ETSI TS 103 624 V1.2.1 (2021-10)

TECHNICAL SPECIFICATION

Characterization Methodology and Requirement Specifications for the ETSI LC3plus speech codec

Reference
RTS/STQ-292

Keywords
codec, listening quality, speech

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 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are 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 Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, 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.

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.

Foreword

This Technical Specification (TS) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

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.

1 Scope

The present document specifies the subjective and objective methodologies developed in cooperation between TC STQ and TC DECT for the characterization of the Low Complexity Communication Codec Plus (LC3plus) speech codec. It describes experimental tests and conditions used for subjective and objective testing. Based on these methodologies the performance requirements for this codec are specified.

The requirements in the present document are specified to characterize a high-quality codec for use in modern telecommunication networks, including but not limited to DECT and VoIP. A special focus is placed on the fact that end-to-end connections are often of hybrid nature concatenating different technologies and thus tandeming (i.e. transcoding) different codecs.

In addition to its speech capabilities, the LC3plus codec has the option for high quality music streaming. This is out of scope of the present document.

The 2021 revision of the present document adds the test plan for the auditory test, see annex B, the results of the subjective characterization test for the ETSI LC3plus codec, see annex C as well as the related electronic attachment(s).

2 References

2.1 Normative references

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

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

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

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

[1] Recommendation ITU-T P.800 (08/1996): "Methods for subjective determination of transmission quality".

[2] Recommendation ITU-T P.863 (03/2018): "Perceptual objective listening quality prediction".

[3] Recommendation ITU-T G.722 (09/2012): "7 kHz audio-coding within 64 kbit/s".

[4] Recommendation ITU-T G.726 (12/1990): "40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)".

[5] ETSI TS 103 634: "Digital Enhanced Cordless Telecommunications (DECT); Low Complexity Communication Codec plus (LC3plus)".

[6] Recommendation ITU-T G.191 (01/2019): "Software tools for speech and audio coding standardization".

[7] ETSI TS 126 442: "Universal Mobile Telecommunications System (UMTS); LTE; Codec for Enhanced Voice Services (EVS); ANSI C code (fixed-point) (3GPP TS 26.442)".

[8] ETSI TS 126 173: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; ANSI-C code for the Adaptive Multi-Rate - Wideband (AMR-WB) speech codec (3GPP TS 26.173)".

[9] ETSI TS 126 073: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; ANSI-C code for the Adaptive Multi Rate (AMR) speech codec (3GPP TS 26.073)".

[10] Recommendation ITU-T G.711 Appendix I (09/1999): "A high quality low-complexity algorithm for packet loss concealment with G.711".

[11] IETF RFC 8251: "Update to the Opus Audio Codec".

[12] Recommendation ITU-T G.711 (11/1988): "Pulse code modulation (PCM) of voice frequencies".

[13] Recommendation ITU-T G.722 Appendix IV (11/2006): "A low-complexity algorithm for packet loss concealment with G.722".

[14] ITU-T Handbook (2011): "Practical procedures for subjective testing".

[15] Recommendation ITU-T G.192 (03/1996): "A common digital parallel interface for speech standardization activities".

[16] Recommendation ITU-T P.56 (12/2011): "Objective measurement of active speech level".

[17] Recommendation ITU-T P.50 (09/1999): "Artificial voices".

[18] ETSI TS 126 441: "Universal Mobile Telecommunications System (UMTS); LTE; Codec for Enhanced Voice Services (EVS); General overview (3GPP TS 26.441)".

[19] Recommendation ITU-T G.722.2 (07/03): "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)".

2.2 Informative references

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

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

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

[i.1] ETSI TR 103 590: "Digital Enhanced Cordless Telecommunications (DECT); Study of Super Wideband Codec in DECT for narrowband, wideband and super-wideband audio communication including options of low delay audio connections (<= 10 ms framing)".

[i.2] IETF RFC 6716: "Definition of the Opus Audio Codec".

[i.3] 3GPP S4-141392: "EVS-7c Processing functions for characterization phase", TSG S4#81.

[i.4] 3GPP S4-141319: "EVS-8b EVS Permanent Document EVS-8b: Test plans for selection phase including lab task specification", TSG S4#81.

[i.5] 3GPP S4-141372: "EVS-8c EVS Permanent Document EVS-8c: Test plans for characterization phase including lab task specification", TSG S4#81.

[i.6] "A method for comparing the performance of EVS and other voice codecs under bursty packet loss", IPTcomm, 2018, IEEE™.

3 Definition of terms, symbols and abbreviations

3.1 Terms

Void.

3.2 Symbols

Void.

3.3 Abbreviations

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

ACR	Absolute Category Rating
AMR	Adaptive MultiRate speech codec
AMR-NB	Adaptive Multirate speech codec Narrow Band
AMR-WB	Adaptive Multirate speech codec Wide Band
ASL	Active Speech Level
BER	Bit Error Rate
BT	Better Than
CBR	Constant Bitrate
CELT	Constrained Energy Lapped Transform
CI95	Confidence Interval at 95 % probability level
CuT	Codec under Test
DCR	Degradation Category Rating
DECT	Digital Enhanced Cordless Telecommunications
DP	DECT Profile (e.g. DP0, DP1 etc.)
EID	Error Insertion Device
EN	English
EP	Error Protection
EPF	Error Protection File
EVS	codec for Enhanced Voice Services
EVS-WB	EVS WideBand
FB MB	Full Band Mixed Band
FB	FullBand
FEC	Forward Error Correction
FER	Frame Error Rate
FP	Fixed Part
LC3plus	Low Complexity Communication Codec Plus
MAN	Mandarin
MB	Mixed Band
MNRU	Modulated Noise Reference Unit
MOS	Mean Opinion Score
MOS-LQS	Mean Opinion Score – Listening Quality (Subjective)
MP3	MPEG Layer 3
MPEG	Motion Picture Experts Group
NB	Narrow Band
NWT	Not Worse Than
PCM	Pulse Code Modulation
PLC	Packet Loss Concealment
PLP	Packet Loss Profile
PLR	Packet Loss Rate
PP	Portable Part
RF	Radio Frequency
RSSI	Received Signal Strength Indicator
RTP	RealTime Protocol
SPL	Sound Pressure Level
SPL(A)	Sound pressure level, A-weighting
STD	Standard Deviation
STL	Software Tools Library
SWB	Super WideBand
URL	Uniform Recource Locator
VAD	Voice Activity Detection
VoIP	Voice over IP
WAV	WAVeform audio file
WB	WideBand

4 Introduction

The present document defines characterization methodologies as well as the performance requirements to be evaluated for the ETSI Low Complexity Communication Codec Plus (LC3plus) [5]. The performance of the codec was initially studied by the TC DECT group in ETSI TR 103 590 [i.1] which is considered as qualification of the codec.

The purpose of the characterization phase experiments is to demonstrate the performance of the codec over a set of conditions and the following use cases:

- Voice services in DECT and VoIP
- Interworking VoIP scenarios between different networks

The characterization utilizes the set of characterization methodologies and configurations of subjective and objective experiments defined in clause 5. The experiments are designed in order to evaluate whether LC3Plus achieves the following codec objectives:

- Introduction of Super-Wideband (SWB) quality in voice services
- Increased capacity of DECT systems when compared to legacy DECT codecs
- Improved robustness for packet loss and bit errors
- Ensure suitable performance in case of transcoding or self-tandeming conditions

All details on the definition of codec objectives for DECT and VoIP and the derived performance requirements and performance objectives are specified in clause 6.

Clause 7 defines the statistical analysis to be conducted on the subjective results to verify that the performance of the Codec under Test (CuT) is sufficient in comparison to the specified performance requirement or performance objectives. In the present document, CuT always means ETSI LC3plus [5].

5 Characterization methodologies

5.1 Overview

The present clause describes the experiment design and the subjective and objective methodologies. The aim of the characterization test is to assess the clean channel performance, self-tandeming capabilities, cross-tandeming, as well as rate switching conditions and variation of the input speech level.

The characterization tests shall be conducted in a similar fashion as the 3GPP EVS selection/characterization process 3GPP S4-141319 [i.4] and 3GPP S4-141372 [i.5].

5.2 Experiments

All test conditions shall be separated according to the category audio bandwidth and channel conditions. This results in six experiments, i.e. 3x audio bandwidth times 2x channel conditions. Additionally, one multi-bandwidth experiment shall be conducted in order to provide a quality overview. Such experiment is subdivided in order to allow for a test in different languages.

Each experiment is evaluated using subjective and objective methodologies described in clauses 5.4 and 5.5.

Table 1 outlines the experiment setup.

Table 1: Experiment overview

Experiment number	Experiment label	Max. bandwidth of input	Channel conditions	Estimated number of conditions
1	NB clean	4 000 Hz	No error	40
2	NB error	4 000 Hz	Bit error & packet loss	32
3	WB clean	8 000 Hz	No error	60
4	WB error	8 000 Hz	Bit error & packet loss	32
5	SWB clean	16 000 Hz	No error	40
6	SWB error	16 000 Hz	Bit error & packet loss	35
7a	MB fullscale short American English (see note)	20 000 Hz	No error & bit error & packet loss	30
7b	MB fullscale short Mandarin (see note)	20 000 Hz	No error & bit error & packet loss	30
NOTE: The MB fullscale experiment contains all bandwidth conditions to span the complete P.800 quality range [1].				

