN terminals when used to provide narrowband, wideband, super-wideband or fullband telephony to be assessed.

NOTE For 3G, LTE, NR and WLAN, acoustic requirements are specified in TS 26.131, test methods are specified in TS 26.132. For GSM, most acoustic requirements are specified in TS 43.050, test methods are specified in TS 51.010. These specifications are in many cases harmonized with or even refer to TS 26.131 and TS 26.132. See TS 43.050 and TS 51.010 for details. The reason for including GSM, UMTS, LTE, NR and WLAN terminals within the scope of the present specification is to avoid, whenever possible, duplication of test method descriptions for terminals supporting multiple access technologies..

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

- [1] 3GPP TS 26.131: "Terminal Acoustic Characteristics for Telephony; Requirements".
- [2] ITU-T Recommendation B.12 (1988): "Use of the decibel and the neper in telecommunications".
- [3] ITU-T Recommendation G.103 (1998): "Hypothetical reference connections".
- [4] ITU-T Recommendation G.111 (1993): "Loudness ratings (LRs) in an international connection".
- [5] ITU-T Recommendation G.121 (1993): "Loudness ratings (LRs) of national systems".
- [6] ITU-T Recommendation G.122 (1993): "Influence of national systems on stability and talker echo in international connections".
- [7] Void.
- [8] ITU-T Recommendation P.11 (1993): "Effect of transmission impairments".
- [9] ITU-T Recommendation P.38 (1993): "Transmission characteristics of operator telephone systems (OTS)".
- [10] ITU-T Recommendation P.50 (1993): "Artificial voices".
- [11] 3GPP TS 43.058 : "Digital Cellular Telecommunications System Characterization test methods and quality assessment for hands-free mobiles".
- [12] IEC Publication 60651: "Sound Level Meters".
- [13] ITU-T Recommendation P.51 (1996): "Artificial mouth".
- [14] ITU-T Recommendation P.57 (12/2011): "Artificial ears".
- [15] ITU-T Recommendation P.58 (05/2013): "Head and torso simulator for telephonometry."

- [16] ITU-T Recommendation P.79 (11/2007) with Annex A: "Calculation of loudness ratings for telephone sets."
- [17] 3GPP TS 46.077 : "Minimum Performance Requirements for Noise Suppressor Application to the AMR Speech Encoder".
- [18] ITU-T Recommendation P.64 (11/2007): "Determination of sensitivity/frequency characteristics of local telephone systems".
- [19] ITU-T Recommendation P.581 (02/2014): "Use of head and torso simulator (HATS) for hands-free and handset terminal testing".
- [20] ITU-T Recommendation P.340 (05/2000): "Transmission characteristics and speech quality parameters of hands-free terminals".
- [21] ITU-T Recommendation G.712 (11/2001): "Transmission performance characteristics of pulse code modulation channels".
- [22] ITU-T Recommendation P.501 (06/2015): "Test signals for use in telephony".
- [23] ITU-T Recommendation O.41 (10/1994): "Psophometer for use on telephone-type circuits".
- [24] ITU-T Recommendation O.131 (11/1988): "Quantizing distortion measuring equipment using a pseudo-random noise test signal".
- [25] Void.
- [26] ISO 3745: "Acoustics - Determination of sound power levels of noise sources using sound pressure - Precision methods for anechoic and hemi-anechoic rooms".
- [27] ITU-T Recommendation O.132 (11/1988): "Quantizing distortion measuring equipment using a sinusoidal test signal".
- [28] ETSI TS 103 737 (2010-08) V1.1.2: "Transmission requirements for narrowband wireless terminals (handset and headset) from a QoS perspective as perceived by the user".
- [29] ETSI TS 103 738 (2010-09) V1.1.2: "Transmission requirements for narrowband wireless terminals (handsfree) from a QoS perspective as perceived by the user".
- [30] ETSI TS 103 739 (2010-09) V1.1.2: "Transmission requirements for wideband wireless terminals (handset and headset) from a QoS perspective as perceived by the user".
- [31] ETSI TS 103 740 (2010-09) V1.1.2: "Transmission requirements for wideband wireless terminals (handsfree) from a QoS perspective as perceived by the user".
- [32] ITU-T Recommendation P.380 (11/2003): "Electro-acoustic measurements on headsets".
- [33] ITU-T Recommendation P.501 Amendment 1 (2012): "Test signals for use in telephony".
- [34] ETSI TS 103 106(2013-03) V1.2.1: "Speech Quality performance in the presence of background noise: Background noise transmission of mobile terminals-Objective test methods".
- [35] ETSI ES 202 396-1 (2012-10) V1.4.1: "Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database".
- [36] ETSI EG 202 396-3 (2011-02) V1.3.1: "Speech quality performance in the presence of background noise; Part 3: *Background noise transmission – objective test methods*: Background noise simulation technique and background noise database".
- [37] ITU-T Recommendation P.56 (12/2011): "Objective measurement of active speech level".
- [38] IEC 61672: "Electroacoustics – sound level meters - part 1: specifications".
- [39] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia Telephony; Media handling and interaction".

- [40] 3GPP TS 43.050: "Transmission planning aspects of the speech service in the GSM Public Land Mobile Network (PLMN) system".
- [41] 3GPP TS 51.010: "Mobile Station (MS) conformance specification; Part 1: Conformance specification".
- [42] 3GPP TS 23.203: "Policy and charging control architecture".
- [43] ETSI TS 103 224 (2015-08): V1.2.1 "A sound field reproduction method for terminal testing including a background noise database".
- [44] ITU-T Recommendation P.863 (09/2014): "Perceptual objective listening quality assessment".
- [45] ITU-T Recommendation P.863.1 (09/2014): "Application guide for Recommendation ITU-T P.863".
- [46] 3GPP TS 36.521-1: "User Equipment (UE) conformance specification Radio transmission and reception; Part 1: Conformance Testing".
- [47] 3GPP TR 21.905: "Vocabulary for 3GPP specifications".
- [48] 3GPP TS 23.402: "Architecture enhancements for non-3GPP accesses".
- [49] 3GPP TS 24.302: "Access to the 3GPP Evolved Packet Core (EPC) via non-3GPP access networks; Stage 3".
- [50] ETSI TS 103 281(2017-04) V1.1.1: "Speech quality in the presence of background noise: Objective test methods for super-wideband and fullband terminals".
- [51] ITU-T Recommendation P.10/G.100 (06/2019).
- [52] 3GPP TR 26.921: "Investigations on ambient noise reproduction systems for acoustic testing of terminals".

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of the present document the terms *narrowband*, *wideband*, *super-wideband* and *fullband* refer to signals associated with the corresponding operating modes of the speech codecs specified in 5.2.

For the purposes of the present document, the terms dB, dBr, dBm0, dBm0p and dBA, shall be interpreted as defined in ITU-T Recommendation G.100 [51]; the term dBPa shall be interpreted as the sound pressure level relative to 1 pascal expressed in dB (0 dBPa is equivalent to 94 dB SPL).

A 3GPP softphone is a telephony system running on a general purpose computer or PDA complying with the 3GPP terminal acoustic requirements (TS 26.131 and 26.132).

For the purposes of the present document the term *clock skew* is defined as the difference between the clock of the device under test (C_{DUT}) and the clock of the reference client (C_{REF}). The skew of C_{DUT} relative to C_{REF} is defined in parts per million (PPM) as: $(C_{DUT} - C_{REF}) \cdot 10^6 / C_{REF}$.

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [47] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [47].

5GC	5G Core Network
ADC	Analogue to Digital Converter

AMR	Adaptive Multi Rate
CSS	Composite Source Signal
DAC	Digital to Analogue Converter
DRP	Eardrum Reference Point
DTX	Discontinuous Transmission
EEC	Electrical Echo Control
EEP	Ear Entrance Point
EL	Echo Loss
ERP	Ear Reference Point
EVS	Enhanced Voice Services
FFT	Fast Fourier Transform
G-MOS-LQO _n	Global (Overall) - Mean Opinion Score - Listening Quality Objective - Narrowband
G-MOS-LQO _w	Global (Overall) - Mean Opinion Score - Listening Quality Objective - Wideband
G-MOS-LQO _{fb}	Global (Overall) - Mean Opinion Score - Listening Quality Objective - Fullband
HATS	Head and Torso Simulator
IMS	IP Multimedia Subsystem
LSTR	Listener Sidetone Rating
LTE	Long Term Evolution
MRP	Mouth Reference Point
MS	Mobile Station
MTSI	Multimedia Telephony Service for IMS
N-MOS-LQO _n	Noise (Background) - Mean Opinion Score Listening Quality Objective - Narrowband
N-MOS-LQO _w	Noise (Background) - Mean Opinion Score Listening Quality Objective - Wideband
N-MOS-LQO _{fb}	Noise (Background) - Mean Opinion Score Listening Quality Objective - Fullband
NR	New Radio
OLR	Overall Loudness Rating
PCM	Pulse Code Modulation
PDA	Personal Digital Assistant
POI	Point of Interconnection (with PSTN)
PSTN	Public Switched Telephone Network
RLR	Receive Loudness Rating
RMC	Reference Measurement Channel
RMS	Root Mean Squared
SLR	Send Loudness Rating
S-MOS-LQO _n	Speech Signal Quality - Mean Opinion Score - Listening Quality Objective - Narrowband
S-MOS-LQO _w	Speech Signal Quality - Mean Opinion Score - Listening Quality Objective - Wideband
S-MOS-LQO _{fb}	Speech Signal Quality - Mean Opinion Score - Listening Quality Objective - Fullband
SS	System Simulator
STMR	Sidetone Masking Rating
SS	System Simulator
TX	Transmission
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
WLAN	Wireless Local Area Network

4 Interfaces

The interfaces required to define terminal acoustic characteristics are the acoustic interfaces, the air interface and the point of interconnect (POI), see Figure 1.

The Air Interfaces for GSM, 3G, LTE and NR are specified by GSM 05, 3GPP 45, 3GPP 25, 3GPP 36 and 3GPP 38 series specifications, and the Air Interface for WLAN access to EPC is specified by WLAN access to EPC as defined in 3GPP TS 23.402 [48] and TS 24.302 [49]. MTSI speech aspects are specified by TS 26.114 [17].

Measurements can be made using a system simulator (SS) comprising the appropriate radio terminal equipment and speech transcoder. The losses and gains introduced by the test equipment shall be accounted for.

The POI with the public switched telephone network (PSTN) is considered to have a relative level of 0 dB.

Five classes of acoustic interface are considered in this specification:

- Handset UE including softphone UE used as a handset;
- Headset UE including softphone UE used with headset;
- Vehicle Mounted Hands-free UE including softphone UE mounted in a vehicle;
- Desktop-mounted hands-free UE including softphone UE with external loudspeaker(s) used in hands-free mode;
- Hand-held hands-free UE including softphone UE with internal loudspeaker(s) used in hands-free mode.

(See definition of softphone in Clause 3.1)

NOTE: The test setup for a softphone UE shall be derived according to the following rules:

- When using a softphone UE as a handset: the test setup shall correspond to handset mode.
- When using a softphone UE with headset: the test setup shall correspond to headset mode.
- When a softphone UE is mounted in a vehicle: the test setup shall correspond to vehicle-mounted hands-free mode.
- When using a softphone UE in hands-free mode:
 - When using internal loudspeaker(s), the test setup shall correspond to hand-held hands-free.
 - When using external loudspeaker(s), the test setup shall correspond to desktop-mounted hands-free.

5 Test configurations

This section describes the test setups for terminal acoustic testing.

The configurations are shown for one-channel (mono) operation, configurations for two-channel (stereo) operation is for further study.

NOTE: If the terminal has several mechanical configurations (e.g., sliding design open or closed), all manufacturer-defined configurations shall be tested.

5.1 Setup for terminals

The general access to terminals is described in figure 1. The preferred acoustic access to GSM, 3G, LTE, NR and WLAN terminals is the most realistic simulation of the "average" subscriber. This can be made by using HATS (head and torso simulator), with appropriate ear simulation and appropriate mountings of handset terminals to the HATS in a realistic but reproducible way. Hands-free terminals shall use the HATS or free field microphone techniques in a realistic but reproducible way.

HATS is described in ITU-T Recommendation P.58 [15], appropriate ears are described in ITU-T Recommendation P.57 [14] (Type 3.3), proper positioning of handsets in realistic conditions is found in ITU-T Recommendation P.64, and the test setups for various types of hands-free terminals can be found in ITU-T Recommendation P.581.

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

The preferred way of testing is the connection of a terminal to the system simulator with exact defined settings and access points. The test sequences are fed in either electrically using a reference codec, using the direct signal processing approach, or acoustically using ITU-T specified devices.

The system simulator shall simulate the access network and core network including the speech encoding/decoding specified for the test (e.g. AMR-NB or AMR-WB) but excluding further transcoding beyond linear PCM, see Figure 1.

Unless specified otherwise for the respective test, the radio conditions on the air interface shall have a block error rate of 0% and the jitter in the IP transport for MTSI-based speech shall be ≤ 1 ms.

NOTE 1: For WLAN connections, an RF shielded room may be one way to achieve these requirements on block error rate and jitter. Otherwise, care should be taken with potential sources of radio interference and their impact.

In case of MTSI-based speech, the reference client shall allow to synchronize to the clock of the device under test and include a de-jitter buffer to equalize possible jitter in the signal received from the UE.

When operating with synchronized clock, the de-jitter buffer shall be a static de-jitter buffer and the jitter buffer management shall not compensate for clock skew. The reference client shall not lose or discard packets, shall not trigger retransmission, and shall not use error concealment or time-warping. The initial jitter buffer size (filling level) shall be higher than the maximum expected network jitter and the maximum jitter buffer size shall be at least twice the initial size. During jitter buffer reset, the de-jitter buffer shall be emptied/filled to the initial buffer size. In case of buffer over- or underruns, the reference client shall give a warning and it shall be reported.

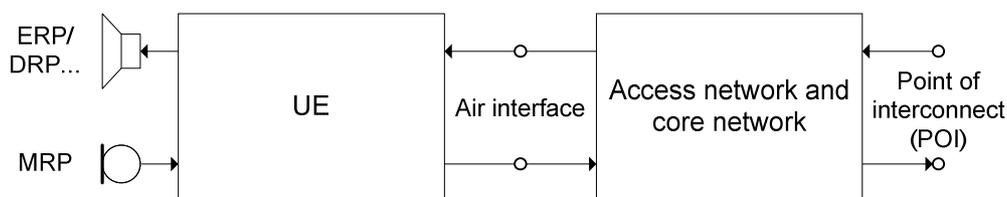
NOTE 2: A static de-jitter buffer is a first-in-first-out (FIFO) buffer which at the beginning buffers packets until a given initial buffer size is reached. Due to changing network delays the filling level of the de-jitter buffer can change, but the sum of network delay and jitter buffer delay is constant (as opposed to an adaptive jitter buffer management). The filling level of the de-jitter buffer represents the de-jitter buffer delay.

For measurements with unsynchronized clock e.g. the measurement of clock skew, jitter buffer over- and underruns cannot be avoided due to the unsynchronized clocks. Under the assumption of jitter-free condition the initial jitter buffer size (filling level) shall be chosen such that the maximum clock skew can be compensated without any loss of packets for a given time. For the measurement of clock skews the jitter buffer size should be chosen such that for clock skew of up to 100ppm no loss of packets due to buffer over- or under-run shall occur for a sequence of 160s.

For LTE and NR connections, the system simulator shall be configured for FDD operation, with a default or dedicated bearer (EPC) and QoS flow (5GC) and reference measurement channel scheduling that provides enough resource block allocation for transmitting a full speech packet within a transmission time interval of 1ms. No HARQ re-transmissions shall occur. TDD operation, TTI bundling, connected DRX and other forms of scheduling (e.g. SPS) are for further study.

The test setup has to ensure proper clock synchronization of the test equipment to the UE. Clock skew shall be negligible and packet loss shall not occur during the test.

NOTE 3: Any clock skew may result in improper delay calculation or in wrong positioning of the analysis window.



NOTE: Connection to PSTN should include electrical echo control (EEC).

Figure 1: Interfaces (MRP, ERP/DRP..., Air interface and Point of interconnect) for specification of terminal acoustic characteristics

5.1.1 Setup for handset terminals

When using a handset UE, the handset is placed on HATS as described in ITU-T Recommendation P.64 Annex E [18]. A suitable position shall be defined for each handset UE and documented in the test report. The artificial mouth shall conform to ITU-T Recommendation P.58 [15]. The artificial ear shall conform to ITU-T Recommendation P.57 [14]. Type 3.3 ear shall be used and positioned on HATS according to ITU-T Recommendation P.58 [15].

Position and calibration of HATS

The sending and receiving characteristics shall be tested with the HATS. It shall be indicated what application force was used. If not stated otherwise in TS 26.131, an application force of 8 ± 2 N shall be used.

The horizontal positioning of the HATS reference plane shall be guaranteed within $\pm 2^\circ$.

5.1.2 Setup for headset terminals

Recommendations for the setup and positioning of headsets are given in ITU-T Recommendation P.380. If not stated otherwise, headsets shall be placed in their recommended wearing position. Some insert earphones might not fit properly in Type 3.3 ear simulators. For such insert type headsets, an ITU-T Recommendation P.57 [14] Type 2 ear simulator may be used in conjunction with the HATS mouth simulator. The HATS should be equipped with two artificial ears as specified in ITU-T Recommendation P.57 [14]. For binaural headsets two artificial ears are required.

5.1.3 Setup for hands-free terminals

5.1.3.1 Vehicle-mounted hands-free

If not stated otherwise, the artificial head (HATS – head and torso simulator, according to ITU-T Recommendation P.58 [15]) is positioned in the driver's seat for the measurement as shown in figure 3a. The position has to be in line with the average users' position; therefore, all positions and sizes of users have to be taken into account. Typically, all except the tallest 5% and the shortest 5% of the driving population have to be considered. The size of these persons can be derived, e.g., from the 'anthropometric data set' for the corresponding year (e.g., based on data used by car manufacturers). The position of the HATS (mouth/ears) within the positioning arrangement is given individually by each car manufacturer. The position used has to be reported in detail in the test report. If no requirements for positioning are given the distance from the microphone to the MRP is defined by the test lab.

By using suitable measures (e.g., marks in the car, relative position to A-pillar, B-pillar, height from the floor, etc.) an exact reproduction of the artificial head position must be possible at any later time.

NOTE – Different positions of the artificial head may greatly influence the test results. Depending on the application, different positions of the artificial head may be chosen for the tests. It is recommended to check the worst-case position, e.g., those positions where the SNR and/or the speech quality in send may be worst.

Figure 2: void

Figure 3: void

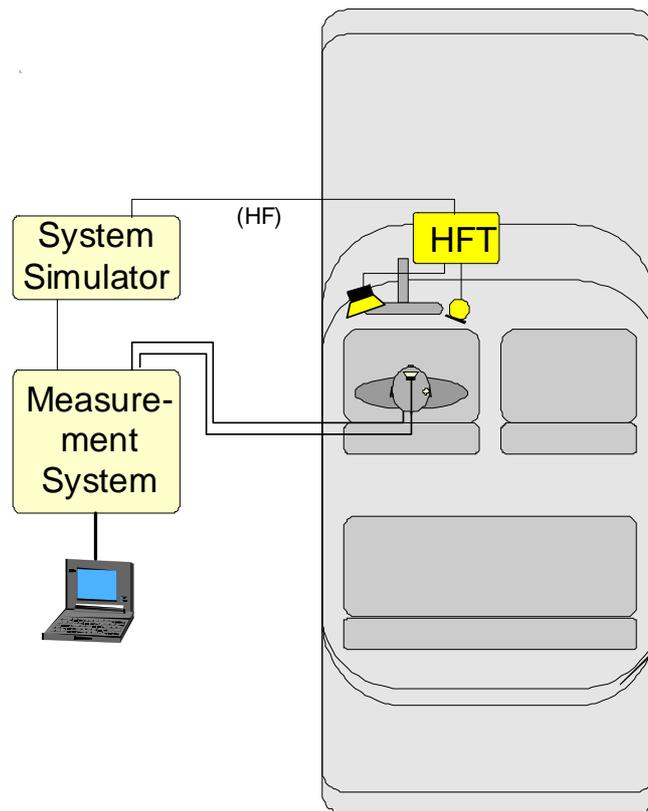


Figure 3a: Test Configuration for vehicle mounted hands-free, using HATS

5.1.3.2 Desktop hands-free

For HATS test equipment, the definition of hands-free terminals and setup for desktop hands-free terminals can be found in ITU-T Recommendation P.581. Measurement setup using a free-field microphone and a discrete P.51 [13] artificial mouth for desktop hands-free terminals can be found in ITU-T Recommendation P.340. The positioning for different types of desktop hands-free terminals is given in ETSI TS 103 738 and ETSI TS 103 740.

5.1.3.3 Hand-held hands-free

Either HATS or a free-field microphone with a discrete P.51 [13] artificial mouth may be used to measure a hand-held hands-free type UE.

If HATS measurement equipment is used, it shall be configured to the hand-held hands-free UE according to figure 4. The HATS should be positioned so that the HATS Reference Point is at a distance d_{HF} from the centre point of the visual display of the Mobile Station. The distance d_{HF} is specified by the manufacturer. A vertical angle θ_{HF} may be specified by the manufacturer. Where it is not specified, the nominal distance d_{HF} shall be 42 cm and θ_{HF} shall be 0° .

NOTE: The nominal distance of 42 cm corresponds to the distance between the HATS reference point and lip-plane (12 cm) with an additional 30 cm giving a realistic figure as a reference usage of hand-held terminals.

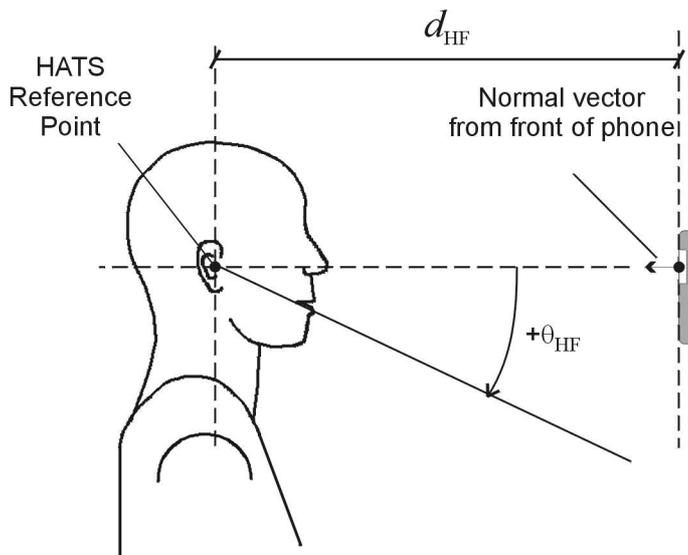


Figure 4: Configuration of hand-held hands-free UE relative to the HATS

If a free-field microphone and a discrete P.51 [13] mouth are used, they shall be configured to the hand-held hands-free UE according to figure 5 for receiving measurements and figure 6 for sending measurements. The microphone should be located at a distance d_{HF} from the centre of the visual display of the UE. The mouth simulator should be located at a distance $d_{HF}-12$ cm from the centre of the visual display of the UE. The distance d_{HF} is specified by the manufacturer. Where it is not specified the nominal distance d_{HF} shall be 42 cm.

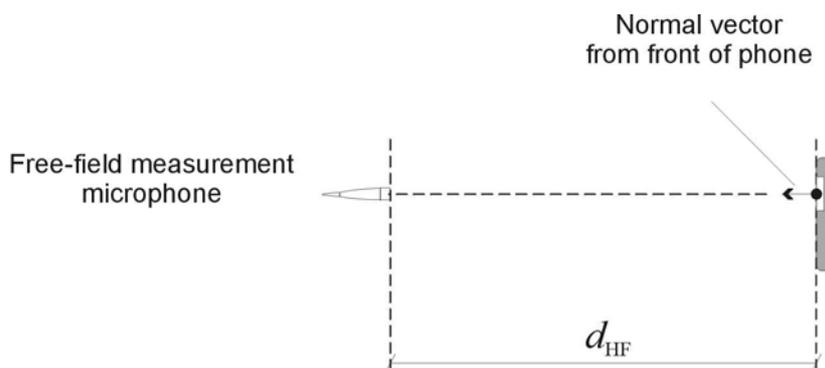


Figure 5: Configuration of hand-held hands-free UE; free-field microphone for receiving measurements

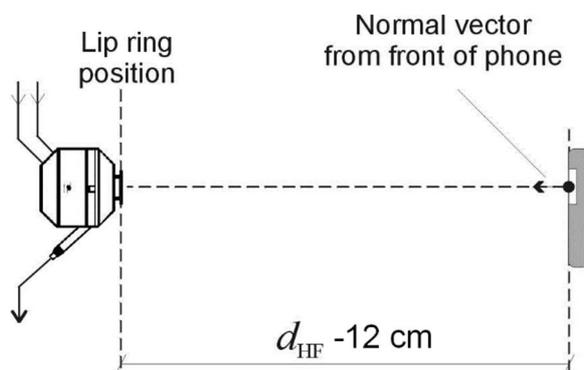


Figure 6: Configuration of hand-held hands-free UE; discrete P.51 artificial mouth for sending measurements

5.1.3.4 Softphone including speakers and microphone

This test setup is applicable to laptop computers or similar devices as seen in figure 7 through figure 11.

Where the manufacturer gives conditions of use, these will apply for testing. If the manufacturer gives no other requirement, the softphone will be positioned according the following conditions:

Measurement with artificial ear and microphone:

Artificial mouth (for sending tests)

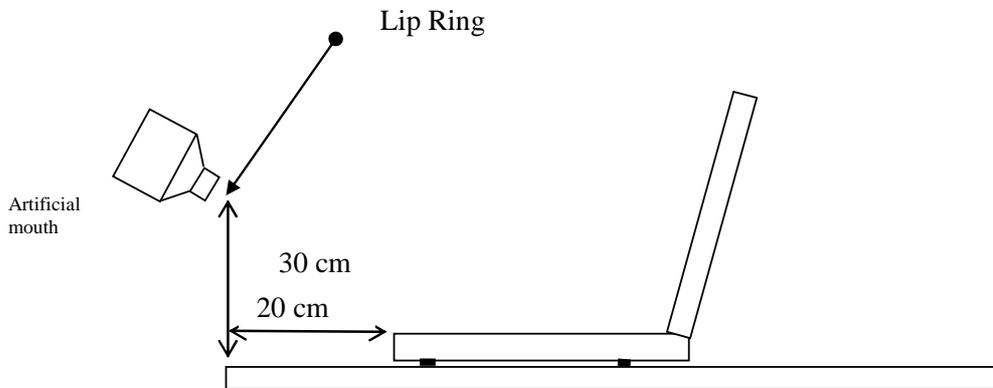


Figure 7: Configuration of a softphone relative to the artificial mouth side view

Free field microphone (for receiving):

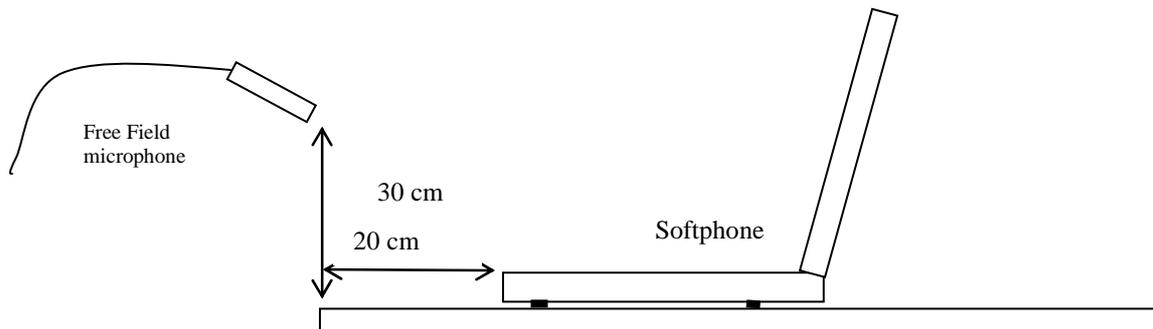


Figure 8: Configuration of a softphone relative to the free field microphone side view

Position of a softphone on the table:

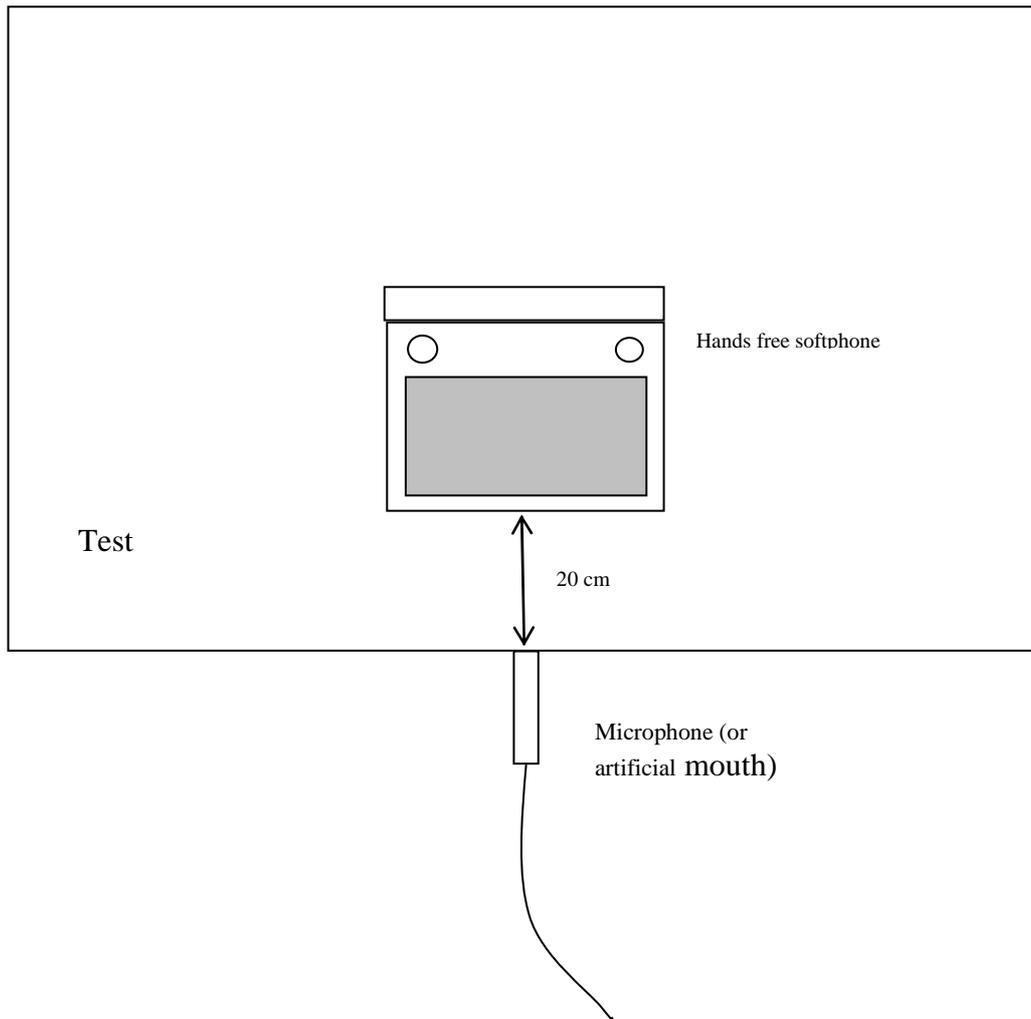


Figure 9: Configuration of a softphone relative to the free-field microphone or artificial mouth viewed from above

Measurement with HATS:

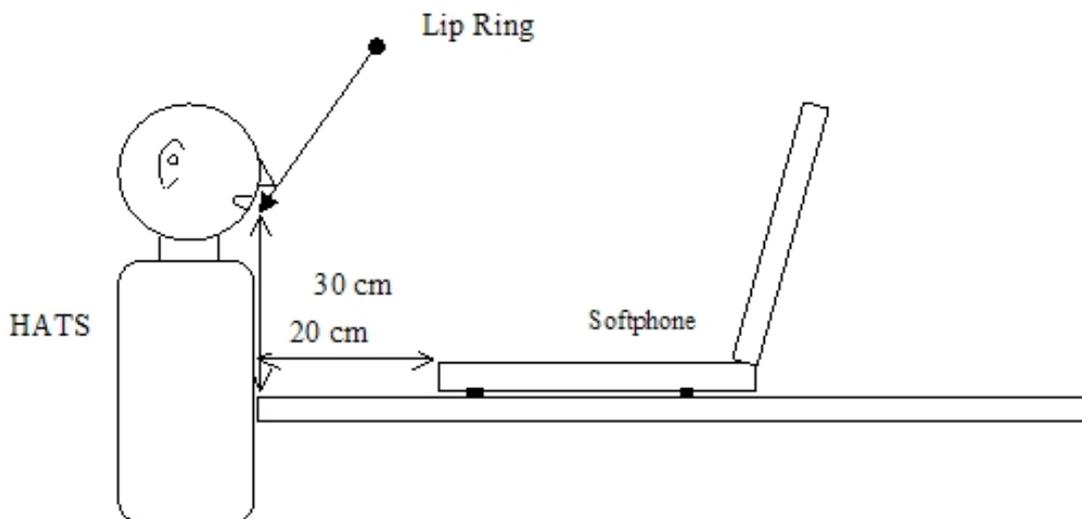


Figure 10: Configuration of a softphone relative to the HATS side view

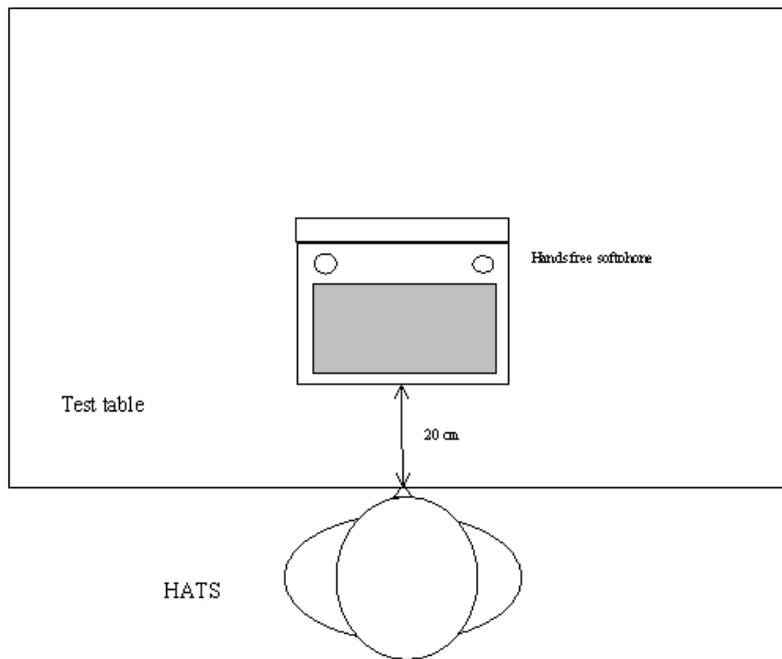


Figure 11: Configuration of a softphone relative to the HATS viewed from above

5.1.3.5 Softphone with separate speakers

This test setup is applicable to laptop computers or similar devices as seen in figure 12 through figure 15.

Where the manufacturer gives conditions of use, these will apply for testing. If the manufacturer gives no other requirement, the softphone will be positioned according to the following conditions:

Where separate loudspeakers are used, the system will be positioned as in figure 12 or figure 13.

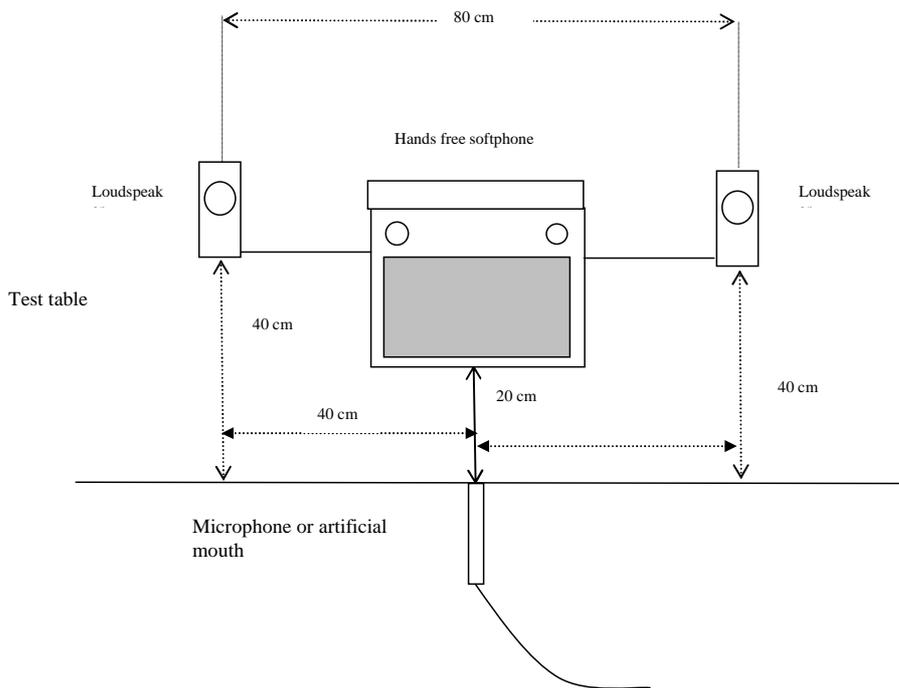


Figure 12: Configuration of a softphone using external speakers relative to microphone or artificial mouth viewed from above

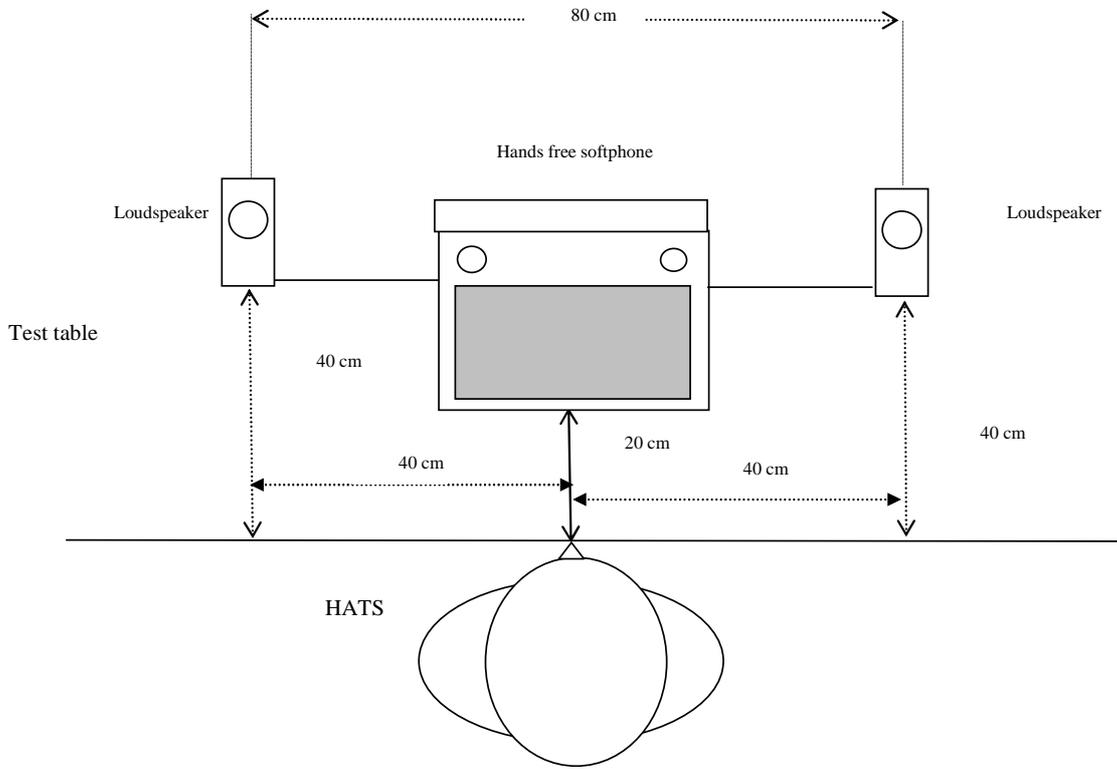


Figure 13: Configuration of a softphone using external speakers relative to the HATS viewed from above

Where an external microphone and speakers are used, the system will be positioned as in figure 14 or figure 15.

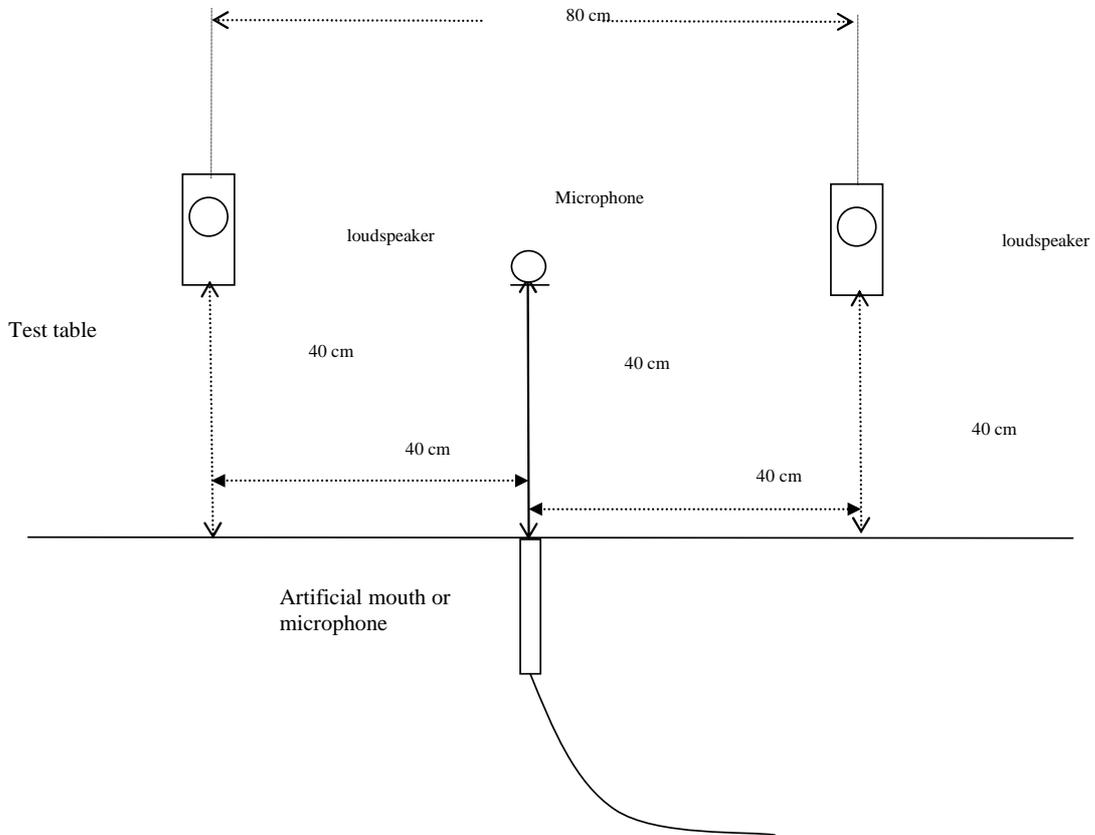


Figure 14: Configuration of a softphone using external speakers and a microphone relative to microphone or artificial mouth viewed from above

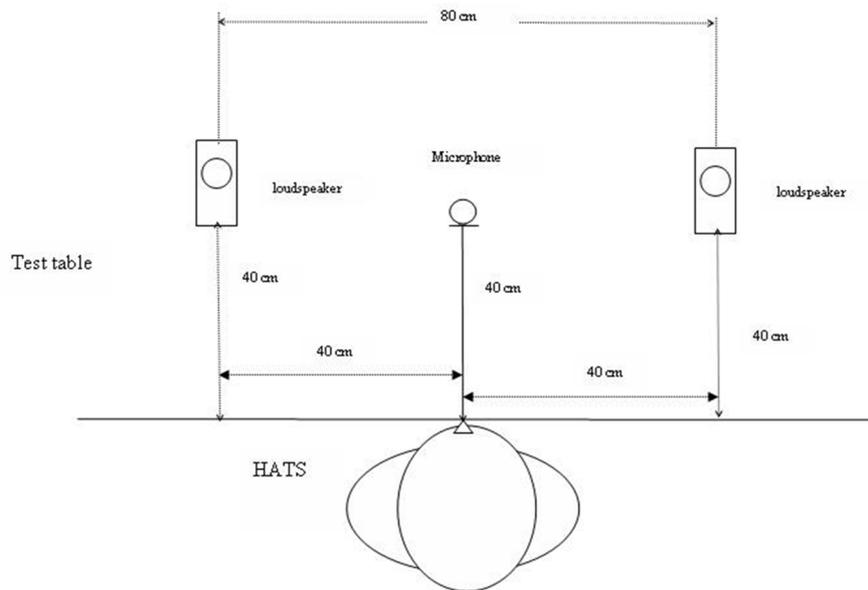


Figure 15: Configuration of a softphone using external speakers and a microphone relative to the HATS viewed from above

5.1.4 Position and calibration of HATS

The horizontal positioning of the HATS reference plane shall be guaranteed within $\pm 2^\circ$ for testing hands-free equipment.

The HATS shall be equipped with a Type 3.3 Artificial Ear. For hands-free measurements the HATS shall be equipped with two artificial ears. The pinnae are specified in Recommendation P.57 [14] for Type 3.3 artificial ears. The pinnae shall be positioned on HATS according to ITU-T Recommendation P.58 [15].

The exact calibration and equalization procedures as well as how to combine the two ear signals for the purpose of measurements can be found in ITU-T Recommendation P.581.

If not stated otherwise, the HATS shall be diffuse-field equalized using values from ITU-T Recommendation P.58 [15]. For 1/3-octave band measurements, the inverse of the nominal diffuse field curve in P.58 table 3 shall be used. For 1/12-octave band measurements, the inverse of the nominal diffuse field curve in P.58 Annex A shall be used. For measurements requiring diffuse-field correction values for frequencies other than those used in the P.58 tables, linear interpolation on a log frequency scale from the P.58 Annex A values shall be used.

For hand-held hands-free UE, the setup corresponding to 'portable hands-free' in ITU-T Recommendation P.581 should be used.

5.1.5 Test setup for quality in the presence of ambient noise measurements

5.1.5.1 Handset

The setup for simulating realistic ambient noises and the positioning of the HATS in a lab-type environment is described in ETSI ES 202 396-1 [35].

ETSI ES 202 396-1 [35] contains a description of the recording arrangement for realistic ambient noises, a description of the setup for a loudspeaker arrangement suitable to simulate an ambient noise field in a lab-type environment and a database of realistic ambient noises, part of which is used for testing the terminal performance with a variety of conditions.

Two different methods (Method 1 and Method 2) for equalization and calibration of the loudspeakers are allowed and specified below. The method selected for the test setup (Method 1 or Method 2) shall be reported. Method 2 is preferred over Method 1 due to the improved cross-lab reproduction accuracy reported in clause 5 of TR 26.921 [52].

Method 1:

The equalization and calibration procedure for this test setup given in ETSI ES 202 396-1 [35] shall be used.

This method works semi-automated and requires manual adjustment steps during the equalization procedure. The equalization is carried out in the frequency range between 50 Hz and 10 kHz.

Method 2:

The equalization and calibration procedure for this test setup given in Clause 7 of ETSI TS 103 224 [43] shall be used. During the equalization procedure, a so-called “reference sound” is needed for verification purposes. In this case, the noise type “TrainStation_bin” of Clause 8 of ETSI TS 103 224 [43] (time range 15.0 to 22.5s) shall be used.

NOTE: Clause 7 of ETSI TS 103 224 [43] describes recording and reproduction arrangement for an (almost) arbitrary number of microphones and loudspeakers. In order to produce ambient noise field compatible to Method 1, a flexible setup with two recording/playback channels (binaural noise signals) and a 4.0 (or 4.1) loudspeaker configuration is used.

In contrast to the Method 1 according to ETSI ES 202 396-1 [35], a fully automated equalization procedure including reduction of cross-talk is provided by Method 2. The equalization is carried out in the frequency range between 50 Hz and 20 kHz.

5.1.5.2 Hand-held hands-free

For hand-held hands-free UE, the test setup for simulating realistic ambient noises and the position of the HATS in a lab-type environment is described in ETSI TS 103 224 [43]. The setup in ETSI ES 202 396-1 [35] may also be used as described below.

ETSI TS 103 224 [43] contains a description of the recording arrangement for realistic ambient noises, a description of the setup for loudspeaker arrangement suitable to simulate an ambient noise field in a lab type environment, and a database of realistic ambient noises, part of which is used for testing the hands-free UE performance. Note that TS 103 224 [43] provides improved reproduction accuracy in terms of frequency and spatial characteristics, and so is recommended for testing hands-free UE. The equalization setup and the test setup is shown in Figs. 15a1 – 15a5.

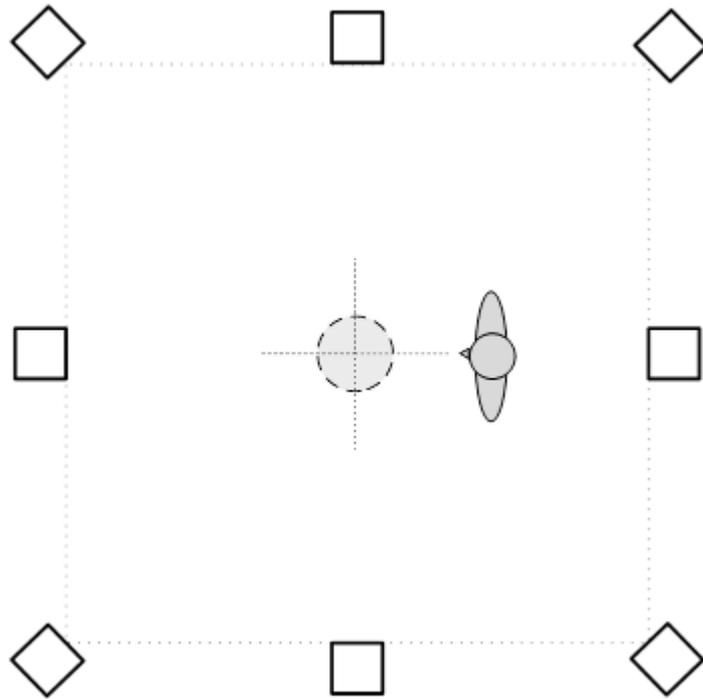


Figure 15a1: Equalization using the TS 103 224 method, the circle indicates the microphone array used for the equalization

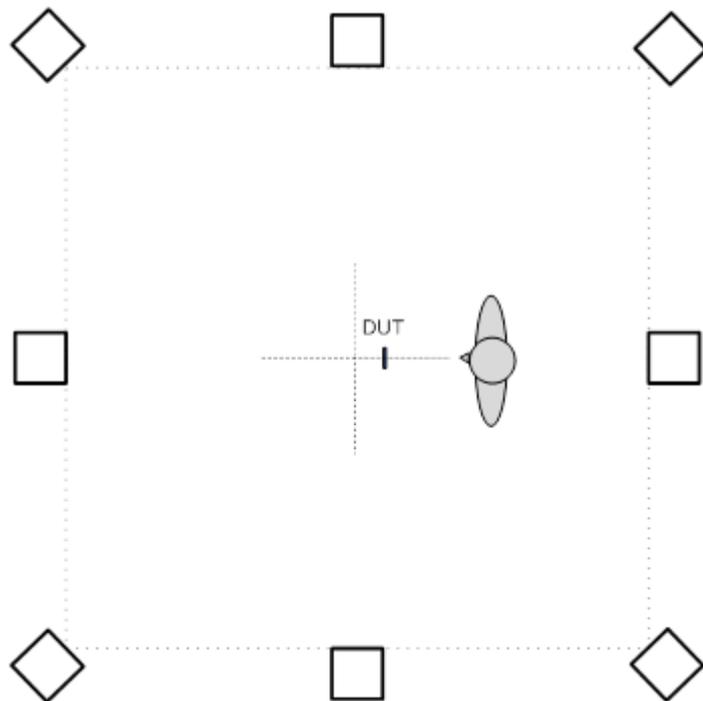


Figure 15a2: Measurements using the TS 103 224 method

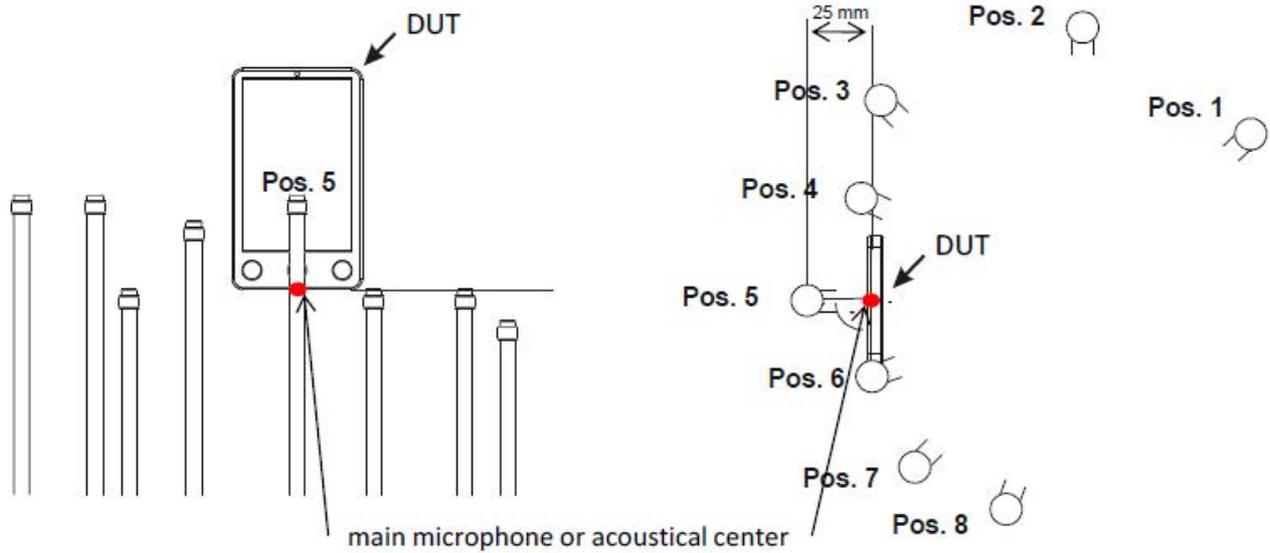


Figure 15a3: Detailed positioning of the hands-free UE

The setup in ETSI ES 202 396-1 [35] may be used for testing hands-free UE, provided that after the equalization and calibration procedure for handset UE are completed, the HATS is removed from the calibration location in the test room, and the hands-free UE under test is placed at the calibration location. The HATS is then placed in a location with respect to the hands-free UE under test, as described in Clause 5.1.3.2 and 5.1.3.3. The equalization setup and the test setup is shown in Figs. 15a4 and 15a5. Each ambient noise in the database of ETSI TS 103 224 [43] is available in a version compatible with reproduction using ETSI ES 202 396-1 [35].

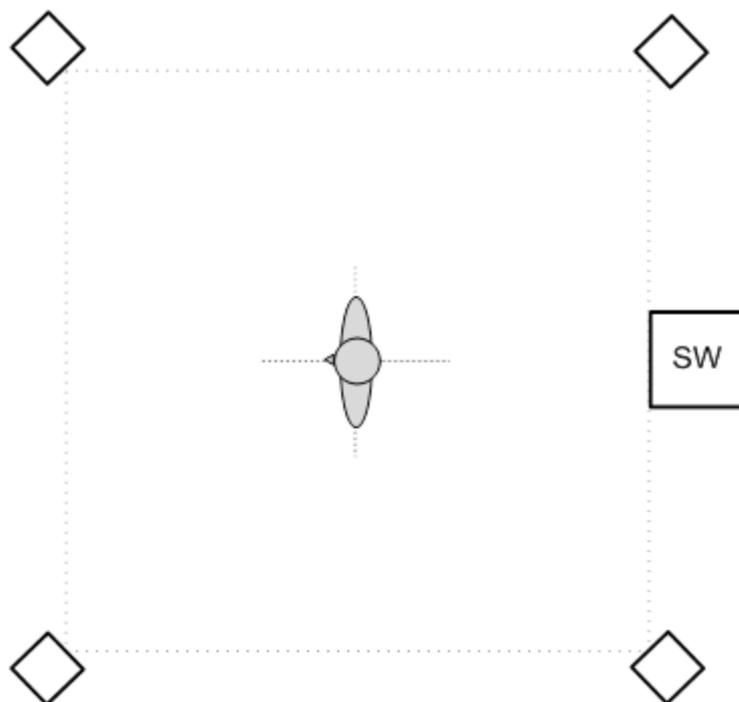


Figure 15a4: Equalization using the ES 202 396-1 method

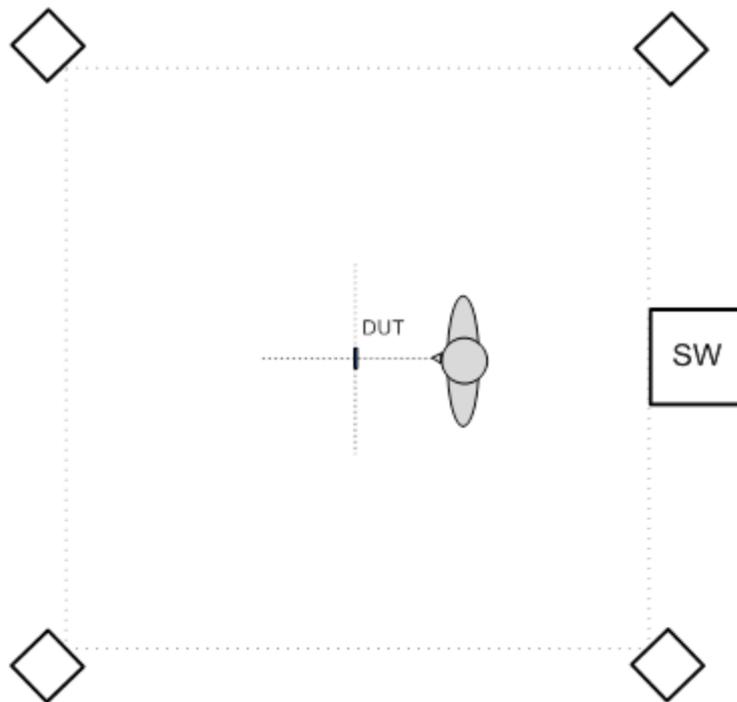


Figure 15a5: Measurements using the ES 202 396-1 method

5.2 Setup of the electrical interfaces

5.2.1 Codec approach and specification

In this approach, a codec is used to convert the digital input/output bit-stream of the system simulator to the equivalent analogue values. With this approach a system simulator simulating the radio link to the terminal under controlled and error-free conditions is required, unless otherwise specified for the respective test. The system simulator has to be equipped with a high-quality codec with characteristics as close as possible to ideal.

Definition of 0 dB_r point:

D/A converter - a Digital Test Sequence (DTS) representing the codec equivalent of an analogue sinusoidal signal with an RMS value of 3,14 dB below the maximum full-load capacity of the codec shall generate 0 dBm across a 600 ohm load;

A/D converter - a 0 dBm signal generated from a 600 ohm source shall give the digital test sequence (DTS) representing the codec equivalent of an analogue sinusoidal signal with an RMS value of 3,14 dB below the maximum full-load capacity of the codec.

Narrowband telephony testing

For testing of a GSM, 3G, LTE, NR or WLAN terminal supporting narrowband telephony, the system simulator shall use the AMR speech codec as defined in the 3GPP TS 26 series of specifications, at the source coding bit-rate of 12,2 kbit/s.

Wideband telephony testing

For testing of a GSM, 3G, LTE, NR or WLAN terminal supporting wideband telephony, the system simulator shall use the AMR-WB speech codec as defined in 3GPP TS 26 series of specifications, at the source coding bit-rate of 12,65 kbit/s.

Super-wideband telephony testing

For testing of a LTE, NR or WLAN terminal supporting super-wideband telephony, the system simulator shall use the EVS speech codec as defined in 3GPP TS 26 series of specifications at 32 kHz sampling rate, mono, at the source coding bit-rate of 24,4 kbit/s unless specified otherwise for the respective test case. The system simulator should support all operation modes of EVS speech codec, the complete EVS RTP payload format, and all SDP attributes as specified in TS 26.445 Annex A; it should be possible to control all the SDP attributes declared and negotiated by the system simulator.

Fullband telephony testing

For testing of a LTE, NR or WLAN terminal supporting fullband telephony, the system simulator shall use the EVS speech codec as defined in 3GPP TS 26 series of specifications at 48 kHz sampling rate, mono, at the source coding bit-rate of 24,4 kbit/s.

5.2.2 Direct digital processing approach

In this approach, the digital input/output bit-stream of the terminal connected through the radio link to the system simulator is operated upon directly.

Narrowband telephony testing

For testing of a GSM, 3G, LTE, NR or WLAN terminal supporting narrowband telephony, the system simulator shall use the AMR speech codec as defined in the 3GPP TS 26 series of specifications, at the source coding bit-rate of 12,2 kbit/s.

Wideband telephony testing

For testing of a GSM, 3G, LTE, NR or WLAN terminal supporting wideband telephony, the system simulator shall use the AMR-WB speech codec as defined in the 3GPP TS 26 series of specifications, at the source coding bit rate of 12,65 kbit/s.

Super-wideband telephony testing

For testing of a LTE, NR or WLAN terminal supporting super-wideband telephony, the system simulator shall use the EVS speech codec as defined in 3GPP TS 26 series of specifications in super-wideband mode, at 32 kHz sampling rate, mono, at the source coding bit-rate of 24,4 kbit/s. The system simulator should support all operation modes of EVS speech codec, the complete EVS RTP payload format, and all SDP attributes as specified in TS 26.445 Annex A; it should be possible to control all the SDP attributes declared and negotiated by the system simulator.

Fullband telephony testing

For testing of a LTE, NR or WLAN terminal supporting fullband telephony, the system simulator shall use the EVS speech codec as defined in 3GPP TS 26 series of specifications in fullband mode, at 48 kHz sampling rate, mono, at the source coding bit-rate of 24,4 kbit/s.

5.3 Accuracy of test equipment

Unless specified otherwise, the accuracy of measurements made by test equipment shall exceed the requirements defined in table 1a.

Table 1a: Test equipment measurement accuracy

Item	Accuracy
Electrical Signal Power	$\pm 0,2$ dB for levels ≥ -50 dBm
	$\pm 0,4$ dB for levels < -50 dBm
Sound pressure	$\pm 0,7$ dB
Time	$\pm 5\%$
Frequency	$\pm 0,2\%$
Clock (only for UE clock accuracy measurement)	$[\pm 5]$ PPM
NOTE: The values in [] are provisional pending some confirmation of test equipment compliance in near future	

Unless specified otherwise, the accuracy of the signals generated by the test equipment shall exceed the requirements defined in table 1b.

Table 1b: Test equipment signal generation accuracy

Quantity	Accuracy
Sound pressure level at MRP[, in 1/3 rd octave bands]	± 3 dB for 100 Hz to 200 Hz
	± 1 dB for 200 Hz to 8 kHz
	± 3 dB for 8 kHz to 20 kHz (see note 3)
Mouth simulator equalization	The flatness of the mouth simulator transfer characteristics after equalization, measured in 1/3rd octave bands with the signal used for equalization, shall be within ± 1 dB from 100 Hz to 200 Hz and shall be within ± 0.5 dB above 200 Hz (see note 3).
Electrical excitation levels	$\pm 0,4$ dB (see note 1)
Frequency generation	$\pm 2\%$ (see note 2)
NOTE 1: Across the whole frequency range. NOTE 2: When measuring sampled systems, it is advisable to avoid measuring at sub-multiples of the sampling frequency. There is a tolerance of $\pm 2\%$ on the generated frequencies, which may be used to avoid this problem, except for 4 kHz where only the -2% tolerance may be used. NOTE 3: Not all mouth simulators can be successfully equalized up to 20 kHz; in this case the upper frequency shall be reported. The validity of the equalization, especially with respect to super-wideband and fullband, shall be checked.	

The measurements' results shall be corrected for the measured deviations from the nominal level.
 The sound level measurement equipment shall conform to IEC 60651 Type 1.

5.4 Test signals

Unless stated otherwise, appropriate test signals for GSM/3G/LTE/NR/WLAN acoustic tests are generally described and defined in ITU-T Recommendation P.501 [22].

More information can be found in the test procedures described below.

For testing the narrowband telephony service provided by the UE, the test signal used shall be band limited between 100 Hz and 4 kHz with a bandpass filter providing a minimum of 24 dB/oct. filter roll-off, when feeding into the receiving direction.

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

For testing the super-wideband telephony service provided by the UE, the test signal used shall be band limited between 50 Hz and 16 kHz with a bandpass filter providing a minimum of 24 dB/oct. filter roll-off, when feeding into the receiving direction.

For testing the fullband telephony service provided by the UE, the test signal used shall be band limited between 20 Hz and 20 kHz with a bandpass filter providing a minimum of 24 dB/oct. filter roll-off, when feeding into the receiving direction.

The test signal levels are referred to the average level of the (band limited in receiving direction) test signal, averaged over the complete test sequence, unless specified otherwise. For real speech, the test signal levels are referred to the ITU-T P.56 [37] active speech level of the (band limited in receiving direction) test signal, calculated over the complete test sequence.

5.5 Void

5.5.1 Void

5.5.2 Void

6 Test conditions

6.1 Environmental conditions

6.1.1 Handset and headset terminals

For handset and headset measurements the test room shall be practically free-field down to a lowest frequency of 275 Hz; the handset or headset, including the HATS, shall be totally within this free-field volume.

Qualification of the test room may be performed using the method and limits for deviation from ideal free-field conditions described in either ISO 3745 Annex A (Table A.2), or ITU-T P. 340 §5.4 (Table 1).

Alternatively, a test room may be used which meets the following two criteria:

1. The relationship between the pressure at the mouth opening and that at 5,0 cm, 7,5 cm and 10 cm in front of the centre of the lip ring is within ± 0.5 dB of that which exists in a known acoustic free-field.
2. The relationship between the pressure at the mouth opening and that at the Ear canal Entrance Point (EEP) at both the left and right ears of the HATS does not differ by more than ± 1 dB from that which exists in a known free-field.

The ambient noise level shall be less than -30 dBPa(A); for idle channel noise measurements the ambient noise level shall be less than -64dBPa(A).

Echo measurements shall be conducted in realistic rooms with an ambient noise level \leq -64 dBPa(A).

6.1.2 Hands-free terminals

Hands-free terminals should generally be tested in their typical environment of application. Care must be taken that, e.g., noise levels are sufficiently low in order not to interfere with the measurements.

For desktop hands-free terminals the appropriate requirements shall be taken from ITU-T Recommendation P.340.

The broadband noise level shall not exceed -70 dBPa(A). The octave band noise level shall not exceed the values specified in Table 2.

Table 2: P.340 Noise level

Center frequency (Hz)	Octave band pressure level (dBPa)
63	-45
125	-60
250	-65
500	-65
1 000	-65
2 000	-65
4 000	-65
8 000	-65

Echo measurements shall be conducted in realistic rooms with an ambient noise level \leq -70 dBPa(A).

6.2 System simulator conditions

Unless otherwise specified for the respective test, the system simulator should provide an error-free radio connection to the UE under test. The speech codec and bitrate shall be as defined in clause 5. Discontinuous Transmission (DTX) silence suppression shall be disabled in uplink and downlink for the purposes of GSM acoustic testing by appropriate configuration of the system simulator. For 3G, LTE, NR and WLAN, DTX shall be disabled for the downlink path by appropriate configuration of the system simulator, unless otherwise stated for the respective test. If DTX is enabled for the downlink path in a respective test using the EVS codec, the system simulator shall use DTX/CNG operation for EVS Primary with a SID update rate of once every 8 frames as described in clause 4.4.3 of TS 26.445.

NOTE: For LTE/NR/WLAN with AMR and AMR-WB codecs and for 3G, the UE DTX for the uplink path cannot be controlled from the system simulator.

7 Narrowband telephony transmission performance test methods

7.1 Applicability

The test methods in this clause shall apply when testing a UE that is used to provide narrowband telephony, either as a stand-alone service, or as part of a multimedia service.

7.2 Overall loss/loudness ratings

7.2.1 General

The SLR and RLR values for GSM, 3G, LTE, NR or WLAN networks apply up to the POI. However, the main determining factors are the characteristics of the UE, including the analogue to digital conversion (ADC) and digital to analogue conversion (DAC). In practice, it is convenient to specify loudness ratings to the Air Interface. For the normal case, where the GSM, 3G, LTE, NR or WLAN network introduce no additional loss between the Air Interface and the POI, the loudness ratings to the PSTN boundary (POI) will be the same as the loudness ratings measured at the Air Interface.

7.2.2 Connections with handset UE

7.2.2.1 Sending loudness rating (SLR)

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be -4,7 dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence.
- b) The handset terminal is setup as described in clause 5. The sending sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17. For the calculation, the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.
- c) The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23b), over bands 4 to 17, using $m = 0,175$ and the sending weighting factors from ITU-T Recommendation P.79 [16], table 1.

7.2.2.2 Receiving loudness rating (RLR)

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The handset terminal is setup as described in clause 5. The receiving sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23c), over bands 4 to 17, using $m = 0,175$ and the receiving weighting factors from table 1 of ITU-T Recommendation P.79 [16].
- d) DRP-ERP correction is used. No leakage correction shall be applied.

7.2.2.3 Receiving loudness rating (RLR) in the presence of background noise

This test shall be performed with the volume control at the maximum setting.

- a) The downlink test signal to be used for the measurements shall be composed from the British-English material of clause 7.3 in ITU-T Recommendation P.501 [22]. The signal is a concatenation of the following parts (see also Figure 15a6):
 - 1) The complete female short conditioning sequence (which includes a leading 5 s silence which is here used for conditioning the UE under noise-only conditions). The level shall be -16 dBm0 active speech level (ITU-T P.56) measured at the digital reference point or the equivalent analogue point.
 - 2) Pause 1 sec.
 - 3) The first 5 s of the male short conditioning sequence. The level shall be -16 dBm0 active speech level (ITU-T P.56) measured at the digital reference point or the equivalent analogue point.
 - 4) The complete single-talk sequence. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete single-talk test sequence.
- b) The background noise playback shall be set up and calibrated according to handset related descriptions in ETSI TS 103 224 or ES 202 396-1. The "Pub Noise (Pub)" recording is used. This scenario is available for both playback systems and is described in sub-clause 8.2 of TS 103 224 (8-channel recording) and sub-clause 8.2 of ES 202 396-1 (binaural recording). The type of background noise reproduction used for the test shall be reported.
- c) The handset terminal is setup as described in clause 5 and the relevant noise playback specification, TS 103 224 or ES 202 396-1. Set the volume control to maximum.
- d) The start of the recording of the artificial ear signal and the start of the background noise playback shall be synchronized to ensure consistent timing of repeated recordings, and allow background noise cancellation by subtraction in the time domain as described below. The noise playback shall be looped with a 20 ms crossfade at the loop point.
- e) The background noise and downlink speech stimulus are applied and a recording of the artificial ear DRP signal is performed. The sound file is saved (file1). The playback is stopped.
- f) The recording described in e) is repeated, this time with muted downlink signal. The soundfile is saved (file2).
- g) A subtraction of the samples is performed (file1-file2). The resulting time signal contains the output of the UE to the artificial ear, where the background noise has been essentially eliminated. For further analysis, the duration of the conditioning sequence and the receive delay from POI to DRP is discarded, resulting in a time signal containing only the single-talk sequence.
- h) The RLR is calculated according to sub-clause 7.2.2.2.

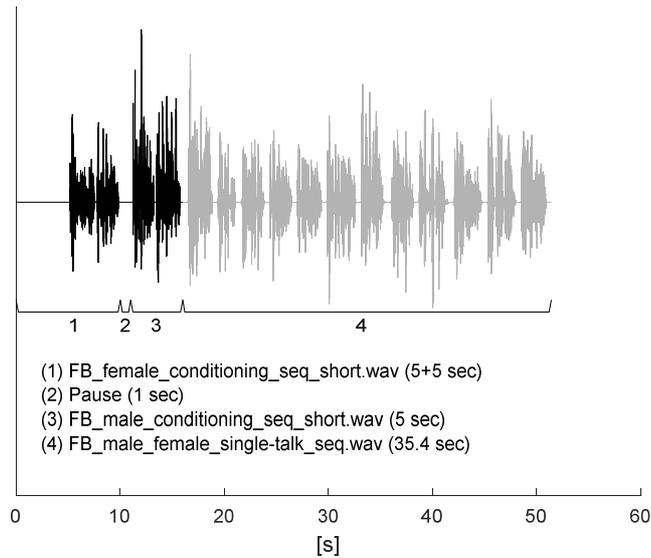


Figure 15a6 Downlink test signal

7.2.3 Connections with desktop and vehicle-mounted hands-free UE

Vehicle-mounted hands-free UE should be tested within the vehicle (for totally integrated vehicle hands-free systems) or in a vehicle simulator, as described in 3GPP TS 03.58 [11].

Free-field measurements for vehicle-mounted hands-free are for further study.

7.2.3.1 Sending loudness rating (SLR)

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be $-4,7$ dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence. The broadband signal level is then adjusted to $-28,7$ dBPa at the HFRP or the HATS HFRP (as defined in ITU-Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as references to determine the sending sensitivity S_{mJ} .

- b) The hands-free terminal is setup as described in clause 5. The sending sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17. For the calculation, the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.
- c) The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23b), over bands 4 to 17, using $m = 0,175$ and the sending weighting factors from ITU-T Recommendation P.79 [16], table 1.

7.2.3.2 Receiving Loudness Rating (RLR)

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations

or measurements. The receiving sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17.

For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

- c) The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23c), over bands 4 to 17, using $m = 0,175$ and the receiving weighting factors from table 1 of ITU-T Recommendation P.79 [16].
- d) No leakage correction shall be applied. The hands-free correction, as described in ITU-T Recommendation P.340 shall be applied. To compute the receiving loudness rating (RLR) for a hands-free terminal (see also ITU-T Recommendation P.340), when using the combination of left and right artificial ear signals from the HATS, the HFL_E has to be 8 dB instead of 14 dB. For further information see ITU-T Recommendation P.581.

7.2.4 Connections with hand-held hands-free UE

7.2.4.1 Sending loudness rating (SLR)

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be -4,7 dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence. The broadband signal level is then adjusted to -28,7 dBPa at the HFRP or the HATS HFRP (as defined in ITU-T Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as references to determine the sending sensitivity S_{mJ} .

- b) The hands-free terminal is setup as described in clause 5. The sending sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17. For the calculation, the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.
- c) The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23b), over bands 4 to 17, using $m = 0,175$ and the sending weighting factors from ITU-T Recommendation P.79 [16], table 1.

7.2.4.2 Receiving loudness rating (RLR)

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. The receiving sensitivity shall be calculated from each band of the 14 frequencies given in table 1 of ITU-T Recommendation P.79 [16], bands 4 to 17.

For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

- c) The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23c), over bands 4 to 17, using $m = 0,175$ and the receiving weighting factors from table 1 of ITU-T Recommendation P.79 [16].
- d) No leakage correction shall be applied. The hands-free correction as described in ITU-T Recommendation P.340 shall be applied. To compute the receiving loudness rating (RLR) for hands-free terminals (see also ITU-T Recommendation P.340), when using the combination of left and right artificial ear signals from the HATS, the HFL_E has to be 8 dB instead of 14 dB. For further information see ITU-T Recommendation P.581.

7.2.5 Connections with headset UE

Same as for handset.

7.3 Idle channel noise (handset and headset UE)

For idle noise measurements in sending and receiving directions, care should be taken that only the noise is windowed out by the analysis and the result is not impaired by any remaining reverberation or by noise and/or interference from various other sources. Some examples are air-conducted or vibration-conducted noise from sources inside or outside the test chamber, disturbances from lights and regulators, mains supply induced noise including grounding issues, test system and system simulator inherent noise as well as radio interference from the UE to test equipment such as ear simulators, microphone amplifiers, etc.

7.3.1 Sending

The terminal should be configured to the test equipment as described in subclause 5.1.

The environment shall comply with the conditions described in subclause 6.1.

The noise level at the output of the SS is measured with psophometric weighting. The psophometric weighting filter is described in ITU-T Recommendation O.41.

A test signal may have to be intermittently applied to prevent 'silent mode' operation of the MS. This is for further study.

The measured part of the noise shall be 170,667 ms (which equals 8192 samples in a 48 kHz sample rate test system). The spectral distribution of the noise is analyzed with an 8k FFT using windowing with ≤ 0.1 dB leakage for non bin-centered signals. This can be achieved with a window function commonly known as a "flat top window". Within the specified frequency range, the FFT bin that has the highest level is searched for; the level of this bin is the maximum level of a single frequency disturbance.

To improve repeatability, the test sequence (optional activation followed by the noise level measurement) may be contiguously repeated one or more times.

The total noise powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

The single frequency maximum powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

7.3.2 Receiving

The terminal should be configured to the test equipment as described in subclause 5.1.

The environment shall comply with the conditions described in subclause 6.1.

A test signal may have to be intermittently applied to prevent 'silent mode' operation of the MS. This is for further study.

The noise level shall be measured with A-weighting at the DRP with diffuse-field correction. The A-weighting filter is described in IEC 60651.

The measured part of the noise shall be 170,667 ms (which equals 8192 samples in a 48 kHz sample rate test system). The spectral distribution of the noise is analyzed with an 8k FFT using windowing with ≤ 0.1 dB leakage for non bin-centred signals. This can be achieved with a window function commonly known as a "flat top window". Within the specified frequency range, the FFT bin that has the highest level is searched for; the level of this bin is the maximum level of a single frequency disturbance.

To improve repeatability, considering the test sequence (optional activation followed by the noise level measurement) may be contiguously repeated one or more times.

The total noise powers obtained from such repeats shall be averaged. The total result shall be $10 \cdot \log_{10}$ of this average in dB.

The single frequency maximum powers obtained from such repeats shall be averaged. The total result shall be $10 \cdot \log_{10}$ of this average in dB.

7.4 Sensitivity/frequency characteristics

7.4.1 Handset and headset UE sending

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be -4,7 dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence.
- b) The handset terminal is setup as described in clause 5. Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 4 kHz inclusive. For the calculation, the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.
- c) The sensitivity is expressed in terms of dBV/Pa.

7.4.2 Handset and headset UE receiving

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The handset terminal is setup as described in clause 5. Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 4 kHz inclusive. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The HATS is diffuse-field equalized. The sensitivity is expressed in terms of dBPa/V. Information about correction factors is available in subclause 5.1.4.

Optionally, the measurements may be repeated with a 2 N and 13 N application force. For these test cases no normative values apply.

7.4.3 Desktop and vehicle-mounted hands-free UE sending

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be -4,7 dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence. The broadband signal level is then adjusted to -28,7 dBPa at the HFRP or the HATS HFRP (as defined in ITU-T Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as references to determine the sending sensitivity S_{mJ} .

- b) The hands-free terminal is setup as described in clause 5. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 4 kHz inclusive. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The sensitivity is expressed in terms of dBV/Pa.

7.4.4 Desktop and vehicle-mounted hands-free UE receiving

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 4 kHz inclusive. For the calculation the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The sensitivity is expressed in terms of dBPa/V.

7.4.5 Hand-held hands-free UE sending

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be -4,7 dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence. The broadband signal level then is adjusted to -28,7 dBPa at the HFRP or the HATS HFRP (as defined in ITU-T Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as reference to determine the sending sensitivity S_{mJ} .

- b) The hands-free terminal is setup as described in clause 5. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 4 kHz inclusive. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The sensitivity is expressed in terms of dBV/Pa.

7.4.6 Hand-held hands-free UE receiving

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave band frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 4 kHz inclusive. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The sensitivity is expressed in terms of dBPa/V.

7.5 Sidetone characteristics

7.5.1 Connections with handset UE

The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be -4,7 dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence.

7.5.1.1 void

7.5.1.2 Connections with handset UE – HATS method

The handset UE is setup as described in clause 5. The application force shall be 13 N on the Type 3.3 artificial ear.

Where a user operated volume control is provided, the measurements shall be carried out at the nominal setting of the volume control. In addition the measurement is repeated at the maximum volume control setting. It is expected that for other positions of the volume control setting a consistent behaviour to that of the nominal and maximum settings should be observed. Additional measurements for these positions are not required.

Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 8 kHz inclusive. The averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

The sidetone path loss (L_{meST}), as expressed in dB, shall be calculated from each 1/3rd-octave band (ITU-T Recommendation P.79 [16], table B.1, bands 4 to 17). The Sidetone Masking Rating (STMR), expressed in dB, shall be calculated from formula B-4 of ITU-T Recommendation P.79 [16], using $m = 0.225$ and the weighting factors in table B.2 (unsealed condition) of ITU-T Recommendation P.79 [16]. No leakage correction (L_E) shall be applied. DRP-ERP correction is used.

In case the STMR is below the limit, the measurement shall be repeated with the electrical sidetone path disabled and both sets of results shall be reported. In case the STMR is below the limit also with the electrical sidetone path disabled, the result shall not be regarded as a failure. Disconnecting the call is normally disabling the electrical sidetone path; otherwise the UE can be switched off to enter the wanted state.

7.5.2 Headset UE

The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be -4,7 dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence.

Where a user operated volume control is provided, the measurements shall be carried out at the nominal setting of the volume control. In addition the measurement is repeated at the maximum volume control setting. It is expected that for other positions of the volume control setting a consistent behaviour to that of the nominal and maximum settings should be observed. Additional measurements for these positions are not required.

Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 8 kHz inclusive. The averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

The sidetone path loss (L_{meST}), as expressed in dB, shall be calculated from each 1/3rd-octave band (ITU-T Recommendation P.79 [16], table B.1, bands 4 to 17). The STMR (in dB) shall be calculated from formula B-4 of ITU-T Recommendation P.79 [16], using $m = 0.225$ and the weighting factors in table B.2 (unsealed condition) of ITU-T Recommendation P.79 [16]. No leakage correction (L_E) shall be applied. DRP-ERP correction is used.

In case the STMR is below the limit, the measurement shall be repeated with the electrical sidetone path disabled and both sets of results shall be reported. In case the STMR is below the limit also with the electrical sidetone path disabled, the result shall not be regarded as a failure. Disconnecting the call is normally disabling the electrical sidetone path; otherwise the UE can be switched off to enter the wanted state.

7.5.3 Hands-free UE (all categories)

No requirement other than echo control.