A complete list of all experiments and conditions describing the exact configuration for each condition and the relevant comparison points are contained in archive ts_103624v010201p0.zip which accompanies the present document.

NOTE: The experiment labelled "FB M_fullscale" in the attachment is not actually used in the test, since it has been replaced by the experiment labelled "FB M_fullscale_short" which will be conducted in two languages.

5.3 Item processing

The test items shall be processed according to the EVS processing plan [i.3]. For transcoding, no frame synchronization between the codecs shall be applied. The frequency masks used by 3GPP EVS characterization tests shall be applied to the input signals. The items shall be processed and prepared for the experiments using the tools provided in Recommendation ITU-T G.191 [6].

5.4 Subjective methodologies

All subjective experiments shall be conducted using the Recommendation ITU-T P.800 [1] procedure using speech material. Subjects shall be naïve listeners and native speakers. Experiments should be conducted in different languages and labs.

Table 2 shows the Recommendation ITU-T P.800 [1] experiment configurations.

Table 2: P.800 experiment configuration

Parameter	Experiment							
	1	2	3	4	5	6	7a	7b
Rating scale	ACR	ACR	ACR	ACR	DCR	ACR	ACR	ACR
Minimum number of listeners	24	24	24	24	24	24	24	24
Minimum number of talkers	4	4	4	4	4	4	4	4
Minimum number of samples per talker	6	6	6	6	6	6	6	6
Minimum number of votes per sample	4	4	4	4	4	4	4	4
Minimum number of votes per condition	96	96	96	96	96	96	96	96
Estimated test duration in minutes (see note)	47	41	63	41	47	38	29	29
NOTE: Estimation calculation contained in archive ts_103624v010201p0.zip.								

5.5 Objective methodologies

All experiments listed in table 1 shall be assessed by the objective quality evaluation using the perceptual objective listening quality prediction tool standardized by ITU-T also known as Recommendation ITU-T P.863 [2].

Tests shall be run in the full band mode with full band reference files and appropriate degraded files.

6 Characterization test plan

6.1 Testing Conventions

6.1.1 Introduction

The following clauses specify performance requirements and conditions to be evaluated for the following use cases:

- DECT with clean channel conditions.
- DECT with error prone channel conditions.
- VoIP without packet loss conditions.
- VoIP including packet loss conditions.

Besides performance requirements, performance objectives are specified. The performance objectives are only foreseen as informative comparison conditions.

6.1.2 Software versions

The following software versions for the different codecs shall be used:

- G.711 A-law: Recommendation ITU-T G.711 [12] and G.711 Appendix I (PLC) [10].
- IETF RFC 8251 [11] OPUS: V1.3.0 (latest), fix-point.

NOTE: OPUS is a codec in accordance with IETF RFC 6716 [i.2] and IETF RFC 8251 [11].

- EVS: EVS Codec ETSI TS 126 442 [7].
- LC3plus: Latest.
- G.722: Recommendation ITU-T G.722 [3] and G.722 Appendix IV [13].
- AMR-WB (G.722.2): ETSI TS 126 173 [8].
- AMR-NB: ETSI TS 126 073 [9].
- G.726: Recommendation ITU-T G.726 [4].

6.1.3 Test condition numbering

The test conditions are numbered according the scheme given in table 3.

Table 3: Test condition numbering

NB	WB	SWB
DECT with clean channel conditions		
1xx	2xx	3xx
DECT with error prone channel conditions		
4xx	5xx	6xx
VoIP without packet loss conditions		
7xx	8xx	9xx
VoIP including packet loss conditions		
10xx	11xx	12xx

6.2 Characterization test plan for clean channels with application in DECT scenarios

6.2.1 Overview

CuT in DECT shall provide the same or better voice quality than the VoIP network provides and guarantees higher efficiency than DECT audio codecs used today, meaning same quality at lower bit rates to allow better DECT slot exploitation in conjunction with channel coding to provide better protection for bit errors and packet loss concealment.

As network interworking scenarios, the following cases shall be evaluated:

- Voice calls from legacy VoIP to DECT
- Voice calls from DECT to legacy VoIP
- Voice calls from DECT over legacy VoIP to DECT

DECT uses today G.726 (NB) and G.722 (WB). Today's VoIP terminals utilize G.711 (NB) and G.722 (WB).

6.2.2 NB conditions

The test shall verify the performance of the CuT in NB mode. Speech coding for narrowband speech connections using a normal 32 kbit/s payload DECT RF slot shall not be worse than what is achieved by Recommendation ITU-T G.726 [4]. The CuT shall enable the same range where communication is possible between DECT PP and FP as achieved by DECT-G.726 connections of the previous technology.

The voice quality by transcoding between VoIP G.711 to/from CuT shall not be worse than connections between VoIP-G711 and DECT-G.726.

Additional performance objectives should be defined in comparison to OPUS (CELT mode, constant bitrate mode (CBR), 32 kbit/s, complexity=0, FEC off, NB mode, 10 ms framing).

The following NB conditions shall be included into the test (Input speech levels to be applied are -16 dBov, -26 dBov, -36 dBov):

100. Direct reference conditions with limited audio bandwidth (cut off frequency of 4 kHz) but no speech coding.

CuT:

101. LC3plus 32 kbit/s, 10 ms framing.

Requirement:

102. G.726, 32 kbit/s with G.711 Appendix I PLC [10].

Performance objective:

103. OPUS, CELT mode, CBR, 32 kbit/s, complexity = 0, FEC off, NB mode, 10 ms framing.

The following transcoding scenarios shall be tested:

CuT:

104. G.711 -> LC3plus (32 kbit/s).
105. LC3plus (32 kbit/s) -> G.711.
106. LC3plus (32 kbit/s) -> G.711 -> LC3plus (32 kbit/s).

Requirement:

107. G.711 -> G.726 (32 kbit/s).
108. G.726 (32 kbit/s) -> G.711.
109. G.726 (32 kbit/s) -> G.711 ->G.726 (32 kbit/s).

Performance objective:

110. G.711 -> OPUS (32 kbit/s).
111. OPUS (32 kbit/s) -> G.711.
112. OPUS (32 kbit/s) -> G.711 -> OPUS (32 kbit/s).

The following codecs shall be tested for self-tandeming (double and triple):

113. LC3plus (32 kbit/s).
114. G.726 (32 kbit/s).
115. OPUS (32 kbit/s).
116. G.711 (64 kbit/s).

6.2.3 WB conditions

The test shall verify the performance of the candidate codec in WB mode for DECT scenarios. Speech coding for wideband speech connections using a 32 kbit/s payload for normal DECT RF slots shall not be worse than what is achieved recently by Recommendation ITU-T G.722 [3] using a 64 kbit/s payload for long DECT RF slots. The DECT evolution RF connection shall enable at least the same range where communication is possible between DECT PP and FP compared to previous G.722 DECT connections. It is envisioned that the range can be further extended.

The voice quality by transcoding between VoIP networks using G.722 to/from the DECT evolution speech codec shall not be worse than connections between VoIP-G.722 and DECT-G.722.

Additional performance objectives should be defined in comparison to OPUS (CELT mode, CBR, 32 kbit/s, complexity = 0, FEC off, WB mode, 10 ms framing).

The following WB conditions shall be included into the test (input speech levels which shall be used are -16 dBov, -26 dBov and -36 dBov):

200. Direct reference condition with limited audio bandwidth with cut off frequency of 8 kHz, but no speech coding.

CuT:

201. LC3plus, 32 kbit/s, 16 kHz, 10 ms framing, 16 bits per audio sample.

Requirement:

202. G.722, 64 kbit/s.

Performance objective:

203. OPUS, CELT mode, Constant BitRate (CBR): 32 kbit/s, complexity=0, FEC off, WB mode, 10 ms framing.

To be characterized:

204. LC3plus for bitrates: 32 kbit/s, 48 kbit/s. Sampling rate of 16 kHz and nominal speech level. Short frame size (5 ms frame size) against regular frame size LC3plus 32 kbit/s codec (10 ms frame size).
205. LC3plus for bitrates: 64 kbit/s, 96 kbit/s. Sampling rate of 16 kHz and nominal speech level. Short frame size (2,5 ms frame size) against regular frame size LC3plus 32 kbit/s codec (10 ms frame size).

The following transcoding scenarios shall be tested:

CuT:

206. LC3plus (32 kbit/s) -> G.722 (64 kbit/s).
207. G.722 (64 kbit/s) -> LC3plus (32 kbit/s).
208. LC3plus (32 kbit/s) ->G.722 (64 kbit/s) -> LC3plus (32 kbit/s).

Requirement:

209. G.722 (64 kbit/s) -> G.722 (64 kbit/s).
210. G.722 (64 kbit/s) -> G.722 (64 kbit/s) -> G.722 (64 kbit/s).

Performance objective:

211. OPUS (32 kbit/s) -> G.722 (64 kbit/s).
212. G.722 (64 kbit/s) -> OPUS (32 kbit/s).
213. OPUS (32 kbit/s) -> G.722 (64 kbit/s) -> OPUS (32 kbit/s).

The following codecs shall be tested for self-tandeming (double, triple):

214. LC3plus (32 kbit/s).
215. OPUS (32 kbit/s).
216. G.722 (64 kbit/s).

6.2.4 SWB conditions

Speech coding for super-wideband speech connections using a long 64 kbit/s payload DECT RF slot shall not be worse than what is achieved by EVS-SWB at 13,2 kbit/s and better than what is achieved by Recommendation ITU-T G.722 [3] at 64 kbit/s. The DECT evolution RF connection shall enable the same range where communication is possible between DECT PP and FP as achieved today by G.722 DECT connections.

The voice quality degradation by transcoding between VoIP networks using OPUS (fullband mode) or EVS to/from DECT evolution speech codec shall be characterized.

Additional objectives should be defined in comparison to OPUS (CELT mode, CBR, 64 kbit/s, complexity=0, FEC off, FB mode, 10 ms framing).

The following conditions shall be tested (Input speech levels to be applied are -16 dBov, -26 dBov, -36 dBov):

300. Direct reference conditions with limited audio bandwidth but no speech coding. Lowpass cut-off frequency of 16 kHz shall be used.

CuT:

301. LC3plus, 64 kbit/s at sampling rate of 32 kHz. 10 ms framing, 16 bits per audio sample.

Requirement:

302. G.722 (64 kbit/s, WB, with Appendix IV PLC).
303. EVS (SWB at 13,2 kbit/s). No channel aware mode used.

Performance Objective:

304. OPUS, CELT mode, CBR, 64 kbit/s, complexity = 0, FEC off, fullband, mode, 10 ms framing.

To be characterized:

305. LC3plus for bitrates: 64 kbit/s, 96 kbit/s. Sampling rate of 32 kHz and nominal speech level. Short frame size (5 ms frame size) against regular frame size LC3plus 64 kbit/s codec (10 ms frame size).
306. LC3plus for bitrates: 96 kbit/s, 128 kbit/s. Sampling rate of 16 kHz and nominal speech level. Short frame size (2,5 ms frame size) against regular frame size LC3plus 64 kbit/s codec (10 ms frame size).

The following transcoding scenarios shall be tested:

307. EVS (SWB at 13,2 kbit/s) -> LC3plus (64 kbit/s).

308. LC3plus (64 kbit/s) -> EVS (SWB at 13,2 kbit/s).
309. LC3plus (64 kbit/s) -> EVS (SWB at 13,2 kbit/s) -> LC3plus (64 kbit/s).

310. EVS (SWB at 13,2 kbit/s) -> OPUS (64 kbit/s).
311. OPUS (64 kbit/s) -> EVS (SWB at 13,2 kbit/s).
312. OPUS (64 kbit/s) -> EVS (SWB at 13,2 kbit/s) -> OPUS (64 kbit/s).

313. LC3plus (64 kbit/s) -> OPUS (64 kbit/s).
314. OPUS (64 kbit/s) -> LC3plus (64 kbit/s).
315. LC3plus (64 kbit/s) -> OPUS (64 kbit/s) -> LC3plus (64 kbit/s).

The following codec shall be tested for self-tandeming (double, triple and quadruple):

316. LC3plus (64 kbit/s).
317. Opus (64 kbit/s).
318. EVS (SWB at 13,2 kbit/s).

6.3 Characterization plan for error prone channels with application in DECT scenarios

6.3.1 Overview

The packet loss concealment performance of CuT shall be evaluated compared to G.726 in NB and G.722 in WB and to SWB codec.

Table 4: DECT Packet loss and bit error rates

Packet Loss Profile	Normalized Averaged signal strength (RSSI)	PLR [%]	BER rounded [%]
DP0	1 (136 dB)	0	0
DP1	0,41 (56 dB)	0,99	0,01
DP2	0,35 (48 dB)	0,88	0,31
DP3	0,29 (40 dB)	7,39	2,92

The CuT shall be compared to the requirement condition under four typical signal strengths representing DECT packet loss and bit error profiles for a 10 ms framing. The configurations of the codecs include a specific setup to adapt to the DECT channel characteristic, e.g. configuration of the channel coder or addition of parity bit. The DECT error profiles are labelled DP0, DP1, DP2 and DP3 (see table 4).

Switching of channel coder configurations (including rate switching) shall be tested as well. Switching within all possible channel configurations shall not be worse than only operating in the mode with highest protection.

For completeness, random patterns of 3 % and 6 % for a 10 ms frame size shall be tested. The random patterns are labelled as FER 3 % and 6 %.

6.3.2 NB conditions

The error profiles DP0, DP1, DP2, DP3 (see table 5) and random FER 3 % and 6 % shall be applied for:

CuT:

400. LC3plus, 32 kbit/s, sampling rate 8 kHz, 10 ms framing.

Requirement:

401. G.726 32 kbit/s with G.711 Appendix I PLC [10].

6.3.3 WB conditions

The error profiles DP0, DP1, DP2, DP3 (see table 5) and random FER 3 % and 6 % shall be applied to:

CuT:

500. LC3plus, 32 kbit/s, sampling rate 16 kHz, 10 ms framing.

Requirement:

501. G.722 64 kbit/s with G.722 Appendix IV PLC.

6.3.4 SWB conditions

The error profiles DP0, DP1, DP2, DP3 (see table 5) and random FER 3 % and 6 % shall be applied for:

CuT:

600. LC3plus, bitrate 64 kbit/s, sampling rate 32 kHz, 10 ms framing.

Requirement:

601. G.722 64 kbit/s with G.722 Appendix IV PLC (G.722 conditions required to check that SWB connections achieve the same DECT distance of portable and fix part compared to legacy WB connections).

6.4 Characterization test plan for clean channels with application in VoIP scenarios

6.4.1 Overview

VoIP networks today use mainly G.711 for NB, G.722 for WB and OPUS for SWB. As SWB services are currently deployed in VoLTE, EVS-SWB may serve as the alternative reference point. However, OPUS is used in the following proposal.

A new ETSI VoIP codec shall provide the same or better speech quality than previous VoIP networks provide. It is envisioned that the new codec provides a Packet Loss Concealment (PLC) better than the PLC provided for G.711 and G.722 for narrowband and wideband calls.

For VoIP, the network interworking scenario with mobile phones shall be of main focus, leading to the transcoding conditions: VoIP to mobile terminals and vice versa.

As relevant mobile codecs AMR-NB, AMR-WB, EVS-WB and EVS-SWB shall be considered operating at the most commonly used configurations as outlined in the following clauses 6.4.2 to 6.4.4.

6.4.2 NB conditions

For narrowband VoIP network speech coding connections using up to 64 kbit/s payload, the CuT shall not be worse than G.711 coding.

The following conditions shall be included into the test (Input speech levels to be applied are -16 dBov, -26 dBov, -36 dBov):

CuT:

700. LC3plus, bitrate of 32 kbit/s, sampling rate 8 kHz, 10 ms framing, 16 bits per audio sample.

Requirement:

701. G.711 64 kbit/s, with Appendix I PLC.

Performance objective:

702. OPUS, CELT mode, CBR, 32 kbit/s, complexity = 0, FEC off, NB mode, 10 ms framing.
703. AMR-NB 12,2 kbit/s.

The following transcoding scenarios shall be tested:

CuT:

704. AMR-NB (12,2 kbit/s) -> LC3plus (32 kbit/s).
705. LC3plus (32 kbit/s) -> AMR-NB (12,2 kbit/s).

Requirement:

706. AMR-NB (12,2 kbit/s) -> G.711.
707. G.711 -> AMR-NB (12,2 kbit/s).

Performance objective:

708. AMR-NB (12,2 kbit/s) -> OPUS (32 kbit/s).
709. OPUS (32 kbit/s) -> AMR-NB (12,2 kbit/s).

6.4.3 WB conditions

VoIP network speech coding done by CuT for wideband speech connections using up to 64 kbit/s payload shall not be worse than G.722 coding.

The following WB conditions shall be included into the test (Input speech levels to be applied are -16 dBov, -26 dBov, -36 dBov):

CuT:

800. LC3plus, 32 kbit/s, sampling rate 16 kHz, 10 ms framing, 16 bits per audio sample.

Requirement:

801. G.722, 64 kbit/s with Appendix IV PLC will be used.

Performance objective:

802. OPUS, CELT mode, CBR, 32 kbit/s, complexity=0, FEC off, WB mode, 10 ms framing.

The following transcoding scenarios shall be tested:

CuT:

803. AMR-WB (23,85 kbit/s) -> LC3plus (32 kbit/s).
804. LC3plus (32 kbit/s) -> AMR-WB (23,85 kbit/s).
805. AMR-WB (12,65 kbit/s) -> LC3plus (32 kbit/s).
806. LC3plus (32 kbit/s) -> AMR-WB (12,65 kbit/s).
807. EVS-WB (24,4 kbit/s) -> LC3plus (32 kbit/s).
808. LC3plus (32 kbit/s) -> EVS-WB (24,4 kbit/s).
809. EVS-WB (13,2 kbit/s) -> LC3plus (32 kbit/s).
810. LC3plus (32 kbit/s) -> EVS-WB (13,2 kbit/s).