7.5.4 Sidetone delay for handset or headset

The handset or headset terminal is setup as described in clause 5.

The test signal is a CS-signal complying with ITU-T Recommendation P.501 using a PN-sequence with a length, T , of 4 096 points (for a 48 kHz sample rate test system). The duration of the complete test signal is as specified in ITU-T Recommendation P.501. The level of the signal shall be -4,7 dBPa at the MRP.

The cross-correlation function $\Phi_{xy}(\tau)$ between the input signal $S_x(t)$ generated by the test system in send direction and the output signal $S_y(t)$ measured at the artificial ear is calculated in the time domain:

$$\Phi_{xy}(\tau) = \frac{1}{T} \int_{t=-\frac{T}{2}}^{\frac{T}{2}} S_x(t) \cdot S_y(t + \tau)$$

The measurement window, T , shall be identical to the test signal period, T , with the measurement window synchronized to the PN-sequence of the test signal.

The sidetone delay is calculated from the envelope $E(\tau)$ of the cross-correlation function $\Phi_{xy}(\tau)$. The first maximum of the envelope function occurs in correspondence with the direct sound produced by the artificial mouth; the second one occurs with a possible delayed sidetone signal. The difference between the two maxima corresponds to the sidetone delay. The envelope $E(\tau)$ is calculated by the Hilbert transformation $H\{xy(\tau)\}$ of the cross-correlation:

$$H\{xy(\tau)\} = \sum_{u=-\infty}^{+\infty} \frac{\Phi_{xy}(u)}{\pi(\tau - u)}$$

$$E(\tau) = \sqrt{[\Phi_{xy}(\tau)]^2 + [H\{xy(\tau)\}]^2}$$

It is assumed that the measured sidetone delay is less than $T/2$.

7.6 Stability loss

Where a user-controlled volume control is provided it is set to maximum.

Handset UE: The handset is placed on a hard plane surface with the earpiece facing the surface.

Headset UE: The requirement applies for the closest possible position between microphone and headset receiver within the intended wearing position.

NOTE: Depending on the type of headset it may be necessary to repeat the measurement in different positions.

Hands-free UE (all categories): No requirement other than echo loss.

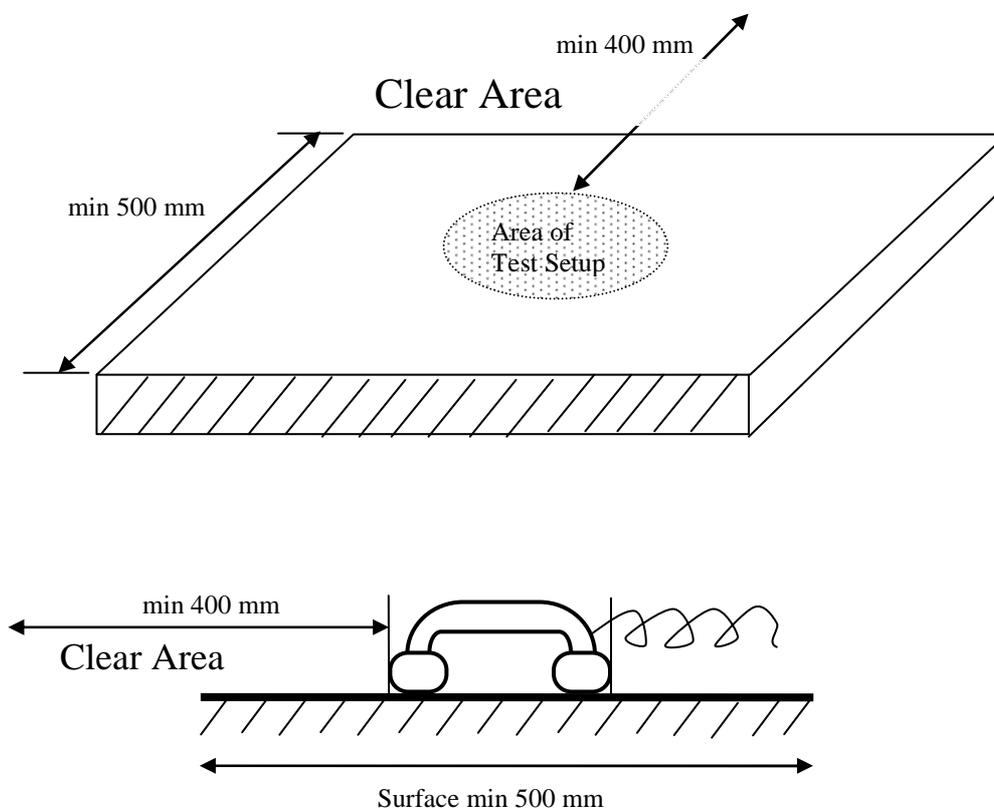
Before the actual test a training sequence consisting of the British-English single talk sequence described in ITU-T Recommendation P.501 [22] is applied. The training sequence level shall be -16 dBm0 in order to not overload the codec.

The test signal is a PN-sequence complying with ITU-T Recommendation P.501 with a length of 4 096 points (for a 48 kHz sampling rate system) and a crest factor of 6 dB instead of 11 dB. The PN-sequence is generated as described in P.501 with $W(k)$ constant within the frequency range 200-4000 Hz and zero outside this range. The duration of the test signal is 250 ms. With an input signal of -3 dBm0, the attenuation from input to output of the system simulator shall be measured under the following conditions:

- The handset or the headset, with the transmission circuit fully active, shall be positioned on a hard plane surface with at least 400 mm free space in all directions; the earpiece shall face towards the surface as shown in figure 15c;
- The headset microphone is positioned as close as possible to the receiver(s) within the intended wearing position;
- For a binaural headset, the receivers are placed symmetrically around the microphone.

Figure 15a. Void

Figure 15b. Void



NOTE: All dimensions in mm.
Figure 15c. Test configuration for stability loss measurement on handset or headset UE

The attenuation from input to output of the system simulator shall be measured in the frequency range from 200 Hz to 4 kHz. The spectral distribution of the output signal is analysed with a 4k FFT (for a 48 kHz sample rate test system), thus the measured part of the output signal is 85.333 ms. To avoid leakage effects, the frequency resolution of the FFT must be the same as the frequency spacing of the PN-sequence.

7.7 Acoustic echo control

7.7.1 General

The echo loss (EL) presented by the GSM, 3G, LTE or NR networks at the POI should be at least 46 dB during single talk. This value takes into account the fact that UE is likely to be used in a wide range of noise environments.

7.7.2 Acoustic echo control in a hands-free UE

The hands-free UE is setup in a room with acoustic properties similar to a typical "office-type" room; a vehicle-mounted hands-free UE should be tested in a vehicle or vehicle simulator, as specified by the UE manufacturer (see also 3GPP TS 03.58 [11]). The ambient noise level ≤ 70 dBPa(A). The attenuation from reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, clause B.4 (trapezoidal rule). For the calculation, the averaged measured echo level at each frequency band is referred to the averaged test signal level measured in each frequency band. The first 17,0 s of the test signal (6 sentences) are discarded from the analysis to

allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences).

The test signal level shall be -10 dBm0.

7.7.3 Acoustic echo control in handset UE

The handset is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A). The attenuation from the reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, clause B.4 (trapezoidal rule). For the calculation, the averaged measured echo level at each frequency band is referred to the averaged test signal level measured in each frequency band. The first 17,0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences).

The test signal level shall be -10 dBm0.

7.7.4 Acoustic echo control in a headset UE

The headset is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A). The attenuation from reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, clause B.4 (trapezoidal rule). For the calculation, the averaged measured echo level at each frequency band is referred to the averaged test signal level measured in each frequency band. The first 17,0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences).

The test signal level shall be -10 dBm0.

7.8 Distortion

7.8.1 Sending distortion

The handset, headset, or hands-free UE is setup as described in clause 5.

The signal used is a sine-wave signal with a frequency of 1020 Hz. The sine-wave signal level shall be calibrated to the following RMS levels at the MRP: 5, 0, -4,7, -10, -15, -20 dBPa. The test signals have to be applied in this sequence, i.e., from high levels down to low levels.

The duration of the sine-wave signal is recommended to be 360 ms. The manufacturer shall be allowed to request tone lengths up to 1 s. The measured part of the signal shall be 170.667 ms (which equals $2 * 4096$ samples in a 48 kHz sample rate test system). The times are selected to be relatively short in order to reduce the risk that the test tone is treated as a stationary signal.

It is recommended that an optional activation signal be presented immediately preceding each test signal to ensure that the UE is in a typical state during measurement. An appropriate speech or speech-like activation signal shall be chosen from ITU-T Recommendations P.501 or P.50 [10]. A recommendation for the use of an activation signal as part of the measurement is defined in figure 16. The RMS level of the active parts of this activation signal is recommended to be equal to the subsequent test tone RMS level. In practice, certain types of processing may be impacted due to the introduction of the activation signal. The manufacturer shall be allowed to specify disabling of the activation signal. It shall be reported whether an activation signal was used or not, along with the characteristics of the activation signal, as specified by the manufacturer.

The ratio of the signal to total distortion power of the signal output of the SS shall be measured with the psophometric noise weighting (see ITU-T Recommendations G.712, O.41 and 0.132). The psophometric filter shall be normalized

(0 dB gain) at 800 Hz as specified in ITU-T Recommendation O.41. The weighting function shall be applied to the total distortion component only (not to the signal component).

For measurement of the total distortion component an octave-wide band-stop filter shall be applied to the signal to suppress the sine-wave signal and associated coding artefacts. The filter shall have a lower passband ending at $0.7071 * f_s$, and an upper passband starting at $1.4142 * f_s$, where f_s is the frequency of the sine-wave signal. The passband ripple of the filter shall be ≤ 0.2 dB. The attenuation of the band-stop filter at the sine-wave frequency shall be ≥ 60 dB. Alternatively, the described characteristics can be implemented by an appropriate weighting on the spectrum obtained from an FFT. The total distortion component is defined as the measured signal within the frequency range 200 Hz to 4 kHz, after applying psophometric and stop filters (hence no correction for the lost power due to the stop filter, known as "bandwidth correction", shall be applied).

To improve repeatability, considering the variability introduced by speech coding and voice processing, the test sequence (activation signal followed by the test signal) may be contiguously repeated one or more times.. The single signal-to-total-distortion power ratios obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

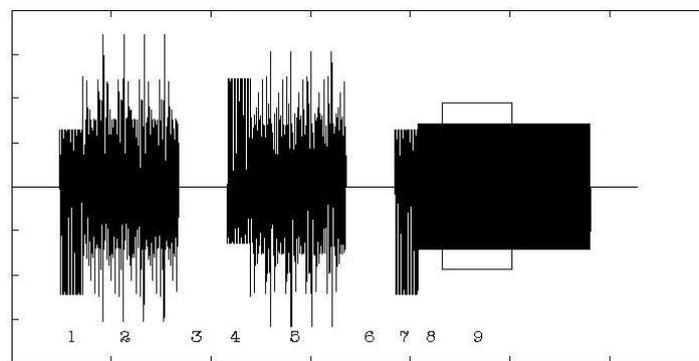


Figure 16: Recommended activation sequence and test signal.

The activation signal consists of a "Bandlimited composite source signal with speech-like power density spectrum" signal according to ITU-T Recommendation P.501 with 48,62 ms voiced part (1), 200 ms unvoiced part (2) and 101,38 ms pause (3), followed by the same signal but polarity inverted (4, 5, 6), followed by the voiced part only (7). The pure test tone is applied and after 50 ms settling time (8), the analysis is made over the following 170,667 ms (9).

NOTE 1: Void.

NOTE 2: In order to ensure that the correct part of the signal is analyzed, the total delay of the terminal and SS may have to be determined prior to the measurement.

NOTE 3: For hands-free terminals tested in environments defined in subclause 6.1.2, care should be taken that the reverberation in the test room, caused by the activation signal, does not affect the test results to an unacceptable degree, referring to subclause 5.3.

7.8.2 Receiving

The handset, headset, or hands-free UE is setup as described in clause 5.

The signal used is a sine-wave signal with frequencies specified in clause 5.8 of 3GPP TS 26.131. The signal level shall be -16 dBm0, except for the sine-wave signal with a frequency 1020 Hz that shall be applied at the signal input of the SS at the following levels: 0, -3, -10, -16, -20, -30, -40, -45 dBm0. The test signals have to be applied in this sequence, i.e., from high levels down to low levels.

The duration of the sine-wave signal is recommended to be 360 ms. The manufacturer shall be allowed to request tone lengths up to 1 s. The measured part of the signal shall be 170.667 ms (which equals $2 * 4096$ samples in a 48 kHz sample rate test system). The times are selected to be relatively short in order to reduce the risk that the test tone is treated as a stationary signal.

It is recommended that an optional activation signal be presented immediately preceding each test signal to ensure that the UE is in a typical state during measurement. An appropriate speech or speech-like activation signal shall be chosen

from ITU-T Recommendations P.501 or P.50 [10]. A recommendation for the use of an activation signal as part of the measurement is defined in figure 17. The RMS level of the active parts of this activation signal is recommended to be equal to the subsequent test tone RMS level for low and medium test levels. To avoid saturation of the SS speech encoder, it is recommended for high test levels that the activation signal level be adjusted such that its peak level equals the peak level of the test tone. In practice, certain types of processing may be impacted due to the introduction of the activation signal. The manufacturer shall be allowed to specify disabling of the activation signal. It shall be reported whether an activation signal was used or not, along with the characteristics of the activation signal, as specified by the manufacturer.

The ratio of the signal to total distortion power shall be measured at the applicable acoustic measurement point (DRP with diffuse-field correction for handset and headset modes; free field for hands-free modes) with psophometric noise weighting (see ITU-T Recommendations G.712, O.41 and O.132). The psophometric filter shall be normalized to have 0 dB gain at 800 Hz as specified in ITU-T Recommendation O.41. The weighting function shall be applied to the total distortion component only (not to the signal component).

For measurement of the total distortion component an octave-wide band-stop filter shall be applied to the signal to suppress the sine-wave signal and associated coding artefacts. The filter shall have a lower passband ending at $0,7071 * f_s$, and an upper passband starting at $1,4142 * f_s$, where f_s is the frequency of the sine-wave signal. The passband ripple of the filter shall be ≤ 0.2 dB. The attenuation of the band-stop filter at the sine-wave frequency shall be ≥ 60 dB. Alternatively, the described characteristics can be implemented by an appropriate weighting on the spectrum obtained from an FFT. The total distortion component is defined as the measured signal within the frequency range 200 Hz to 4 kHz, after applying psophometric and stop filters (hence no correction for the lost power due to the stop filter, known as "bandwidth correction", shall be applied).

To improve repeatability, considering the variability introduced by speech coding and voice processing, the test sequence (activation signal followed by the test signal) may be contiguously repeated one or more times. The single signal-to-total-distortion power ratios obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

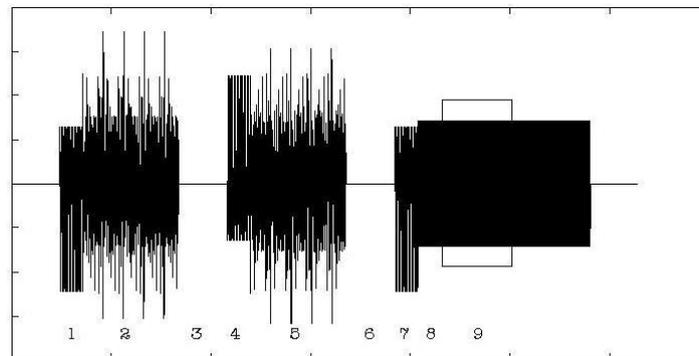


Figure 17: Recommended activation sequence and test signal.

The activation signal consists of a "Bandlimited composite source signal with speech-like power density spectrum" signal according to ITU-T Recommendation P.501 with 48,62 ms voiced part (1), 200 ms unvoiced part (2) and 101,38 ms pause (3), followed by the same signal but polarity inverted (4, 5, 6), followed by the voiced part only (7). The pure test tone is applied and after 50 ms settling time (8), the analysis is made over the following 170,667 ms (9).

NOTE 1: Void.

NOTE 2: In order to ensure that the correct part of the signal is analyzed, the total delay of the terminal and SS may have to be determined prior to the measurement.

NOTE 3: For hands-free terminals tested in environments defined in subclause 6.1.2, care should be taken that the reverberation in the test room, caused by the activation signal, does not affect the test results to an unacceptable degree, referring to subclause 5.3.

7.9 Void

7.10 Delay

7.10.0 UE Delay Measurement Methodologies

For UMTS circuit-switched operation and MTSI-based speech with LTE, NR or WLAN access in error and jitter free conditions, the sum of the UE delays in the sending and receiving directions (T_S+T_R) shall be measured according to the methods described in clauses 7.10.1 and 7.10.2. In the event that the delays of the test equipment in send and/or receive directions are not stable between calls or cannot be accurately determined, the alternative method described in clause 7.10.3 may be used to obtain (T_S+T_R) and the measured instability or inaccuracy observed when the methods described in 7.10.1 and 7.10.2 were performed shall be recorded in the test report. The test method(s) used and all results obtained shall also be recorded in the test report.

For MTSI-based speech with LTE, NR or WLAN access in conditions with simulated packet arrival time variations, the sum of the UE delays in the sending and receiving directions ($T_S+T_{R-jitter}$) and the objective speech quality in the receive direction shall be measured according to the method described in clause 7.10.4.

For MTSI-based speech with LTE, NR or WLAN access, prior to the actual measurements, the clock skew between UE and reference client shall be compensated by adjusting the clock of the reference client to match the clock of the UE (as stated in clause 5). The inaccuracy of the clock skew adjustment shall be less than 1ppm measured according to the procedure in Annex D.

It shall be ensured that the packet generation by the reference client and the packet treatment of the test equipment are free of jitter.

7.10.1 Delay in sending direction (Handset UE)

The handset terminal is setup as described in clause 5.1.1.

The UE delay in the sending direction is obtained by measuring the delay between MRP and the electrical access point of the test equipment and subtracting the delays introduced by the test equipment from the measured value.

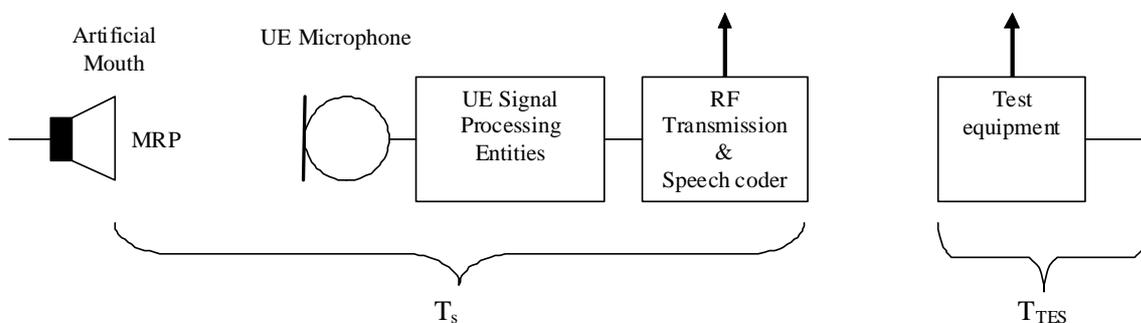


Figure 17b1: Different entities when measuring the delay in sending direction

The delay measured from MRP to the electrical access point of the test equipment is $T_S + T_{TES}$.

T_{TES} : The delay between the last bit of a speech frame at the system simulator antenna and the first electrical event at the electrical access point of the test equipment.

1. For the measurements, a Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] is used. The pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 32 k samples (with 48 kHz sampling rate). The test signal level is -4,7 dBPa at the MRP.
2. The reference signal is the original signal (test signal). The setup of the handset/headset terminal is made corresponding to clause 5.1.

3. The delay is determined by cross-correlation analysis between the measured signal at the electrical access point and the original signal. The measurement is corrected by subtracting the test equipment delay T_{TES} .
4. The delay is measured in ms and the maximum of the cross-correlation envelope is used for the determination.

For MTSI-based speech with LTE, NR or WLAN access, a variability of up to 20ms may be expected between different calls due to the synchronization between the speech frame processing in the sending UE and the bits of the speech frames at the UE antenna. This synchronization is attributed to the UE sending delay according to the definition of the UE delay reference points. Hence, the maximum value of the UE sending delay obtained from at least 5 individual calls shall be reported as the UE delay in the sending direction. All values shall be reported in the test report.

A further variability of up to 20ms may be expected between different calls due to the synchronization between the speech frames at the UE antenna and the speech frame processing in the receiving reference client of the test system. In an end-to-end call this synchronization of the frames will only take place at the receiver, and this variability of the measurement shall be deduced from the UE sending delay. Hence, if the reference client of the test equipment does not adjust for the effect of the speech frame synchronization (as specified by the manufacturer of the reference client), this maximum uncertainty shall be subtracted from measured maximum value reported as the UE sending delay in order to compensate for the uncertainty of the test equipment. This correction value (i.e. maximum uncertainty) shall be reported in the test report.

7.10.1a Delay in sending direction (headset UE)

The UE delay in the sending direction is obtained by measuring the delay between MRP and the electrical access point of the test equipment and subtracting the delays introduced by the test equipment from the measured value.

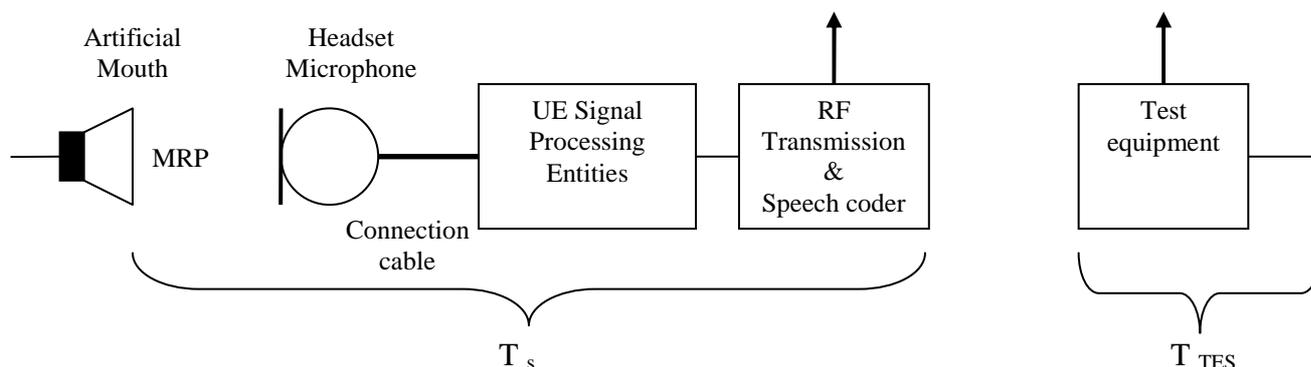


Figure 17b2: Different entities when measuring the delay in sending direction with a headset connected via cable

NOTE: The test setup only applies to headsets connected by wire. Wireless headsets (e.g. connected by Bluetooth) are currently out of scope.

The test method is the same as for handset UE (clause 7.10.1).

7.10.2 Delay in receiving direction (handset UE)

The handset terminal is setup as described in clause 5.

The UE delay in the receiving direction is obtained by measuring the delay between the electrical access point of the test equipment and the DRP and subtracting the delays introduced by the test equipment from the measured value.

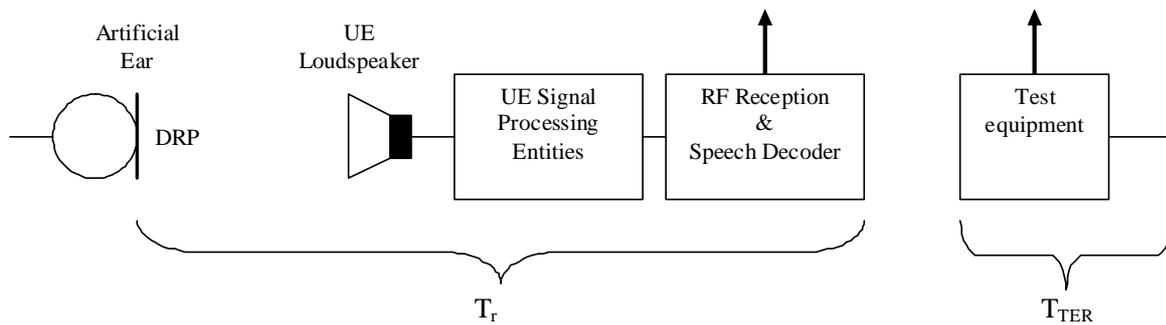


Figure 17b3: Different entities when measuring the delay in receiving direction

The delay measured from the electrical access point of the test equipment to DRP is $T_R + T_{TER}$.

T_{TER} : The delay between the first electrical event at the electrical access point of the test equipment and the first bit of the corresponding speech frame at the system simulator antenna.

Before the actual test for MTSI-based speech with LTE, NR or WLAN access a conditioning sequence consisting of the British-English single talk sequence described in ITU-T Recommendation P.501 [22] is applied for convergence of the jitter buffer management of the UE. The conditioning sequence level shall be -16 dBm0 in order to not overload the codec.

1. For the measurements a Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] is used. The pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 32 k samples (with 48 kHz sampling rate). The test signal level is -16 dBm0 measured at the digital reference point or the equivalent analogue point.
2. The reference signal is the original signal (test signal). The setup of the handset/headset terminal is in correspondence to clause 5.1.
3. The delay is determined by cross-correlation analysis between the measured signal at the electrical access point and the original signal. The measurement is corrected by subtracting the test equipment delay T_{TER} .
4. The delay is measured in ms and the maximum of the cross-correlation envelope is used for the determination.

For MTSI-based speech with LTE, NR or WLAN access, a variability of up to 20ms may be expected between different calls due to the synchronization between the bits of the speech frames at the UE antenna and the speech frame processing in the receiving UE. This synchronization is attributed to the UE receiving delay according to the definition of the UE delay reference points. Hence, the maximum value of the UE receiving delay obtained from at least 5 individual calls shall be reported as the UE delay in the receiving direction. All values shall be reported in the test report.

7.10.2a Delay in receiving direction (headset UE)

The UE delay in the receiving direction is obtained by measuring the delay between the electrical access point of the test equipment and the DRP and subtracting the delays introduced by the test equipment from the measured value.

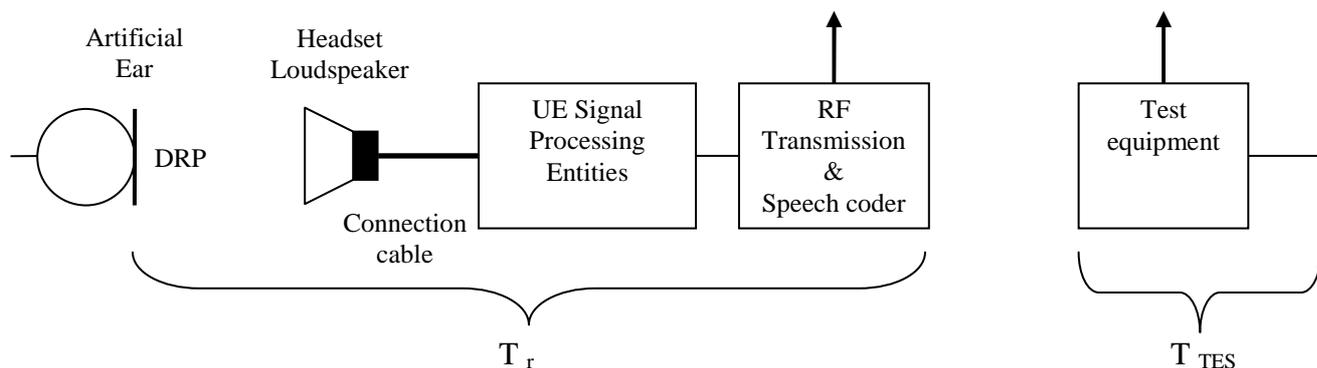


Figure 17b4: Different entities when measuring the delay in receiving direction with a headset connected via cable

NOTE: The test setup only applies to headsets connected by wire. Wireless headsets (e.g. connected by Bluetooth) are currently out of scope.

The test method is the same as for handset UE (clause 7.10.2).

7.10.3 Delay in sending + receiving direction using "echo" method (handset UE)

The UE delay is obtained by measuring the delay between the MRP and the DRP and subtracting the delays introduced by the test equipment from the measured value.

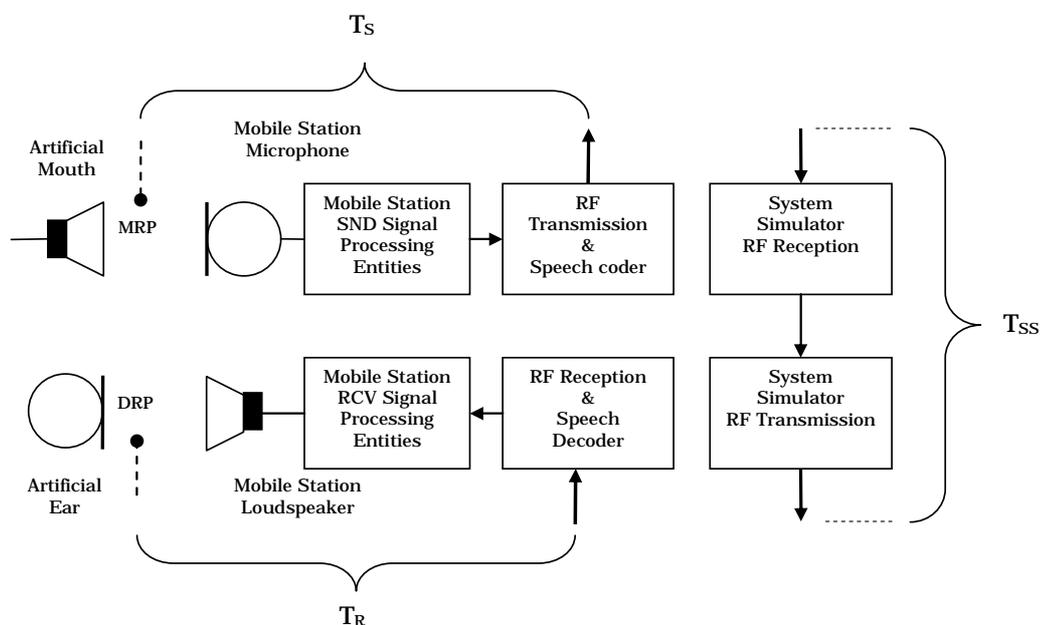


Figure 17b4bis: Different entities when measuring the delay in sending + receiving direction

The delay measured from MRP to DRP is $(T_s + T_R + T_{SS})$.

T_{SS} : The delay between the last bit of a speech frame at the system simulator antenna and the first bit of the looped back speech frame at the system simulator antenna.

Before the actual test for MTSI-based speech with LTE, NR or WLAN access a conditioning sequence consisting of the British-English single talk sequence described in ITU-T Recommendation P.501 [22] is applied for convergence of the jitter buffer management of the UE. The conditioning sequence level shall be -16 dBm0 in order to not overload the codec.

1. For the measurements a Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] is used. It is recommended to use a pn sequence of 32 k samples (with 48 kHz sampling rate). The test signal level is -4.7 dBPa at the MRP.
2. The system simulator is configured for "loopback" or "echo" operation with the additional loopback delay as specified below when applicable. In "loopback" or "echo" operation, the packets in the sending direction are routed to the receiving direction by the system simulator.
3. The reference signal is the original signal (test signal). The setup of the mobile station is in correspondence to clause 5.1.
4. The mouth-to-ear delay is determined by cross-correlation analysis between the measured signal at DRP and the original signal. The analysis window for the cross-correlation shall start at an instant $T > 50\text{ms}$ in order to discard the cross-correlation peaks corresponding to the direct acoustic path from mouth to ear and possible delayed sidetone signal. The measurement is corrected by subtracting the system simulator delay T_{SS} to obtain the $T_S + T_R$ delay.
5. The delay is measured in ms and the maximum of the cross-correlation envelope is used for the determination.

For MTSI-based speech with LTE, NR or WLAN access, a variability of the UE delay with up to 20ms in the respective sending and receiving direction may be expected due to the synchronization of the speech frame processing in the UE to the bits of the speech frame on the UE antenna. This synchronization is attributed to the UE delay according to the definition of the UE delay reference points. Hence, the UE delay shall be reported as the maximum value from at least 5 separate calls each with a different loopback delay T_{SS} in at least 5 steps of 4ms in the full range from 0 to 16ms. All values shall be reported in the test report.

7.10.3a Delay in sending + receiving direction using "echo" method (headset UE)

The UE delay is obtained by measuring the delay between the MRP and the DRP and subtracting the delays introduced by the test equipment, T_{SS} , from the measured value.

The test method is the same as for handset UE (clause 7.10.3).

7.10.4 Delay and speech quality in conditions with packet arrival time variations and packet loss

7.10.4.1 Delay in sending direction

The UE delay in the sending direction, T_S , shall be measured in jitter and error free conditions according to clause 7.10.0.

7.10.4.2 Delay in receiving direction

For this test it shall be ensured that the call is originated from the mobile terminal (MO).

NOTE 1: Differences have been observed between mobile originated call and mobile terminated call. For better consistency MO calls are used.

The test signal consists of 3 repeats of the Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] followed by a speech signal of 160s. During the first two CSS signals the terminal can adapt its jitter buffer. The

third CSS is used for measuring the delay in constant-delay condition, and the speech signal is used for delay and quality measurement in the packet impairment condition.

Constant delay T_c corresponding to the minimum delay of the profile (i.e. the compensation value for the profile) shall be added at the beginning of the different delay/loss profiles, to avoid unnecessary delay jumps between the two measurement phases and realistic conditions for the second measurement test phase.

In receiving direction, the delay between the electrical access point of the test equipment and the DRP, $T_{\text{TEAP-DRP}}(t) = T_{\text{R-jitter}}(t) + T_{\text{TER}}$, is measured in two successive phases:

- 1) First the delay in constant-delay condition $T_{\text{TEAP-DRP-constant}}$ is measured as described in steps 1 to 4, clause 7.10.2, using the third CSS signal. The constant delay T_c is subtracted from $T_{\text{TEAP-DRP-constant}}$ to obtain $T_{\text{R-constant}}$.
- 2) Then the delay with packet impairment $T_{\text{R-jitter}}(t)$ is measured continuously for a speech signal during the inclusion of packet delay and loss profiles in the receiving direction RTP voice stream.

Packet impairments shall be applied between the reference client and system simulator eNodeB. Separate calls shall be established for each packet impairment condition.

The start of the delay profiles must be synchronized with the start of the downlink speech material reproduction (compensated by the delay between reproduction and the point of impairment insertion, i.e. the delay of the reference client) in order to ensure a repeatable application of impairments to the test speech signal. Tests shall be performed with DTX enabled in the reference client.

NOTE 2: RTP packet impairments representing packet delay variations and loss in LTE transmission scenarios are specified in Annex E. These LTE jitter/loss profiles are reused also for tests with WLAN and NR access. Care must be taken that the system simulator uses a dedicated bearer with no buffering/scheduling of packets for transmission.

For the CSS signal repeated 3 times, the pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 32 k samples (with 48 kHz sampling rate). The test signal level is -16 dBm0 measured at the digital reference point or the equivalent analogue point.

For the speech signal, 8 English test sentences according to ITU-T P.501 Annex C.2.3, normalized to an active speech level of -16dBm0, are used (2 male, 2 female speakers). The sequences are concatenated in such a way that all sentences are centered within a 4.0s time window, which results in an overall duration of 32.0s. The sequences are repeated 5 times, resulting in a test file 160.0s long. The first 2 sentences are used for convergence of the UE jitter buffer manager and are discarded from the analysis. Equivalent implementations of the concatenation by repeating the test sentences in sequence may be used.

For the delay calculation with the speech signal, a cross-correlation with a rectangular window length of 4s, centered at each sentence of the stimulus file, is used. The process is repeated for each sample. For each cross correlation, the maximum of the envelope is obtained producing one delay value per sentence.

The UE delay in the receive direction, $T_{\text{R-jitter}}(t)$, is obtained by subtracting the delay introduced by the test equipment and the simulated transport network packet delay introduced by the delay and loss profile (as specified for the respective profile in Annex E) from the first electrical event at the electrical access point of the test equipment to the first bit of the corresponding speech frame at the system simulator antenna, T_{TER} , from the measured $T_{\text{TEAP-DRP}}(t)$.

The difference D_T between maximum receiving delay obtained with at least 5 individual calls (see clause 7.10.2) and the delay $T_{\text{R-constant}}$ measured for the CSS signal in constant delay condition is calculated. The quantity "Call-to-Call Variability Adjustment" (CCVA) = $\max(0, D_T)$ shall be added to the obtained delay for the speech signal $T_{\text{R-jitter}}(t)$.

For stationary packet delay variation test conditions (test condition 1 and 2), the first 2 sentences are used for convergence of the jitter buffer management and are discarded from the analysis. The CCVA-adjusted UE delay ($T_{\text{R-CCVA}}(t) = T_{\text{R-jitter}}(t) + \text{CCVA}$) in the receiving direction shall be reported as the maximum value excluding the two largest values of the remaining sequence of the 38 sentence delay values, i.e. the 95-percentile value of $T_{\text{R-CCVA}}(t)$. The $T_{\text{R-CCVA}}$ values for all 40 sentences shall be reported in the test report.

NOTE 3: The synchronization of the speech frame processing in the UE to the bits of the speech frames at the UE antenna may lead to a variability of up to 20 ms of the measured UE receive delay between different calls. This synchronization is attributed to the UE receiving delay according to the definition of the UE delay reference points. The effect of this possible call-to-call variation is taken into account with the $\text{CCVA} = \max(0, D_T)$ value.

7.10.4.3 Speech quality loss in conditions with packet arrival time variations and packet loss

For the evaluation of speech quality loss in conditions with packet arrival time variations and packet loss, the test signal described in clause 7.10.4.2 shall be used. The first 2 sentences are used for convergence of the UE jitter buffer manager and are discarded from the analysis. Two 48 kHz recordings are used to produce the speech quality loss metric:

- A recording obtained in jitter and error free conditions with the test signal described in clause 7.10.4.2 (reference condition)
- A recording obtained during the application of packet arrival time variations and packet loss as described in clause 7.10.4.2 (test condition)

The speech quality of the signal is estimated using the measurement algorithm described in ITU-T Recommendation P.863 [44]. Level pre-alignment to -26 dBov of recordings shall be used – see P.863.1 clause 10.2 [45].

NOTE: The setup for acoustical measurement described in P.863 [44] is used. P.863 needs the signal at DRP with diffuse-field equalization.

A score shall be computed for each 8s speech sentence pair and averaged to produce a mean MOS-LQO value for the reference and test conditions.

$$\text{MOS-LQO}_{\text{REF}} = \left(\sum_{i=1}^{19} \text{MOS} - \text{LQO}_{\text{reference}}(i) \right) / 19$$

$$\text{MOS-LQO}_{\text{TEST}} = \left(\sum_{i=1}^{19} \text{MOS} - \text{LQO}_{\text{testcondition}}(i) \right) / 19$$

NOTE: This evaluation of the speech quality requirement is only applicable to test conditions with a stationary statistic of the packet delay variation. Evaluation of the speech quality for a test condition with non-stationary packet delay variations is for further study.

The synchronization between stimuli and degraded condition shall be done by the test system before applying the P.863 algorithm on each sentence pair.

7.10.5 UE send clock accuracy

The UE clock accuracy in send direction shall be measured according to Annex D.

NOTE1: For this specific measurement, care should be taken about the clock accuracy of the test equipment. See Table 1a.

NOTE2: As required in clause 5, prior to the actual measurements for MTSI-based speech with LTE, NR or WLAN access, the clocks of the reference client and the UE have to be synchronized. This measurement of UE send clock accuracy does not need to be repeated and can be obtained from this setup procedure.

7.10.6 UE receiving with clock skew

For further study.

7.11 Echo control characteristics

7.11.1 Test set-up and test signals

The device is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A).

The test shall be performed with the British-English "long" double-talk and conditioning speech sequences from ITU-T Recommendation P.501 [22], with the signals in the receiving direction band limited according to clause 5.4.

A description of the test stimuli is presented in Table 2a and Table 2b. The test sequence is composed of an initial conditioning sequence of 23,5 s and a double talk sequence of 35 s. For the analysis, the double talk sequence is divided into two segments, a first double-talk sequence with single short near-end words (0 – 20 s), and a second double-talk sequence with continuous double talk (20 – 35 s).

The sending speech during double-talk and the "near-end speech only" are recorded individually, with the "near-end speech only" sequence recorded with silence in the receiving direction. The time-alignment of the two recorded sequences is performed off-line during the analysis.

Table 2a: Test stimuli for recording of Echo Canceller operation

	Conditioning	Single words (segment 1) and full sentence (segment 2) double talk
Far-end signal	FB_female_conditioning_seq_long.wav	FB_male_female_single-talk_seq.wav
Artificial mouth signal	FB_male_conditioning_seq_long.wav	FB_male_female_double-talk_seq.wav

Table 2b: Test stimuli for reference "near-end speech only" recording.

	Conditioning	Single words (segment 1) and full sentence (segment 2) double talk
Far-end signal	FB_female_conditioning_seq_long.wav	silence
Artificial mouth signal	FB_male_conditioning_seq_long.wav	FB_male_female_double-talk_seq.wav

The level of the signal of the artificial mouth shall be -4,7 dBPa measured at the MRP. In order to obtain a reproducible time alignment as seen by the UE, the artificial mouth signal shall be delayed by the amount of the receiving direction delay. For the purpose of this alignment, the receiving direction delay for handset and headset modes is defined from the system simulator input to the artificial ear. For hands-free modes, the downlink delay is defined from the system simulator input to the acoustic output from the UE loudspeaker.