Requirement:

811. AMR-WB (23,85 kbit/s) -> G.722 (64 kbit/s).
812. G.722 (64 kbit/s) -> AMR-WB (23,85 kbit/s).
813. AMR-WB (12,65 kbit/s) -> G.722 (64 kbit/s).
814. G.722 (64 kbit/s) -> AMR-WB (12,65 kbit/s).
815. EVS-WB (24,4 kbit/s) -> G.722 (64 kbit/s).
816. G.722 (64 kbit/s) -> EVS-WB (24,4 kbit/s).
817. EVS-WB (13,2 kbit/s) -> G.722 (64 kbit/s).
818. G.722 (64 kbit/s) -> EVS-WB (13,2 kbit/s).

Performance objective:

819. AMR-WB (23,85 kbit/s) -> OPUS (32 kbit/s).
820. OPUS (32 kbit/s) -> AMR-WB (23,85 kbit/s).
821. AMR-WB (12,65 kbit/s) -> OPUS (32 kbit/s).
822. OPUS (32 kbit/s) -> AMR-WB (12,65 kbit/s).
823. EVS-WB (24,4 kbit/s) -> OPUS (32 kbit/s).
824. OPUS (32 kbit/s) -> EVS-WB (24,4 kbit/s).
825. EVS-WB (13,2 kbit/s) -> OPUS (32 kbit/s).
826. OPUS (32 kbit/s) -> EVS-WB (13,2 kbit/s).

The following codecs shall be tested for self-tandeming (double, triple):

827. LC3plus (32 kbit/s).
828. OPUS (32 kbit/s).
829. EVS-WB (13,2 kbit/s).

6.4.4 SWB conditions

VoIP network speech coding for super wideband speech connection typically uses a payload of 64 kbit/s . The CuT shall not be worse than OPUS ((CELT mode, constant bitrate mode (CBR), 64 kbit/s, complexity = 0, FEC off, fullband mode, 10 ms framing) for coding.

The following SWB conditions shall be included into the test (Input speech levels to be applied are -16 dBov, -26 dBov, -36 dBov):

CuT:

900. LC3plus, bitrate 64 kbit/s, sampling rate 32 kHz, 10 ms framing, 16 bits per audio sample.

Requirement:

901. OPUS, CELT mode, CBR, 64 kbit/s, complexity = 0, FEC off, fullband mode, 10 ms framing.

The following transcoding scenarios shall be tested:

902. EVS-SWB (24,4 kbit/s) -> LC3plus (64 kbit/s).
903. LC3plus (64 kbit/s) -> EVS-SWB (24,4 kbit/s).
904. EVS-SWB (13,2 kbit/s) -> LC3plus (64 kbit/s).
905. LC3plus (64 kbit/s) -> EVS-SWB (13,2 kbit/s).
906. EVS-SWB (24,4 kbit/s) -> OPUS (64 kbit/s).
907. OPUS (64 kbit/s) -> EVS-SWB (24,4 kbit/s).
908. EVS-SWB (13,2 kbit/s) -> OPUS (64 kbit/s).
909. OPUS (64 kbit/s) -> EVS-SWB (13,2 kbit/s).

The following codecs shall be tested for self-tandeming (double, triple):

910. LC3plus (64 kbit/s).
911. OPUS (64 kbit/s).
912. EVS-WB (13,2 kbit/s).

6.5 Characterization plan for Packet Loss Concealment (PLC) with application in VoIP scenarios

6.5.1 Overview

The following Packet Loss Profiles (PLPs) shall be applied.

Table 5: Packet Loss Profiles for VoIP application testing

N	Packet Loss Profile	Burst Length (No. Packets)	Average Loss Rate (%)	Comments
1	PLP0	0	0	No loss
2	PLP1	1	1,43	Uniform loss
3	PLP2	2	2,87	Synchronized with PLP1
4	PLP4	4	5,74	Synchronized with PLP1
5	PLP7	7	10,05	Synchronized with PLP1

The PLPs have been designed to test the impact of different bursts of losses on the quality of voice when different codecs and packet loss mitigation and resilience mechanisms are used (such as Packet Loss Concealment and Forward Error Correction). As shown in table 5, each PLP has a distinct length of loss burst.

PLP1 was the first profile to be developed and it was designed by distributing individual losses (i.e. burst length of 1) randomly along the whole PLP. It was used to design the rest of the PLPs by increasing the size of the burst. For example, if an extract of packet statuses from PLP1 was 111110111111 (where 1 indicates a successfully received packet and 0 indicates a lost packet) PLP4 is then developed by switching the status of the 3 packets succeeding the lost one from 1 to 0 (i.e. 111110111111 becomes 111110000111). As a result, all PLP types have bursts that begin at the same location in the speech sample (see figure 1 for an example applied to a 12 s speech sample consisting of 3 sentences).

NOTE: This only represents one approach to testing conditions with bursty packet loss and the applied loss profiles can be changed (e.g. to replicate the characteristics of particular networks and network conditions).

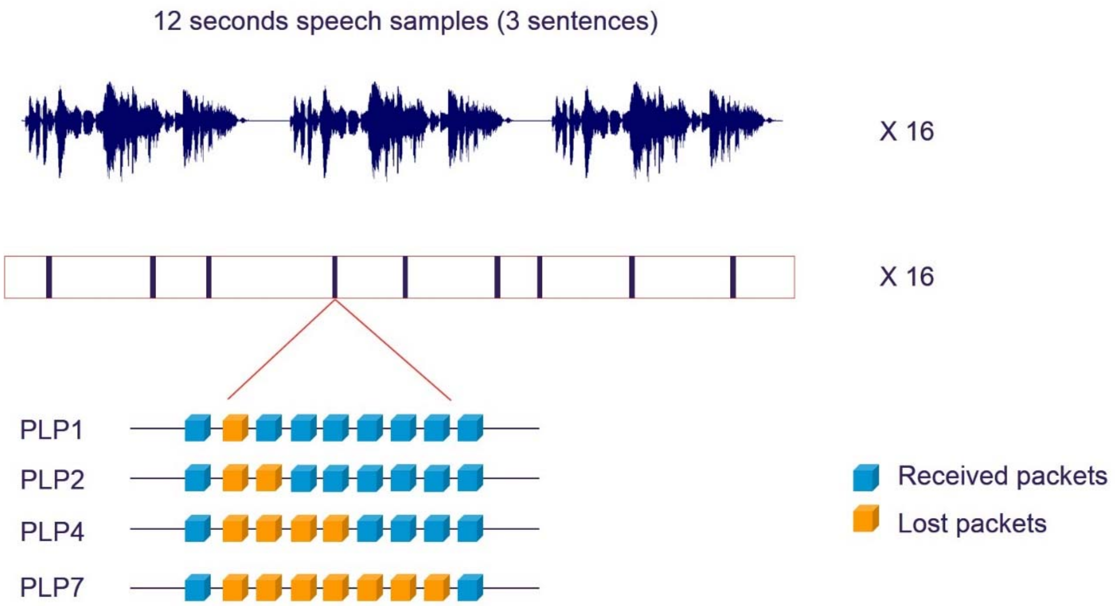

Figure 1

PLC frames should be aligned to the same frames for all speech input frames; For VoIP, only 20 ms PLC frames will be triggered. This means for 10 ms frame codec always 2 frames of PLC indication are in a row.

For completeness, random pattern of 3 % and 6 % for a 20 ms frame size shall be tested. The random patterns are labelled as FER 3 % and 6 %.

As LC3plus may not use the full bit rate of 64 kbit/s of the VoIP transmission, RTP based redundancy modes should be tested as well. Here, besides the main LC3plus frame of the current frame, an additional redundant LC3plus frame with an offset of X packets is transmitted in the same RTP payload. In this scenario, the playout is delayed by X packet, but the redundant LC3plus frame can be used to handle packet losses. The assessment described in the IEEE paper [i.6] mentioned above uses an offset of 3 packets for the EVS codec. The same value should be used for LC3plus. The processed condition may be based on simulations on bit stream level.

It is envisioned that the CuT provides a Packet Loss Concealment (PLC) better than the currently provided PLC for G.711, G.722 and OPUS for NB, WB and SWB calls respectively.

6.5.2 NB conditions

The CuT PLC shall be as good as, or better than G.711 appendix I for random packet losses at 20 ms packet size.

The Packet Loss Profiles PLP0, PLP1, PLP2, PLP4 and PLP7 (see table 5) and FER 3 % and 6 % shall be applied for:

CuT:

1000. LC3plus, bitrate of 32 kbit/s, sampling rate 8 kHz, 10 ms framing, 16 bits per audio sample.
1001. LC3plus, 64 kbit/s with RTP redundancy mode, sampling rate 8 kHz, 10 ms framing, 16 bits per audio sample (only relevant for PLP4 and PLP7; due to a redundancy configuration with offset=3, single and double losses in PLP1 and PLP2 can be completely compensated).

Requirement:

1002. G.711 64 kbit/s, with Appendix I PLC (Packet Loss Concealment).

6.5.3 WB conditions

The CuT PLC shall be as good as or better than G.722 Appendix IV for random packet losses at 20 ms packet size.

The Packet Loss Profiles PLP0, PLP1, PLP2, PLP4 and PLP7 (see table 5) and FER 3 % and 6 % shall be applied for:

CuT:

1100. LC3plus, bitrate 32 kbit/s, sampling rate 16 kHz, 10 ms framing, 16 bits per audio sample.
1101. LC3plus, bitrate 64 kbit/s with RTP redundancy mode, sampling rate 16 kHz, 10 ms framing, 16 bits per audio sample (only relevant for PLP4 and PLP7, due to a redundancy configuration with offset=3, single and double losses in PLP1 and PLP2 can be completely compensated).

Requirement:

1102. G.722 64 kbit/s with Appendix IV PLC (Packet Loss Concealment).

6.5.4 SWB conditions

The Packet Loss Profiles PLP0, PLP1, PLP2, PLP4 and PLP7 (see table 5) and FER 3 % and 6 % shall be applied for:

CuT:

1200. LC3plus, bitrate 64 kbit/s, sampling rate 32 kHz, 10 ms framing, 16 bits per audio sample.
1201. LC3plus, bitrate 64 kbit/s with RTP redundancy mode, sampling rate 32 kHz, 10 ms framing, 16 bits per audio sample (relevant for all PLPs).

Requirement:

1202. OPUS, CELT mode, CBR, 64 kbit/s, complexity=0, FEC off, FB mode, 10 ms framing.

7 Requirement verification

7.1 Requirement verification for subjective tests

Each condition in clauses 6.1 to 6.4 labelled as requirement shall be compared to the corresponding CuT condition using a Student's Dependent Groups t-test (single-sided at 95 % confidence level) on the subjective scores. This data will be the base for verifying that the CuT meets or exceeds the requirement. Additionally, the same rule shall be applied to the performance objectives, whereas the comparison between CuT and performance objective have only informative character.

A complete report for each requirement and objective will be provided in a future revision of the present document.

7.2 Requirement verification for objective tests

Each condition in clauses 6.2 to 6.5 labelled as requirement shall be compared to the corresponding CuT condition using a Student's Dependent Groups t-test (single-sided at 95 % confidence level) on the objective scores generated as described in clause 5.3. Additionally, the same rule shall be applied to the performance objectives.

The objective requirement verification is only for information. A complete report for each requirement and objective in comparison to the subjective results will be provided in a future revision of the present document.

8 Performance summary

The electronic attachment of the present document "Summary_TTF005_green_v2" in the folder "Raw voting data along with their statistical processing" provides an overview of all statistical results of the LC3plus codec in comparison to requirement and objective conditions based on the subjective data.

The LC3plus passes all statistical tests (zero fails) where LC3plus is compared to the defined requirement and objective reference conditions. Altogether, 153 statistical tests were conducted. In 61 cases, the LC3plus passes the test by being "not worse than" (NWT) the reference condition. In 92 cases, the LC3plus passes the test by being "better than" (BT) the reference condition. This means in 60% of all tests, LC3plus exceeds the performance reference point defined for DECT and VoIP applications.

9 Conclusion

LC3plus is recommended as the ETSI codec for global deployment in DECT and VoIP applications.

Annex A (normative): Conditions for the P.800 experiments

The conditions for the seven P.800 experiments are contained in subfolder "Conditions for the P.800 experiments" in archive ts_103624v010201p0.zip which accompanies the present document.

Compared to ts_103624v010101p0.zip which was part of V1.1.1 of the present document, the experiment FB MB_fullscale_short is added as smaller version of FB MB_fullscale in order to be realizable in two languages.

Annex B (normative): Subjective characterization test plan for the ETSI LC3plus codec

B.1 Introduction

This annex contains the subjective test plan for the ETSI LC3plus codec.

Test methodologies used for all planned subjective tests are based on Recommendation ITU-T P.800 [1], the ITU-T Handbook [14] and the guidelines provided in the main part of the present document.

The following steps are part of the test plan:

- Test samples are delivered in "ready-to-listen-to" format, means cropped and windowed, at a level corresponding to the final playout listening level, and grouped by listening panel.
- The randomization are arranged different for each listening panel. The randomization tables are contained subfolder "Randomization tables" in archive ts_103624v010201p0.zip which accompanies the present document.
- The raw voting data along with their statistical processing (MOS calculation and conditional STD and CI95) are contained in subfolder "Raw voting data along with their statistical processing" in archive ts_103624v010201p0.zip which accompanies the present document.
- No sample name blinding or anonymization will be applied, as it is principally not needed.

B.2 Description of the subjective experiments

B.2.1 General considerations

- Four talkers (two males, two females) are used in each language.
- 24 subjects for each experiment, minimum 4 (2M+2F) talkers, minimum 6 samples per talker, minimum 4 votes per sample, minimum 96 votes per condition, each panel with an independent randomization.
- Preliminary or training conditions are selected from the material available by the listening lab.
- Randomizations are constructed under "partially-balanced/randomized blocks" experimental design described in the ITU-T Handbook [14].
- Test duration: maximum 2 hours per listening panel. Test duration comprises 50 % of actual listening time and 50 % test overhead including administration, initial briefing, preliminaries, and breaks.
- Listening level -21 dB Pa (73 dB SPL) equals to -26 dBov.
- Files are played back with diffuse-field equalized headphones and diotic presentation.
- Six (6) experiments are performed using ACR methodology of Recommendation ITU-T P.800 [1] and one experiment is performed using DCR methodology [1]. Overall, 7 experiments are performed.
- Listening environment: All tests are performed in an acoustically treated critical listening room that conforms to Recommendation ITU-T P.800 [1] requirements in full.

Details of the listening test instructions are provided in clause B.3.

B.2.2 Speech Material

Each source speech file contains a sentence pair and lasts exactly 8 s. Each sentence is centred inside a 4 s time window. Leading and trailing silence parts are longer than 0,5 s. The sentences are simple meaningful sentences, similar to those described in annex B.1.4 of Recommendation ITU-T P.800 [1].

The test languages will be American English (Experiment 1 to 6 and Experiment 7a) and Mandarin (Experiment 7b) as outlined in table B.1.

B.2.3 Test samples

Each experiment sample database consists of a number of samples to be calculated by: nTalker(4) x nSamples(6) x nConditions (see table B.1). The test sample are processed according to clause B.4.

B.2.4 Experiments

Table B.1 outlines the basic parameters of the experiments. A detailed list of the exact conditions is provided in clause B.2.5.

Table B.1: Experiment parameters

Experiment number	Experiment label (Excel sheet)	Experiment designator (file naming)	Max. audio bandwidth	Channel conditions	Number of conditions (nConditions)	Rating Scale	Language
1	NB clean	n1	4 000 Hz	No error	40	ACR	English
2	NB error	n2	4 000 Hz	Bit error & packet loss	32	ACR	English
3	WB clean	w1	8 000 Hz	No error	60	ACR	English
4	WB error	w2	8 000 Hz	Bit error & packet loss	32	ACR	English
5	SWB clean	s1	16 000 Hz	No error	40	DCR	English
6	SWB error	s2	16 000 Hz	Bit error & packet loss	35	ACR	English
7a	FB MB_fullscale short	f1	20 000 Hz	No error & packet loss & bit error	30	ACR	English
7b	FB MB_fullscale short	f2	20 000 Hz	No error & packet loss & bit error	30	ACR	Mandarin
NOTE: The FB MB_fullscale_short fullscale experiment contains all bandwidth conditions to span the complete quality range of Recommendation ITU-T P.800 [1].							

B.2.5 Test randomizations

For presentation to the subjects, the samples are grouped into 3 or 6 panels according to the "partially-balanced/randomized blocks" experimental design.

Table B.2 visualizes the items to be presented for each panel. The item selection follows the partially-balanced/randomized blocks design. The list are defined for 6 or 3 panels. For 3 panel presentation, panel 1 & 2, 3 & 4 and 5 & 6 need to be combined.

Table B.2: Experimental design of panels

	Block	Seq	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5	Panel 6
0	1	1	an1f1s5.c34	an1f1s3.c13	an1f2s4.c16	an1f1s4.c28	an1f2s4.c39	an1f2s4.c26
1	1	2	an1m1s2.c13	an1m1s1.c33	an1m2s4.c26	an1m2s1.c31	an1m1s1.c01	an1m1s3.c37
2	1	3	an1f2s2.c05	an1f1s6.c29	an1f1s2.c23	an1f1s3.c33	an1f1s6.c07	an1f2s3.c38
...								
...								

The complete randomization tables are contained in subfolder "Randomization tables" in archive ts_103624v010201p0.zip which accompanies the present document.

B.3 Instructions for the subjective tests

B.3.1 P.800 ACR test instructions in English

In this experiment, you will be listening to short samples via headphones, and giving your opinion of the speech you hear.

Follow the instructions on the touchscreen in front of you, and listening to each sample, press the appropriate button to indicate your opinion on the following scale.