The level of the downlink signal shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point.

7.11.2 Test method

The test method measures the duration of any level difference between the sending signal of a double-talk sequence (where the echo canceller has been exposed to simultaneous echo and near-end speech) and the sending signal of the same near-end speech only. The level difference is classified into eight categories according to Figure 17b5 and Table 2c, representing various degrees of "Full duplex operation", "Near-end clipping", and "Residual echo".

NOTE 1: The limits for specifying the categories in Figure 17b5 and Table 2c are provisional pending further analysis and validation.

NOTE 2: The categories in Figure 17b5 and Table 2c are labelled in a functional order and the subjective impression of the respective categories is for further study.

NOTE 3: To reduce potential issues associated with low-frequency test room noise, a [4th]-order high-pass filter with a cut-off frequency of [100] Hz can be applied before the level computation.

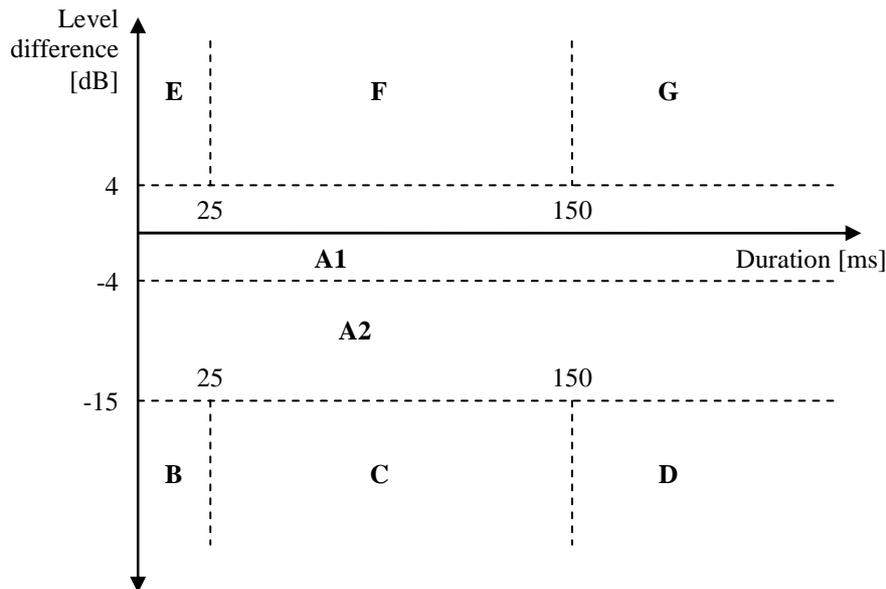


Figure 17b5: Classification of echo canceller performance

Table 2c: Categories for echo canceller performance classification

Category	Level difference (ΔL)	Duration (D)	Description
A1	$-4 \text{ dB} \leq \Delta L < 4 \text{ dB}$		Full-duplex and full transparency
A2	$-15 \text{ dB} \leq \Delta L < -4 \text{ dB}$		Full-duplex with level loss in Tx
B	$\Delta L < -15 \text{ dB}$	$D < 25 \text{ ms}$	Very short clipping
C	$\Delta L < -15 \text{ dB}$	$25 \text{ ms} \leq D < 150 \text{ ms}$	Short clipping resulting in loss of syllables
D	$\Delta L < -15 \text{ dB}$	$D \geq 150 \text{ ms}$	Clipping resulting in loss of words
E	$\Delta L \geq 4 \text{ dB}$	$D < 25 \text{ ms}$	Very short residual echo
F	$\Delta L \geq 4 \text{ dB}$	$25 \text{ ms} \leq D < 150 \text{ ms}$	Echo bursts
G	$\Delta L \geq 4 \text{ dB}$	$D \geq 150 \text{ ms}$	Continuous echo

A pseudo-code reference of the test method including test scripts and test-vectors is presented in clause C.3 and outlined in the following sub clauses.

7.11.2.1 Signal alignment

For the analysis of the signal level difference, the send signal during double-talk and the near-end only signal are aligned using a correlation analysis as described in clause C.3.2.

7.11.2.2 Signal level computation and frame classification

The analysis is based on the digital level measured with a meter according to IEC 61672 [38] with a time constant of 12,5 ms, sampled at 5 ms intervals corresponding to the evaluated frames.

The "double-talk" frames are defined as the frames where both the far-end (receiving direction) signal includes active speech (extended with a hang-over period of 200 ms) and the near-end signal is composed of active speech. Active speech is defined to be detected using a speech level meter according to ITU-T P.56, and frames within -15.9 dB from the active speech level are classified as active speech frames.

The "far-end single-talk adjacent to double-talk" frames are similarly defined using a speech level meter according to ITU-T P.56 as the frames with active far-end speech (extended with a hang-over period of 200 ms) and no active near-end speech (extended with a hang-over period of 200 ms).

A reference implementation of the signal level computation and frame classification is presented in clause C.3.3.

7.11.2.3 Classification into categories

The analysis and classification into the categories according to Figure 17b5 and Table 2c is performed according to the reference implementation described in clause C.3.4 and C.3.4.

The frames are first categorized according to the level categories defined in Table 2c. To determine the durations, the amount of adjacent frames falling into the same level category is determined.

The classification is then performed individually for the following situations:

- frames classified as "double-talk" from segment 1 of the double-talk sequence (see clause 7.11.1)
- frames classified as "far-end single-talk adjacent to double-talk" from segment 1 of the double-talk sequence
- frames classified as "double-talk" from segment 2 of the double-talk sequence
- frames classified as "far-end single-talk adjacent to double-talk" from segment 2 of the double-talk sequence

To determine the percentage values for each category (A1, A2, B, C, D, E, F, and G) within each situation, the number of frames falling into the respective category is divided by the total number of frames within the situation in question.

To determine the averaged level difference of the frames for each category (A1, A2, B, C, D, E, F, and G) within each situation, the sum of the level difference (in dB) of the frames falling into the respective category is divided by the total number of frames within the situation in question.

7.12 Quality (speech quality, noise intrusiveness) in the presence of ambient noise

7.12.1 Handset

The speech quality in sending for narrowband systems is tested based on ETSI TS 103 106 [34]. This test method leads to three MOS-LQOn quality numbers:

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

S-MOS-LQOn: Transmission quality of the speech

G-MOS-LQOn: Overall transmission quality

The test arrangement is given in clause 5.1.5. The measurement is conducted for 8 noise conditions as described in Table 2d. The measurements should be made in the same unique and dedicated call. The noise types shall be presented according to the order specified in Table 2d.

Table 2d: Noise conditions used for ambient noise simulation in handset mode as specified in ES 202 396-1 [35]

Description	File name	Duration	Level	Type
Recording in pub	Pub_Noise_binaural_V2	30 s	L: 75,0 dB(A) R: 73,0 dB(A)	Binaural
Recording at pavement	Outside_Traffic_Road_binaural	30 s	L: 74,9 dB(A) R: 73,9 dB(A)	Binaural
Recording at pavement	Outside_Traffic_Crossroads_binaural	20 s	L: 69,1 dB(A) R: 69,6 dB(A)	Binaural
Recording at departure platform	Train_Station_binaural	30 s	L: 68,2 dB(A) R: 69,8 dB(A)	Binaural
Recording at the drivers position	Fullsize_Car1_130Kmh_binaural	30 s	L: 69,1 dB(A) R: 68,1 dB(A)	Binaural
Recording at sales counter	Cafeteria_Noise_binaural	30 s	L: 68,4 dB(A) R: 67,3 dB(A)	Binaural
Recording in a cafeteria	Mensa_binaural	22 s	L: 63,4 dB(A) R: 61,9 dB(A)	Binaural
Recording in business office	Work_Noise_Office_Callcenter_binaural	30 s	L: 56,6 dB(A) R: 57,8 dB(A)	Binaural

- 1) Before starting the measurements a proper conditioning sequence shall be used. The conditioning sequence shall be comprised of the four additional sentences 1- 4 described in ETSI TS 103 106 [34], applied to the beginning of the 16-sentence test sequence.

NOTE: The sequence of speech samples concatenated for the test signal, consisting of alternating talkers in the sending direction, reduces the overall test time but may represent an unrealistic behaviour for certain voice enhancement technologies. Alternative concatenations are for further study.

- 2) The send speech signal consists of the 16 sentences of speech as described in ETSI TS 103 106 [34]. The test signal level is -1,7 dBPa at the MRP, measured as the active speech level according to ITU-T P.56 [37]. Three signals are required for the tests:
- The clean speech signal is used as the undisturbed reference (see ETSI TS 103 106 [34], ETSI EG 202 396-3 [36]).
 - The speech plus undisturbed background noise signal is recorded at the terminal's microphone position using an omnidirectional measurement microphone with a linear frequency response between 50 Hz and 12 kHz.
 - The send signal is recorded at the POI.
- 3) N-MOS-LQOn, S-MOS-LQOn and G-MOS-LQOn are calculated as described in ETSI TS 103 106 [34] on a per sentence basis and averaged over all 16 sentences. The results shall be reported as average and standard deviation.
- 4) The measurement is repeated for each ambient noise condition described in Table 2d.
- 5) The average of the results derived from all ambient noise types is calculated.

7.12.2 Hand-held hands-free

The speech quality in sending for narrowband systems is tested based on ETSI TS 103 106 [34]. This test method leads to three MOS-LQOn quality numbers:

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

S-MOS-LQOn: Transmission quality of the speech

G-MOS-LQOn: Overall transmission quality

The test arrangement is given in clause 5.1.5.

When using the simulation method described in TS 103 224 [43], the measurement is conducted for 5 noise conditions as described in Table 2d2. When using the ES 202 396-1 method, the equivalent binaurally recorded noises described in Table 2d2, and available in the source file directory of TS 103 224 [43], are used.

Table 2d2: Noise conditions used for ambient noise simulation in hand-held hands-free mode as specified in TS 103 224 [43], A-weighted

Name	Description	Length	Hands-free Levels	Binaural L	Binaural R
Full-size car 130 km/h (FullSizeCar_130)	HATS and microphone array at co-drivers position	30 s	1: 69,5 dB 2: 68,6 dB 3: 68,6 dB 4: 68,7 dB 5: 68,8 dB 6: 68,8 dB 7: 69,2 dB 8: 69,7 dB	68.7 dB	70.7 dB
Crossroadnoise (Crossroadnoise)	HATS and microphone array standing outside near a crossroad	30 s	1: 69,9 dB 2: 69,6 dB 3: 69,6 dB 4: 69,9 dB 5: 69,6 dB 6: 69,5 dB 7: 69,6 dB 8: 69,7 dB	70.8 dB	71.6 dB
Cafeteria (Cafeteria)	HATS and microphone array inside a cafeteria	30 s	1: 69,0 dB 2: 69,7 dB 3: 69,6 dB 4: 69,8 dB 5: 69,5 dB 6: 69,5 dB 7: 69,7 dB 8: 70,0 dB	69.8 dB	70.3 dB
Sales Counter (SalesCounter)	HATS and microphone array in a supermarket	30 s	1: 65,5 dB 2: 65,3 dB 3: 65,2 dB 4: 65,5 dB 5: 65,6 dB 6: 65,3 dB 7: 65,2 dB 8: 65,3 dB	66.7 dB	66.6 dB
Callcenter 2 (Callcenter)	HATS and microphone array in business office	30 s	1: 59,3 dB 2: 59,3 dB 3: 59,5 dB 4: 59,6 dB 5: 59,4 dB 6: 59,3 dB 7: 59,3 dB 8: 59,5 dB	60,2 dB	60,0 dB

- 1) Before starting the measurements a proper conditioning sequence shall be used. The conditioning sequence shall be comprised of the four additional sentences 1- 4 described in ETSI TS 103 106 [34], applied to the beginning of the 16-sentence test sequence. The conditioning signal level is +1.3 dBPa at the MRP, measured as active speech level according to ITU-T P.56 [37].

NOTE: The sequence of speech samples concatenated for the test signal, consisting of alternating talkers in the sending direction, reduces the overall test time but may represent an unrealistic behaviour for certain voice enhancement technologies. Alternative concatenations are for further study.

- 2) The send speech signal consists of the 16 sentences of speech as described in ETSI TS 103 106 [34]. The test signal level is +1.3dBPa at the MRP, measured as active speech level according to ITU-T P.56 [37]. Three signals are required for the tests:
 - The clean speech signal is used as the undisturbed reference (see ETSI TS 103 106 [34], ETSI EG 202 396-3 [36]).
 - The speech plus undisturbed background noise signal is recorded at the terminal's microphone position using an omnidirectional measurement microphone with a linear frequency response between 50 Hz and 12 kHz.
 - The send signal is recorded at the POI.
- 3) N-MOS-LQOn, S-MOS-LQOn and G-MOS-LQOn are calculated as described in ETSI TS 103 106 [34] on a per sentence basis and averaged over all 16 sentences. The results shall be reported as average and standard deviation.

- 4) The measurement is repeated for each ambient noise condition described in Table 2d2.
- 5) The average of the results derived from all ambient noise types is calculated.

7.13 Jitter buffer management behaviour

7.13.0 General

For speech-only with LTE, NR or WLAN access, a jitter buffer is used in receiving to handle the variation in packet receiver timing. To minimize the additional latency introduced by the jitter buffer, adaptation is used to minimize delay while preventing packet losses due to packet delivery timing variations. See clause 8 of TS 26.114 [39] for the definition of jitter buffer and minimum performance requirements on JBM.

The test method is used to characterize different possible strategies and trade-offs in the design of JBM implementations used in MTSI terminals.

7.13.1 Delay histogram

For this test it shall be ensured that the call is originated from the mobile terminal (MO).

NOTE 1: Differences have been observed between mobile originated call and mobile terminated call. For better consistency MO calls are used.

The test signal consists of 3 repeats of the Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] followed by a speech signal of 160s. During the first two CSS signals the terminal can adapt its jitter buffer. The third CSS is used for measuring the delay in constant-delay condition, and the speech signal is used for delay and quality measurement in the packet impairment condition.

Constant delay T_c corresponding to the minimum delay of the profile (i.e. the compensation value for the profile) shall be added at the beginning of the different delay/loss profiles, to avoid unnecessary delay jumps between the two measurement phases and realistic conditions for the second measurement test phase.

In receiving direction, the delay between the electrical access point of the test equipment and the DRP, $T_{\text{TEAP-DRP}}(t) = T_{\text{R-jitter}}(t) + T_{\text{TER}}$, is measured in two successive phases:

- 1) First the delay in constant-delay condition $T_{\text{TEAP-DRP-constant}}$ is measured as described in steps 1 to 4, clause 7.10.2, using the third CSS signal. The constant delay T_c is subtracted from $T_{\text{TEAP-DRP}}$ to obtain $T_{\text{R-constant}}$.
- 2) Then the delay with packet impairment $T_{\text{R-jitter}}(t)$ is measured continuously for a speech signal during the inclusion of packet delay and loss profiles in the receiving direction RTP voice stream.

Packet impairments shall be applied between the reference client and system simulator eNodeB. Separate calls shall be established for each packet impairment condition.

The start of the delay profiles must be synchronized with the start of the downlink speech material reproduction (compensated by the delay between reproduction and the point of impairment insertion, i.e. the delay of the reference client) in order to ensure a repeatable application of impairments to the test speech signal. Tests shall be performed with DTX enabled in the reference client.

NOTE 2: RTP packet impairments representing packet delay variations and loss are specified in Annex F. Care must be taken that the system simulator uses a dedicated bearer with no buffering/scheduling of packets for transmission.

For the CSS signal repeated 3 times, the pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 32 k samples (with 48 kHz sampling rate). The test signal level is -16 dBm0 measured at the digital reference point or the equivalent analogue point.

For the speech signal, 8 English test sentences according to ITU-T P.501 Annex C.2.3, normalized to an active speech level of -16dBm0, are used (2 male, 2 female speakers). The sequences are concatenated in such a way that all sentences are centered within a 4.0s time window, which results in an overall duration of 32.0s. The sequences are repeated 5 times, resulting in a test file 160.0s long. The first 2 sentences are used for convergence of the UE jitter

buffer manager and are discarded from the analysis. Equivalent implementations of the concatenation by repeating the test sentences in sequence may be used.

For the delay calculation with the speech signal, a cross-correlation with a rectangular window length of 4s, centered at each sentence of the stimulus file, is used. The process is repeated for each sample. For each cross correlation, the maximum of the envelope is obtained producing one delay value per sentence.

The UE delay in the receive direction, $T_{R-jitter}(t)$, is obtained by subtracting the delay introduced by the test equipment and the simulated transport network packet delay introduced by the delay and loss profile (as specified for the respective profile in Annex F) from the first electrical event at the electrical access point of the test equipment to the first bit of the corresponding speech frame at the system simulator antenna, T_{TER} , from the measured $T_{TEAP-DRP}(t)$.

The difference D_T between maximum receiving delay obtained with at least 5 individual calls (see clause 7.10.2) and the delay $T_{R-constant}$ measured for the CSS signal in constant delay condition is calculated. The quantity "Call-to-Call Variability Adjustment" (CCVA) = $\max(0, D_T)$ shall be added to the obtained delay for the speech signal $T_{R-jitter}(t)$.

The UE delay in the receiving direction shall be reported in the form of an histogram covering the range of measured CCVA-adjusted values ($T_{R-CCVA}(t) = T_{R-jitter}(t) + CCVA$) with a step of 20 ms. The following pseudo code provides an example implementation for the histogram:

```
lo=min(floor( $T_{R-CCVA}(t=1..40)/20$ )*20)
hi=max(ceil( $T_{R-CCVA}(t=1..40)/20$ )*20)
[n,x]=hist( $T_{R-CCVA}(t=1..40)$ ,lo:20:hi)
bar(x,n)
```

The T_{R-CCVA} values for all 40 sentences shall also be reported in the test report.

NOTE 3: The synchronization of the speech frame processing in the UE to the bits of the speech frames at the UE antenna may lead to a variability of up to 20 ms of the measured UE receive delay between different calls. This synchronization is attributed to the UE receiving delay according to the definition of the UE delay reference points. The effect of this possible call-to-call variation is taken into account with the CCVA = $\max(0, D_T)$ value.

7.13.2 Speech quality loss histogram

For the evaluation of speech quality loss in conditions with packet arrival time variations and packet loss, the speech test signal described in clause 7.13.1 shall be used. Two 48 kHz recordings are used to produce the speech quality loss metric:

- A recording obtained in jitter and error free conditions with the test signal described in clause 7.13.1 (reference condition)
- A recording obtained during the application of packet arrival time variations and packet loss as described in clause 7.13.1 (test condition)

The speech quality of the signal is estimated using the measurement algorithm described in ITU-T Recommendation P.863 [44]. Level pre-alignment to -26 dBov of recordings shall be used – see P.863.1 clause 10.2 [45].

NOTE: The setup for acoustical measurement described in P.863 [44] is used. P.863 needs the signal at DRP with diffuse-field equalization.

A score shall be computed for each 8s speech sentence pair. The MOS-LQO values for the reference and test conditions shall be reported in the form of an histogram covering the range of measured values with a step of 0.1 and the values for all 20 sentences pairs shall also be reported in the test report. The following pseudo code provides an example implementation for the histogram:

```
lo=min(floor( $MOS-LQO_{test\ condition}(i=1..20)/0.1$ )*0.1)
hi=max(ceil( $MOS-LQO_{test\ condition}(i=1..20)/0.1$ )*0.1)
[n,x]=hist( $MOS-LQO_{test\ condition}(i=1..20)$ ,lo:0.1:hi)
```

$\text{bar}(x, n)$

The synchronization between stimuli and degraded condition shall be done by the test system before applying the P.863 algorithm on each sentence pair.

8 Wideband telephony transmission performance test methods

8.1 Applicability

The test methods in this clause shall apply when testing a UE that is used to provide wideband telephony, either as a stand-alone service, or as part of a multimedia service.

The application force used to apply the handset against the artificial ear shall be 8 ± 2 N. For the headset case, the application of the headset shall comply with ITU-T Recommendation P.57 [14].

8.2 Overall loss/loudness ratings

8.2.1 General

The SLR and RLR values for GSM, 3G, LTE, NR or WLAN networks apply up to the POI. However, the main determining factors are the characteristics of the UE, including the analogue to digital conversion (ADC) and digital to analogue conversion (DAC). In practice, it is convenient to specify loudness ratings to the Air Interface. For the normal case, where the GSM, 3G, LTE, NR or WLAN network introduce no additional loss between the Air Interface and the POI, the loudness ratings to the PSTN boundary (POI) will be the same as the loudness ratings measured at the Air Interface.

8.2.2 Connections with handset UE

8.2.2.1 Sending loudness rating (SLR)

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be $-4,7$ dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence.
- b) The handset terminal is setup as described in clause 5. The sending sensitivity shall be calculated from each band of the 20 frequencies given in table G.1 of ITU-T Recommendation P.79 Annex A [16], bands 1 to 20. For the calculation, the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.
- c) The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23b), over bands 1 to 20, using $m = 0,175$ and the sending weighting factors from ITU-T Recommendation P.79 Annex A [16], table A2.

8.2.2.2 Receiving loudness rating (RLR)

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The handset terminal is setup as described in clause 5. The receiving sensitivity shall be calculated from each band of the 20 frequencies given in table A.2 of ITU-T Recommendation P.79 Annex A [16], bands 1 to 20. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

- c) The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23c), over bands 1 to 20, using $m = 0,175$ and the receiving weighting factors from table A.2 of ITU-T Recommendation P.79 Annex A [16].
- d) DRP-ERP correction is applied. No leakage correction shall be applied.

8.2.2.3 Receiving loudness rating (RLR) in the presence of background noise

This test shall be performed with the volume control at the maximum setting.

- a) The downlink test signal to be used for the measurements shall be composed from the British-English material of clause 7.3 in ITU-T Recommendation P.501 [22]. The signal is a concatenation of the following parts (see also Figure 15a6):
 - 1) The complete female short conditioning sequence (which includes a leading 5 s silence which is here used for conditioning the UE under noise-only conditions). The level shall be -16 dBm0 active speech level (ITU-T P.56) measured at the digital reference point or the equivalent analogue point.
 - 2) Pause 1 sec.
 - 3) The first 5 s of the male short conditioning sequence. The level shall be -16 dBm0 active speech level (ITU-T P.56) measured at the digital reference point or the equivalent analogue point.
 - 4) The complete single-talk sequence. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete single-talk test sequence.
- b) The background noise playback shall be set up and calibrated according to handset related descriptions in ETSI TS 103 224 or ES 202 396-1. The "Pub Noise (Pub)" recording is used. This scenario is available for both playback systems and is described in sub-clause 8.2 of TS 103 224 (8-channel recording) and sub-clause 8.2 of ES 202 396-1 (binaural recording). The type of background noise reproduction used for the test shall be reported.
- c) The handset terminal is setup as described in clause 5 and the relevant noise playback specification, TS 103 224 or ES 202 396-1. Set the volume control to maximum.
- d) The start of the recording of the artificial ear signal and the start of the background noise playback shall be synchronized to ensure consistent timing of repeated recordings, and allow background noise cancellation by subtraction in the time domain as described below. The noise playback shall be looped with a 20 ms crossfade at the loop point.
- e) The background noise and downlink speech stimulus are applied and a recording of the artificial ear DRP signal is performed. The sound file is saved (file1). The playback is stopped.
- f) The recording described in e) is repeated, this time with muted downlink signal. The soundfile is saved (file2).
- g) A subtraction of the samples is performed (file1-file2). The resulting time signal contains the output of the UE to the artificial ear, where the background noise has been essentially eliminated. For further analysis, the duration of the conditioning sequence and the receive delay from POI to DRP is discarded, resulting in a time signal containing only the single-talk sequence.
- h) The RLR is calculated according to sub-clause 8.2.2.2.

8.2.3 Connections with desktop and vehicle-mounted hands-free UE

Vehicle-mounted hands-free UE should be tested within the vehicle (for the totally integrated vehicle hands-free systems) or in a vehicle simulator, as described in 3GPP TS 03.58 [11].

Free-field measurements for vehicle-mounted hands-free are for further study.

8.2.3.1 Sending loudness rating (SLR)

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be $-4,7$ dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence. The broadband signal level then is adjusted to $-28,7$ dBPa at the HFRP or the HATS HFRP (as defined in ITU-Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as references to determine the sending sensitivity S_{mJ} .

- b) The hands-free terminal is setup as described in clause 5. The sending sensitivity shall be calculated from each band of the 20 frequencies given in table A.2 of ITU-T Recommendation P.79 Annex A [16], bands 1 to 20. For the calculation, the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.
- c) The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23b), over bands 1 to 20, using $m = 0,175$ and the sending weighting factors from ITU-T Recommendation P.79 Annex A [16], table A.2.

8.2.3.2 Receiving loudness rating (RLR)

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. The receiving sensitivity shall be calculated from each band of the 20 frequencies given in table A.2 of ITU-T Recommendation P.79 Annex A [16], bands 1 to 20.

For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

- c) The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23c), over bands 1 to 20, using $m = 0,175$ and the receiving weighting factors from table A.2 of ITU-T Recommendation P.79 Annex A [16].
- d) No leakage correction shall be applied. The hands-free correction as described in ITU-T Recommendation P.340 shall be applied. To compute the receiving loudness rating (RLR) for a hands-free terminal (see also ITU-T Recommendation P.340), when using the combination of left and right artificial ear signals from the HATS, the HFL_E has to be 8 dB instead of 14 dB. For further information see ITU-T Recommendation P.581.

8.2.4 Connections with hand-held hands-free UE

8.2.4.1 Sending loudness rating (SLR)

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be $-4,7$ dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence. The broadband signal level then is adjusted to $-28,7$ dBPa at the HFRP or the HATS HFRP (as defined in P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as reference to determine the sending sensitivity S_{mJ} .

- b) The hands-free terminal is setup as described in clause 5. The sending sensitivity shall be calculated from each band of the 20 frequencies given in table A.2 of ITU-T Recommendation P.79 Annex A [16], bands 1 to 20. For

the calculation the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.

- c) The sensitivity is expressed in terms of dBV/Pa and the SLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23b), over bands 1 to 20, using $m = 0,175$ and the sending weighting factors from ITU-T Recommendation P.79 Annex A [16], table A.2.

8.2.4.2 Receiving loudness rating (RLR)

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. The receiving sensitivity shall be calculated from each band of the 20 frequencies given in table A.2 of ITU-T Recommendation P.79 Annex A [16], bands 1 to 20.

For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

- c) The sensitivity is expressed in terms of dBPa/V and the RLR shall be calculated according to ITU-T Recommendation P.79 [16], formula (A-23c), over bands 1 to 20, using $m = 0,175$ and the receiving weighting factors from table A.2 of ITU-T Recommendation P.79 Annex A [16].
- d) No leakage correction shall be applied. The hands-free correction as described in ITU-T Recommendation P.340 shall be applied. To compute the receiving loudness rating (RLR) for hands-free terminals (see also ITU-T Recommendation P.340) when using the combination of left and right artificial ear signals from the HATS the HFL_E has to be 8 dB, instead of 14 dB. For further information see ITU-T Recommendation P.581.

8.2.5 Connections with headset UE

Same as for handset.

8.3 Idle channel noise (handset and headset UE)

For idle noise measurements in sending and receiving directions, care should be taken that only the noise is windowed out by the analysis and the result is not impaired by any remaining reverberation or by noise and/or interference from various other sources. Some examples are air-conducted or vibration-conducted noise from sources inside or outside the test chamber, disturbances from lights and regulators, mains supply induced noise including grounding issues, test system and system simulator inherent noise as well as radio interference from the UE to test equipment such as ear simulators, microphone amplifiers, etc.

8.3.1 Sending

The terminal should be configured to the test equipment as described in subclause 5.1.

The environment shall comply with the conditions described in subclause 6.1.

The noise level at the output of the SS is measured with A-weighting. The A-weighting filter is described in IEC 60651.

A test signal may have to be intermittently applied to prevent 'silent mode' operation of the MS. This is for further study.

The measured part of the noise shall be 170,667 ms (which equals 8192 samples in a 48 kHz sample rate test system). The spectral distribution of the noise is analyzed with an 8k FFT using windowing with $\leq 0,1$ dB leakage for non bin-centered signals. This can be achieved with a window function commonly known as a "flat top window". Within the specified frequency range, the FFT bin that has the highest level is searched for; the level of this bin is the maximum level of a single frequency disturbance.

To improve repeatability, the test sequence (optional activation followed by the noise level measurement) may be contiguously repeated one or more times.

The total noise powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

The single frequency maximum powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

8.3.2 Receiving

The terminal should be configured to the test equipment as described in subclause 5.1.

The environment shall comply with the conditions described in subclause 6.1.

A test signal may have to be intermittently applied to prevent 'silent mode' operation of the MS. This is for further study.

The noise shall be measured with A-weighting at the DRP with diffuse-field correction. The A-weighting filter is described in IEC 60651.

The measured part of the noise shall be 170,667 ms (which equals 8192 samples in a 48 kHz sample rate test system). The spectral distribution of the noise is analyzed with an 8k FFT using windowing with $\leq 0,1$ dB leakage for non bin-centered signals. This can be achieved with a window function commonly known as a "flat top window". Within the specified frequency range the FFT bin that has the highest level is searched for; the level of this bin is the maximum level of a single frequency disturbance.

To improve repeatability, the test sequence (optional activation followed by the noise level measurement) may be contiguously repeated one or more times.

The total noise powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

The single frequency maximum powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

8.4 Sensitivity/frequency characteristics

8.4.1 Handset and headset UE sending

The headset case is similar to the handset one, except for the application force.

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be $-4,7$ dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence.
- b) The handset terminal is setup as described in clause 5. Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 8 kHz inclusive. For the calculation, the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.
- c) The sensitivity is expressed in terms of dBV/Pa.

8.4.2 Handset and headset UE receiving

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The handset terminal is setup as described in clause 5. Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 8 kHz inclusive. For the

calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

- c) The HATS is diffuse-field equalized. The sensitivity is expressed in terms of dBPa/V. Information about correction factors is available in subclause 5.1.4.

Optionally, the measurements may be repeated with 2 N and 13 N application force. For these test cases no normative values apply.

8.4.3 Desktop and vehicle-mounted hands-free UE sending

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be $-4,7$ dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence. The broadband signal level is then adjusted to $-28,7$ dBPa at the HFRP or the HATS HFRP (as defined in ITU-T Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as references to determine the sending sensitivity S_{mJ} .

- b) The hands-free terminal is setup as described in clause 5. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 8 kHz inclusive. For the calculation the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The sensitivity is expressed in terms of dBV/Pa.

8.4.4 Desktop and vehicle-mounted hands-free UE receiving

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 8 kHz inclusive. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The sensitivity is expressed in terms of dBPa/V.

8.4.5 Hand-held hands-free UE sending

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be $-4,7$ dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence. The broadband signal level is then adjusted to $-28,7$ dBPa at the HFRP or the HATS HFRP (as defined in ITU-T Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as reference to determine the sending sensitivity S_{mJ} .

- b) The hands-free terminal is setup as described in clause 5.1.3.3. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 8 kHz inclusive. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The sensitivity is expressed in terms of dBV/Pa.

8.4.6 Hand-held hands-free UE receiving

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave band frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. Measurements shall be made at 1/3-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 8 kHz inclusive. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The sensitivity is expressed in terms of dBPa/V.

8.5 Sidetone characteristics

8.5.1 Connections with handset UE

The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal shall be produced by the HATS. The test signal level shall be -4,7 dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence.

The handset UE is set up as described in clause 5. The application force shall be 13 N on the Type 3.3 artificial ear.

Where a user operated volume control is provided, the measurements shall be carried out at the nominal setting of the volume control. In addition the measurement is repeated at the maximum volume control setting. It is expected that for other positions of the volume control setting a consistent behaviour to that of the nominal and maximum settings should be observed. Additional measurements for these positions are not required,

Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 8 kHz inclusive. The averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

The sidetone path loss (L_{meST}), as expressed in dB, shall be calculated from each 1/3rd-octave band (ITU-T Recommendation P.79 [16], table B.1, bands 1 to 20). The Sidetone Masking Rating (STMR), expressed in dB, shall be calculated from formula B-4 of ITU-T Recommendation P.79 [16], using $m = 0.225$ and the weighting factors in table B2 (unsealed condition) of ITU-T Recommendation P.79 [16]. No leakage correction (L_E) shall be applied. DRP-ERP correction is used.

In case the STMR is below the limit, the measurement shall be repeated with the electrical sidetone path disabled and both sets of results shall be reported. In case the STMR is below the limit also with the electrical sidetone path disabled, the result shall not be regarded as a failure. Disconnecting the call is normally disabling the electrical sidetone path; otherwise the UE can be switched off to enter the wanted state.

8.5.2 Headset UE

The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be -4,7 dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence.

Where a user operated volume control is provided, the measurements shall be carried out at the nominal setting of the volume control. In addition the measurement is repeated at the maximum volume control setting. It is expected that for other positions of the volume control setting a consistent behaviour to that of the nominal and maximum settings should be observed. Additional measurements for these positions are not required,

Measurements shall be made at 1/12-octave intervals as given by the R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 8 kHz inclusive. The averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

The sidetone path loss (L_{meST}), as expressed in dB, shall be calculated from each 1/3rd-octave band (ITU-T Recommendation P.79 [16], table B.1, bands 1 to 20). The STMR (in dB) shall be calculated from formula B-4 of ITU-T Recommendation P.79 [16], using $m = 0.225$ and the weighting factors in table B.2 (unsealed condition) of ITU-T Recommendation P.79 [16]. No leakage correction (L_E) shall be applied. DRP-ERP correction is used.

In case the STMR is below the limit, the measurement shall be repeated with the electrical sidetone path disabled and both sets of results shall be reported. In case the STMR is below the limit also with the electrical sidetone path disabled, the result shall not be regarded as a failure. Disconnecting the call is normally disabling the electrical sidetone path; otherwise the UE can be switched off to enter the wanted state.

8.5.3 Hands-free UE (all categories)

No requirement other than echo control.

8.5.4 Sidetone delay for handset or headset

The handset or headset terminal is setup as described in clause 5.

The test signal is a CS-signal complying with ITU-T Recommendation P.501 using a PN-sequence with a length, T , of 4 096 points (for a 48 kHz sample rate test system). The duration of the complete test signal is as specified in ITU-T Recommendation P.501. The level of the signal shall be -4,7 dBPa at the MRP.

The cross-correlation function $\Phi_{xy}(\tau)$ between the input signal $S_x(t)$ generated by the test system in send direction and the output signal $S_y(t)$ measured at the artificial ear is calculated in the time domain:

$$\Phi_{xy}(\tau) = \frac{1}{T} \int_{t=-\frac{T}{2}}^{\frac{T}{2}} S_x(t) \cdot S_y(t + \tau) \quad (1)$$

The measurement window, T , shall be identical to the test signal period, T , with the measurement window synchronized to the PN-sequence of the test signal.

The sidetone delay is calculated from the envelope $E(\tau)$ of the cross-correlation function $\Phi_{xy}(\tau)$. The first maximum of the envelope function occurs in correspondence with the direct sound produced by the artificial mouth; the second one occurs with a possible delayed sidetone signal. The difference between the two maxima corresponds to the sidetone delay. The envelope $E(\tau)$ is calculated by the Hilbert transformation $H\{xy(\tau)\}$ of the cross-correlation:

$$H\{xy(\tau)\} = \sum_{u=-\infty}^{+\infty} \frac{\Phi_{xy}(u)}{\pi(\tau - u)} \quad (2)$$

$$E(\tau) = \sqrt{[\Phi_{xy}(\tau)]^2 + [H\{xy(\tau)\}]^2} \quad (3)$$

It is assumed that the measured sidetone delay is less than $T/2$.

8.6 Stability loss

Where a user-controlled volume control is provided it is set to maximum.

Handset UE: The handset is placed on a hard plane surface with the earpiece facing the surface.

Headset UE: The requirement applies for the closest possible position between microphone and headset receiver within the intended wearing position.

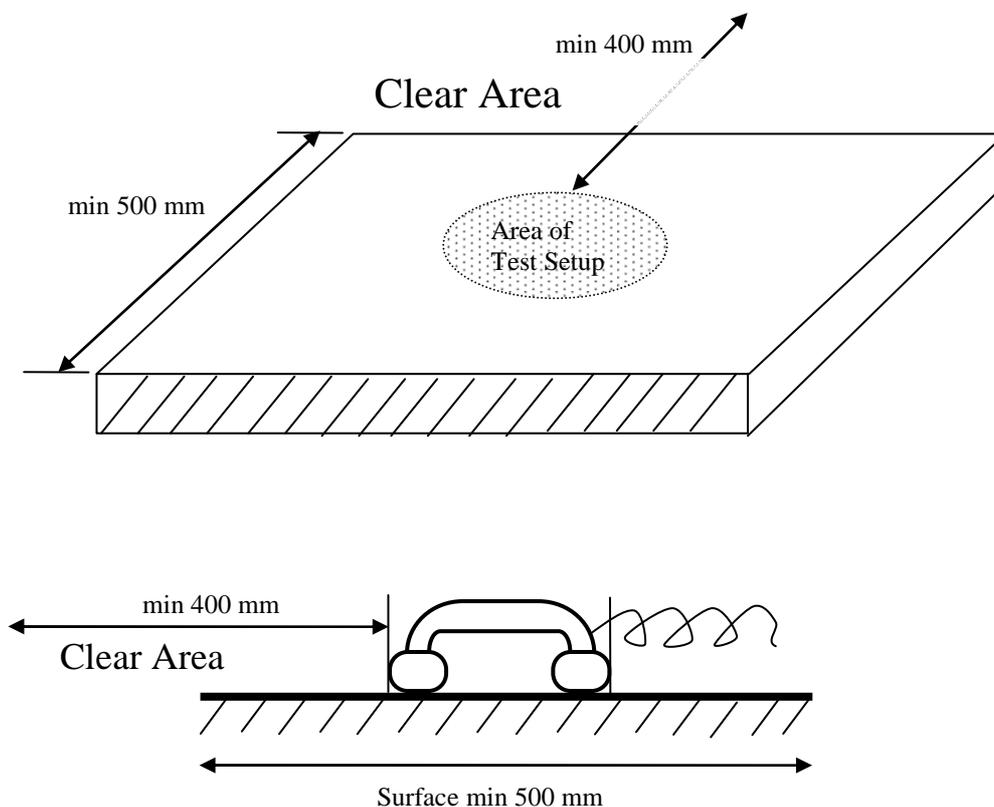
NOTE: Depending on the type of headset it may be necessary to repeat the measurement in different positions.

Hands-free UE (all categories): No requirement other than echo loss.

Before the actual test a training sequence consisting of the British-English single talk sequence described in ITU-T Recommendation P.501 [22] is applied. The training sequence level shall be -16 dBm0 in order to not overload the codec.

The test signal is a PN-sequence complying with ITU-T Recommendation P.501 with a length of 4 096 points (for a 48 kHz sampling rate system) and a crest factor of 6 dB instead of 11 dB. The PN-sequence is generated as described in P.501 with $W(k)$ constant within the frequency range 100-8000 Hz and zero outside this range. The duration of the test signal is 250 ms. With an input signal of -3 dBm0, the attenuation from input to output of the system simulator shall be measured under the following conditions:

- The handset or the headset, with the transmission circuit fully active, shall be positioned on a hard plane surface with at least 400 mm free space in all directions. The earpiece shall face towards the surface as shown in figure 17c;
- The headset microphone is positioned as close as possible to the receiver(s) within the intended wearing position;
- For a binaural headset, the receivers are placed symmetrically around the microphone.



NOTE: All dimensions in mm.

Figure 17c. Test configuration for stability loss measurement on handset or headset UE

The attenuation from input to output shall be measured in the frequency range from 100 Hz to 8 kHz. The spectral distribution of the output signal is analysed with a 4k FFT (for a 48 kHz sample rate test system), thus the measured part of the output signal is 85,333 ms. To avoid leakage effects the frequency resolution of the FFT must be the same as the frequency spacing of the PN-sequence.

8.7 Acoustic echo control

8.7.1 General

The echo loss (EL) presented by the GSM, 3G, LTE, NR or WLAN networks at the POI should be at least 46 dB during single talk. This value takes into account the fact that UE is likely to be used in a wide range of noise environments.

8.7.2 Acoustic echo control in a hands-free UE

The hands-free UE is setup in a room with acoustic properties similar to a typical "office-type" room; a vehicle-mounted hands-free UE should be tested in a vehicle or vehicle simulator, as specified by the UE manufacturer (see also 3GPP TS 03.58 [11]). The ambient noise level shall be ≤ -70 dBPa(A). The attenuation from reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, clause B.4 (trapezoidal rule) but using the frequency range of 300 Hz to 6 700 Hz (instead of 300 Hz to 3 400 Hz). For the calculation, the averaged measured echo level at each frequency band is referred to the averaged test signal level measured in each frequency band. The first 17,0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences).

The test signal level shall be -10 dBm0.

8.7.3 Acoustic echo control in a handset UE

The handset is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A). The attenuation from the reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, clause B.4 (trapezoidal rule) but using the frequency range of 300 Hz to 6 700 Hz (instead of 300 Hz to 3 400 Hz). For the calculation, the averaged measured echo level at each frequency band is referred to the averaged test signal level measured in each frequency band. The first 17,0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences).

The test signal level shall be -10 dBm0.

8.7.4 Acoustic echo control in a headset UE

The headset is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A). The attenuation from the reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, clause B.4 (trapezoidal rule) but using the frequency range of 300 Hz to 6 700 Hz (instead of 300 Hz to 3 400 Hz). For the calculation, the averaged measured echo level at each frequency band is referred to the averaged test signal level measured in each frequency band. The first 17,0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences).

The test signal level shall be -10 dBm0.

8.8 Distortion

8.8.1 Sending distortion

The handset, headset, or hands-free UE is setup as described in clause 5.

The signal used is a sine-wave signal with frequencies specified in clause 6.8 of 3GPP TS 26.131. The sine-wave signal level shall be calibrated to -4,7 dBPa at the MRP for all frequencies, except for the sine-wave with a frequency 1020 Hz

which shall be applied at the following levels at the MRP: 5, 0, -4,7, -10, -15, -20 dBPa. The test signals have to be applied in this sequence, i.e., from high levels down to low levels.

The duration of the sine-wave signal is recommended to be 360 ms. The manufacturer shall be allowed to request tone lengths up to 1 s. The measured part of the signal shall be 170,667 ms (which equals $2 * 4096$ samples in a 48 kHz sample rate test system). The times are selected to be relatively short in order to reduce the risk that the test tone is treated as a stationary signal.

It is recommended that an optional activation signal be presented immediately preceding each test signal to ensure that the UE is in a typical state during measurement (see Note 1.). An appropriate speech or speech-like activation signal shall be chosen from ITU-T Recommendations P.501 or P.50 [10]. A recommendation for the use of an activation signal as part of the measurement is defined in figure 18. The RMS level of the active parts of this activation signal is recommended to be equal to the subsequent test tone RMS level. In practice, certain types of processing may be impacted due to the introduction of the activation signal. The manufacturer shall be allowed to specify disabling of the activation signal. It shall be reported whether an activation signal was used or not, along with the characteristics of the activation signal, as specified by the manufacturer.

The ratio of the signal to total distortion power of the signal output of the SS shall be measured with the psophometric noise weighting (see ITU-T Recommendations G.712, O.41 and O.132). The psophometric filter shall be normalized (0 dB gain) at 800 Hz as specified in ITU-T Recommendation O.41. The weighting function shall be applied to the total distortion component only (not to the signal component).

For measurement of the total distortion component an octave-wide band-stop filter shall be applied to the signal to suppress the sine-wave signal and associated coding artefacts. The filter shall have a lower passband ending at $0.7071 * f_s$, and an upper passband starting at $1.4142 * f_s$, where f_s is the frequency of the sine-wave signal. The passband ripple of the filter shall be $\leq 0,2$ dB. The attenuation of the band-stop filter at the sine-wave frequency shall be ≥ 60 dB. Alternatively, the described characteristics can be implemented by an appropriate weighting on the spectrum obtained from an FFT. The total distortion component is defined as the measured signal within the frequency range 100 Hz to 6 kHz, after applying psophometric and stop filters (hence no correction for the lost power due to the stop filter, known as "bandwidth correction", shall be applied).

To improve repeatability, considering the variability introduced by speech coding and voice processing, the test sequence (activation signal followed by the test signal) may be contiguously repeated one or more times. The single signal-to-total-distortion power ratios obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

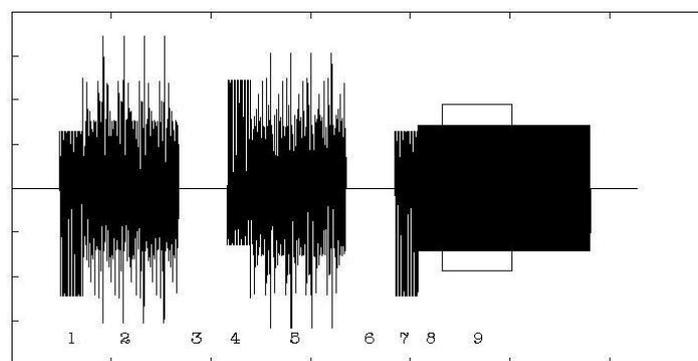


Figure 18: Recommended activation sequence and test signal.

The activation signal consists of a "Bandlimited composite source signal with speech-like power density spectrum" signal according to ITU-T Recommendation P.501 with 48,62 ms voiced part (1), 200 ms unvoiced part (2) and 101,38 ms pause (3), followed by the same signal but polarity inverted (4, 5, 6), followed by the voiced part only (7). The pure test tone is applied and after 50 ms settling time (8), the analysis is made over the following 170,667 ms (9).

NOTE 1: Depending on the type of codec the test signal used may need to be adapted. If a sine-wave is not usable, an alternative test signal could be a band-limited noise signal centered on the above frequencies.

NOTE 2: Void.

NOTE 3: Void.

NOTE 4: In order to ensure that the correct part of the signal is analyzed, the total delay of the terminal and SS may have to be determined prior to the measurement.

NOTE 5: For hands-free terminals tested in environments defined in subclause 6.1.2, care should be taken that the reverberation in the test room, caused by the activation signal, does not affect the test results to an unacceptable degree, referring to subclause 5.3.

8.8.2 Receiving

The handset, headset, or hands-free UE is setup as described in clause 5.

The signal used is a sine-wave signal with frequencies specified in clause 6.8 of 3GPP TS 26.131. The signal level shall be -16 dBm₀, except for the sine-wave signal with a frequency 1020 Hz that shall be applied at the signal input of the SS at the following levels: 0, -3, -10, -16, -20, -30, -40, -45 dBm₀. The test signals have to be applied in this sequence, i.e., from high levels down to low levels.

The duration of the sine-wave signal is recommended to be 360 ms. The manufacturer shall be allowed to request tone lengths up to 1 s. The measured part of the signal shall be 170,667 ms (which equals $2 * 4096$ samples in a 48 kHz sample rate test system). The times are selected to be relatively short in order to reduce the risk that the test tone is treated as a stationary signal.

It is recommended that an optional activation signal be presented immediately preceding each test signal to ensure that the UE is in a typical state during measurement (see Note 1.). An appropriate speech or speech-like activation signal shall be chosen from ITU-T Recommendations P.501 or P.50 [10]. A recommendation for the use of an activation signal as part of the measurement is defined in figure 19. The RMS level of the active parts of this activation signal is recommended to be equal to the subsequent test tone RMS level for low and medium test levels. To avoid saturation of the SS speech encoder, it is recommended for high test levels that the activation signal level is adjusted so that its peak level equals the peak level of the test tone. In practice, certain types of processing may be impacted due to the introduction of the activation signal. The manufacturer shall be allowed to specify disabling of the activation signal. It shall be reported whether an activation signal was used or not, along with the characteristics of the activation signal, as specified by the manufacturer.

The ratio of the signal to total distortion power shall be measured at the applicable acoustic measurement point (DRP with diffuse-field correction for handset and headset modes; free field for hands-free modes) with the psophometric noise weighting (see ITU-T Recommendations G.712, O.41 and O.132). The psophometric filter shall be normalized to have 0 dB gain at 800 Hz as specified in ITU-T Recommendation O.41. The weighting function shall be applied to the total distortion component only (not to the signal component).

For measurement of the total distortion component an octave-wide band-stop filter shall be applied to the signal to suppress the sine-wave signal and associated coding artefacts. The filter shall have a lower passband ending at $0,7071 * f_s$, and an upper passband starting at $1,4142 * f_s$, where f_s is the frequency of the sine-wave signal. The passband ripple of the filter shall be $\leq 0,2$ dB. The attenuation of the band stop filter at the sine-wave frequency shall be ≥ 60 dB. Alternatively the described characteristics can be implemented by an appropriate weighting on the spectrum obtained from an FFT. The total distortion component is defined as the measured signal within the frequency range 100 Hz to 6 kHz, after applying psophometric and stop filters (hence no correction for the lost power due to the stop filter, known as "bandwidth correction", shall be applied).

To improve repeatability, considering the variability introduced by speech coding and voice processing, the test sequence (activation signal followed by the test signal) may be contiguously repeated one or more times. The single signal-to-total-distortion power ratios obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

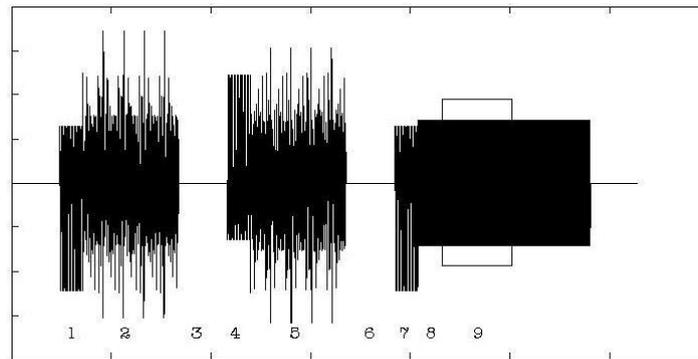


Figure 19: Recommended activation sequence and test signal.

The activation signal consists of a "Bandlimited composite source signal with speech-like power density spectrum" signal according to ITU-T Recommendation P.501 with 48,62 ms voiced part (1), 200 ms unvoiced part (2) and 101,38 ms pause (3), followed by the same signal but polarity inverted (4, 5, 6), followed by the voiced part only (7). The pure test tone is applied and after 50 ms settling time (8), the analysis is made over the following 170,667 ms (9).

NOTE 1: Void.

NOTE 2: Void.

NOTE 3: In order to ensure that the correct part of the signal is analyzed, the total delay of the terminal and SS may have to be determined prior to the measurement.

NOTE 4: For hands-free terminals tested in environments defined in subclause 6.1.2, care should be taken that the reverberation in the test room, caused by the activation signal, does not affect the test results to an unacceptable degree, referring to subclause 5.3.

8.9 Void

8.10 Delay

8.10.0 UE Delay Measurement Methodologies

For UMTS circuit-switched operation and MTSI-based speech with LTE, NR or WLAN access in error and jitter free conditions, the sum of the UE delays in the sending and receiving directions (T_S+T_R) shall be measured according to the methods described in clauses 8.10.1 and 8.10.2. In the event that the delays of the test equipment in send and/or receive directions are not stable between calls or cannot be accurately determined, the alternative method described in clause 8.10.3 may be used to obtain (T_S+T_R) and the measured instability or inaccuracy observed when the methods described in 8.10.1 and 8.10.2 were performed shall be recorded in the test report. The test method(s) used and all results obtained shall also be recorded in the test report.

For MTSI-based speech with LTE, NR or WLAN access in conditions with simulated packet arrival time variations, the sum of the UE delays in the sending and receiving directions ($T_S+T_{R-jitter}$) and the objective speech quality in the receive direction shall be measured according to the method described in clause 8.10.4.

For MTSI-based speech with LTE, NR or WLAN access, prior to the actual measurements, the clock skew between UE and reference client shall be compensated by adjusting the clock of the reference client to match the clock of the UE (as stated in clause 5). The inaccuracy of the clock skew adjustment shall be less than 1ppm measured according to the procedure in Annex D.

It shall be ensured that the packet generation by the reference client and the packet treatment of the test equipment are free of jitter.

8.10.1 Delay in sending direction (handset UE)

The handset terminal is setup as described in clause 5.1.1.

The UE delay in the sending direction is obtained by measuring the delay between MRP and the electrical access point of the test equipment and subtracting the delays introduced by the test equipment from the measured value.

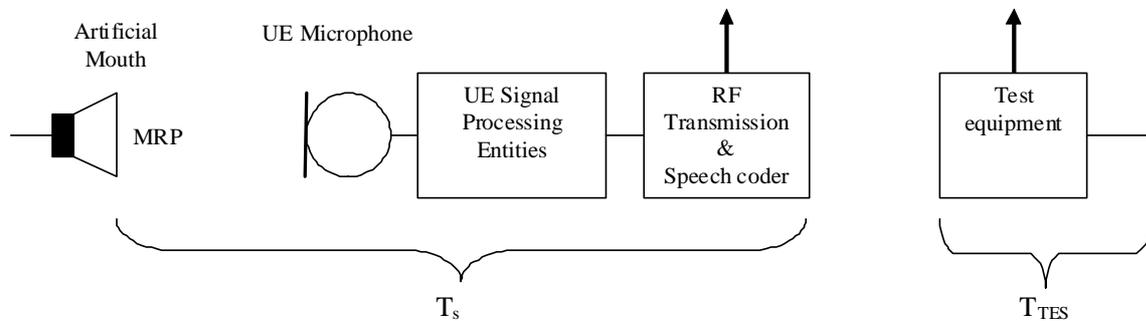


Figure 19b1: Different entities when measuring the delay in sending direction

The delay measured from MRP to the electrical access point of the test equipment is $T_s + T_{TES}$.

T_{TES} : The delay between the last bit of a speech frame at the system simulator antenna and the first electrical event at the electrical access point of the test equipment.

1. For the measurements, a Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] is used. The pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 32 k samples (with 48 kHz sampling rate). The test signal level is -4,7 dBPa at the MRP.
2. The reference signal is the original signal (test signal). The setup of the handset/headset terminal is made corresponding to clause 5.1.
3. The delay is determined by cross-correlation analysis between the measured signal at the electrical access point and the original signal. The measurement is corrected by subtracting the test equipment delay T_{TES} .
4. The delay is measured in ms and the maximum of the cross-correlation function is used for the determination.

For MTSI-based speech with LTE, NR or WLAN access, a variability of up to 20ms may be expected between different calls due to the synchronization between the speech frame processing in the sending UE and the bits of the speech frames at the UE antenna. This synchronization is attributed to the UE sending delay according to the definition of the UE delay reference points. Hence, the maximum value of the UE sending delay obtained from at least 5 individual calls shall be reported as the UE delay in the sending direction. All values shall be reported in the test report.

A further variability of up to 20ms may be expected between different calls due to the synchronization between the speech frames at the UE antenna and the speech frame processing in the receiving reference client of the test system. In an end-to-end call this synchronization of the frames will only take place at the receiver, and this variability of the measurement shall be deduced from the UE sending delay. Hence, if the reference client of the test equipment does not adjust for the effect of the speech frame synchronization (as specified by the manufacturer of the reference client), this maximum uncertainty shall be subtracted from measured maximum value reported as the UE sending delay in order to compensate for the uncertainty of the test equipment. This correction value (i.e. maximum uncertainty) shall be reported in the test report.

8.10.1a Delay in sending direction (headset UE)

The UE delay in the sending direction is obtained by measuring the delay between MRP and the electrical access point of the test equipment and subtracting the delays introduced by the test equipment from the measured value.

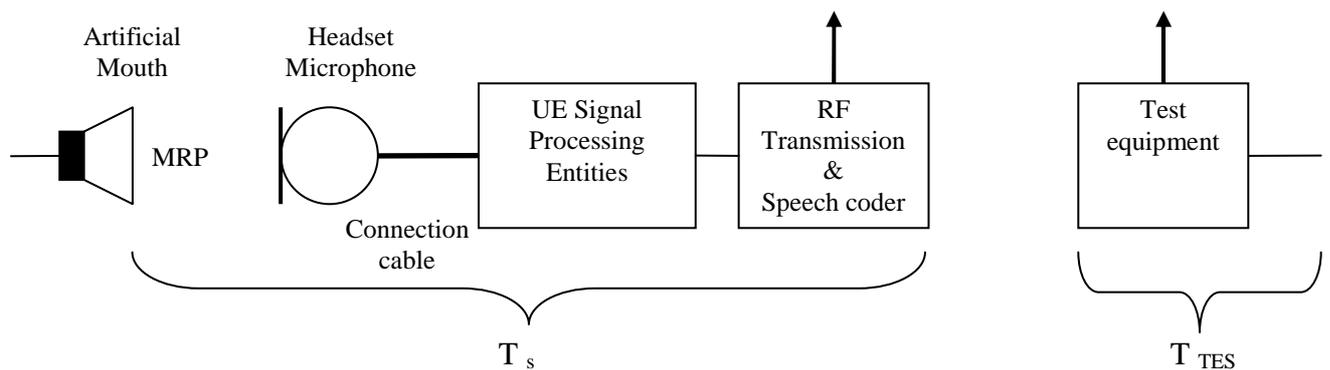


Figure 19b2: Different entities when measuring the delay in sending direction with a headset connected via cable

NOTE: The test setup only applies to headsets connected by wire. Wireless headsets (e.g. connected by Bluetooth) are currently out of scope.

The test method is the same as for handset UE (clause 8.10.1).

8.10.2 Delay in receiving direction (handset UE)

The handset terminal is setup as described in clause 5.

The UE delay in the receiving direction is obtained by measuring the delay between the electrical access point of the test equipment and the DRP and subtracting the delays introduced by the test equipment from the measured value.

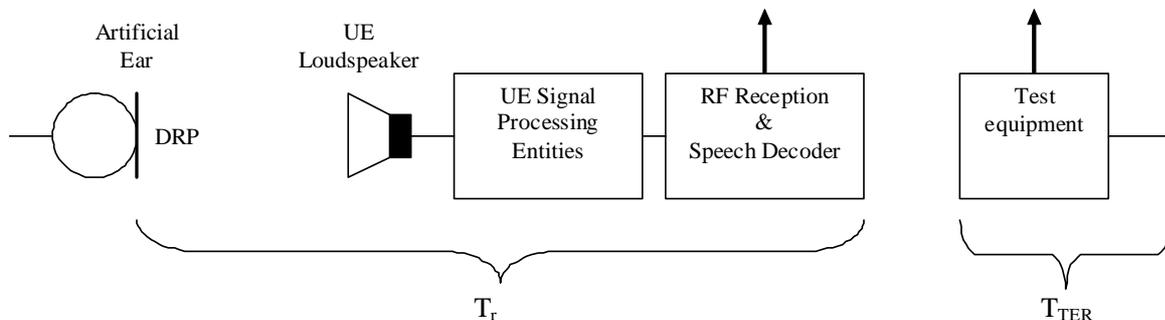


Figure 19b3: Different entities when measuring the delay in receiving direction

The delay measured from the electrical access point of the test equipment to DRP is $T_R + T_{TER}$.

T_{TER} : The delay between the first electrical event at the electrical access point of the test equipment and the first bit of the corresponding speech frame at the system simulator antenna.

Before the actual test for MTSI-based speech with LTE, NR or WLAN access a conditioning sequence consisting of the British-English single talk sequence described in ITU-T Recommendation P.501 [22] is applied for convergence of the jitter buffer management of the UE. The conditioning sequence level shall be -16 dBm0 in order to not overload the codec.

1. For the measurements, a Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] is used. The pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is

recommended to use a pn sequence of 32 k samples (with 48 kHz sampling rate). The test signal level is -16 dBm0 measured at the digital reference point or the equivalent analogue point.

2. The reference signal is the original signal (test signal). The setup of the handset/headset terminal is in correspondence to clause 5.1.
3. The delay is determined by cross-correlation analysis between the measured signal at the electrical access point and the original signal. The measurement is corrected by subtracting the test equipment delay T_{TER} .
4. The delay is measured in ms and the maximum of the cross-correlation function is used for the determination.

For MTSI-based speech with LTE, NR or WLAN access, a variability of up to 20ms may be expected between different calls due to the synchronization between the bits of the speech frames at the UE antenna and the speech frame processing in the receiving UE. This synchronization is attributed to the UE receiving delay according to the definition of the UE delay reference points. Hence, the maximum value of the UE receiving delay obtained from at least 5 individual calls shall be reported as the UE delay in the receiving direction. All values shall be reported in the test report.

8.10.2a Delay in receiving direction (headset UE)

The UE delay in the receiving direction is obtained by measuring the delay between the electrical access point of the test equipment and the DRP and subtracting the delays introduced by the test equipment from the measured value.

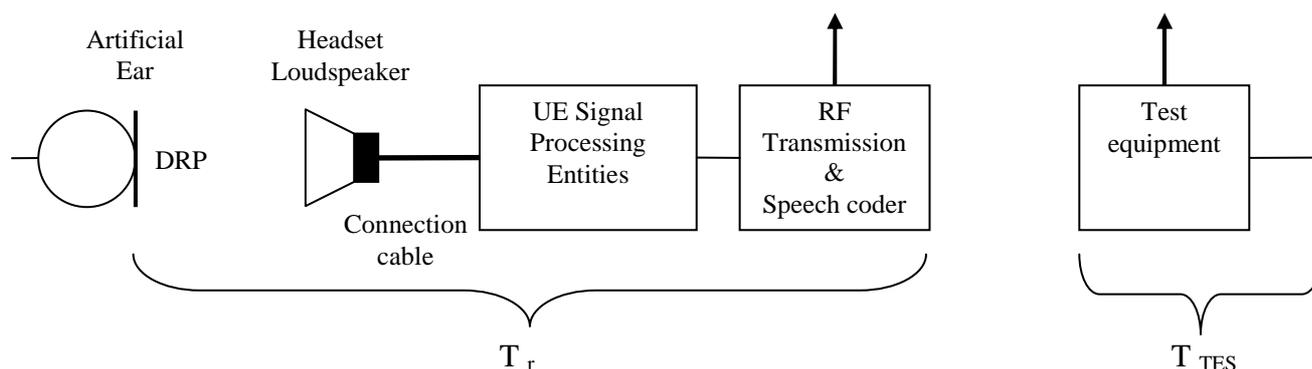


Figure 19b4: Different entities when measuring the delay in receiving direction with a headset connected via cable

NOTE: The test setup only applies to headsets connected by wire. Wireless headsets (e.g. connected by Bluetooth) are currently out of scope.

The test method is the same as for handset UE (subclause 8.10.2).

8.10.3 Delay in sending + receiving direction using "echo" method (handset UE)

The UE delay is obtained by measuring the delay between the MRP and the DRP and subtracting the delays introduced by the test equipment from the measured value.

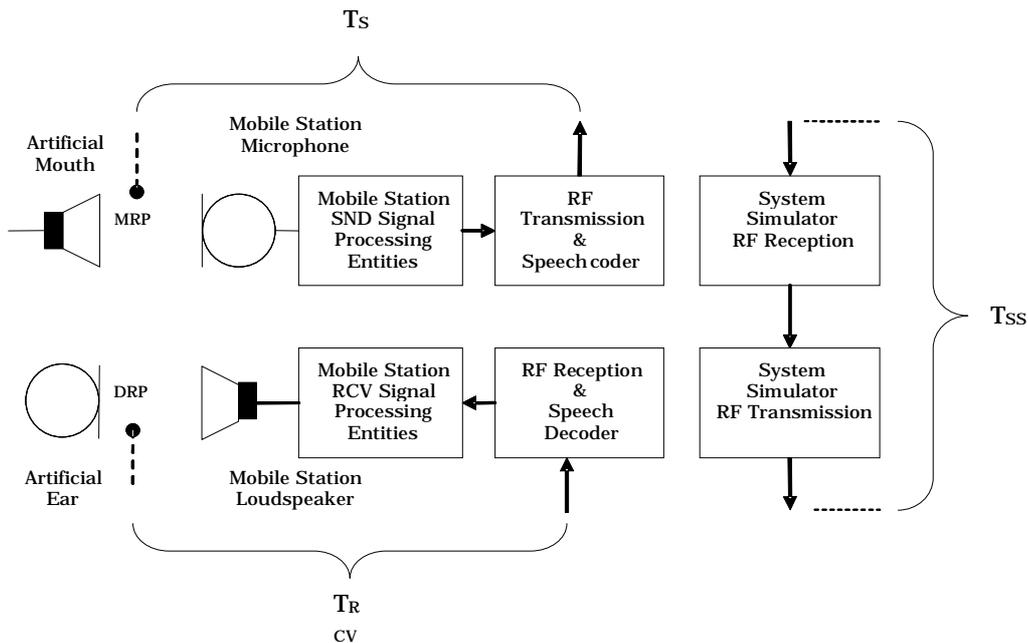


Figure 19b4bis: Different entities when measuring the delay in sending + receiving direction

The delay measured from MRP to DRP is $(T_S + T_R + T_{SS})$.

T_{SS} : The delay between the last bit of a speech frame at the system simulator antenna and the first bit of the looped back speech frame at the system simulator antenna.

Before the actual test for MTSI-based speech with LTE, NR or WLAN access a conditioning sequence consisting of the British-English single talk sequence described in ITU-T Recommendation P.501 [22] is applied for convergence of the jitter buffer management of the UE. The conditioning sequence level shall be -16 dBm0 in order to not overload the codec.

1. For the measurements a Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] is used. It is recommended to use a pn sequence of 32 k samples (with 48 kHz sampling rate). The test signal level is -4.7 dBPa at the MRP.
2. The system simulator is configured for "loopback" or "echo" operation with the additional loopback delay as specified below when applicable. In "loopback" or "echo" operation, the packets in the sending direction are routed to the receiving direction by the system simulator.
3. The reference signal is the original signal (test signal). The setup of the mobile station is in correspondence to clause 5.1.
4. The mouth-to-ear delay is determined by cross-correlation analysis between the measured signal at DRP and the original signal. The analysis window for the cross-correlation shall start at an instant $T > 50\text{ms}$ in order to discard the cross-correlation peaks corresponding to the direct acoustic path from mouth to ear and possible delayed sidetone signal. The measurement is corrected by subtracting the system simulator delay T_{SS} to obtain the $T_S + T_R$ delay.
5. The delay is measured in ms and the maximum of the cross-correlation envelope is used for the determination.

For MTSI-based speech with LTE, NR or WLAN access, a variability of the UE delay with up to 20ms in the respective sending and receiving direction may be expected due to the synchronization of the speech frame processing in the UE to the bits of the speech frame on the UE antenna. This synchronization is attributed to the UE delay according to the definition of the UE delay reference points. Hence, the UE delay shall be reported as the maximum value from at least 5

separate calls each with a different loopback delay TSS in at least 5 steps of 4ms in the full range from 0 to 16ms. All values shall be reported in the test report.

8.10.3a Delay in sending + receiving direction using "echo" method (headset UE)

The UE delay is obtained by measuring the delay between the MRP and the DRP and subtracting the delays introduced by the test equipment, T_{SS} , from the measured value.

The test method is the same as for handset UE (clause 8.10.3).

8.10.4 Delay and speech quality in conditions with packet arrival time variations and packet loss

8.10.4.1 Delay in sending direction

The UE delay in the sending direction, T_s , shall be measured in jitter and error free conditions according to clause 8.10.0.

8.10.4.2 Delay in receiving direction

For this test it shall be ensured that the call is originated from the mobile terminal (MO).

NOTE 1: Differences have been observed between mobile originated call and mobile terminated call. For better consistency MO calls are used.

The test signal consists of 3 repeats of the Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] followed by a speech signal of 160s. During the first two CSS signals the terminal can adapt its jitter buffer. The third CSS is used for measuring the delay in constant-delay condition, and the speech signal is used for delay and quality measurement in the packet impairment condition.

Constant delay T_c corresponding to the minimum delay of the profile (i.e. the compensation value for the profile) shall be added at the beginning of the different delay/loss profiles, to avoid unnecessary delay jumps between the two measurement phases and realistic conditions for the second measurement test phase.

In receiving direction, the delay between the electrical access point of the test equipment and the DRP, $T_{TEAP-DRP}(t) = T_{R-jitter}(t) + T_{TER}$, is measured in two successive phases:

- 1) First the delay in constant-delay condition $T_{TEAP-DRP-constant}$ is measured as described in steps 1 to 4, clause 7.10.2, using the third CSS signal. The constant delay T_c is subtracted from $T_{TEAP-DRP}$ to obtain $T_{R-constant}$.
- 2) Then the delay with packet impairment $T_{R-jitter}(t)$ is measured continuously for a speech signal during the inclusion of packet delay and loss profiles in the receiving direction RTP voice stream.

Packet impairments shall be applied between the reference client and system simulator eNodeB. Separate calls shall be established for each packet impairment condition.

The start of the delay profiles must be synchronized with the start of the downlink speech material reproduction (compensated by the delay between reproduction and the point of impairment insertion, i.e. the delay of the reference client) in order to ensure a repeatable application of impairments to the test speech signal. Tests shall be performed with DTX enabled in the reference client.

NOTE 2: RTP packet impairments representing packet delay variations and loss in LTE transmission scenarios are specified in Annex E. These LTE jitter/loss profiles are reused also for tests with WLAN and NR access. Care must be taken that the system simulator uses a dedicated bearer with no buffering/scheduling of packets for transmission.

For the CSS signal repeated 3 times, the pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 32 k samples (with 48 kHz sampling rate). The test signal level is -16 dBm0 measured at the digital reference point or the equivalent analogue point.

For the speech signal, 8 English test sentences according to ITU-T P.501 Annex C.2.3, normalized to an active speech level of -16dBm0, are used (2 male, 2 female speakers). The sequences are concatenated in such a way that all sentences are centered within a 4.0s time window, which results in an overall duration of 32.0s. The sequences are repeated 5 times, resulting in a test file 160.0s long. The first 2 sentences are used for convergence of the UE jitter buffer manager and are discarded from the analysis. Equivalent implementations of the concatenation by repeating the test sentences in sequence may be used.

For the delay calculation with the speech signal, a cross-correlation with a rectangular window length of 4s, centered at each sentence of the stimulus file, is used. The process is repeated for each sample. For each cross correlation, the maximum of the envelope is obtained producing one delay value per sentence.

The UE delay in the receive direction, $T_{R-jitter}(t)$, is obtained by subtracting the delay introduced by the test equipment and the simulated transport network packet delay introduced by the delay and loss profile (as specified for the respective profile in Annex E) from the first electrical event at the electrical access point of the test equipment to the first bit of the corresponding speech frame at the system simulator antenna, T_{TER} , from the measured $T_{TEAP-DRP}(t)$.

The difference D_T between maximum receiving delay obtained with at least 5 individual calls (see clause 7.10.2) and the delay $T_{R-constant}$ measured for the CSS signal in constant delay condition is calculated. The quantity "Call-to-Call Variability Adjustment" (CCVA) = $\max(0, D_T)$ shall be added to the obtained delay for the speech signal $T_{R-jitter}(t)$.

For stationary packet delay variation test conditions (test condition 1 and 2), the first 2 sentences are used for convergence of the jitter buffer management and are discarded from the analysis. The CCVA-adjusted UE delay ($T_{R-CCVA}(t) = T_{R-jitter}(t) + CCVA$) in the receiving direction shall be reported as the maximum value excluding the two largest values of the remaining sequence of the 38 sentence delay values, i.e. the 95-percentile value of $T_{R-CCVA}(t)$. The T_{R-CCVA} values for all 40 sentences shall be reported in the test report.

NOTE 3: The synchronization of the speech frame processing in the UE to the bits of the speech frames at the UE antenna may lead to a variability of up to 20 ms of the measured UE receive delay between different calls. This synchronization is attributed to the UE receiving delay according to the definition of the UE delay reference points. The effect of this possible call-to-call variation is taken into account with the $CCVA = \max(0, D_T)$ value.

8.10.4.3 Speech quality loss in conditions with packet arrival time variations and packet loss

For the evaluation of speech quality loss in conditions with packet arrival time variations and packet loss, the test signal described in clause 8.10.4.2 shall be used. The first 2 sentences are used for convergence of the UE jitter buffer manager and are discarded from the analysis. Two 48 kHz recordings are used to produce the speech quality loss metric:

- A recording obtained in jitter and error free conditions with the test signal described in clause 8.10.4.2 (reference condition)
- A recording obtained during the application of packet arrival time variations and packet loss as described in clause 8.10.4.2 (test condition)

The speech quality of the signal is estimated using the measurement algorithm described in ITU-T Recommendation P.863 [44]. Level pre-alignment to -26 dBov of recordings shall be used – see P.863.1 clause 10.2 [45].

NOTE: The setup for acoustical measurement described in P.863 [44] is used. P.863 needs the signal at DRP with diffuse-field equalization.

A score shall be computed for each 8s speech sentence pair and averaged to produce a mean MOS-LQO value for the reference and test conditions.

$$MOS-LQO_{REF} = \left(\sum_{i=1}^{19} MOS - LQO_{reference}(i) \right) / 19$$

$$MOS-LQO_{TEST} = \left(\sum_{i=1}^{19} MOS - LQO_{testcondition}(i) \right) / 19$$

NOTE: This evaluation of the speech quality requirement is only applicable to test conditions with a stationary statistic of the packet delay variation. Evaluation of the speech quality for a test condition with non-stationary packet delay variations is for further study.

The synchronization between stimuli and degraded condition shall be done by the test system before applying the P.863 algorithm on each sentence pair.

8.10.5 UE send clock accuracy

The UE clock accuracy in send direction shall be measured according to Annex D.

NOTE1: For this specific measurement, care should be taken about the clock accuracy of the test equipment. See Table 1a.

NOTE2: As required in clause 5, prior to the actual measurements for MTSI-based speech with LTE, NR or WLAN access, the clocks of the reference client and the UE have to be synchronized. This measurement of UE send clock accuracy does not need to be repeated and can be obtained from this setup procedure.

8.10.6 UE receiving with clock skew

For further study.

8.11 Echo control characteristics

8.11.1 Test set-up and test signals

The device is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A).

The test shall be performed with the British-English "long" double-talk and conditioning speech sequences from ITU-T Recommendation P.501 [22], with the signals in the receiving direction band limited according to clause 5.4.

A description of the test stimuli is presented in Table 2e and Table 2f. The test sequence is composed of an initial conditioning sequence of 23,5 s and a double talk sequence of 35 s. For the analysis, the double talk sequence is divided into two segments, a first double-talk sequence with single short near-end words (0 – 20 s), and a second double-talk sequence with continuous double talk (20-35 s).

The sending speech during double-talk and the "near-end speech only" are recorded individually, with the "near-end speech only" sequence recorded with silence in the receiving direction. The time-alignment of the two recorded sequences is performed off-line during the analysis.

Table 2e: Test stimuli for recording of Echo Canceller operation

	Conditioning	Single words (segment 1) and full sentence (segment 2) double talk
Far-end signal	FB_female_conditioning_seq_long.wav	FB_male_female_single-talk_seq.wav
Artificial mouth signal	FB_male_conditioning_seq_long.wav	FB_male_female_double-talk_seq.wav

Table 2f: Test stimuli for reference "near-end speech only" recording.

	Conditioning	Single words (segment 1) and full sentence (segment 2) double talk
Far-end signal	FB_female_conditioning_seq_long.wav	silence
Artificial mouth signal	FB_male_conditioning_seq_long.wav	FB_male_female_double-talk_seq.wav

The level of the signal of the artificial mouth shall be - 4.7 dBPa measured at the MRP. In order to obtain a reproducible time alignment as seen by the UE, the artificial mouth signal shall be delayed by the amount of the receiving direction delay. For the purpose of this alignment, the receiving direction delay for handset and headset modes is defined from

the system simulator input to the artificial ear. For handsfree modes, the downlink delay is defined from the system simulator input to the acoustic output from the UE loudspeaker.

The level of the downlink signal shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point.

8.11.2 Test method

The test method measures the duration of any level difference between the sending signal of a double-talk sequence (where the echo canceller has been exposed to simultaneous echo and near-end speech) and the sending signal of the same near-end speech only. The level difference is classified into eight categories according to Figure 19b5 and Table 2g, representing various degrees of "Full duplex operation", "Near-end clipping", and "Residual echo".

NOTE: The limits for specifying the categories in Figure 19b5 and Table 2g are provisional pending further analysis and validation.

NOTE: The categories in Figure 19b5 and Table 2g are labelled in a functional order and the subjective impression of the respective categories is for further study.

NOTE: To reduce potential issues associated with low-frequency test room noise, a [4th]-order high-pass filter with a cut-off frequency of [100] Hz can be applied before the level computation.

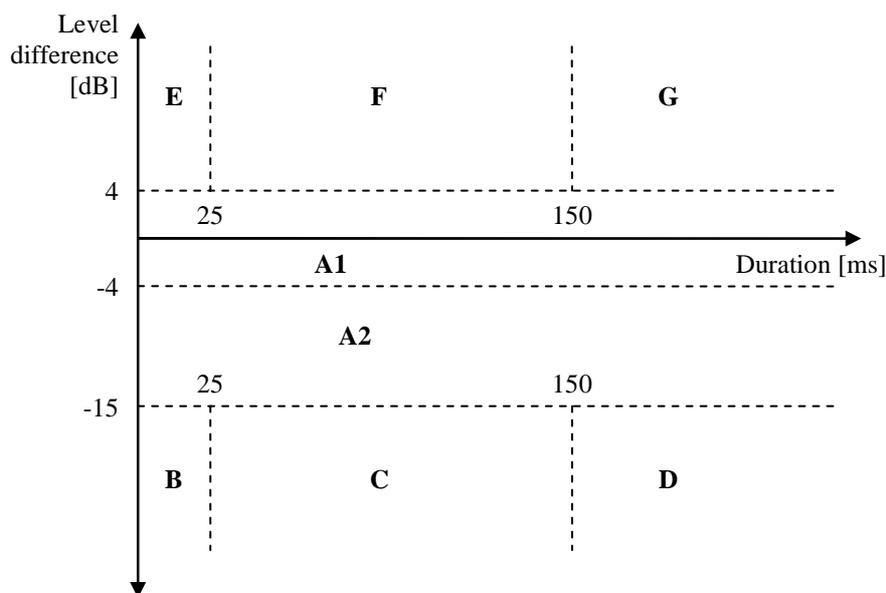


Figure 19b5: Classification of echo canceller performance

Table 2g: Categories for echo canceller performance classification

Category	Level difference (ΔL)	Duration (D)	Description
A1	$-4 \text{ dB} \leq \Delta L < 4 \text{ dB}$		Full-duplex and full transparency
A2	$-15 \text{ dB} \leq \Delta L < -4 \text{ dB}$		Full-duplex with level loss in Tx
B	$\Delta L < -15 \text{ dB}$	$D < 25 \text{ ms}$	Very short clipping
C	$\Delta L < -15 \text{ dB}$	$25 \text{ ms} \leq D < 150 \text{ ms}$	Short clipping resulting in loss of syllables
D	$\Delta L < -15 \text{ dB}$	$D \geq 150 \text{ ms}$	Clipping resulting in loss of words
E	$\Delta L \geq 4 \text{ dB}$	$D < 25 \text{ ms}$	Very short residual echo
F	$\Delta L \geq 4 \text{ dB}$	$25 \text{ ms} \leq D < 150 \text{ ms}$	Echo bursts
G	$\Delta L \geq 4 \text{ dB}$	$D \geq 150 \text{ ms}$	Continuous echo

A pseudo-code reference of the test method including test scripts and test-vectors is presented in Clause C.3 and outlined in the following sub clauses.

8.11.2.1 Signal alignment

For the analysis of the signal level difference, the send signal during double-talk and the near-end only signal are aligned using a correlation analysis as described in Clause C.3.2.

8.11.2.2 Signal level computation and frame classification

The analysis is based on the digital level measured with a meter according to IEC 61672 [38] with a time constant of 12.5 ms, sampled at 5 ms intervals corresponding to the evaluated frames.

The "double-talk" frames are defined as the frames where both the far-end (receiving direction) signal includes active speech (extended with a hang-over period of 200 ms) and the near-end signal is composed of active speech. Active speech is defined to be detected using a speech level meter according to ITU-T P.56, and frames within -15.9 dB from the active speech level are classified as active speech frames.

The "far-end single-talk adjacent to double-talk" frames are similarly defined using a speech level meter according to ITU-T P.56 as the frames with active far-end speech (extended with a hang-over period of 200 ms) and no active near-end speech (extended with a hang-over period of 200 ms).

A reference implementation of the signal level computation and frame classification is presented in Clause C.3.3.

8.11.2.3 Classification into categories

The analysis and classification into the categories according to Figure 19b5 and Table 2g is performed according to the reference implementation described in Clause C.3.4 and C.3.4.

The frames are first categorized according to the level categories defined in Table 2g. To determine the durations, the amount of adjacent frames falling into the same level category is determined.

The classification is then performed individually for the following situations:

- frames classified as "double-talk" from segment 1 of the double-talk sequence (see 8.11.1)
- frames classified as "far-end single-talk adjacent to double-talk" from segment 1 of the double-talk sequence
- frames classified as "double-talk" from segment 2 of the double-talk sequence
- frames classified as "far-end single-talk adjacent to double-talk" from segment 2 of the double-talk sequence

To determine the percentage values for each category (A1, A2, B, C, D, E, F, and G) within each situation, the number of frames falling into the respective category is divided by the total number of frames within the situation in question.

To determine the averaged level difference of the frames for each category (A1, A2, B, C, D, E, F, and G) within each situation, the sum of the level difference (in dB) of the frames falling into the respective category is divided by the total number of frames within the situation in question.

8.12 Quality (speech quality, noise intrusiveness) in the presence of ambient noise

8.12.1 Handset

The speech quality in sending for wideband systems is tested based on ETSI TS 103 106 [34]. This test method leads to three MOS-LQOw quality numbers:

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

S-MOS-LQOw: Transmission quality of the speech

G-MOS-LQOw: Overall transmission quality

The test arrangement is given in clause 5.1.5. For connections with handset UE, the measurement is conducted for 8 noise conditions as described in Table 2h. The measurements should be made in the same unique and dedicated call. The noise types shall be presented according to the order specified in Table 2h.