WHAT WAS THE QUALITY OF THE SAMPLE YOU HAVE JUST HEARD?

5 Excellent

4 Good

3 Fair

2 Poor

1 Bad

B.3.2 P.800 DCR test instructions in English

In this experiment, you will be listening to pairs of speech samples via headphones, and giving your opinion on the quality of the second sample compared to the quality of the first sample.

Follow the instructions on the touchscreen in front of you, and listening to each sample pairs, press the appropriate button to indicate your opinion on the following scale:

WHAT WAS THE QUALITY DEGRADATION OF THE SECOND SAMPLE COMPARED TO THE FIRST SAMPLE IN THE PAIR YOU HAVE JUST HEARD?

5 Degradation is inaudible

4 Degradation is audible but not annoying

3 Degradation is slightly annoying

2 Degradation is annoying

B.3.3 P.800 ACR test instructions in Mandarin Chinese

ACR主观测试示例说明

在本实验中，我们将对用于电信业务的语音频编解码系统进行评估。

实验包含若干个小节。在每一小节中，您将通过耳机听到一个音频。每个样本包含 同一个人说的两句话。请您认真听完整个样本，然后根据您对样本整体音质的感受，按下面的5分制进行打分：

5 非常好

4 好

3 一般

2 较差

1 很差

请注意：您的任务是对语音样本的整体音质进行评价。

在您听完一个样本后，请您对刚才听到的样本的整体音质按上述的评分方法进行打分。

您打完分之后，会有一个短暂的停顿，然后播放下一个样本。

请您在实验中不要和其他测听者讨论您打的分数。

B.3.4 P.800 DCR test instructions in Mandarin Chinese

DCR主观测试示例说明

在本实验中，我们将对用于电信业务中的语音频编解码系统进行评估。

实验包含若干个小节。每一小节中包含两个样本，每个样本为一个音频样本。每个小节中的第一个样本是参考样本，第二个样本的内容与第一个样本的内容完全相同，但它是经过了一个电信系统后得到的。请您认真听完每一小节中的两个样本，然后根据您察觉到的第二个样本相对于第一个样本的整体音质失真程度，按下面的5分制进行打分：

5 听不出音质失真

4 能听到音质失真，但不令人厌烦

3 能听到音质失真，且有一点令人厌烦

2 能听到音质失真，且令人厌烦

1 能听到音质失真，且非常令人厌烦

请注意：您的任务是对您所察觉到的第二个待测样本相对于第一个参考样本的失真程度给出相应的分数。请您在听完两个样本之后，按上述的评分方法进行打分。

请您在实验中不要和其他测听者讨论您打的分数。

B.4 Processing functions for the ETSI LC3plus codec characterization phase

B.4.1 Introduction

This clause and the following clauses define how audio material shall be prepared for the characterization tests of the ETSI LC3plus codec and provides details on the software modules and files required.

B.4.2 Definitions and formats

The following filename convention is used for source material:

Input<Fs> Fs (in kHz) is the input sampling frequency to a processing module (or sequence of modules). Fs is either 8, 16, 32 or 48.

Output<Fs> Fs (in kHz) is the output sampling frequency from a processing module (or sequence of modules). Fs is either 8, 16, 32 or 48.

The format of the above files is headerless PCM with samples stored in 16-bit 2's complements little endian format. Other command line filenames and variables are described in the relevant part of the present document.

The codecs use the Recommendation ITU-T G.192 [15] format as a common bit-stream interface, if available.

All executables are in the Win32 format, i.e. are native binaries for the 32-bit Microsoft Windows™ platform.

B.4.3 Processing Stages for the LC3plus codec characterization phase

B.4.3.1 Introduction

This clause and the following clauses define, in the form of diagrams, the processing stages required by the LC3plus codec under test. The latest version of the ITU-T Software Tool Library as provided in Recommendation ITU-T G.191 [6] is used for this processing.

The source material shall be 48 kHz sampled with 16 bit resolution. The format shall be headerless PCM, little endian. For full-band experiments the source material shall have frequency components above 16 kHz.

B.4.3.2 Source material pre-processing requirements

B.4.3.2.1 Input speech file naming

The filenames of the input speech samples are represented by:

aeegysz.48k

where:

- *ee* stands for the experiment number, e.g. n1 (see table B.3)
- *g* is gender of talker (i.e. f for female and m for male) and *y* is the talker number: 1, 2, 3
- *s* stands for sample and z is the sample number; 1, 2, 3, 4, 5, .., x

B.4.3.2.2 Processed speech file naming

The filenames of the processed speech samples are represented by:

aeegysz.cnn

where:

- *ee* stands for the experiment designator, e.g. n1 (see table 1)
- *g* is gender of talker (i.e. f for female and m for male) and *y* is the talker number: 1, 2, 3
- *s* stands for sample and z is the sample number; 1, 2, 3, 4, 5, .., x
- *c* stands for the condition with the number nn = 01, 02, .. as specified in this annex.

B.4.3.3 File concatenation, separation, sequences and module initialization

B.4.3.3.1 Concatenated sequences processing

In all experiments, the pre-processed material will be processed in concatenated files comprising a preamble and a series of sentences for speech. The preamble will be 10 s long.

B.4.3.3.2 Preamble definition

The preamble in the speech path shall not be digital silence (i.e. 16 bit samples equal to zero), but a low-level random noise with amplitude between +4 and -4.

B.4.3.3.3 Frame error application

In frame erasure conditions, the erasures shall affect the same segments of speech signal for the all codecs in the test. This shall be done by compensating for all encoder-side delays (or for all encoder-side delays of the core layer in case of embedded codecs). Some codecs, e.g. EVS or LC3plus, have encoder and decoder delay compensation implemented in the executable. For all other codecs, delay shall be compensated prior to the processing by the reference encoder, as specified in table B.4.

NOTE: Exact delay compensation might not be possible.

B.4.3.3.4 File naming for error patterns

Random frame erasure files will have the name "*patterns\eeEPFt1_rr_t2.g192*", where:

- *ee* stands for the experiment designator, e.g. n1 (see table B.3)
- *t1* stands for original frame size of pattern

- *rr* stands for the frame loss rate in per cent, e.g. 06, or 10
- *t2* stands for adapted frame size of pattern, e.g. duplicated to align 20 ms to 10 ms frame size

Bursty frame erasure files will have the name "*patterns\eePLP_rr_t1.g192*", where:

- *ee* stands for the experiment designator, e.g. n1 (see table 1)
- *rr* stands for the frame loss rate frame in a row, i.e. 1, 2, 4 or 7
- *T1* stands for frame size of pattern

B.4.3.4 File naming for rate switching profiles

Rate switching profiles will have the name "*patterns\eersx*", where:

- *ee* stands for the experiment designator, e.g. n1 (see table B.3)
- *rs* stands for rate switching file
- *x* stands for the sequence in the experiment, i.e. 1, 2, etc.

B.4.3.5 Concatenation setup

Audio files are concatenated after level adjustment. For concatenation order, see table B.3 below where all samples are concatenated from left to right and line by line. For speech files, the concatenation order is therefore *m1s1*, *f1s1*, *m2s1*, *f2s1*, … *mYs1*, *fYs1*, *m1s2*, *f1s2*, etc.. The final concatenated file consists of the preamble and the audio file.

Table B.3: Concatenation order of speech files

Sample	**Talker sequence in concatenated files**							
1	m1	f1	m2	f2	..	mY	fY
2	m1	f1	m2	f2	..	mY	fY
…							
x							

For all experiments, the processed concatenated files should be divided into separated audio samples and named with the according file extension as specified in the test plan. This shall be performed before final up-sampling stage, and a cosine (Hanning) window of duration 100 ms shall be applied to the start and end of the separated files.