Table 2h: Noise conditions used for ambient noise simulation in handset mode as specified in ES 202 396-1 [35]

Description	File name	Duration	Level	Type
Recording in pub	Pub_Noise_binaural_V2	30 s	L: 75,0 dB(A) R: 73,0 dB(A)	Binaural
Recording at pavement	Outside_Traffic_Road_binaural	30 s	L: 74,9 dB(A) R: 73,9 dB(A)	Binaural
Recording at pavement	Outside_Traffic_Crossroads_binaural	20 s	L: 69,1 dB(A) R: 69,6 dB(A)	Binaural
Recording at departure platform	Train_Station_binaural	30 s	L: 68,2 dB(A) R: 69,8 dB(A)	Binaural
Recording at the drivers position	Fullsize_Car1_130Kmh_binaural	30 s	L: 69,1 dB(A) R: 68,1 dB(A)	Binaural
Recording at sales counter	Cafeteria_Noise_binaural	30 s	L: 68,4 dB(A) R: 67,3 dB(A)	Binaural
Recording in a cafeteria	Mensa_binaural	22 s	L: 63,4 dB(A) R: 61,9 dB(A)	Binaural
Recording in business office	Work_Noise_Office_Callcenter_binaural	30 s	L: 56,6 dB(A) R: 57,8 dB(A)	Binaural

- 1) Before starting the measurements a proper conditioning sequence shall be used. The conditioning sequence shall be comprised of the four additional sentences 1-4 described in ETSI TS 103 106 [34], applied to the beginning of the 16-sentence test sequence.

NOTE: The sequence of speech samples concatenated for the test signal, consisting of alternating talkers in the sending direction, reduces the overall test time but may represent an unrealistic behaviour for certain voice enhancement technologies. Alternative concatenations are for further study.

- 2) The send speech signal consists of the 16 sentences of speech as described in ETSI TS 103 106 [34]. The test signal level is - 1.7 dBPa at the MRP, measured as active speech level according to ITU-T P.56 [37]. Three signals are required for the tests:
 - The clean speech signal is used as the undisturbed reference (see ETSI TS 103 106 [34], ETSI EG 202 396-3 [36]).
 - The speech plus undisturbed background noise signal is recorded at the terminal's microphone position using an omnidirectional measurement microphone with a linear frequency response between 50 Hz and 12 kHz.
 - The send signal is recorded at the POI.
- 3) N-MOS-LQOw, S-MOS-LQOw and G-MOS-LQOw are calculated as described in ETSI TS 103 106 [34] on a per sentence basis and averaged over all 16 sentences. The results shall be reported as average and standard deviation.
- 4) The measurement is repeated for each ambient noise condition described in Table 2h.
- 5) The average of the results derived from all ambient noise types is calculated.

8.12.2 Hand-held hands-free

For connections with hand-held hands-free UE, when using the simulation method described in TS 103 224 [43], the measurement is conducted for 5 noise conditions as described in Table 2h2. When using the ES 202 396-1 method, the equivalent binaurally recorded noises described in Table 2h2, and available in the source file directory of TS 103 224 [43], are used.

Table 2h2: Noise conditions used for ambient noise simulation in hand-held hands-free mode as specified in TS 103 224 [43], A-weighted

Name	Description	Length	Hands-free Levels	Binaural L	Binaural R
Full-size car 130 km/h (FullSizeCar_130)	HATS and microphone array at co-drivers position	30 s	1: 69,5 dB 2: 68,6 dB 3: 68,6 dB 4: 68,7 dB 5: 68,8 dB 6: 68,8 dB 7: 69,2 dB 8: 69,7 dB	68.7 dB	70.7 dB
Crossroadnoise (Crossroadnoise)	HATS and microphone array standing outside near a crossroad	30 s	1: 69,9 dB 2: 69,6 dB 3: 69,6 dB 4: 69,9 dB 5: 69,6 dB 6: 69,5 dB 7: 69,6 dB 8: 69,7 dB	70.8 dB	71.6 dB
Cafeteria (Cafeteria)	HATS and microphone array inside a cafeteria	30 s	1: 69,0 dB 2: 69,7 dB 3: 69,6 dB 4: 69,8 dB 5: 69,5 dB 6: 69,5 dB 7: 69,7 dB 8: 70,0 dB	69.8 dB	70.3 dB
Sales Counter (SalesCounter)	HATS and microphone array in a supermarket	30 s	1: 65,5 dB 2: 65,3 dB 3: 65,2 dB 4: 65,5 dB 5: 65,6 dB 6: 65,3 dB 7: 65,2 dB 8: 65,3 dB	66.7 dB	66.6 dB
Callcenter 2 (Callcenter)	HATS and microphone array in business office	30 s	1: 59,3 dB 2: 59,3 dB 3: 59,5 dB 4: 59,6 dB 5: 59,4 dB 6: 59,3 dB 7: 59,3 dB 8: 59,5 dB	60,2 dB	60,0 dB

- 1) Before starting the measurements a proper conditioning sequence shall be used. The conditioning sequence shall be comprised of the four additional sentences 1-4 described in ETSI TS 103 106 [34], applied to the beginning of the 16-sentence test sequence.

NOTE: The sequence of speech samples concatenated for the test signal, consisting of alternating talkers in the sending direction, reduces the overall test time but may represent an unrealistic behaviour for certain voice enhancement technologies. Alternative concatenations are for further study.

- 2) The send speech signal consists of the 16 sentences of speech as described in ETSI TS 103 106 [34]. For connections with a hand-free UE, the test signal level is +1.3dBPa at the MRP, measured as active speech level according to ITU-T P.56 [37]. Three signals are required for the tests:

- The clean speech signal is used as the undisturbed reference (see ETSI TS 103 106 [34], ETSI EG 202 396-3 [36]).
 - The speech plus undisturbed background noise signal is recorded at the terminal's microphone position using an omnidirectional measurement microphone with a linear frequency response between 50 Hz and 12 kHz.
 - The send signal is recorded at the POI.
- 3) N-MOS-LQOw, S-MOS-LQOw and G-MOS-LQOw are calculated as described in ETSI TS 103 106 [34] on a per sentence basis and averaged over all 16 sentences. The results shall be reported as average and standard deviation.
 - 4) The measurement is repeated for each ambient noise condition described in Table 2h2.
 - 5) The average of the results derived from all ambient noise types is calculated.

8.13 Jitter buffer management behaviour

8.13.0 General

For speech-only with LTE, NR or WLAN access, a jitter buffer is used in receiving to handle the variation in packet receiver timing. To minimize the additional latency introduced by the jitter buffer, adaptation is used to minimize delay while preventing packet losses due to packet delivery timing variations. See clause 8 of TS 26.114 [39] for the definition of jitter buffer and minimum performance requirements on JBM.

The test method is used to characterize different possible strategies and trade-offs in the design of JBM implementations used in MTSI terminals.

8.13.1 Delay histogram

For this test it shall be ensured that the call is originated from the mobile terminal (MO).

NOTE 1: Differences have been observed between mobile originated call and mobile terminated call. For better consistency MO calls are used.

The test signal consists of 3 repeats of the Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] followed by a speech signal of 160s. During the first two CSS signals the terminal can adapt its jitter buffer. The third CSS is used for measuring the delay in constant-delay condition, and the speech signal is used for delay and quality measurement in the packet impairment condition.

Constant delay T_c corresponding to the minimum delay of the profile (i.e. the compensation value for the profile) shall be added at the beginning of the different delay/loss profiles, to avoid unnecessary delay jumps between the two measurement phases and realistic conditions for the second measurement test phase.

In receiving direction, the delay between the electrical access point of the test equipment and the DRP, $T_{TEAP-DRP}(t) = T_{R-jitter}(t) + T_{TER}$, is measured in two successive phases:

- First the delay in constant-delay condition $T_{TEAP-DRP-constant}$ is measured as described in steps 1 to 4, clause 8.10.2, using the third CSS signal. The constant delay T_c is subtracted from $T_{TEAP-DRP}$ to obtain $T_{R-constant}$.
- Then the delay with packet impairment $T_{R-jitter}(t)$ is measured continuously for a speech signal during the inclusion of packet delay and loss profiles in the receiving direction RTP voice stream.

Packet impairments shall be applied between the reference client and system simulator eNodeB. Separate calls shall be established for each packet impairment condition.

The start of the delay profiles must be synchronized with the start of the downlink speech material reproduction (compensated by the delay between reproduction and the point of impairment insertion, i.e. the delay of the reference client) in order to ensure a repeatable application of impairments to the test speech signal. Tests shall be performed with DTX enabled in the reference client.

NOTE 2: RTP packet impairments representing packet delay variations and loss are specified in Annex F. Care must be taken that the system simulator uses a dedicated bearer with no buffering/scheduling of packets for transmission.

For the CSS signal repeated 3 times, the pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 32 k samples (with 48 kHz sampling rate). The test signal level is -16 dBm0 measured at the digital reference point or the equivalent analogue point.

For the speech signal, 8 English test sentences according to ITU-T P.501 Annex C.2.3, normalized to an active speech level of -16dBm0, are used (2 male, 2 female speakers). The sequences are concatenated in such a way that all sentences are centered within a 4.0s time window, which results in an overall duration of 32.0s. The sequences are repeated 5 times, resulting in a test file 160.0s long. The first 2 sentences are used for convergence of the UE jitter buffer manager and are discarded from the analysis. Equivalent implementations of the concatenation by repeating the test sentences in sequence may be used.

For the delay calculation with the speech signal, a cross-correlation with a rectangular window length of 4s, centered at each sentence of the stimulus file, is used. The process is repeated for each sample. For each cross correlation, the maximum of the envelope is obtained producing one delay value per sentence.

The UE delay in the receive direction, $T_{R-jitter}(t)$, is obtained by subtracting the delay introduced by the test equipment and the simulated transport network packet delay introduced by the delay and loss profile (as specified for the respective profile in Annex F) from the first electrical event at the electrical access point of the test equipment to the first bit of the corresponding speech frame at the system simulator antenna, T_{TER} , from the measured $T_{TEAP-DRP}(t)$.

The difference D_T between maximum receiving delay obtained with at least 5 individual calls (see clause 7.10.2) and the delay $T_{R-constant}$ measured for the CSS signal in constant delay condition is calculated. The quantity "Call-to-Call Variability Adjustment" (CCVA) = $\max(0, D_T)$ shall be added to the obtained delay for the speech signal $T_{R-jitter}(t)$.

The UE delay in the receiving direction shall be reported in the form of an histogram covering the range of measured CCVA-adjusted values ($T_{R-CCVA}(t) = T_{R-jitter}(t) + CCVA$) with a step of 20 ms. The following pseudo code provides an example implementation for the histogram:

```
lo=min(floor( $T_{R-CCVA}(t=1..40)/20$ )*20)
hi=max(ceil( $T_{R-CCVA}(t=1..40)/20$ )*20)
[n,x]=hist( $T_{R-CCVA}(t=1..40)$ ,lo:20:hi)
bar(x,n)
```

The T_{R-CCVA} values for all 40 sentences shall also be reported in the test report.

NOTE 3: The synchronization of the speech frame processing in the UE to the bits of the speech frames at the UE antenna may lead to a variability of up to 20 ms of the measured UE receive delay between different calls. This synchronization is attributed to the UE receiving delay according to the definition of the UE delay reference points. The effect of this possible call-to-call variation is taken into account with the $CCVA = \max(0, D_T)$ value.

8.13.2 Speech quality loss histogram

For the evaluation of speech quality loss in conditions with packet arrival time variations and packet loss, the speech test signal described in clause 7.13.1 shall be used. Two 48 kHz recordings are used to produce the speech quality loss metric:

- A recording obtained in jitter and error free conditions with the test signal described in clause 7.13.1 (reference condition)
- A recording obtained during the application of packet arrival time variations and packet loss as described in clause 7.13.1 (test condition)

The speech quality of the signal is estimated using the measurement algorithm described in ITU-T Recommendation P.863 [44]. Level pre-alignment to -26 dBov of recordings shall be used – see P.863.1 clause 10.2 [45].

NOTE: The setup for acoustical measurement described in P.863 [44] is used. P.863 needs the signal at DRP with diffuse-field equalization.

A score shall be computed for each 8s speech sentence pair. The MOS-LQO values for the reference and test conditions shall be reported in the form of an histogram covering the range of measured values with a step of 0.1 and the values for all 20 sentences pairs shall also be reported in the test report. The following pseudo code provides an example implementation for the histogram:

```
lo=min(floor(MOS-LQOtest condition(i=1...20)/0.1)*0.1)
hi=max(ceil(MOS-LQOtest condition(i=1...20)/0.1)*0.1)
[n,x]=hist(MOS-LQOtest condition(i=1...20),lo:0.1:hi)
bar(x,n)
```

The synchronization between stimuli and degraded condition shall be done by the test system before applying the P.863 algorithm on each sentence pair.

9 Super-wideband telephony transmission performance test methods

9.1 Applicability

The test methods in this clause shall apply when testing a UE that is used to provide super-wideband telephony, either as a stand-alone service, or as part of a multimedia service.

The application force used to apply the handset against the artificial ear shall be 8 ± 2 N. For the headset case, the application of the headset shall comply with ITU-T Recommendation P.57 [14].

9.2 Overall loss/loudness ratings

9.2.1 General

The SLR and RLR values for GSM, 3G, LTE, NR or WLAN networks apply up to the POI. However, the main determining factors are the characteristics of the UE, including the analogue to digital conversion (ADC) and digital to analogue conversion (DAC). In practice, it is convenient to specify loudness ratings to the Air Interface. For the normal case, where the GSM, 3G, LTE, NR or WLAN network introduce no additional loss between the Air Interface and the POI, the loudness ratings to the PSTN boundary (POI) will be the same as the loudness ratings measured at the Air Interface.

9.2.2 Connections with handset UE

9.2.2.1 Sending loudness rating (SLR)

The test method is the same as for wideband (see sub-clause 8.2.2.1).

9.2.2.2 Receiving loudness rating (RLR)

The test method is the same as for wideband (see sub-clause 8.2.2.2, observing the signal properties for super-wideband described in sub-clause 5.4).

9.2.2.3 Receiving loudness rating (RLR) in the presence of background noise

The test method is the same as for wideband (see sub-clause 8.2.2.3, observing the signal properties for super-wideband described in sub-clause 5.4).

9.2.3 Connections with desktop and vehicle-mounted hands-free UE

The description is the same as for wideband (see sub-clause 8.2.3).

9.2.3.1 Sending loudness rating (SLR)

The test method is the same as for wideband (see sub-clause 8.2.3.1).

9.2.3.2 Receiving loudness rating (RLR)

The test method is the same as for wideband (see sub-clause 8.2.3.2, observing the signal properties for super-wideband described in sub-clause 5.4).

9.2.4 Connections with hand-held hands-free UE

9.2.4.1 Sending loudness rating (SLR)

The test method is the same as for wideband (see sub-clause 8.2.4.1).

9.2.4.2 Receiving loudness rating (RLR)

The test method is the same as for wideband (see sub-clause 8.2.4.2, observing the signal properties for super-wideband described in sub-clause 5.4).

9.2.5 Connections with headset UE

The description is the same as for wideband (see sub-clause 8.2.5).

9.3 Idle channel noise (handset and headset UE)

For idle noise measurements in sending and receiving directions, care should be taken that only the noise is windowed out by the analysis and the result is not impaired by any remaining reverberation or by noise and/or interference from various other sources. Some examples are air-conducted or vibration-conducted noise from sources inside or outside the test chamber, disturbances from lights and regulators, mains supply induced noise including grounding issues, test system and system simulator inherent noise as well as radio interference from the UE to test equipment such as ear simulators, microphone amplifiers, etc.

9.3.1 Sending

The terminal should be configured to the test equipment as described in subclause 5.1.

The environment shall comply with the conditions described in subclause 6.1.

The noise level at the output of the SS is measured from 100 Hz to 16 kHz with A-weighting. The A-weighting filter is described in IEC 60651.

A test signal may have to be intermittently applied to prevent 'silent mode' operation of the MS. This is for further study.

The measured part of the noise shall be 170,667 ms (which equals 8192 samples in a 48 kHz sample rate test system). The spectral distribution of the noise is analyzed with an 8k FFT using windowing with $\leq 0,1$ dB leakage for non bin-centered signals. This can be achieved with a window function commonly known as a "flat top window". Within the specified frequency range, the FFT bin that has the highest level is searched for; the level of this bin is the maximum level of a single frequency disturbance.

To improve repeatability, the test sequence (optional activation followed by the noise level measurement) may be contiguously repeated one or more times.

The total noise powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

The single frequency maximum powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

9.3.2 Receiving

The terminal should be configured to the test equipment as described in subclause 5.1.

The environment shall comply with the conditions described in subclause 6.1.

A test signal may have to be intermittently applied to prevent 'silent mode' operation of the MS. This is for further study.

The noise shall be measured from 100 Hz to 20 kHz with A-weighting at the DRP with diffuse-field correction. The A-weighting filter is described in IEC 60651.

The measured part of the noise shall be 170,667 ms (which equals 8192 samples in a 48 kHz sample rate test system). The spectral distribution of the noise is analyzed with an 8k FFT using windowing with $\leq 0,1$ dB leakage for non bin-centered signals. This can be achieved with a window function commonly known as a "flat top window". Within the specified frequency range the FFT bin that has the highest level is searched for; the level of this bin is the maximum level of a single frequency disturbance.

To improve repeatability, the test sequence (optional activation followed by the noise level measurement) may be contiguously repeated one or more times.

The total noise powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

The single frequency maximum powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

9.4 Sensitivity/frequency characteristics

9.4.0 General

The sensitivity/frequency characteristics are checked against the tolerances at the center frequencies of the fractional octave bands used in the respective test after interpolating the tolerances according to TS 26.131.

9.4.1 Handset and headset UE sending

9.4.1.1 Handset UE sending

The headset case is similar to the handset one, except for the application force.

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be $-4,7$ dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence.
- b) The handset terminal is setup as described in clause 5. Measurements shall be made at both 1/3-octave and 1/12-octave intervals as given by the R.10 and R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 16 kHz inclusive. For the calculation, the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.
- c) The sensitivity is expressed in terms of dBV/Pa.

9.4.1.2 Headset UE sending

The headset case is similar to the handset one, except for the application force and measurement intervals (only 1/3-octave intervals are used).

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be $-4,7$ dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence.
- b) The handset terminal is setup as described in clause 5. Measurements shall be made at 1/3-octave intervals as given by the R.10 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 16 kHz inclusive. For the calculation, the averaged measured level at the electrical reference point for each frequency band is referred to the averaged test signal level measured in each frequency band at the MRP.
- c) The sensitivity is expressed in terms of dBV/Pa.

9.4.2 Handset and headset UE receiving

9.4.2.1 Handset UE receiving

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The handset terminal is setup as described in clause 5. Measurements shall be made at both 1/3-octave and 1/12-octave intervals as given by the R.10 and R.40 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 16 kHz inclusive. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The HATS is diffuse-field equalized. The sensitivity is expressed in terms of dBPa/V. Information about correction factors is available in subclause 5.1.4.

Optionally, the measurements may be repeated with 2 N and 13 N application force. For these test cases no normative values apply.

9.4.2.2 Headset UE receiving

The headset case is similar to the handset one, except for the measurement intervals (only 1/3-octave intervals are used).

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The handset terminal is setup as described in clause 5. Measurements shall be made at 1/3-octave intervals as given by the R.10 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 16 kHz inclusive. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The HATS is diffuse-field equalized. The sensitivity is expressed in terms of dBPa/V. Information about correction factors is available in subclause 5.1.4.

Optionally, the measurements may be repeated with 2 N and 13 N application force. For these test cases no normative values apply.

9.4.3 Desktop and vehicle-mounted hands-free UE sending

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be $-4,7$ dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence. The broadband signal level is then

adjusted to $-28,7$ dBPa at the HFRP or the HATS HFRP (as defined in ITU-T Recommendation P.581) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as references to determine the sending sensitivity S_{mJ} .

- b) The hands-free terminal is setup as described in clause 5. Measurements shall be made at 1/3-octave intervals as given by the R.10 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 16 kHz inclusive. For the calculation the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The sensitivity is expressed in terms of dBV/Pa.

9.4.4 Desktop and vehicle-mounted hands-free UE receiving

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. Measurements shall be made at 1/3-octave intervals as given by the R.10 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 16 kHz inclusive. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The sensitivity is expressed in terms of dBPa/V.

9.4.5 Hand-held hands-free UE sending

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The spectrum of the acoustic signal produced by the artificial mouth is calibrated under free-field conditions at the MRP. The test signal level shall be $-4,7$ dBPa measured at the MRP. The test signal level is calculated over the complete test signal sequence. The broadband signal level is then adjusted to $-28,3$ dBPa at the HFRP or the HATS HFRP (as defined in subclause 9.2.3.1) and the spectrum is not altered.

The spectrum at the MRP and the actual level at the MRP (measured in 1/3-octaves) are used as reference to determine the sending sensitivity S_{mJ} .

- b) The hands-free terminal is setup as described in clause 5.1.3.3. Measurements shall be made at 1/3-octave intervals as given by the R.10 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 16 kHz inclusive. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.
- c) The sensitivity is expressed in terms of dBV/Pa.

9.4.6 Hand-held hands-free UE receiving

- a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [22]. The test signal level shall be -16 dBm0 measured at the digital reference point or the equivalent analogue point. The test signal level is calculated over the complete test signal sequence.
- b) The hands-free terminal is setup as described in clause 5. If a HATS is used, then it is free-field equalized as described in ITU-T Recommendation P.581. The equalized output signal of each artificial ear is power-averaged over the total duration of the analysis; the right and left artificial ear signals are voltage-summed for each 1/3-octave band frequency band; these 1/3-octave band data are considered as the input signal to be used for calculations or measurements. Measurements shall be made at 1/3-octave intervals as given by the R.10 series of preferred numbers in ISO 3 for frequencies from 100 Hz to 16 kHz inclusive. For the calculation, the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

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

9.5 Sidetone characteristics

9.5.1 Connections with handset UE

The test method is the same as for wideband (see sub-clause 8.5.1).

9.5.2 Headset UE

The test method is the same as for wideband (see sub-clause 8.5.2).

9.5.3 Hands-free UE (all categories)

No requirement other than echo control.

9.5.4 Sidetone delay for handset or headset

The test method is the same as for wideband (see sub-clause 8.5.4).

9.6 Stability loss

Where a user-controlled volume control is provided it is set to maximum.

Handset UE: The handset is placed on a hard plane surface with the earpiece facing the surface.

Headset UE: The requirement applies for the closest possible position between microphone and headset receiver within the intended wearing position.

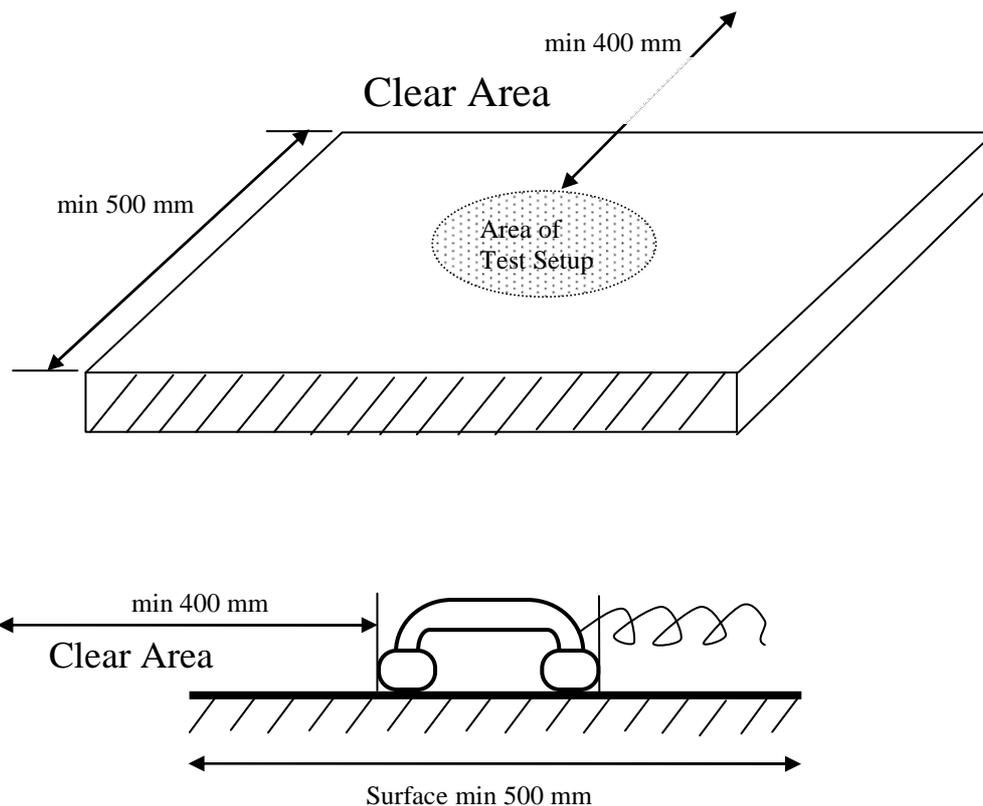
NOTE: Depending on the type of headset it may be necessary to repeat the measurement in different positions.

Hands-free UE (all categories): No requirement other than echo loss.

Before the actual test a training sequence consisting of the British-English single talk sequence described in ITU-T Recommendation P.501 [22] is applied. The training sequence level shall be -16 dBm₀ in order to not overload the codec.

The test signal is a PN-sequence complying with ITU-T Recommendation P.501 with a length of 4 096 points (for a 48 kHz sampling rate system) and a crest factor of 6 dB instead of 11 dB. The PN-sequence is generated as described in P.501 with $W(k)$ constant within the frequency range 100-16000 Hz and zero outside this range. The duration of the test signal is 250 ms. With an input signal of -3 dBm₀, the attenuation from input to output of the system simulator shall be measured under the following conditions:

- a) The handset or the headset, with the transmission circuit fully active, shall be positioned on a hard plane surface with at least 400 mm free space in all directions. The earpiece shall face towards the surface as shown in figure 20;
- b) The headset microphone is positioned as close as possible to the receiver(s) within the intended wearing position;
- c) For a binaural headset, the receivers are placed symmetrically around the microphone.



NOTE: All dimensions in mm.

Figure 20. Test configuration for stability loss measurement on handset or headset UE

The attenuation from input to output shall be measured in the frequency range from 100 Hz to 16 kHz. The spectral distribution of the output signal is analysed with a 4k FFT (for a 48 kHz sample rate test system), thus the measured part of the output signal is 85,333 ms. To avoid leakage effects the frequency resolution of the FFT must be the same as the frequency spacing of the PN-sequence.

9.7 Acoustic echo control

9.7.1 General

The echo loss (EL) presented by the GSM, 3G, LTE, NR or WLAN networks at the POI should be at least 46 dB during single talk. This value takes into account the fact that UE is likely to be used in a wide range of noise environments.

NOTE: A test method fully adapted to super-wideband acoustic echo control is for further study

9.7.2 Acoustic echo control in a hands-free UE

The hands-free UE is setup in a room with acoustic properties similar to a typical "office-type" room; a vehicle-mounted hands-free UE should be tested in a vehicle or vehicle simulator, as specified by the UE manufacturer (see also 3GPP TS 03.58 [11]). The ambient noise level shall be ≤ -70 dBPa(A). The attenuation from reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, clause B.4 (trapezoidal rule) but using the frequency range of 300 Hz to 6 700 Hz (instead of 300 Hz to 3 400 Hz). For the calculation, the averaged measured echo level at each frequency band is referred to the averaged test signal level measured in each frequency

band. The first 17,0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences).

The test signal level shall be -10 dBm0.

9.7.3 Acoustic echo control in a handset UE

The handset is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A). The attenuation from the reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, clause B.4 (trapezoidal rule) but using the frequency range of 300 Hz to 6 700 Hz (instead of 300 Hz to 3 400 Hz). For the calculation, the averaged measured echo level at each frequency band is referred to the averaged test signal level measured in each frequency band. The first 17,0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences).

The test signal level shall be -10 dBm0.

9.7.4 Acoustic echo control in a headset UE

The headset is set up according to clause 5. The ambient noise level shall be ≤ -64 dBPa(A). The attenuation from the reference point input to reference point output shall be measured using the compressed real speech signal described in clause 7.3.3 of ITU-T P.501 Amendment 1 [33].

The TCLw is calculated according to ITU-T Recommendation G.122 [8], annex B, clause B.4 (trapezoidal rule) but using the frequency range of 300 Hz to 6 700 Hz (instead of 300 Hz to 3 400 Hz). For the calculation, the averaged measured echo level at each frequency band is referred to the averaged test signal level measured in each frequency band. The first 17,0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences).

The test signal level shall be -10 dBm0.

9.8 Distortion

9.8.1 Sending distortion

The test method is the same as for wideband (see sub-clause 8.8.1).

9.8.2 Receiving

The test method is the same as for wideband (see sub-clause 8.8.2, observing the signal properties for super-wideband described in sub-clause 5.4).

9.9 Void

9.10 Delay

9.10.0 UE Delay Measurement Methodologies

The test method is the same as in wideband (see clause 8.10.0).

9.10.1 Delay in sending direction (handset UE)

The test method is the same as in wideband (see clause 8.10.1).

9.10.1a Delay in sending direction (headset UE)

The test method is the same as in wideband (see clause 8.10.1a).

9.10.2 Delay in receiving direction (handset UE)

The test method is the same as in wideband (see clause 8.10.2, observing the test signal properties defined for super-wideband described in clause 5.4).

9.10.2a Delay in receiving direction (headset UE)

The test method is the same as in wideband (see clause 8.10.2a, observing the test signal properties for super-wideband described in clause 5.4).

9.10.3 Delay in sending + receiving direction using "echo" method (handset UE)

The test method is the same as in wideband (see clause 8.10.3, observing the test signal properties for super-wideband described in clause 5.4).

9.10.3a Delay in sending + receiving direction using "echo" method (headset UE)

The test method is the same as in wideband (see clause 8.10.3a, observing the test signal properties for super-wideband in clause 5.4).

9.10.4 Delay and speech quality in conditions with packet arrival time variations and packet loss

9.10.4.1 Delay in sending direction

The test method is the same as in wideband (see clause 8.10.4.1).

9.10.4.2 Delay in receiving direction

For this test it shall be ensured that the call is originated from the mobile terminal (MO).

NOTE 1: Differences have been observed between mobile originated call and mobile terminated call. For better consistency MO calls are used.

The test signal consists of 3 repeats of the Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] followed by a speech signal of 160s. During the first two CSS signals the terminal can adapt its jitter buffer. The third CSS is used for measuring the delay in constant-delay condition, and the speech signal is used for delay and quality measurement in the packet impairment condition.

Constant delay T_c corresponding to the minimum delay of the profile (i.e. the compensation value for the profile) shall be added at the beginning of the different delay/loss profiles, to avoid unnecessary delay jumps between the two measurement phases and realistic conditions for the second measurement test phase.

In receiving direction, the delay between the electrical access point of the test equipment and the DRP, $T_{\text{TEAP-DRP}}(t) = T_{\text{R-jitter}}(t) + T_{\text{TER}}$, is measured in two successive phases:

- First the delay in constant-delay condition $T_{\text{TEAP-DRP-constant}}$ is measured as described in steps 1 to 4, clause 9.10.2, using the third CSS signal. The constant delay T_c is subtracted from $T_{\text{TEAP-DRP}}$ to obtain $T_{\text{R-constant}}$.
- Then the delay with packet impairment $T_{\text{R-jitter}}(t)$ is measured continuously for a speech signal during the inclusion of packet delay and loss profiles in the receiving direction RTP voice stream.

Packet impairments shall be applied between the reference client and system simulator eNodeB. Separate calls shall be

established for each packet impairment condition.

The start of the delay profiles must be synchronized with the start of the downlink speech material reproduction (compensated by the delay between reproduction and the point of impairment insertion, i.e. the delay of the reference client) in order to ensure a repeatable application of impairments to the test speech signal. Tests shall be performed with DTX enabled in the reference client.

NOTE 2: RTP packet impairments representing packet delay variations and loss in LTE transmission scenarios are specified in Annex E. These LTE jitter/loss profiles are reused also for tests with WLAN and NR access. Care must be taken that the system simulator uses a dedicated bearer with no buffering/scheduling of packets for transmission.

For the CSS signal repeated 3 times, the pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 32 k samples (with 48 kHz sampling rate). The test signal level is -16 dBm0 measured at the digital reference point or the equivalent analogue point.

For the speech signal, 8 English test sentences according to ITU-T P.501 Annex C.2.3, normalized to an active speech level of -16dBm0, are used (2 male, 2 female speakers). The sequences are concatenated in such a way that all sentences are centered within a 4.0s time window, which results in an overall duration of 32.0s. The sequences are repeated 5 times, resulting in a test file 160.0s long. The first 2 sentences are used for convergence of the UE jitter buffer manager and are discarded from the analysis. Equivalent implementations of the concatenation by repeating the test sentences in sequence may be used.

For the delay calculation with the speech signal, a cross-correlation with a rectangular window length of 4s, centered at each sentence of the stimulus file, is used. The process is repeated for each sample. For each cross correlation, the maximum of the envelope is obtained producing one delay value per sentence.

The UE delay in the receive direction, $T_{R-jitter}(t)$, is obtained by subtracting the delay introduced by the test equipment and the simulated transport network packet delay introduced by the delay and loss profile (as specified for the respective profile in Annex E) from the first electrical event at the electrical access point of the test equipment to the first bit of the corresponding speech frame at the system simulator antenna, T_{TER} , from the measured $T_{TEAP-DRP}(t)$.

The difference D_T between maximum receiving delay obtained with at least 5 individual calls (see clause 7.10.2) and the delay $T_{R-constant}$ measured for the CSS signal in constant delay condition is calculated. The quantity "Call-to-Call Variability Adjustment" ($CCVA = \max(0, D_T)$) shall be added to the obtained delay for the speech signal $T_{R-jitter}(t)$.

For stationary packet delay variation test conditions (test condition 1 and 2), the first 2 sentences are used for convergence of the jitter buffer management and are discarded from the analysis. The CCVA-adjusted UE delay ($T_{R-CCVA}(t) = T_{R-jitter}(t) + CCVA$) in the receiving direction shall be reported as the maximum value excluding the two largest values of the remaining sequence of the 38 sentence delay values, i.e. the 95-percentile value of $T_{R-CCVA}(t)$. The T_{R-CCVA} values for all 40 sentences shall be reported in the test report.

NOTE 3: The synchronization of the speech frame processing in the UE to the bits of the speech frames at the UE antenna may lead to a variability of up to 20 ms of the measured UE receive delay between different calls. This synchronization is attributed to the UE receiving delay according to the definition of the UE delay reference points. The effect of this possible call-to-call variation is taken into account with the $CCVA = \max(0, D_T)$ value.

9.10.4.3 Speech quality loss in conditions with packet arrival time variations and packet loss

The test method is the same as in wideband (see clause 8.10.4.3, observing the test signal properties for super-wideband described in clause 5.4).

9.10.5 UE send clock accuracy

The UE clock accuracy in send direction shall be measured according to Annex D.

NOTE1: For this specific measurement, care should be taken about the clock accuracy of the test equipment. See Table 1a.

NOTE2: As required in clause 5, prior to the actual measurements for MTSI-based speech with LTE, NR or WLAN access, the clocks of the reference client and the UE have to be synchronized. This measurement of UE send clock accuracy does not need to be repeated and can be obtained from this setup procedure.

9.10.6 UE receiving with clock skew

For further study.

9.11 Echo control characteristics

9.11.1 Test set-up and test signals

The test method is the same as for wideband (see sub-clause 8.11.1, observing the signal properties for super-wideband described in sub-clause 5.4).

9.11.2 Test method

The test method is the same as for wideband (see sub-clause 8.11.2, observing the signal properties for super-wideband described in sub-clause 5.4).

9.11.2.1 Signal alignment

The test method is the same as for wideband (see sub-clause 8.11.2.1).

9.11.2.2 Signal level computation and frame classification

The test method is the same as for wideband (see sub-clause 8.11.2.2).

9.11.2.3 Classification into categories

The test method is the same as for wideband (see sub-clause 8.11.2.3).

9.12 Quality (speech quality, noise intrusiveness) in the presence of ambient noise

9.12.1 Handset

The speech quality in sending for super-wideband systems is tested based on ETSI TS 103 281 [50]. This test method leads to three MOS-LQO_{fb} quality numbers:

- N-MOS-LQO_{fb}: Transmission quality of the background noise
- S-MOS-LQO_{fb}: Transmission quality of the speech
- G-MOS-LQO_{fb}: Overall transmission quality

The test arrangement is given in clause 5.1.5. For connections with handset UE, the measurement is conducted for 8 noise conditions as described in Table 2i. The measurements should be made in the same unique and dedicated call. The noise types shall be presented according to the order specified in Table 2i.

Table 2i: Noise conditions used for ambient noise simulation in handset mode as specified in ES 202 396-1 [35]

Description	File name	Duration	Level	Type
Recording in pub	Pub_Noise_binaural_V2	30 s	L: 75,0 dB(A) R: 73,0 dB(A)	Binaural
Recording at pavement	Outside_Traffic_Road_binaural	30 s	L: 74,9 dB(A) R: 73,9 dB(A)	Binaural
Recording at pavement	Outside_Traffic_Crossroads_binaural	20 s	L: 69,1 dB(A) R: 69,6 dB(A)	Binaural
Recording at departure platform	Train_Station_binaural	30 s	L: 68,2 dB(A) R: 69,8 dB(A)	Binaural
Recording at the drivers position	Fullsize_Car1_130Kmh_binaural	30 s	L: 69,1 dB(A) R: 68,1 dB(A)	Binaural
Recording at sales counter	Cafeteria_Noise_binaural	30 s	L: 68,4 dB(A) R: 67,3 dB(A)	Binaural
Recording in a cafeteria	Mensa_binaural	22 s	L: 63,4 dB(A) R: 61,9 dB(A)	Binaural
Recording in business office	Work_Noise_Office_Callcenter_binaural	30 s	L: 56,6 dB(A) R: 57,8 dB(A)	Binaural

- 1) Before starting the measurements, the calibration procedure described in clause 9.5 of ETSI TS 103 281 [50] shall be performed with the UE in handset mode. Also, a proper conditioning sequence shall be used. The conditioning sequence shall be comprised of the four additional sentences 1-4 described in ETSI TS 103 281 [50], applied to the beginning of the 16-sentence test sequence.

NOTE: The sequence of speech samples concatenated for the test signal, consisting of alternating talkers in the sending direction, reduces the overall test time but may represent an unrealistic behaviour for certain voice enhancement technologies. Alternative concatenations are for further study.

- 2) The send speech signal consists of the 16 sentences of speech as described in ETSI TS 103 281 [50] The test signal level is - 1.7 dBPa at the MRP, measured as active speech level per ITU-T P.56 [37]. Two signals are required for the tests:

- The clean speech signal is used as the undisturbed reference (see ETSI TS 103 281 [50])
- The send signal is recorded at the POI.

- 3) N-MOS-LQO_{fb}, S-MOS-LQO_{fb} and G-MOS-LQO_{fb} are calculated according to the Model A objective predictor described in ETSI TS 103 281[50] on a per sentence basis and averaged over all 16 sentences. The final results are derived as follows:

- $S\text{-MOS-LQO}_{fb} = S\text{-MOS-LQO}_{fb_modelA}$
- $N\text{-MOS-LQO}_{fb} = 1.438 * N\text{-MOS-LQO}_{fb_modelA} - 1.959$
- $G\text{-MOS-LQO}_{fb} = G\text{-MOS-LQO}_{fb_modelA}$

- 4) The measurement is repeated for each ambient noise condition described in Table 2i.

- 5) The average of the results derived from all ambient noise types is calculated.

NOTE: Recent investigations indicated an improved prediction performance when combining both models A and B. The usage of model B according to ETSI TS 103 281 [50] is for further study, pending a commercially available implementation.

9.12.2 Hand-held hands-free

For connections with hand-held hands-free UE, when using the simulation method described in TS 103 224 [43], the measurement is conducted for 5 noise conditions as described in Table 2i2. When using the ES 202 396-1 method, the equivalent binaurally recorded noises described in Table 2i2, and available in the source file directory of TS 103 224 [43], are used.

Table 2i2: Noise conditions used for ambient noise simulation in hand-held hands-free mode as specified in TS 103 224 [43], A-weighted

Name	Description	Length	Hands-free Levels	Binaural L	Binaural R
Full-size car 130 km/h (FullSizeCar_130)	HATS and microphone array at co-drivers position	30 s	1: 69,5 dB 2: 68,6 dB 3: 68,6 dB 4: 68,7 dB 5: 68,8 dB 6: 68,8 dB 7: 69,2 dB 8: 69,7 dB	68.7 dB	70.7 dB
Crossroadnoise (Crossroadnoise)	HATS and microphone array standing outside near a crossroad	30 s	1: 69,9 dB 2: 69,6 dB 3: 69,6 dB 4: 69,9 dB 5: 69,6 dB 6: 69,5 dB 7: 69,6 dB 8: 69,7 dB	70.8 dB	71.6 dB
Cafeteria (Cafeteria)	HATS and microphone array inside a cafeteria	30 s	1: 69,0 dB 2: 69,7 dB 3: 69,6 dB 4: 69,8 dB 5: 69,5 dB 6: 69,5 dB 7: 69,7 dB 8: 70,0 dB	69.8 dB	70.3 dB
Sales Counter (SalesCounter)	HATS and microphone array in a supermarket	30 s	1: 65,5 dB 2: 65,3 dB 3: 65,2 dB 4: 65,5 dB 5: 65,6 dB 6: 65,3 dB 7: 65,2 dB 8: 65,3 dB	66.7 dB	66.6 dB
Callcenter 2 (Callcenter)	HATS and microphone array in business office	30 s	1: 59,3 dB 2: 59,3 dB 3: 59,5 dB 4: 59,6 dB 5: 59,4 dB 6: 59,3 dB 7: 59,3 dB 8: 59,5 dB	60,2 dB	60,0 dB

- 1) Before starting the measurements, the calibration procedure described in clause 9.5 of ETSI TS 103 281 [50] shall be performed with the UE in hands-free mode. Also, a proper conditioning sequence shall be used. The conditioning sequence shall be comprised of the four additional sentences 1-4 described in ETSI TS 103 281 [50], applied to the beginning of the 16-sentence test sequence.

NOTE: The sequence of speech samples concatenated for the test signal, consisting of alternating talkers in the sending direction, reduces the overall test time but may represent an unrealistic behaviour for certain voice enhancement technologies. Alternative concatenations are for further study.

- 2) The send speech signal consists of the 16 sentences of speech as described in ETSI TS 103 281 [34]. For connections with a hand-free UE, the test signal level is +1.3dBPa at the MRP, measured as active speech level according to ITU-T P.56 [37]. Two signals are required for the tests:
 - The clean speech signal is used as the undisturbed reference (see ETSI TS 103 281 [50])
 - The send signal is recorded at the POI.
- 3) N-MOS-LQO_{fb}, S-MOS-LQO_{fb} and G-MOS-LQO_{fb} are calculated according to the Model A objective predictor described in ETSI TS 103 281 [50] on a per sentence basis and averaged over all 16 sentences. The final results are derived as follows:
 - S-MOS-LQO_{fb} = S-MOS-LQO_{fb_modelA}
 - N-MOS-LQO_{fb} = 1.438*N-MOS-LQO_{fb_modelA} - 1.959
 - G-MOS-LQO_{fb} = G-MOS-LQO_{fb_modelA}

- 4) The measurement is repeated for each ambient noise condition described in Table 2i2.
- 5) The average of the results derived from all ambient noise types is calculated.

NOTE: Recent investigations indicated an improved prediction performance when combining both models A and B. The usage of model B according to ETSI TS 103 281 [50] is for further study, pending a commercially available implementation.

9.13 Jitter buffer management behaviour

9.13.0 General

For MTSI-based speech-only with LTE, NR or WLAN access, a jitter buffer is used in receiving to handle the variation in packet receiver timing. To minimize the additional latency introduced by the jitter buffer, adaptation is used to minimize delay while preventing packet losses due to packet delivery timing variations. See clause 8 of TS 26.114 [39] for the definition of jitter buffer and minimum performance requirements on JBM.

The test method is used to characterize different possible strategies and trade-offs in the design of JBM implementations used in MTSI terminals.

9.13.1 Delay histogram

For this test it shall be ensured that the call is originated from the mobile terminal (MO).

NOTE 1: Differences have been observed between mobile originated call and mobile terminated call. For better consistency MO calls are used.

The test signal consists of 3 repeats of the Composite Source Signal (CSS) according to ITU-T Recommendation P.501 [22] followed by a speech signal of 160s. During the first two CSS signals the terminal can adapt its jitter buffer. The third CSS is used for measuring the delay in constant-delay condition, and the speech signal is used for delay and quality measurement in the packet impairment condition.

Constant delay T_c corresponding to the minimum delay of the profile (i.e. the compensation value for the profile) shall be added at the beginning of the different delay/loss profiles, to avoid unnecessary delay jumps between the two measurement phases and realistic conditions for the second measurement test phase.

In receiving direction, the delay between the electrical access point of the test equipment and the DRP, $T_{TEAP-DRP}(t) = T_{R-jitter}(t) + T_{TER}$, is measured in two successive phases:

- First the delay in constant-delay condition $T_{TEAP-DRP-constant}$ is measured as described in steps 1 to 4, clause 7.10.2, using the third CSS signal. The constant delay T_c is subtracted from $T_{TEAP-DRP}$ to obtain $T_{R-constant}$.
- Then the delay with packet impairment $T_{R-jitter}(t)$ is measured continuously for a speech signal during the inclusion of packet delay and loss profiles in the receiving direction RTP voice stream.

Packet impairments shall be applied between the reference client and system simulator eNodeB. Separate calls shall be established for each packet impairment condition.

The start of the delay profiles must be synchronized with the start of the downlink speech material reproduction (compensated by the delay between reproduction and the point of impairment insertion, i.e. the delay of the reference client) in order to ensure a repeatable application of impairments to the test speech signal. Tests shall be performed with DTX enabled in the reference client.

NOTE 2: RTP packet impairments representing packet delay variations and loss are specified in Annex F. Care must be taken that the system simulator uses a dedicated bearer with no buffering/scheduling of packets for transmission.

For the CSS signal repeated 3 times, the pseudo random noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 32 k samples (with 48 kHz sampling rate). The test signal level is -16 dBm0 measured at the digital reference point or the equivalent analogue point.

For the speech signal, 8 English test sentences according to ITU-T P.501 Annex C.2.3, normalized to an active speech level of -16dBm0, are used (2 male, 2 female speakers). The sequences are concatenated in such a way that all sentences are centred within a 4.0s time window, which results in an overall duration of 32.0s. The sequences are repeated 5 times, resulting in a test file 160.0s long. The first 2 sentences are used for convergence of the UE jitter buffer manager and are discarded from the analysis. Equivalent implementations of the concatenation by repeating the test sentences in sequence may be used.

For the delay calculation with the speech signal, a cross-correlation with a rectangular window length of 4s, centered at each sentence of the stimulus file, is used. The process is repeated for each sample. For each cross correlation, the maximum of the envelope is obtained producing one delay value per sentence.

The UE delay in the receive direction, $T_{R-jitter}(t)$, is obtained by subtracting the delay introduced by the test equipment and the simulated transport network packet delay introduced by the delay and loss profile (as specified for the respective profile in Annex F) from the first electrical event at the electrical access point of the test equipment to the first bit of the corresponding speech frame at the system simulator antenna, T_{TER} , from the measured $T_{TEAP-DRP}(t)$.

The difference D_T between maximum receiving delay obtained with at least 5 individual calls (see clause 7.10.2) and the delay $T_{R-constant}$ measured for the CSS signal in constant delay condition is calculated. The quantity "Call-to-Call Variability Adjustment" (CCVA) = $\max(0, D_T)$ shall be added to the obtained delay for the speech signal $T_{R-jitter}(t)$.

The UE delay in the receiving direction shall be reported in the form of an histogram covering the range of measured CCVA-adjusted values ($T_{R-CCVA}(t) = T_{R-jitter}(t) + CCVA$) with a step of 20 ms. The following pseudo code provides an example implementation for the histogram:

```
lo=min(floor( $T_{R-CCVA}(t=1..40)/20$ )*20)
hi=max(ceil( $T_{R-CCVA}(t=1..40)/20$ )*20)
[n,x]=hist( $T_{R-CCVA}(t=1..40)$ ,lo:20:hi)
bar(x,n)
```

The T_{R-CCVA} values for all 40 sentences shall also be reported in the test report.

NOTE 3: The synchronization of the speech frame processing in the UE to the bits of the speech frames at the UE antenna may lead to a variability of up to 20 ms of the measured UE receive delay between different calls. This synchronization is attributed to the UE receiving delay according to the definition of the UE delay reference points. The effect of this possible call-to-call variation is taken into account with the CCVA = $\max(0, D_T)$ value.

9.13.2 Speech quality loss histogram

For the evaluation of speech quality loss in conditions with packet arrival time variations and packet loss, the speech test signal described in clause 7.13.1 shall be used. Two 48 kHz recordings are used to produce the speech quality loss metric:

- A recording obtained in jitter and error free conditions with the test signal described in clause 7.13.1 (reference condition)
- A recording obtained during the application of packet arrival time variations and packet loss as described in clause 7.13.1 (test condition)

The speech quality of the signal is estimated using the measurement algorithm described in ITU-T Recommendation P.863 [44]. Level pre-alignment to -26 dBov of recordings shall be used – see P.863.1 clause 10.2 [45].

NOTE: The setup for acoustical measurement described in P.863 [44] is used. P.863 needs the signal at DRP with diffuse-field equalization.

A score shall be computed for each 8s speech sentence pair. The MOS-LQO values for the reference and test conditions shall be reported in the form of an histogram covering the range of measured values with a step of 0.1 and the values for all 20 sentences pairs shall also be reported in the test report. The following pseudo code provides an example implementation for the histogram:

```
lo=min(floor(MOS-LQOtest condition(i=1...20)/0.1)*0.1)
hi=max(ceil(MOS-LQOtest condition(i=1...20)/0.1)*0.1)
[n,x]=hist(MOS-LQOtest condition(i=1...20),lo:0.1:hi)
bar(x,n)
```

The synchronization between stimuli and degraded condition shall be done by the test system before applying the P.863 algorithm on each sentence pair.

10 Fullband telephony transmission performance test methods

10.1 Applicability

The test methods in this clause shall apply when testing a UE that is used to provide fullband telephony, either as a stand-alone service, or as part of a multimedia service.

The application force used to apply the handset against the artificial ear shall be 8 ± 2 N. For the headset case, the application of the headset shall comply with ITU-T Recommendation P.57 [14].

10.2 Overall loss/loudness ratings

10.2.1 General

The test method is the same as for super-wideband (see sub-clause 9.2.1).

10.2.2 Connections with handset UE

The description is the same as for super-wideband (see sub-clause 9.2.2).

10.2.2.1 Sending loudness rating (SLR)

The test method is the same as for super-wideband (see sub-clause 9.2.2.1).

10.2.2.2 Receiving loudness rating (RLR)

The test method is the same as for super-wideband (see sub-clause 9.2.2.2, observing the signal properties for fullband described in sub-clause 5.4).

10.2.2.3 Receiving loudness rating (RLR) in the presence of background noise

The test method is the same as for super-wideband (see sub-clause 9.2.2.3, observing the signal properties for fullband described in sub-clause 5.4).

10.2.3 Connections with desktop and vehicle-mounted hands-free UE

The description is the same as for super-wideband (see sub-clause 9.2.3).

10.2.3.1 Sending loudness rating (SLR)

The test method is the same as for super-wideband (see sub-clause 9.2.3.1).

10.2.3.2 Receiving loudness rating (RLR)

The test method is the same as for super-wideband (see sub-clause 9.2.3.2, observing the signal properties for fullband described in sub-clause 5.4).

10.2.4 Connections with hand-held hands-free UE

10.2.4.1 Sending loudness rating (SLR)

The test method is the same as for super-wideband (see sub-clause 9.2.4.1).

10.2.4.2 Receiving loudness rating (RLR)

The test method is the same as for super-wideband (see sub-clause 9.2.4.2, observing the signal properties for fullband described in sub-clause 5.4).

10.2.5 Connections with headset UE

The description is the same as for super-wideband (see sub-clause 9.2.5).

10.3 Idle channel noise (handset and headset UE)

The test method is the same as for super-wideband (see sub-clause 9.3).

10.3.1 Sending

The terminal should be configured to the test equipment as described in subclause 5.1.

The environment shall comply with the conditions described in subclause 6.1.

The noise level at the output of the SS is measured from 100 Hz to 20 kHz with A-weighting. The A-weighting filter is described in IEC 60651.

A test signal may have to be intermittently applied to prevent 'silent mode' operation of the MS. This is for further study.

The measured part of the noise shall be 170,667 ms (which equals 8192 samples in a 48 kHz sample rate test system). The spectral distribution of the noise is analyzed with an 8k FFT using windowing with $\leq 0,1$ dB leakage for non bin-centered signals. This can be achieved with a window function commonly known as a "flat top window". Within the specified frequency range, the FFT bin that has the highest level is searched for; the level of this bin is the maximum level of a single frequency disturbance.

To improve repeatability, the test sequence (optional activation followed by the noise level measurement) may be contiguously repeated one or more times.

The total noise powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

The single frequency maximum powers obtained from such repeats shall be averaged. The total result shall be $10 * \log_{10}$ of this average in dB.

10.3.2 Receiving

The test method is the same as for super-wideband (see sub-clause 9.3.2, observing the signal properties for fullband described in sub-clause 5.4).

10.4 Sensitivity/frequency characteristics

10.4.0 General

The sensitivity/frequency characteristics are checked against the tolerances at the center frequencies of the fractional octave bands used in the respective test after interpolating the tolerances according to TS 26.131.

10.4.1 Handset and headset UE sending

The test method is the same as for super-wideband (see sub-clause 9.4.1).

10.4.2 Handset and headset UE receiving

The test method is the same as for super-wideband (see sub-clause 9.4.2, observing the signal properties for fullband described in sub-clause 5.4).

10.4.3 Desktop and vehicle-mounted hands-free UE sending

The test method is the same as for super-wideband (see sub-clause 9.4.3).

10.4.4 Desktop and vehicle-mounted hands-free UE receiving

The test method is the same as for super-wideband (see sub-clause 9.4.4, observing the signal properties for fullband described in sub-clause 5.4).

10.4.5 Hand-held hands-free UE sending

The test method is the same as for super-wideband (see sub-clause 9.4.5, observing the signal properties for fullband described in sub-clause 5.4).

10.4.6 Hand-held hands-free UE receiving

The test method is the same as for super-wideband (see sub-clause 9.4.6, observing the signal properties for fullband described in sub-clause 5.4).

10.5 Sidetone characteristics

10.5.1 Connections with handset UE

The test method is the same as for super-wideband (see sub-clause 9.5.1).

10.5.2 Headset UE

The test method is the same as for super-wideband (see sub-clause 9.5.2).

10.5.3 Hands-free UE (all categories)

No requirement other than echo control.

10.5.4 Sidetone delay for handset or headset

The test method is the same as for super-wideband (see sub-clause 9.5.4).

10.6 Stability loss

Where a user-controlled volume control is provided it is set to maximum.

Handset UE: The handset is placed on a hard plane surface with the earpiece facing the surface.

Headset UE: The requirement applies for the closest possible position between microphone and headset receiver within the intended wearing position.

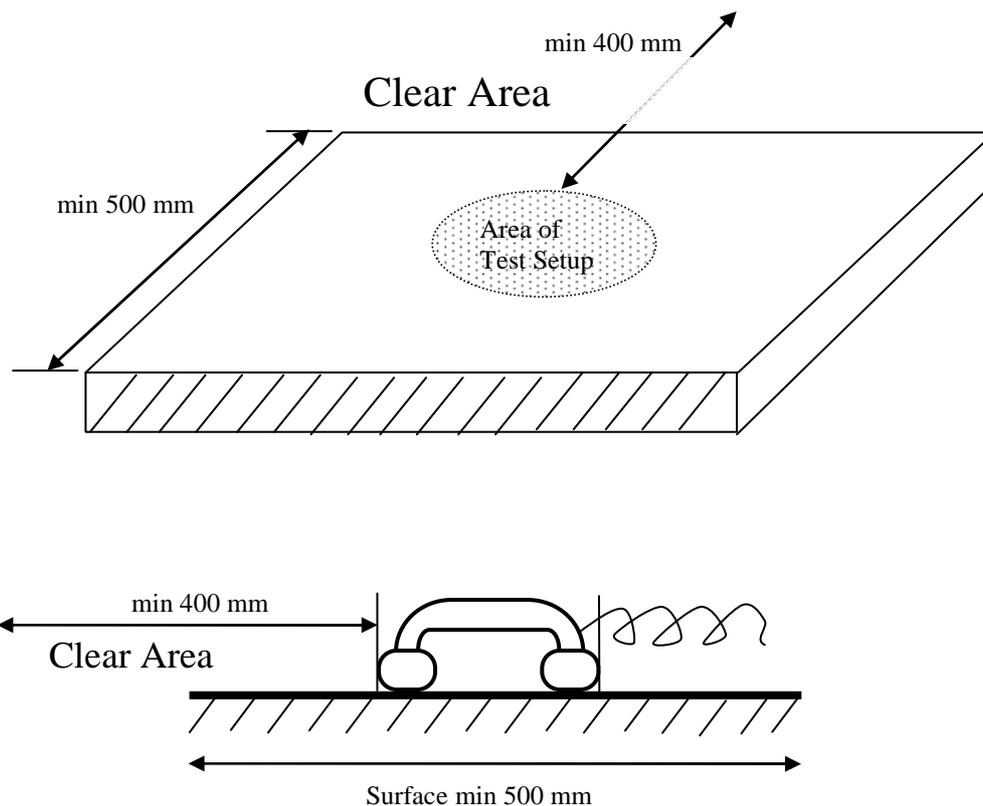
NOTE: Depending on the type of headset it may be necessary to repeat the measurement in different positions.

Hands-free UE (all categories): No requirement other than echo loss.

Before the actual test a training sequence consisting of the British-English single talk sequence described in ITU-T Recommendation P.501 [22] is applied. The training sequence level shall be -16 dBm0 in order to not overload the codec.

The test signal is a PN-sequence complying with ITU-T Recommendation P.501 with a length of 4 096 points (for a 48 kHz sampling rate system) and a crest factor of 6 dB instead of 11 dB. The PN-sequence is generated as described in P.501 with $W(k)$ constant within the frequency range 100-20000 Hz and zero outside this range. The duration of the test signal is 250 ms. With an input signal of -3 dBm0, the attenuation from input to output of the system simulator shall be measured under the following conditions:

- a) The handset or the headset, with the transmission circuit fully active, shall be positioned on a hard plane surface with at least 400 mm free space in all directions. The earpiece shall face towards the surface as shown in figure 21;
- b) The headset microphone is positioned as close as possible to the receiver(s) within the intended wearing position;
- c) For a binaural headset, the receivers are placed symmetrically around the microphone.



NOTE: All dimensions in mm.

Figure 21. Test configuration for stability loss measurement on handset or headset UE

The attenuation from input to output shall be measured in the frequency range from 100 Hz to 20 kHz. The spectral distribution of the output signal is analysed with a 4k FFT (for a 48 kHz sample rate test system), thus the measured part of the output signal is 85,333 ms. To avoid leakage effects the frequency resolution of the FFT must be the same as the frequency spacing of the PN-sequence.

10.7 Acoustic echo control

10.7.1 General

The description is the same as for super-wideband (see sub-clause 9.7.1).

10.7.2 Acoustic echo control in a hands-free UE

The test method is the same as for super-wideband (see sub-clause 9.7.2, observing the signal properties for fullband described in sub-clause 5.4).

10.7.3 Acoustic echo control in a handset UE

The test method is the same as for super-wideband (see sub-clause 9.7.3, observing the signal properties for fullband described in sub-clause 5.4).

10.7.4 Acoustic echo control in a headset UE

The test method is the same as for super-wideband (see sub-clause 9.7.4, observing the signal properties for fullband described in sub-clause 5.4).

10.8 Distortion

10.8.1 Sending distortion

The test method is the same as for super-wideband (see sub-clause 9.8.1).

10.8.2 Receiving

The test method is the same as for super-wideband (see sub-clause 9.8.2, observing the signal properties for fullband described in sub-clause 5.4).

10.9 Void

10.10 Delay

10.10.0 UE Delay Measurement Methodologies

The test method is the same as in super-wideband (see clause 9.10.0).

10.10.1 Delay in sending direction (handset UE)

The test method is the same as in super-wideband (see clause 9.10.1)..

10.10.1a Delay in sending direction (headset UE)

The test method is the same as in super-wideband (see clause 9.10.1a).

10.10.2 Delay in receiving direction (handset UE)

The test method is the same as in super-wideband (see clause 9.10.2, observing the test signal properties for fullband described in clause 5.4).

10.10.2a Delay in receiving direction (headset UE)

The test method is the same as in super-wideband (see clause 9.10.2a, observing the test signal properties for fullband described in clause 5.4).

10.10.3 Delay in sending + receiving direction using "echo" method (handset UE)

The test method is the same as in super-wideband (see clause 9.10.3, observing the test signal properties for fullband described in clause 5.4).

10.10.3a Delay in sending + receiving direction using "echo" method (headset UE)

The test method is the same as in super-wideband (see clause 9.10.3a, observing the test signal properties for fullband described in clause 5.4).

10.10.4 Delay and speech quality in conditions with packet arrival time variations and packet loss

10.10.4.1 Delay in sending direction

The test method is the same as in super-wideband (see clause 9.10.4.1).

10.10.4.2 Delay in receiving direction

The test method is the same as in super-wideband (see clause 9.10.4.2, observing the test signal properties for fullband described in clause 5.4).

10.10.4.3 Speech quality loss in conditions with packet arrival time variations and packet loss

For further study.

10.10.5 UE send clock accuracy

The test method is the same as in super-wideband (see clause 9.10.5, observing the test signal properties for fullband described in clause 5.4).

10.10.6 UE receiving with clock skew

For further study.

10.11 Echo control characteristics

10.11.1 Test set-up and test signals

The test method is the same as for super-wideband (see sub-clause 9.11.1, observing the signal properties for fullband described in sub-clause 5.4).

10.11.2 Test method

The test method is the same as for super-wideband (see sub-clause 9.11.2, observing the signal properties for fullband described in sub-clause 5.4).

10.11.2.1 Signal alignment

The test method is the same as for super-wideband (see sub-clause 9.11.2.1).

10.11.2.2 Signal level computation and frame classification

The test method is the same as for super-wideband (see sub-clause 9.11.2.2).

10.11.2.3 Classification into categories

The test method is the same as for super-wideband (see sub-clause 9.11.2.3).

10.12 Quality (speech quality, noise intrusiveness) in the presence of ambient noise

10.12.1 Handset

The test method is the same as in super-wideband (see sub-clause 9.12.1).

10.12.2 Hand-held hands-free

The test method is the same as in super-wideband (see sub-clause 9.12.2).

10.13 Jitter buffer management behaviour

The test method is the same as in super-wideband (see clause 9.13, observing the test signal properties for fullband described in clause 5.4).

Annex A (informative): Void

Annex B (informative): Reference algorithm for echo control characteristics evaluation.

B.1 General

In this annex, a reference algorithm for evaluation of the echo control characteristics is described in pseudo code. The output of an implementation of the test method with the stimuli from the file "echo_control_reference_files.zip" should equal the results presented in Table 3a and Table 3b. To run the verification, the additional file named "p501-downlink_WB.pcm" in the pseudo code shall be created from the concatenated full band speech samples FB_female_conditioning_seq_long.wav and FB_male_female_single-talk_seq.wav from ITU-T Recommendation P.501, and processed with the following set of commands based on ITU-T Recommendation G.191:

```
filter -down HQ3 far_end_signal_48k.pcm far_end_signal_16k.pcm
filter P341 far_end_signal_16k.pcm p501-downlink_WB.pcm
```

Table 3a: Characterization of segment 1.

Category	Double talk		Single talk	
	Activity	Av. Level [dB]	Activity	Av. Level [dB]
A1	60,8%	-1,2	95,1%	0,1
A2	39,2%	-5,1	1,4%	-4,8
B	0,0%	0	0,0%	0
C	0,0%	0	0,0%	0
D	0,0%	0	0,0%	0
E	0,0%	0	0,3%	9,4
F	0,0%	0	3,2%	8,7
G	0,0%	0	0,0%	0

Table 3b: Characterization of segment 2.

Category	Double talk		Single talk	
	Activity	Av. Level [dB]	Activity	Av. Level [dB]
A1	50.2%	-1.1	93,8%	0,2
A2	40.8%	-7.3	0,3%	-5.6
B	1.2%	-16,9	0,0%	0
C	7.1%	-17,2	0,0%	0
D	0,0%	0	0,0%	0
E	0,0%	0	0,5%	9,5
F	0,7%	4.0	5.5%	6,2
G	0,0%	0	0,0%	0

The pseudo-code reference algorithm produces a text file output, and the implementation of the test method may be tested with the test script on the data in the file "echo_control_reference_files.zip" for which the result shall equal

```
ms01-rec2; segm. 1; Processed signal;
active speech level [dBovl]; -45.8; RMS level [dBovl]; -51.5; speech activity; 0.269
ms01-rec2; segm. 1; Near end signal;
active speech level [dBovl]; -42.6; RMS level [dBovl]; -49.1; speech activity; 0.225
ms01-rec2; segm. 1; Downlink signal;
active speech level [dBovl]; -26.6; RMS level [dBovl]; -27.4; speech activity; 0.823
ms01-rec2; segm. 1; delay 0; DL delay 0;
DT activity 0.100; 0.608; 0.392; 0.000; 0.000; 0.000; 0.000; 0.000; 0.000;
ms01-rec2; segm. 1; delay 0; DL delay 0;
DT level diff; -1.2; -5.1; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0;
ms01-rec2; segm. 1; delay 0; DL delay 0;
ST activity 0.664; 0.951; 0.014; 0.000; 0.000; 0.000; 0.003; 0.032; 0.000;
ms01-rec2; segm. 1; delay 0; DL delay 0;
ST level diff; 0.1; -4.8; 0.0; 0.0; 0.0; 9.4; 8.7; 0.0;
ms01-rec2; segm. 2; Processed signal;
active speech level [dBovl]; -42.0; RMS level [dBovl]; -44.4; speech activity; 0.581
```

```

ms01-rec2; segm. 2; Near end signal;
active speech level [dBovl]; -40.6; RMS level [dBovl]; -42.7; speech activity; 0.625
ms01-rec2; segm. 2; Downlink signal;
active speech level [dBovl]; -26.5; RMS level [dBovl]; -27.2; speech activity; 0.841
ms01-rec2; segm. 2; delay -1; DL delay 0;
DT activity 0.348; 0.502; 0.408; 0.012; 0.071; 0.000; 0.000; 0.007; 0.000;
ms01-rec2; segm. 2; delay -1; DL delay 0;
DT level diff; -1.1; -7.3; -16.9; -17.2; 0.0; 0.0; 4.0; 0.0;
ms01-rec2; segm. 2; delay -1; DL delay 0;
ST activity 0.362; 0.938; 0.003; 0.000; 0.000; 0.000; 0.005; 0.055; 0.000;
ms01-rec2; segm. 2; delay -1; DL delay 0;
ST level diff; 0.2; -5.6; 0.0; 0.0; 0.0; 9.5; 6.2; 0.0;

```

B.2 Test script

```

%
% Set data format
%
fs = 16000;
conditioningTime = 23.5;
downlinkSystemDelay = 0;

%
% Segment the data
%
offsetDoubleTalk = conditioningTime;
offsetNearEnd = conditioningTime;

segmentDoubleTalkIndex(1) = {[0, 20]};
segmentNearEndIndex(1) = {[0, 20]};

segmentDoubleTalkIndex(2) = {[20, 35]};
segmentNearEndIndex(2) = {[20, 35]};

lengthDoubleTalk = max(cell2mat(segmentDoubleTalkIndex(end)));
lengthNearEnd = max(cell2mat(segmentNearEndIndex(end)));

firstSampleDoubleTalk = round(fs*offsetDoubleTalk) + 1;
firstSampleNearEnd = round(fs*offsetNearEnd) + 1;

lastSampleDoubleTalk = round(fs*(offsetDoubleTalk+lengthDoubleTalk));
lastSampleNearEnd = round(fs*(offsetNearEnd+lengthNearEnd));

indexDoubleTalk = [firstSampleDoubleTalk, lastSampleDoubleTalk];
indexNearEnd = [firstSampleNearEnd, lastSampleNearEnd];

%
% Read data from file
%
fid = fopen('ms01_WB_rec2.pcm', 'r');
fseek(fid, 2*round(fs*offsetDoubleTalk), 'bof');
processedData = fread(fid, round(fs*lengthDoubleTalk), 'int16');
fclose(fid);

fid = fopen('ms01_WB_ref.pcm', 'r');
fseek(fid, 2*round(fs*offsetNearEnd), 'bof');
nearendData = fread(fid, round(fs*lengthNearEnd), 'int16');
fclose(fid);

fid = fopen('p501-downlink_WB.pcm', 'r');
fseek(fid, 2*round(fs*offsetDoubleTalk), 'bof');
downlinkData = fread(fid, round(fs*lengthDoubleTalk), 'int16');
fclose(fid);

%
% Evaluate
%
ecEvaluation(processedData, nearendData, downlinkData, ...
    segmentDoubleTalkIndex, segmentNearEndIndex, ...
    'ms01-rec2', downlinkSystemDelay, ...
    fs, 'bitExactTest.txt');

```

B.3 Reference algorithm

B.3.1 Main algorithm

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% processedData:           processed samples
% originalData:           near-end-only samples
% downlinkData:           down-link (loudspeaker) samples
% processedSegmentSet:    set of indices to processed data segments
% originalSegmentSet:     set of indices to original data segments
% PROC_FILE:              name shown in diagrams
% downlinkSystemDelayInMs: delay in DL signal from data to acoustic out
% sampleRate:             sampling frequency of the data
% resultsFile:            output file
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function ecEvaluation(...
    processedData, ...
    nearendData, ...
    downlinkData, ...
    indexProcessed, ...
    indexNearend, ...
    PROC_FILE, ...
    downlinkSystemDelayInMs, ...
    sampleRate, ...
    resultFile)

fid = fopen(resultFile, 'a');

% Define the categories
global D1 D2 D3 D4 L1 L2 L3;
D1 = 25;
D2 = 150;
D3 = 25;
D4 = 150;
L1 = 4;
L2 = -4;
L3 = -15;

global FRAME_LENGTH_MS ...
    MAX_DURATION_MS ...
    MAX_DURATION_FRAMES ...
    MAX_LEVEL_DIFFERENCE ...
    MIN_LEVEL_DIFFERENCE ...
    HISTOGRAM_RESOLUTION_MS

FRAME_LENGTH_MS = 5;
MAX_DURATION_MS = 200;
MAX_DURATION_FRAMES = MAX_DURATION_MS/FRAME_LENGTH_MS;
MAX_LEVEL_DIFFERENCE = 40;
MIN_LEVEL_DIFFERENCE = -40;
HISTOGRAM_RESOLUTION_MS = FRAME_LENGTH_MS;

% Main processing loop
frameLengthInSamples = FRAME_LENGTH_MS*sampleRate/1000; % 5ms frames

for segment = 1:length(indexProcessed)
    % Get the data samples for the segment
    segmentDataProcessed = cell2mat(indexProcessed(segment));
    segmentDataNearend = cell2mat(indexNearend(segment));

    index = (sampleRate*segmentDataProcessed(1)+1):sampleRate*segmentDataProcessed(2);
    x = processedData(index);
    z = downlinkData(index);
    index = (sampleRate*segmentDataNearend(1)+1):sampleRate*segmentDataNearend(2);
    y = nearendData(index);

    % Estimate and compensate for delay between processed and near end
    [x, y, z, delay] = compensateDelay(x, y, z, 0.5*sampleRate);

```

```

% Compute the signal levels and classify the frames
[Rx, Ry, Rz, doubleTalkFrames, singleTalkFrames] = ...
computeSignalLevels(x, y, z, ...
    sampleRate, frameLengthInSamples, ...
    downlinkSystemDelayInMs, ...
    PROC_FILE, segment, fid);

% Evaluate double-talk performance
numberOfDoubleTalkFrames = 0;
% Iterate over blocks of consecutive indices
H_dt = [];
doubleTalkFramesBlocks = findConsecutiveBlocks(doubleTalkFrames);
for i = 1:size(doubleTalkFramesBlocks,1)
    IdxFrom = doubleTalkFramesBlocks(i,1);
    IdxTo = doubleTalkFramesBlocks(i,2);
    currentBlockLength = IdxTo - IdxFrom;
    if currentBlockLength > 1
        [H_dt_Tmp, ld_ax_dt, dur_ax_dt] = levelTimeStatistics(Rx(IdxFFrom:IdxTo), Ry(IdxFFrom:IdxTo));
        if isempty(H_dt)
            H_dt = H_dt_Tmp;
        else
            H_dt = H_dt + H_dt_Tmp;
        end
        numberOfDoubleTalkFrames = numberOfDoubleTalkFrames + currentBlockLength;
    end
end

[C_dt, L_dt] = evaluateHistogram(H_dt, ld_ax_dt, dur_ax_dt, ...
    numberOfDoubleTalkFrames);
activityFactorDoubleTalk = numberOfDoubleTalkFrames/length(Rx);

% Evaluate single-talk performance
numberOfSingleTalkFrames = 0;
% Iterate over blocks of consecutive indices
H_st = [];
singleTalkFramesBlocks = findConsecutiveBlocks(singleTalkFrames);
for i = 1:size(singleTalkFramesBlocks,1)
    IdxFrom = singleTalkFramesBlocks(i,1);
    IdxTo = singleTalkFramesBlocks(i,2);
    currentBlockLength = IdxTo - IdxFrom;
    if currentBlockLength > 1
        [H_st_Tmp, ld_ax_st, dur_ax_st] = levelTimeStatistics(Rx(IdxFFrom:IdxTo), Ry(IdxFFrom:IdxTo));
        if isempty(H_st)
            H_st = H_st_Tmp;
        else
            H_st = H_st + H_st_Tmp;
        end
        numberOfSingleTalkFrames = numberOfSingleTalkFrames + currentBlockLength;
    end
end

[C_st, L_st] = evaluateHistogram(H_st, ld_ax_st, dur_ax_st, ...
    numberOfSingleTalkFrames);
activityFactorSingleTalk = numberOfSingleTalkFrames/length(Rx);

% Save to result file
writeResultsToFile(fid, ...
    PROC_FILE, ...
    segment, ...
    delay, ...
    round(downlinkSystemDelayInMs), ...
    activityFactorDoubleTalk, ...
    activityFactorSingleTalk, ...
    C_dt, ...
    C_st, ...
    L_dt, ...
    L_st);

end

fclose(fid);

```

B.3.2 Delay compensation

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Compensate for delay in processed file
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [x, y, z, delay] = ...
compensateDelay(...
    x, ...
    y, ...
    z, ...
    maxLag)

ii = 1:min(1000000, length(x));

r = xcorr(x(ii), y(ii), maxLag);
[~, delay] = max(abs(r));
delay = delay-maxLag-1;

if (delay > 0)
    x = x((delay+1):end);
    z = z((delay+1):end);
    y = y(1:(end-delay));
elseif (delay < 0)
    y = y((-delay+1):end);
    x = x(1:(end+delay));
    z = z(1:(end+delay));
end;

```

B.3.3 Signal level computation and frame classification

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Determine speech activity and signal levels
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [Rx, Ry, Rz, doubleTalkFrames, singleTalkFrames] = ...
computeSignalLevels(x, y, z, ...
    sampleRate, frameLengthInSamples, ...
    downlinkSystemDelayInMs, ...
    PROC_FILE, segment, fid)

LEVEL_METER_INIT_TIME_MS = 100;
DOWNLINK_HANGOVER_FRAMES = 40;
NEAREND_HANGOVER_FRAMES = 40;

levelMeterInitTime = LEVEL_METER_INIT_TIME_MS*sampleRate/1000;

% Level according to IEC61672
Rx = IEC61672(x, sampleRate, 12.5);
Ry = IEC61672(y, sampleRate, 12.5);
Rz = IEC61672(z, sampleRate, 12.5);

% Correct for system delay
nRz = length(Rz);
minRz = min(Rz(levelMeterInitTime:end));
Rz = [minRz*ones(floor(downlinkSystemDelayInMs*sampleRate/1000), 1); Rz];
Rz = Rz(1:nRz);

% Sub-sample and avoid initialization period of level meter
Rx = Rx(levelMeterInitTime:frameLengthInSamples:end);
Ry = Ry(levelMeterInitTime:frameLengthInSamples:end);
Rz = Rz(levelMeterInitTime:frameLengthInSamples:end);

% Active speech level according to P.56
[activeSpeechLevelProcessed, ...
    longTermLevelProcessed, ...
    activityFactorProcessed] = ...
speechLevelMeter(x, sampleRate);

```

```

[activeSpeechLevelNearend, ...
 longTermLevelNearend, ...
 activityFactorNearend] = ...
speechLevelMeter(y, sampleRate);

[activeSpeechLevelDownlink, ...
 longTermLevelDownlink, ...
 activityFactorDownlink] = ...
speechLevelMeter(z, sampleRate);

% Write active speech levels to file
writeSpeechLevelsToFile(PROC_FILE, segment, fid, ...
 activeSpeechLevelProcessed, ...
 activeSpeechLevelNearend, ...
 activeSpeechLevelDownlink, ...
 longTermLevelProcessed, ...
 longTermLevelNearend, ...
 longTermLevelDownlink, ...
 activityFactorProcessed, ...
 activityFactorNearend, ...
 activityFactorDownlink);

%
% Only evaluate for active downlink/near-end speech including hang-over
%
activeRyFrames = find(Ry > activeSpeechLevelNearend-15.9);
activeRzFrames = find(Rz > activeSpeechLevelDownlink-15.9);

% Downlink with added hangover
activeDownlinkSpeechFrames = zeros(size(Rz));
activeDownlinkSpeechFrames(activeRzFrames) = ones(size(activeRzFrames));

activeDownlinkSpeechFrames = conv(activeDownlinkSpeechFrames, ...
 ones(DOWNLINK_HANGOVER_FRAMES, 1));
activeDownlinkSpeechFrames = activeDownlinkSpeechFrames(1:length(Rz));

% Near-end
activeNearEndSpeechFrames = zeros(size(Ry));
activeNearEndSpeechFrames(activeRyFrames) = ones(size(activeRyFrames));
activeNearEndSpeechHtFrames = conv(activeNearEndSpeechFrames, ...
 ones(NEAREND_HANGOVER_FRAMES, 1));
activeNearEndSpeechHtFrames = activeNearEndSpeechHtFrames(1:length(Rz));

% Only evaluate double talk when both rx+hangover and near-end
doubleTalkSpeechFrames = (activeDownlinkSpeechFrames & ...
 activeNearEndSpeechFrames);
doubleTalkFrames = find(doubleTalkSpeechFrames > 0);

% Single talk defined as rx and no near-end including 200 ms hangover
singleTalkSpeechFrames = (activeDownlinkSpeechFrames & ...
 ~activeNearEndSpeechHtFrames);
singleTalkFrames = find(singleTalkSpeechFrames > 0);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Average speech and noise levels
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [...
 activeSpeechLevel, ...
 longTermLevel, ...
 activityFactor ...
 ] = ...
speechLevelMeter(x, sampleRate)

SPEECH_LEVEL_HANGOVER_TIME_IN_MS = 200;

% Filter data
g = exp(-1/(0.03*sampleRate));
p = filter((1-g), [1, -g], abs(x));
q = filter((1-g), [1, -g], abs(p));

% Add 200ms hangover
hTimeInSamples = SPEECH_LEVEL_HANGOVER_TIME_IN_MS*sampleRate/1000;
qht = q;

```

```

for loop = 1:hTimeInSamples
    qht = max(qht, [zeros(loop, 1); q(1:end-loop)]);
end

% Compute cumulative histogram of signal power with hangover
nData = length(x);
cBins = 2.0.^(0:14)';
histogramCsum = zeros(size(cBins));

for loop = 1:length(cBins)
    histogramCsum(loop) = length(find(qht>cBins(loop)));
end

% Get the levels
sumSquare = sum(x.^2);
refdB = 20*log10(32768);

longTermLevel = 10*log10(sumSquare/nData) - refdB;
A = 10*log10(sumSquare./histogramCsum) - refdB;
C = 20*log10(cBins) - refdB;

Diff = A-C;
if ((A(1) == 0) || ((A(1) - C(1)) <= 15.9))
    activeSpeechLevel = -100;
else
    index = find(Diff <= 15.9, 1, 'first');

    if (Diff(index) == 15.9)
        activeSpeechLevel = A(index);
    else
        C_level = C(index) + ...
            (15.9 - Diff(index))* ...
            (C(index)-C(index-1))/(Diff(index)-Diff(index-1));
        activeSpeechLevel = A(index) + ...
            (C_level - C(index))* ...
            (A(index)-A(index-1))/(C(index)-C(index-1));
    end
end

activityFactor = 10.0^((longTermLevel-activeSpeechLevel)/10);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Speech level meter according to IEC61672
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function Rx = IEC61672(x, sampleRate, tc)
%
%
% This functions computes the power of a sampled signal
% using a discrete filter with time constant equivalent to a first order
% continous time exponential averaging circuit,
%
%

$$Rx = \frac{1/tc}{s + 1/tc} x^2$$

%
% according to IEC 61672 (1993, section 7.2).
%

T = 1/sampleRate;
tc = tc/1000;

%
% Design H by sampling of Hc
%
la = exp(-T/tc);
B = 1-la;
A = [1, -la];

Rx = filter(B, A, x.^2);