B.4.3.6 Processing for all conditions

B.4.3.6.1 Overview

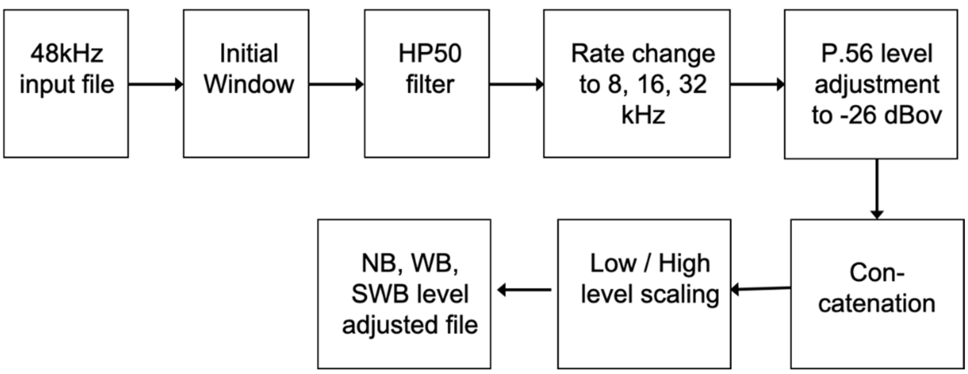

Figure B.1: Processing for all conditions

B.4.3.6.2 General processing stages

B.4.3.6.2.1 Direct conditions

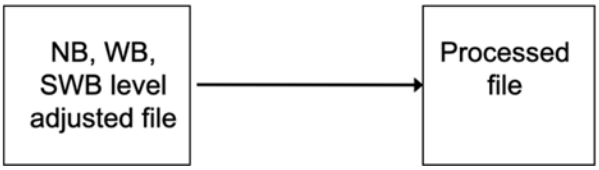

Figure B.2: Processing for direct conditions

B.4.3.6.2.2 Processing for MNRU reference conditions

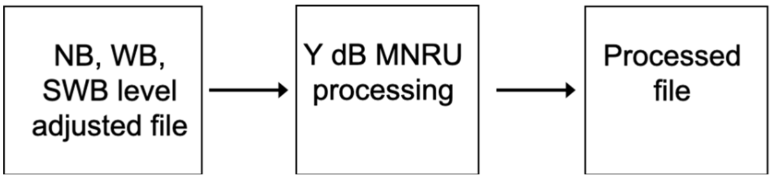

Figure B.3: Processing of MNRU anchors at Y dB

B.4.3.6.2.3 Codec conditions

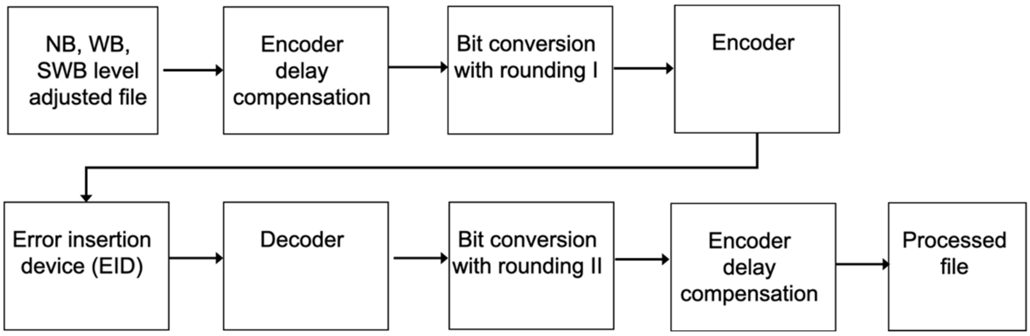

Figure B.4: Processing of codec conditions

Table B.4: Codec parameter table

Codec	Supported input	Encoder delay	Bit conversion	Decoder delay / ms
LC3plus	NB, WB, SWB	0	none	0
G.726	NB	0	13 bits	0
G.711	NB	0	13 bits	0
G.722	WB	11	14 bits	11
AMR-NB	NB	40	13 bits	0
AMR-WB	WB	80	14 bits	15
EVS	NB, WB, SWB	0	none	0
OPUS	NB, WB, SWB	0	none	0
NOTE: OPUS expects 48 kHz as SWB input, see clause B.4.4.3.3.7.				

B.4.3.6.2.4 Tandeming conditions

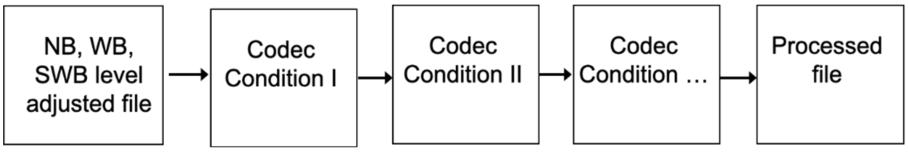

Figure B.5: Processing of tandemed codecs

The two or more codec stages as outlined in clause B.4.3 are processed.

NOTE: No encoder/decoder delay compensations are applied.

B.4.3.7 Post-processing

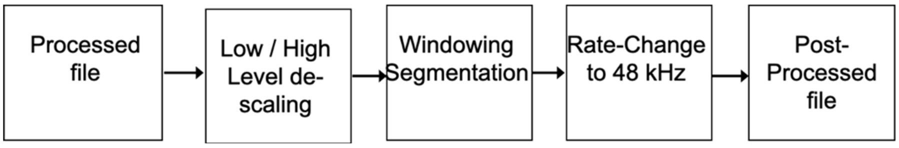

Figure B.6: Post-processing

B.4.3.8 Processing for multi-bandwidth conditions

For experiments containing conditions with multiple audio bandwidths, the relevant pre-processing steps for a dedicated maximum bandwidth are conducted. Some codecs need to be adapted to their native bandwidth by resampling steps.

In the native sample rate domain, the regular encoding, decoding and other processing steps for the codecs can be applied. Other processing steps might be delay compensation or error insertion.

B.4.4 Processing modules

B.4.4.1 Introduction

This clause and the following clauses describe the modules that shall be used in the processing of speech.

B.4.4.2 Pre- and post processing operations

B.4.4.2.1 General delay compensation for the STL filter tool

All filtering steps include a delay compensation step. For preparing the delay compensation, samples of the preamble are added to the end of the input file before applying the filter step. After completion of the filtering step, the samples are to be removed from the beginning of the filtered file.

B.4.4.2.2 Filtering operations

To produce a 50 Hz high pass filtered 48 kHz sampling file use:

`filter.exe HP50_48KHZ` *`Input48 Output48`* `960`

B.4.4.2.3 Recommendation ITU-T P.56 active speech level adjustment

To normalize the P.56 ASL of an 8 kHz sampling file to -26 dBov, use:

`sv56demo.exe -lev -26 -sf 8000` *`Input8 Output8`* `160`

To normalize the P.56 ASL of a 16 kHz sampled file to -26 dBov, use:

```
sv56demo.exe -lev -26 -sf 16000 Input16 Output16 320
```

To normalize the P.56 ASL level of a 32 kHz sampled file to -26 dBov, use:

```
sv56demo.exe -lev -26 -sf 32000 Input32 Output32 640
```

In addition to generate artificially mixed content, to normalize the P.56 ASL level of a 48 kHz sampled file to -26 dBov, use:

```
sv56demo.exe -lev -26 -sf 48000 Input48 Output48 960
```

B.4.4.2.4 Low/high level de-/scaling

In case the speech level shall be -26 dBov, this processing step shall be bypassed. Otherwise, the following gain needs to be applied to the signal:

Table B.5: Scaling and descaling

Speech level	Scaling	Descaling
-16 dBov	G=10	G=-10
-36 dBov	G=-10	G=10

To apply the scaling operation with the gain G, use:

```
scaldemo.exe -dB -gain G -bits 16 -round –nopremask -blk BBB Input Output
```

where `BBB` is 160 for 8 kHz sampled files, 320 for 16 kHz sampled files and 640 for 32 kHz sampled files.

B.4.4.2.5 File concatenation

To concatenate files, the concat command is used:

```
concat.exe –undo undo_concat.txt file1 [file2 file3 …] catfile
```

Where `file1, file2,` … are the files to be concatenated and `catfile` is the concatenated file. The `undo_concat.txt` contains the parameters for segmentation.

In case additional silence sections need to be inserted before concatenation, use the following for all input files:

```
concat.exe silence fileX silence tmp_file
copy tmp_file fileX
```

where the silence file contains 0,2 s of digital silence and `fileX` is one of the input files.

B.4.4.2.6 Audio format conversion

B.4.4.2.6.1 PCM to WAV conversion

To convert a 32 kHz sampled PCM file to a WAV file for LC3plus, use:

```
CopyAudio.exe -F WAVE-NOEX -P integer16,0,32000,native,1,default -I "" input32
output32.wav
```

B.4.4.2.6.2 WAV to PCM conversion

To convert a WAV file to a 32 kHz sampled PCM file for LC3plus, use:

```
CopyAudio.exe -F noheader -D integer16 input32.wav output32
```

B.4.4.2.7 Sampling rate changes

B.4.4.2.7.1 Rate-change from 48 kHz to 8 kHz sampling

To produce an 8 kHz sampling file from a 48 kHz sampling file, use:

```
filter.exe –down SHQ3 input48 tmp16 960
filter.exe –down SHQ2 tmp16 output8 320
```

B.4.4.2.7.2 Rate-change from 48 kHz to 16 kHz sampling

To produce a 16 kHz sampling file from a 48 kHz sampling file, use:

```
filter.exe –down SHQ3 input48 output16 960
```

B.4.4.2.7.3 Rate-change from 48 kHz to 32 kHz sampling

To produce a 32 kHz sampling file from a 48 kHz sampling file, use:

```
filter.exe –up SHQ2 input48 tmp96 960
filter.exe –down SHQ3 tmp96 output32 1920
```

B.4.4.2.7.4 Rate-change from 8 kHz to 48 kHz sampling

To produce a 48 kHz sampling file from an 8 kHz sampling file, use:

```
filter.exe –up SHQ2 input8 tmp16 160
filter.exe –up SHQ3 tmp16 output48 320
```

B.4.4.2.7.5 Rate-change from 16 kHz to 48 kHz sampling

To produce a 48 kHz sampling file from a 16 kHz sampling file, use:

```
filter.exe –up SHQ3 input16 output48 320
```

B.4.4.2.7.6 Rate-change from 32 kHz to 48 kHz sampling

To produce a 48 kHz sampling file from a 32 kHz sampling file, use:

```
filter.exe –up SHQ3 input32 tmp96 640
filter.exe –down SHQ2 tmp96 output48 1920
```