%
% Transform Rx to dBov (square wave),

```

```

%
% 0 dBov <=> power of maximum square wave signal, 32768
%
%      10^0 = 32768^2/X  => X = 32768^2
%
% Avoid log(0) by using log(max(eps, Rx))
%
Rx = 10*log10(max(eps, Rx)/32768/32768);

```

B.3.4 Level vs time computation

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Computation of level and time statistics
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [...
    levelVsDurationHistogram, ...
    levelDifferenceAxis, ...
    durationAxis] = ...
levelTimeStatistics(processedLevel, nearEndLevel)

global MAX_DURATION_FRAMES MAX_LEVEL_DIFFERENCE MIN_LEVEL_DIFFERENCE

FIRST_OCCURENCE = 1;

%
% Compute level difference
%
levelDifference = processedLevel - nearEndLevel;

%
% Only evaluate in integers (rounded towards 0) of dB and limit to max/min difference
%
levelDifference = fix(levelDifference);
levelDifference = min(levelDifference, MAX_LEVEL_DIFFERENCE);
levelDifference = max(levelDifference, MIN_LEVEL_DIFFERENCE);

%
% Produce axis
%
levelDifferenceAxis = MIN_LEVEL_DIFFERENCE:MAX_LEVEL_DIFFERENCE;
durationAxis = 1:(MAX_DURATION_FRAMES+1);

%
% Set initial values for computations and loop through all frames
%
numberOfEvaluatedFrames = length(levelDifference);

levelIncludedInEvaluation = (MAX_LEVEL_DIFFERENCE+1)*...
    ones(numberOfEvaluatedFrames, 1);
levelAndRunLength = zeros(numberOfEvaluatedFrames, 4);
levelVsDurationHistogram = zeros(MAX_LEVEL_DIFFERENCE+ ...
    (-MIN_LEVEL_DIFFERENCE)+1, ...
    MAX_DURATION_FRAMES+1);

previousLevelDifference = 0;

for frame = 1:numberOfEvaluatedFrames-1;
    currentLevelDifference = levelDifference(frame);

    %
    % Evaluate all levels from the previous level up to the current level
    %
    if currentLevelDifference <= 0
        firstEvaluatedLevelDifference = max(min(0, previousLevelDifference), ...
            currentLevelDifference);
        step = -1;
    else
        firstEvaluatedLevelDifference = min(max(0, previousLevelDifference), ...
            currentLevelDifference);
        step = 1;
    end
end

```

```

%
% Loop the levels to be evaluated
%
for evaluatedLevelDifference = ...
    firstEvaluatedLevelDifference:step:currentLevelDifference
    %
    % Check that the current frame is not already included
    % in evaluation for earlier frames
    %
    if (evaluatedLevelDifference ~= levelIncludedInEvaluation(frame))
        if (evaluatedLevelDifference > 0)
            duration = find(levelDifference(frame+1:end) < ...
                evaluatedLevelDifference, FIRST_OCCURENCE);
        else
            duration = find(levelDifference(frame+1:end) > ...
                evaluatedLevelDifference, FIRST_OCCURENCE);
        end

        if (isempty(duration))
            duration = numberOfEvaluatedFrames-frame+1;
        end

        %
        % Set the frames during duration of the level difference
        % as being evaluated
        %
        if (duration > 1)
            levelIncludedInEvaluation(frame:(frame+duration-1)) = ...
                evaluatedLevelDifference*ones(duration, 1);
        end;

        %
        % Add the number of frames in the duration that have
        % absolute level diff greater or equal to evaluatedLevel
        %
        durationIndex = min(duration, MAX_DURATION_FRAMES);
        levelIndex = evaluatedLevelDifference+(-MIN_LEVEL_DIFFERENCE)+1;
        levelVsDurationHistogram(levelIndex, durationIndex) = ...
            levelVsDurationHistogram(levelIndex, durationIndex) + duration;
    end
end

previousLevelDifference = currentLevelDifference;
end

```

B.3.5 Categorization

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Evaluate the histogram data
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [categories, averageLevelsInCategories] = ...
    evaluateHistogram(...
        histogramData, ...
        levelDiff_ax, ...
        duration_ax, ...
        numberOfFrames)

global D1 D2 D3 D4 L1 L2 L3 HISTOGRAM_RESOLUTION_MS;

D1_scaled = D1/HISTOGRAM_RESOLUTION_MS;
D2_scaled = D2/HISTOGRAM_RESOLUTION_MS;
D3_scaled = D3/HISTOGRAM_RESOLUTION_MS;
D4_scaled = D4/HISTOGRAM_RESOLUTION_MS;

levelIndex_L1 = find(levelDiff_ax == L1);
levelIndex_L2 = find(levelDiff_ax == L2);
levelIndex_L3 = find(levelDiff_ax == L3);

duration_A2 = duration_ax;
duration_B = duration_ax<=D1_scaled;

```

```

duration_C = (D1_scaled < duration_ax) & (duration_ax <= D2_scaled);
duration_D = duration_ax > D2_scaled;
duration_E = duration_ax <= D3_scaled;
duration_F = (D3_scaled < duration_ax) & (duration_ax <= D4_scaled);
duration_G = duration_ax > D4_scaled;

framesInCategoryB = sum(histogramData(levelIndex_L3, duration_B));
framesInCategoryC = sum(histogramData(levelIndex_L3, duration_C));
framesInCategoryD = sum(histogramData(levelIndex_L3, duration_D));
framesInCategoryE = sum(histogramData(levelIndex_L1, duration_E));
framesInCategoryF = sum(histogramData(levelIndex_L1, duration_F));
framesInCategoryG = sum(histogramData(levelIndex_L1, duration_G));

framesInCategoryA2 = sum(histogramData(levelIndex_L2, duration_A2));
framesInCategoryA2 = framesInCategoryA2 - ...
    framesInCategoryB - ...
    framesInCategoryC - ...
    framesInCategoryD;

framesInCategoryA1 = numberOfFrames - ...
    framesInCategoryA2 - ...
    framesInCategoryB - ...
    framesInCategoryC - ...
    framesInCategoryD - ...
    framesInCategoryE - ...
    framesInCategoryF - ...
    framesInCategoryG;

categories = [framesInCategoryA1;
    framesInCategoryA2;
    framesInCategoryB;
    framesInCategoryC;
    framesInCategoryD;
    framesInCategoryE;
    framesInCategoryF;
    framesInCategoryG]/numberOfFrames;

averageLevelsInCategories = zeros(8, 1);

% Category A1
index = levelDiff_ax < L1;
index = levelDiff_ax(index) > L2;
weight = levelDiff_ax(index);

duration = duration_ax;
levelTimesDuration = (weight*histogramData(index, duration)).*duration;
nData = sum(histogramData(index, duration)*duration');
if (framesInCategoryA1 > 0)
    averageLevelsInCategories(1) = sum(levelTimesDuration)/nData;
end

% Category A2
index = levelDiff_ax <= L2;
index = levelDiff_ax(index) > L3;
weight = levelDiff_ax(index);

duration = duration_ax;
levelTimesDuration = (weight*histogramData(index, duration)).*duration;
nData = sum(histogramData(index, duration)*duration');
if (framesInCategoryA2 > 0)
    averageLevelsInCategories(2) = sum(levelTimesDuration)/nData;
end

% Category B, C, D
index = find(levelDiff_ax <= L3);
weight = levelDiff_ax(index);

duration = duration_ax(duration_B);
levelTimesDuration = (weight*histogramData(index, duration_B)).*duration;
nData = sum(histogramData(index, duration_B)*duration');
if (framesInCategoryB > 0)
    averageLevelsInCategories(3) = sum(levelTimesDuration)/nData;
end

duration = duration_ax(duration_C);
levelTimesDuration = (weight*histogramData(index, duration_C)).*duration;
nData = sum(histogramData(index, duration_C)*duration');

```

```

if (framesInCategoryC > 0)
    averageLevelsInCategories(4) = sum(levelTimesDuration)/nData;
end

duration = duration_ax(duration_D);
levelTimesDuration = (weight*histogramData(index, duration_D)).*duration;
nData = sum(histogramData(index, duration_D)*duration');
if (framesInCategoryD > 0)
    averageLevelsInCategories(5) = sum(levelTimesDuration)/nData;
end

% Category E, F, G
index = find(levelDiff_ax >= L1);
weight = levelDiff_ax(index);

duration = duration_ax(duration_E);
levelTimesDuration = (weight*histogramData(index, duration_E)).*duration;
nData = sum(histogramData(index, duration_E)*duration');
if (framesInCategoryE > 0)
    averageLevelsInCategories(6) = sum(levelTimesDuration)/nData;
end

duration = duration_ax(duration_F);
levelTimesDuration = (weight*histogramData(index, duration_F)).*duration;
nData = sum(histogramData(index, duration_F)*duration');
if (framesInCategoryF > 0)
    averageLevelsInCategories(7) = sum(levelTimesDuration)/nData;
end

duration = duration_ax(duration_G);
levelTimesDuration = (weight*histogramData(index, duration_G)).*duration;
nData = sum(histogramData(index, duration_G)*duration');
if (framesInCategoryC > 0)
    averageLevelsInCategories(8) = sum(levelTimesDuration)/nData;
end

```

B.3.6 Auxiliary functions for reporting data

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Write the classification to file
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function writeResultsToFile(fid, ...
    PROC_FILE, ...
    segment, ...
    delay, ...
    downlinkSystemDelay, ...
    activityFactorDoubleTalk, ...
    activityFactorSingleTalk, ...
    C_dt, ...
    C_st, ...
    L_dt, ...
    L_st)

str = sprintf('%s; segm. %d; delay %d; DL delay %d; DT activity %1.3f; %1.3f; %1.3f; %1.3f; %1.3f; %1.3f; %1.3f; %1.3f;', ...
    PROC_FILE, segment, delay, downlinkSystemDelay, activityFactorDoubleTalk, ...
    C_dt(1), C_dt(2), C_dt(3), C_dt(4), ...
    C_dt(5), C_dt(6), C_dt(7), C_dt(8));
disp(str);
if (fid > -1)
    fprintf(fid, [str, '\n']);
end;

str = sprintf('%s; segm. %d; delay %d; DL delay %d; DT level diff; %1.1f; %1.1f; %1.1f; %1.1f; %1.1f; %1.1f; %1.1f; %1.1f;', ...
    PROC_FILE, segment, delay, downlinkSystemDelay, ...
    L_dt(1), L_dt(2), L_dt(3), L_dt(4), L_dt(5), L_dt(6), L_dt(7), L_dt(8));
disp(str);
if (fid > -1)
    fprintf(fid, [str, '\n']);
end;

```

```

str = sprintf('%s; segm. %d; delay %d; DL delay %d; ST activity %1.3f; %1.3f; %1.3f; %1.3f; %1.3f;
%1.3f; %1.3f; %1.3f; %1.3f;', ...
    PROC_FILE, segment, delay, downlinkSystemDelay, activityFactorSingleTalk, ...
    C_st(1), C_st(2), C_st(3), C_st(4), ...
    C_st(5), C_st(6), C_st(7), C_st(8));
disp(str);
if (fid > -1)
    fprintf(fid, [str, '\n']);
end;

str = sprintf('%s; segm. %d; delay %d; DL delay %d; ST level diff; %1.1f; %1.1f; %1.1f; %1.1f;
%1.1f; %1.1f; %1.1f; %1.1f;', ...
    PROC_FILE, segment, delay, downlinkSystemDelay, ...
    L_st(1), L_st(2), L_st(3), L_st(4), L_st(5), L_st(6), L_st(7), L_st(8));
disp(str);
if (fid > -1)
    fprintf(fid, [str, '\n']);
end;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Write the signal levels to file
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function writesSpeechLevelsToFile(PROC_FILE, segment, fid, ...
    activeSpeechLevelProcessed, ...
    activeSpeechLevelNearend, ...
    activeSpeechLevelDownlink, ...
    longTermLevelProcessed, ...
    longTermLevelNearend, ...
    longTermLevelDownlink, ...
    activityFactorProcessed, ...
    activityFactorNearend, ...
    activityFactorDownlink)

str = sprintf('%s; segm. %d; Processed signal; active speech level [dBovl]; %3.1f; RMS level
[dBovl]; %3.1f; speech activity; %1.3f', ...
    PROC_FILE, segment, activeSpeechLevelProcessed, ...
    longTermLevelProcessed, activityFactorProcessed);
disp(str);
if (fid > -1)
    fprintf(fid, [str, '\n']);
end;

str = sprintf('%s; segm. %d; Near end signal; active speech level [dBovl]; %3.1f; RMS level [dBovl];
%3.1f; speech activity; %1.3f', ...
    PROC_FILE, segment, activeSpeechLevelNearend, ...
    longTermLevelNearend, activityFactorNearend);
disp(str);
if (fid > -1)
    fprintf(fid, [str, '\n']);
end;

str = sprintf('%s; segm. %d; Downlink signal; active speech level [dBovl]; %3.1f; RMS level [dBovl];
%3.1f; speech activity; %1.3f', ...
    PROC_FILE, segment, activeSpeechLevelDownlink, ...
    longTermLevelDownlink, activityFactorDownlink);
disp(str);
if (fid > -1)
    fprintf(fid, [str, '\n']);
end;

```

B.3.7 Other helper functions

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Find & separate blocks with consecutive indices
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
function [ConsecutiveBlocks] = findConsecutiveBlocks(FrameIndices)
```

```
D = diff(FrameIndices);
```

```
Changes = find(D > 1);
```

```
ConsecutiveBlocks = zeros(length(Changes)+1,2);
```

```
ConsecutiveBlocks(1,1) = FrameIndices(1);
```

```
for i = 1:length(Changes)
```

```
    ConsecutiveBlocks(i,2) = FrameIndices(Changes(i));
```

```
    if i <= length(Changes)
```

```
        ConsecutiveBlocks(i+1,1) = FrameIndices(Changes(i)+1);
```

```
    end
```

```
end
```

```
if ConsecutiveBlocks(end,2) == 0
```

```
    ConsecutiveBlocks(end,2) = FrameIndices(end);
```

```
end
```

Annex C (informative): Measurement method for determining the one way radio delays of LTE radio network simulators

The method described in this Annex can be used to determine or verify the delay introduced by a LTE radio network simulator.

NOTE. There is an inherent uncertainty in the method due to the unknown delay of the modem delay (software stack in the modem). If this delay is known the measurement uncertainty can be reduced.

C.1 Measurement setup

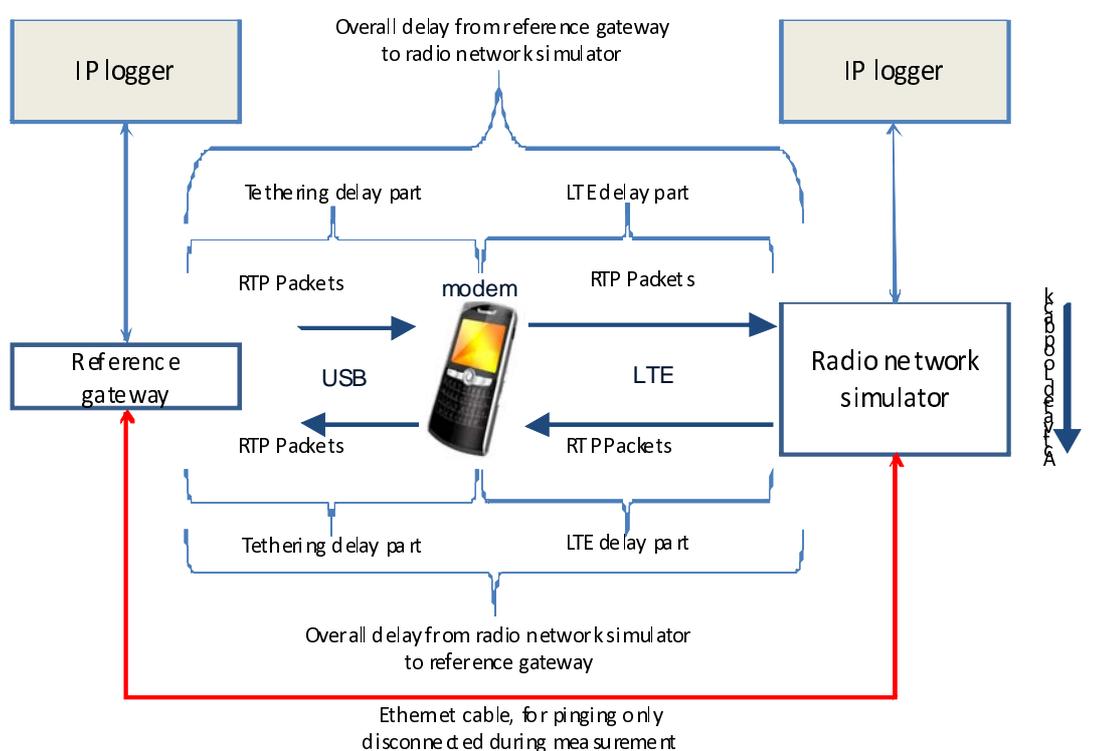


Figure C1: Measurement setup

Annex D (normative): Clock skew measurement

This Annex describes a method to measure the clock skew between the reference client and the device under test.

D.1 Test procedure

As speech test signal, the second sentence of the first female speaker (female1.wav) of the English test sentences according to ITU-T P.501 is used. When measuring in receiving direction the signal is pre-filtered according to the used bandwidth and normalized to an active speech level of -16dBm0. When measuring in sending direction the signal is normalized to an active speech level of -4.7 dBPa. The sequence is centered within a 4.0s time window and repeated 40 times, resulting in a test file of 160.0s length. Alternatively CSS signals may be used with the corresponding adaptation of the step size to 333ms. The test signal used shall be reported.

NOTE– Care should be taken to apply the test procedure to the receive direction, since the measurement could lead to biased results due to jitter buffer adaptation of the DUT during the test.

The clock skew is determined as follows:

- The variable delay per time step is calculated as the time shift of the maximum of the cross correlation between measured and source signal for each time step.
- The step size of the calculation of the variable delay is 4.0s for the above test signal, which corresponds to one single sentence of the source signal.
- This analysis provides the output signal $Y(tn)$ consisting of one single delay value for each time step, i.e., for each sentence of the above source signal.
- A median filter of length 3 is applied to the delay vs. time series $Y(tn)$ yielding the time series $\hat{Y}(tn)$. This smoothes small spikes without soften the delay jumps.
- Delay jumps due to jitter buffer over- or underrun are detected as the time steps where the absolute delay difference $\text{abs}(\hat{Y}(tn+1) - \hat{Y}(tn))$ is larger than a threshold ΔY_{max} , which is appropriate to detect a maximum clock skew of 500 PPM.
- A "continuous" delay vs. time series is generated where the delay jumps are eliminated, e.g., by cutting out the delay jumps and "stitching" together the jump-free segments.
- The clock skew is estimated as the regression coefficient β of the "continuous" delay vs. time series in parts per million (PPM):

$$D = \frac{\beta}{\text{step size}} \cdot 10^6 \text{ PPM.}$$

In case of clock skews higher than 50 PPM the accuracy of the cross-correlation calculation decreases. In such event it is recommended to take a two-step approach. In the first step a coarse skew compensation is performed and the clock of the reference client is adjusted to the PPM derived in the first step. A fine compensation is performed in the second step by applying the same procedure. The final clock skew in PPM is the sum of both clock skews.

D.2 Clock skew estimation algorithm

```
function Skew = EstimatePPM(DelayVsTime, StepSize, DiffMaxPPM=500)

% Difference threshold from which on a difference is assumed to be a packet
% loss or insertion. With larger step sizes, the normal clock skew
% results in higher differences between steps and therefore requires a
% larger threshold.
DelayDiffThreshold = min( max( DiffMaxPPM*1e-6*StepSize, 1e-3 ), 4e-3 );

% Calculate median filter of length 3 to smooth away small spikes without
```

```
% soften the steps.
DelayMedian = DelayVsTime(:);
for i = 2:numel(DelayVsTime)-1
    DelayMedian(i) = median(DelayVsTime(i-1:i+1));
end

% Calculate delay difference per time step.
DelayDiff = diff( DelayMedian );

% Exclude all indices with high delay differences as well as some
% transition range before and after.
Mask = true(size(DelayDiff));
Jumps = find( abs(DelayDiff) > DelayDiffThreshold );
for i = 1:length(Jumps)
    Mask(Jumps(i)-1:Jumps(i)+1) = false;
end

% Generate "continuous" delay vs. time series
Delay = [ 0; cumsum( DelayDiff(Mask) ) ];
Time = ( 0:numel(Delay)-1)';

% Estimate clock skew in PPM as slope of linear regression line
P = polyfit( Time, Delay, 1 );
Skew = P(1) / StepSize * 1e6;
```

Annex E (normative): Packet delay and loss profiles for UE delay testing of MTSI-based speech with LTE access

E.1 General

This Annex provides packet delay and loss profiles to be inserted on the IP packets in the test system when testing the UE delay and speech quality for MTSI-based speech with LTE access in jitter and error conditions. The profiles are based on the assumption that one IP packet corresponds to 20ms of the speech sequence. In order to preserve the synchronization of the elements in the profile and the timing of the speech sequence (e.g. the RTP time stamp) when the speech codec does not produce one packet every 20ms, e.g. during speech codec DTX operation, the profiles need to be sub-sampled in accordance with the rate of production of the packets containing the speech frames.

The variation in the packet arrival time and the loss rate of the speech packets experienced by the receiving UE in an LTE connection will vary depending on several parameters and operating conditions of the LTE network. Standardized characteristics with respect to Packet Delay Budget (PDB) and Packet Error Loss Rate (PELR) for different QoS Class Identifiers (QCI) for an LTE access network are defined in 3GPP TS 23.203 [42]:

- The PDB of a QCI defines an upper bound for the time that a packet may be delayed between the UE and the PCEF. For a certain QCI the value of the PDB is the same in uplink and downlink. The purpose of the PDB is to support the configuration of scheduling and link layer functions (e.g. the setting of scheduling priority weights and HARQ target operating points). The PDB shall be interpreted as a maximum delay with a confidence level of 98 percent.
- The PELR of a QCI defines an upper bound for the rate of SDUs (e.g. IP packets) that have been processed by the sender of a link layer protocol (e.g. RLC in E-UTRAN) but that are not successfully delivered by the corresponding receiver to the upper layer (e.g. PDCP in E-UTRAN). Thus, the PELR defines an upper bound for a rate of non congestion related packet losses. The purpose of the PELR is to allow for appropriate link layer protocol configurations (e.g. RLC and HARQ in E-UTRAN). For a certain QCI the value of the PELR is the same in uplink and downlink.

The stipulated PDB and PELR as defined in 3GPP TS 23.203 for QCI1 intended for Conversational Voice services results in that on each link 99 percent of the packets will be successfully delivered over an LTE air interface with a maximum delay of less than 80ms with a confidence level of 98 percent.

E.2 Simulation model for generating packed delay and loss profiles

This clause describes a simulation model that generates packet arrival time variations and packet loss experienced by the receiving UE for MTSI-based speech with an end-to-end LTE access.

The model is derived solely for the purpose of testing the UE delay for MTSI-based speech with LTE access. As discussed in Section E.1, several LTE network parameters have a significant impact on the packet delay variations experienced by the UE. This model includes the effect of the DRX cycle, the BLER on the MAC/PHY layer, the HARQ re-transmission and RMC scheduler [46]. In addition, jitter in the EPC and the effect of the mis-alignment of the DRX-cycle between the uplink and downlink eNBs are included.

The model is described in Table E.1 and operates on the following input parameters

- Uplink and downlink block error rate (BLER), respectively.
- Maximum number of HARQ re-transmissions on uplink and downlink, respectively.
- DRX cycle length.
- Time difference between the uplink and the downlink eNB DRX cycle.

- Maximum and minimum network delay between the uplink and the downlink eNB.

The random number generator used in the model produces pseudorandom values drawn from the standard uniform distribution on the open interval (0, 1).

The model generates packet delay and loss profiles for two different test system configurations of the UE delay:

- When the system simulator is transparent in the downlink at the MAC/PHY layer and does not operate in DRX. This approach requires the variations of the delay due to downlink HARQ re-transmissions and mapping to the DRX cycle to be simulated. In this case, the model simulates the delay and jitter profiles for the packets from the antenna of the sending UE to the antenna receiving UE (end-to-end simulation). This is the testing condition used in 3GPP TS 26.132.
- When the system simulator is configured for error insertion in the downlink at the MAC/PHY layer and DRX operation, and the system simulator implements the HARQ re-transmissions and the mapping of the packets to the DRX cycle time at the downlink. In this case, the model simulates the delay and loss profiles from the sending UE up to the receiving eNB. This testing condition is currently not used in 3GPP TS 26.132.

It should be noted that the model does not fully utilize the PDB for QCI1 as defined in 3GPP TS 23.203 and does not include temporary variations of the packet arrival time variation and the loss rate that may be experienced during e.g. hand-over or congestion. Hence, the packet delay and loss profiles generated by the model do not fully exercise the conditions that the jitter buffer management of the UE may be exposed to in LTE systems and the profiles generated by the model are only intended for the testing of the UE delay in stationary operating conditions.

Table E.1: Simulation model for generating packet delay and loss profiles for MTSI-based speech with LTE access

```
function [UE1_UE2_dly,UE1_eNB2_dly,plr,comp_e2e,comp] = ...
    VoLTEDelayProfile_vPHY(BLER_tx, BLER_rx, max_tx, max_rx, drx_cycle_length, mis_eNB1_eNB2,
max_net_delay, min_net_delay, nFrames, seed)

% BLER_tx           : The block error rate in uplink.
% BLER_rx           : The block error rate in downlink.
% max_tx            : The maximum number of transmission attempts in uplink.
% max_rx            : The maximum number of transmission attempts in downlink.
% drx_cycle_length  : The length of the DRX cycle
% mis_eNB1_eNB2     : Scheduling time mis-alignment between eNB1 and eNB2
% max_net_delay     : The maximum network delay between eNB1 to eNB2
% min_net_delay     : The minimum network delay between eNB1 to eNB2
% nFrames           : The number of frames for the simulation
% seed              : Random number generator seed
rng(seed);
UE1_UE2_time = zeros(nFrames, 1);
UE1_eNB2_time = zeros(nFrames,1);
eNB1_eNB2_dly = round(min_net_delay + (max_net_delay-min_net_delay).*rand(nFrames,1));
ack1 = zeros(nFrames,1);
ack2 = zeros(nFrames,1);

wall_clock = 20;
frame = 1;
frame_size = 20;
simulationTime = nFrames*frame_size;
% Calculate the delay from UE1 speech encoder delivery to eNB2. If
% transmission to eNB1 is not successful after max_tx attempts, dly = 0 (packet loss)
while (wall_clock<=simulationTime)
    % Set the scheduling time
    if drx_cycle_length == 0
        UE1_scheduling_time=wall_clock;
    else
        UE1_scheduling_time=ceil(wall_clock/drx_cycle_length)*drx_cycle_length;
    end

    % Add the tx effect for the scheduling time
    n=0;
    eNB1_receive_delay = 0;
    while n < max_tx
        if (rand(1) < BLER_tx)
            eNB1_receive_delay = eNB1_receive_delay+8;
            n=n+1;
            ack=0;
        else
            ack=1;
            n=max_tx;
        end
    end
end
```

```

        end
    end

    while (wall_clock<=UE1_scheduling_time)
        UE1_eNB2_time(frame)=ack*(UE1_scheduling_time+eNB1_receive_delay+eNB1_eNB2_dly(frame));
        wall_clock=wall_clock+frame_size;
        ack1(frame)=ack;
        frame=frame+1;
    end;
end

% Translate arrival time to packet delay for UL simulation
wall_clock = frame_size*(1:nFrames)';
UE1_eNB2_dly = max(-1, UE1_eNB2_time-wall_clock);

% Sort for monotonic arrival time to DL for simulation
[UE1_eNB2_time,monotonic_index]=sort(UE1_eNB2_time);

% Calculate the delay from eNB2 to UE2 (only for those packets that
% successfully arrived at the eNB2!). If transmission to UE2 is not
% successful after max_tx attempts, dly = 0; (packet loss)
frame = 1;
UE2_scheduling_time=mis_eNB1_eNB2;

while frame<=nFrames
    % Add the rx effect for the scheduling time
    n=0;
    eNB2_transmit_delay = 0;
    while n < max_rx
        if (rand(1) < BLER_rx)
            eNB2_transmit_delay = eNB2_transmit_delay+8;
            n=n+1;
            ack=0;
        else
            ack=1;
            n=max_rx;
        end
    end
end

while ((frame<=nFrames)&&(UE1_eNB2_time(frame)<UE2_scheduling_time))
    if (UE1_eNB2_time(frame)==-1)
        UE1_UE2_time(frame)=-1;
    else
        UE1_UE2_time(frame)=ack*(UE2_scheduling_time+eNB2_transmit_delay);
    end
    ack2(frame)=ack;
    frame=frame+1;
end

% Update the scheduling time
UE2_scheduling_time=UE2_scheduling_time+drx_cycle_length;
end

% Re-order for original transmit order
UE1_UE2_time(monotonic_index) = UE1_UE2_time;

% Translate arrival time to packet delay
wall_clock = frame_size*(1:nFrames)';
UE1_UE2_dly = max(-1, UE1_UE2_time-wall_clock);

% Set compensation values
if drx_cycle_length==0
    comp_e2e=min_net_delay;
    comp=min_net_delay;
else
    comp_e2e=min(UE1_UE2_dly(UE1_UE2_dly>0));
    comp=min(UE1_eNB2_dly(UE1_eNB2_dly>0));
end

% Calculates the overall packet loss from UE1 to UE2
pl=0;
for frame=1:nFrames
    if UE1_UE2_dly(frame)==-1;
        pl=pl+1;
    end
end
plr=pl/nFrames;

```

E.3 Packet delay and loss profiles for simulated stationary operating conditions with DRX 20 ms and DRX 40 ms

Three delay and loss profiles simulating the packet delay variations and packet loss that a UE may experience in stationary operating conditions with 20ms, 40ms DRX and 40ms DRX with higher BLER, respectively, are generated from the model with the parameter settings as described in Table E.2. The delay profiles are attached as text files and as binary trace files in the form of TCN network emulator.

Table E.2: Parameter settings for packet delay profile generation.

Operating condition	Parameter setting	System simulator operation	Delay and Loss profile
Stationary DRX 20 ms	BLER_tx : 0.1 BLER_rx : 0.1 max_tx : 3 max_rx : 3	Transparent system simulator downlink	dly_profile_20msDRX_10pct_BLER_e2e
	drx_cycle_length : 20 mis_eNB1_eNB2 : 10 max_net_delay : 33 min_net_delay : 27 nFrames : 8000 seed : 0	System simulator generating HARQ re-transmissions and DRX operation on the downlink	dly_profile_20msDRX_10pct_BLER_ue1_to_eNB2 (Not used in 3GPP TS 26.131)
Stationary DRX 40 ms	BLER_tx : 0.1 BLER_rx : 0.1 max_tx : 3 max_rx : 3	Transparent system simulator downlink	dly_profile_40msDRX_10pct_BLER_e2e
	drx_cycle_length : 40 mis_eNB1_eNB2 : 30 max_net_delay : 33 min_net_delay : 27 nFrames : 8000 seed : 0	System simulator generating HARQ re-transmissions and DRX operation on the downlink	dly_profile_40msDRX_10pct_BLER_ue1_to_eNB2 (Not used in 3GPP TS 26.131)
Stationary DRX 40 ms (22% BLER)	BLER_tx : 0.22 BLER_rx : 0.22 max_tx : 3 max_rx : 3 drx_cycle_length : 40 mis_eNB1_eNB2 : 30 max_net_delay : 36 min_net_delay : 24 nFrames : 8000 seed : 0	Transparent system simulator downlink	dly_profile_40msDRX_22pct_BLER_e2e

Since the model simulates the packet transmission end-to-end, the delay profiles include a packet delay component that according to the definition of the UE delay in 3GPP TS 26.131 [1] is not attributed to the UE delay. This delay shall be included in the test equipment delay when deriving the UE delay from the delay measurements based on the cross-correlation of the injected and measured signals. Values for the compensation attributed to the packet delay profiles are shown in Table E.3.

Table E.3: Packet delay profiles compensation values.

Delay and Loss profile	Compensation value
dly_profile_20msDRX_10pct_BLER_e2e	30 ms
dly_profile_40msDRX_10pct_BLER_e2e	30 ms
dly_profile_40msDRX_22pct_BLER_e2e	30 ms

Annex F (normative): Packet delay and loss profiles for jitter buffer management behaviour

F.1 General

This Annex provides packet delay and loss profiles to be inserted on the IP packets in the test system when testing the UE delay and speech quality for speech-only with LTE or WLAN access in jitter and error conditions, to characterise the jitter buffer management behaviour. The profiles are based on the assumption that one IP packet corresponds to 20ms of the speech sequence. In order to preserve the synchronization of the elements in the profile and the timing of the speech sequence (e.g. the RTP time stamp) when the speech codec does not produce one packet every 20ms, e.g. during speech codec DTX operation, the profiles need to be sub-sampled in accordance with the rate of production of the packets containing the speech frames.

F.2 Packet delay and loss profiles

[The profiles listed in Table F.1 are under consideration.]

Table F.1: Packet delay profiles.

Delay and Loss profile	Packet loss (%)	Description
dly_profile_volte	0.425	Real VoLTE capture
dly_ext_profile_error_profile_1	0	MTSI profile #1 [19] extended to 8000 frames
dly_ext_profile_error_profile_2	0.2375	MTSI profile #2 [19] extended to 8000 frames
dly_ext_profile_error_profile_3	0.475	MTSI profile #3 [19] extended to 8000 frames
dly_ext_profile_error_profile_4	2.25	MTSI profile #4 [19] extended to 8000 frames
dly_ext_profile_error_profile_6	0	MTSI profile #6 [19] extended to 8000 frames
NOTE1: MTSI profile #5 is not included because it assumes frame bundling, which may not be supported by test equipments.		
NOTE2: MTSI profiles have a length of 7500 frames; they have been extended to 8000 frames by appending the first 500 values at the end of the profile.		

Since the model simulates the packet transmission end-to-end, the delay profiles include a packet delay component that according to the definition of the UE delay in 3GPP TS 26.131 [1] is not attributed to the UE delay. This delay shall be included in the test equipment delay when deriving the UE delay from the delay measurements based on the cross-correlation of the injected and measured signals. Values for the compensation attributed to the packet delay profiles are shown in Table F.2.

Table F.2: Packet delay profiles compensation values.

Delay and Loss profile	Compensation value
dly_profile_volte	20 ms
dly_profile_error_profile_1	100 ms
dly_profile_error_profile_2	100 ms
dly_profile_error_profile_3	100 ms
dly_profile_error_profile_4	100 ms
dly_profile_error_profile_6	100 ms

]

Annex G (informative): Change history

TSG SA#	TSG doc	Spec	CR	Rev	Cat	Vers	New Vers	Subject
08							3.0.0	Approved
09	SP-000397	26.132	001		F	3.0.0	3.1.0	Handheld hands-free Test Setup
11	SP-010107	26.132	002	1	F	3.1.0	3.2.0	Harmonisation of test methods for acoustics between 3GPP and GSM
11		26.132				3.2.0	4.0.0	Release 4
11	SP-010107	26.132	003	1	B	4.0.0	5.0.0	Compatibility with testing wideband telephony transmission performance
13	SP-010454	26.132	004		B	5.0.0	5.1.0	Extended scope of test signals for Ambient Noise Rejection
13	SP-010454	26.132	006		F	5.0.0	5.1.0	Restricted scope of ITU-T P.501 test signals for 3G acoustic tests
13	SP-010454	26.132	008		A	5.0.0	5.1.0	Bandwidth of test signals for acoustic testing
15	SP-020080	26.132	011	1	A	5.1.0	5.2.0	Correction of references and editorial changes (wrong decimal separators)
16	SP-020435	26.132	016		F	5.2.0	5.3.0	Correction on ANR test for hands-free Ues
21	SP-030445	26.132	026		F	5.3.0	5.4.0	Loudness rating measurements at lower bit rates
25	SP-040649	26.132	028		C	5.4.0	6.0.0	Change of sending distortion test case
35	SP-070026	26.132	0030		F	6.0.0	6.1.0	Reference Update for ITU-T Recommendation P.57 "Artificial Ears"
35	SP-070026	26.132	0031	1	F	6.0.0	6.1.0	Update of reference [16] to P.79-2001 Annex G
35	SP-070026	26.132	0032	1	F	6.0.0	6.1.0	Distinction between narrow-band and wideband telephony in the frequency ranges and loudness rating and STMR weights, and in ANR calculation
36		26.132				6.1.0	7.0.0	Version for Release 7
38	SP-070759	26.132	0034	2	F	7.0.0	7.1.0	Changing the sidetone test to allow type 3.3 or 3.4 artificial ears
42	SP-080674	26.132	0035	1	F	7.1.0	7.2.0	Correction to allow wideband testing for GSM terminals
42	SP-080685	26.132	0037	3	C	7.2.0	8.0.0	Updated test methods for wideband terminal acoustics
43	SP-090015	26.132	0038		F	8.0.0	8.1.0	Clarification on Distortion with psophometric filter
43	SP-090018	26.132	0036	2	C	8.1.0	9.0.0	Speech and video telephony terminal acoustic test
45	SP-090568	26.132	0040	2	A	9.0.0	9.1.0	Correction of STMR calculation
45	SP-090573	26.132	0041	1	F	9.0.0	9.1.0	Handling Acoustic Testing with Noise Suppression Algorithms Employed
47	SP-100021	26.132	0042	1	F	9.1.0	9.2.0	Correction of distortion measurements
51	SP-110042	26.132	0043	5	B	9.2.0	10.0.0	Alignment of 3GPP Audio Test Case Specification
52	SP-110304	26.132	0045	3	C	10.0.0	10.1.0	Remaining modifications to EAAT WI
53	SP-110549	26.132	0047	1	F	10.1.0	10.2.0	Correction to Acoustic Echo Control
54	SP-110793	26.132	0048	1	F	10.2.0	10.3.0	Correction to Ambient Noise Rejection Test Procedure
54	SP-110793	26.132	0049		F	10.2.0	10.3.0	Clarification of Stability loss test signal
55	SP-120022	26.132	0050	1	F	10.3.0	10.4.0	Correction of receiving distortion
57	SP-120501	26.132	0053	2	F	10.4.0	10.5.0	Addition of 1/12 octave diffuse field table values with frequencies according to IEC 1260 (for information)
57	SP-120503	26.132	0052	3	B	10.5.0	11.0.0	Addition of UE delay test method
57	SP-120503	26.132	0054	1	B	10.5.0	11.0.0	Extension of Acoustic Tests
58	SP-120760	26.132	0055	1	F	11.0.0	11.1.0	Correction of references and levels
59	SP-130017	26.132	0056		F	11.1.0	11.2.0	Alignment of free-field definition with ISO 3745 and ITU-T Rec. P.340
59	SP-130017	26.132	0057	1	F	11.1.0	11.2.0	Voiding of ambient noise rejection test cases
60	SP-130185	26.132	0058	1	F	11.2.0	11.3.0	Update of the reference algorithm for echo control characteristics evaluation in Annex B
60	SP-130185	26.132	0061	2	F	11.2.0	11.3.0	UE delay test method
60	SP-130189	26.132	0059		B	11.3.0	12.0.0	Adding receiving distortion tests at frequencies lower than 1020Hz
60	SP-130189	26.132	0060	2	B	11.3.0	12.0.0	Update acoustic test method specification to cover MTSI speech-only services over LTE (narrowband and wideband)
62	SP-130568	26.132	0063		A	12.0.0	12.1.0	Corrections to references
62	SP-130563	26.132	0067	2	A	12.0.0	12.1.0	STMR - adaptation to modern form factors
62	SP-130573	26.132	0068	1	F	12.0.0	12.1.0	Updates of LTE radio conditions for acoustic testing

64	SP-140216	26.132	0069	2	F	12.1.0	12.2.0	Corrections of general test method applicability description and Headset UE sidetone test method
65	SP-140469	26.132	0070	1	B	12.2.0	12.3.0	LTE UE delay test methods
66	SP-140721	26.132	0072		A	12.3.0	12.4.0	Correction to UE receiving loudness rating performance test for wideband telephony
66	SP-140719	26.132	0079	1	A	12.3.0	12.4.0	Correction of broadband signal level at the hands free reference point
66	SP-140731	26.132	0080	2	B	12.3.0	12.4.0	Acoustic test methods for super-wideband and fullband telephony
67	SP-150090	26.132	0082		F	12.4.0	12.5.0	Minor corrections and editorial enhancements to acoustic test method descriptions
68	SP-150209	26.132	0083	2	F	12.5.0	12.6.0	Correcting test conditions for Handset/headset narrowband/wideband STMR
69	SP-150436	26.132	0085		F	12.6.0	12.7.0	Correction of descriptions of diffuse-field correction
69	SP-150445	26.132	0084	1	B	12.7.0	13.0.0	UE delay test methods for MTSI-based services over LTE with the EVS codec
70	SP-150651	26.132	0087	1	B	13.0.0	13.1.0	Additional UE delay test methods for MTSI-based services over LTE with the EVS codec
71	SP-160065	26.132	0089	1	A	13.1.0	13.2.0	Clarification on double measurement for SWB frequency characteristics
71	SP-160067	26.132	0090	1		13.1.0	13.2.0	Addition of test methods for objective speech quality in the presence of background noise for speakerphone (hand-held hands-free) operation

Change history							
Date	Meeting	TDoc	CR	Rev	Cat	Subject/Comment	New version
2016-06	72	SP-160260	0091	1	C	Support of DTX and other corrections	13.3.0
2017-03	75					Versione per Release 14	14.0.0
2017-06	76	SP-170325	0093	2		Extension of UE Delay Test Methods	14.1.0
2017-06	76	SP-170321	0094	2	B	Addition of SWB and FB noise suppression test methods	14.1.0
2018-06	80	SP-180265	0096	1	F	Correction of unclear wording in delay test specification	14.2.0
2018-06	80	SP-180273	0095	3	B	Test method for RLR in the presence of background noise	15.0.0
2018-09	81	SP-180648	0097	-	C	Modification of speech quality assessment method	15.1.0
2019-03	83	SP-190038	0098	2	F	Use of default EVS SID update	16.0.0
2019-09	85	SP-190935	0099	3	B	Support of NR	16.1.0
2020-03	87-e	SP-200103	0102	-	B	Alternative noise field simulation method for terminal testing	16.2.0

History

Document history		
V16.2.0	November 2020	Publication