B.4.4.2.7.7 Rate-change from 32 kHz to 16 kHz sampling

To produce a 16 kHz sampling file from a 32 kHz sampling file, use:

```
filter.exe –down SHQ2 input32 output16 640
```

B.4.4.2.7.8 Rate-change from 32 kHz to 8 kHz sampling

To produce a 8 kHz sampling file from a 32 kHz sampling file, use:

```
filter.exe –down SHQ2 input32 output16 640
filter.exe –down SHQ2 output16 output8 320
```

B.4.4.2.7.9 Rate-change from 16 kHz to 32 kHz sampling

To produce a 32 kHz sampling file from a 16 kHz sampling file, use:

```
filter.exe –up SHQ2 input16 output32 320
```

B.4.4.2.7.10 Rate-change from 8 kHz to 32 kHz sampling

To produce a 32 kHz sampling file from a 8 kHz sampling file, use:

```
filter.exe –up SHQ2 input8 output16 160
filter.exe –up SHQ2 output16 output32 320
```

B.4.4.2.7.11 Rate-change from 16 kHz to 8 kHz sampling

To produce a 8 kHz sampling file from a 16 kHz sampling file, use:

```
filter.exe –down SHQ2 input16 output8 320
```

B.4.4.2.8 Windowing and segmentation

B.4.4.2.8.1 Segmentation for NB conditions

To extract and window an `m` sample long file beginning at sample `s` from a 8 kHz single channel concatenated file, use:

```
astrip.exe –sample –smooth –wlen 800 –start s -n m input8 output8
```

B.4.4.2.8.2 Segmentation for WB conditions

To extract and window an `m` sample long file beginning at sample `s` from a 16 kHz single channel concatenated file, use:

```
astrip.exe –sample –smooth –wlen 1600 –start s -n m input16 output16
```

B.4.4.2.8.3 Segmentation for SWB conditions

To extract and window an `m` sample long file beginning at sample `s` from a 32 kHz single channel concatenated file, use:

```
astrip.exe –sample –smooth –wlen 3200 –start s -n m input32 output32
```

B.4.4.2.8.4 Initial windowing

To apply the initial windowing of a 48 kHz input speech file, use:

```
astrip.exe –sample –smooth –wlen 4800 –start s -n m input48 output48
```

B.4.4.2.9 Bit conversion

B.4.4.2.9.1 Conversion from 16 bit to 13 bit

To convert an 8 kHz narrowband signal from a 16 bit representation to a 13 bit representation including rounding, use:

```
scaldemo.exe -lin -gain 1 -bits 13 -round -nopremask -blk 160 input8 output8
```

B.4.4.2.9.2 Conversion from 16 bit to 14 bit

To convert a 16 kHz wideband signal from a 16 bit representation to a 14 bit representation including rounding, use:

```
scaldemo.exe -lin -gain 1 -bits 14 -round -nopremask -blk 320 input16 output16
```

B.4.4.2.10 Delay compensation for filter operations

The processing steps are delay-compensated in order to apply error insertion on the same parts of the audio signal and to be able to extract the original length and offset for each audio sample used in the tests.

The delay compensation is initialized by concatenating the file to be filtered and the first 960 samples of the preamble. After processing, the delay of the processing operation is compensated and the original file length is restored.

To compensate the delay for filter operations and reference conditions for encoder and decoder in the common scripts, use:

```
astrip –sample –start S+1 –n FILELENGTH input output
```

where `FILELENGTH` denotes the size in samples and the value for S for each filtering operation is given in table B.6.

Table B.6: Delay compensation values for filter operations

Filter operation	Value for delay compensation after filtering operation
-up SHQ2	436
-up SHQ3	436
-down SHQ2	218
-down SHQ3	145
HP50_48KHZ	839
MSIN	92

B.4.4.3 Processing

B.4.4.3.1 Introduction

This clause and the following clauses describe the input file and channel operations processing that shall be used in the preparation of audio material.

B.4.4.3.2 MNRU reference conditions

To generate an MNRU reference at XXX dB for a file *Input8*, use:

```
mnrudemo.exe –Q XXX Input8 Output8 160
```

To generate an MNRU reference at XXX dB for a file *Input16*, use:

```
mnrudemo.exe –Q XXX Input16 Output16 320
```

To generate a P.50 MNRU reference at XXX dB for a file *Input48*, use:

```
p50mnru.exe Input48 Output48 XXX M
```

NOTE: Recommendation ITU-T P.50 [17] MNRU processing for SWB conditions requires rate-change steps between 32 kHz and 48 kHz.

B.4.4.3.3 Encoder and decoder

B.4.4.3.3.1 AMR-NB

To process a file *Input8* through the AMR-NB codec at XXX kbit/s, use:

```
amr_cod_vad2.exe [-dtx] BBB Input8 bitstream
amr_dec.exe bitstream Output8
```

where `BBB` is the bitrate mode corresponding to XXX as given in the following table.

XXX [kbit/s]	4.75	7.4	7.95	10.2	12.2
BBB	MR475	MR74	MR795	MR102	MR122

NOTE: AMR-NB operates at 20 ms frame length and has encoder and decoder delay of 5 ms and 0 ms, respectively.

B.4.4.3.3.2 G.711

To process a file *Input8* through the G.711 codec with 20 ms blocks, use:

```
g711demo.exe A lili Input8 Output8 160
```

where A stands for A-law.

NOTE: G.711 does not have algorithmic delay.

B.4.4.3.3.3 G.726

To process a file *Input8* through the G.726 codec with 10 ms blocks, use:

```
g711demo.exe A lilo Input8 out.g711 80
g726demo.exe A load 32 out.g711 bitstream 80

g726demo.exe A adlo 32 bitstream out.g726 80
g711demo.exe A loli out.g726 Output8 80
```

where A stands for A-law.

B.4.4.3.3.4 AMR-WB

To process a file *Input16* through the AMR-WB codec at XXX kbit/s, use Recommendation ITU-T G.722.2 [19] as:

```
amrwb_cod.exe [–dtx] –itu BBB Input16 bitstream
amrwb_dec.exe -itu bitstream Output16
```

where BBB is the bitrate mode corresponding to XXX as given in the following table.

XXX [kbit/s]	6.6	8.85	12.65	14.25	15.85	18.25	19.85	23.05	23.85
BBB	0	1	2	3	4	5	6	7	8

NOTE: AMR-WB operates at 20 ms frame length and has encoder and decoder delay of 5 ms and 0,9375 ms, respectively.

B.4.4.3.3.5 G.722

To process a file *Input16* through the G.722 codec at XXX kbit/s using 20 ms block size, use:

```
encg722.exe –fsize 320 –mode XXX Input16 bitstream
decg722.exe -fsize 320 –mode XXX bitstream Output16
```

where XXX is either 56 or 64.

NOTE: G.722 operates at 20 ms frame length and has an overall algorithmic delay of 1.625 ms.

B.4.4.3.3.6 EVS

To process a file *Input8* through the EVS codec at XXX bit/s, use:

```
EVS_cod.exe [-dtx] [-no_delay_cmp] XXX/SWF 8 Input8 bitstream
EVS_dec.exe [-no_delay_cmp] 8 bitstream Output8
```

where XXX is one of 5900, 7200, 8000, 9600 or 13200. For rate switching operation, XXX is replaced by a switching file (SWF).

To process a file *Input16* through the EVS codec at XXX bit/s, use:

```
EVS_cod.exe [-dtx] [-no_delay_cmp] XXX/SWF 16 Input16 bitstream
EVS_dec.exe [-no_delay_cmp] 16 bitstream Output16
```

where XXX is one of 5900, 7200, 8000, 9600, 13200, 16400, 24400, 32000, 48000, 64000, 96000 for testing non-IO modes or 6600, 8850, 12650, 14250, 15850, 18250, 19850, 23050, 23850 for testing AMR-WB IO modes. For rate switching operation, XXX is replaced by a switching file (SWF).

To process a file *Input32* through the EVS codec at XXX bit/s, use:

```
EVS_cod.exe [-dtx] [-no_delay_cmp] XXX/SWF 32 Input32 bitstream
EVS_dec.exe [-no_delay_cmp] 32 bitstream Output32
```

where XXX is one of 13200, 16400, 24400, 32000, 48000, 64000, 96000, 128000. For rate switching operation, XXX is replaced by a switching file (SWF).

To process a file *Input48* through the EVS codec at XXX bit/s, use:

```
EVS_cod.exe [-dtx] [-no_delay_cmp] –max_band FB XXX/SWF 48 Input48 bitstream
EVS_dec.exe [-no_delay_cmp] 48 bitstream Output48
```

where XXX is one of 16400, 24400, 32000, 48000, 64000, 96000, 128000. For rate switching operation, XXX is replaced by a switching file (SWF).

The switching file consists of XXX values indicating the bit rate for each frame in bit/s. These values are stored in binary format using 4 byte per value.

NOTE: EVS encoder and decoder provide delay compensated output files except if used for tandem conditions, where the option "-no_delay_cmp" is to be enabled.

B.4.4.3.3.7 OPUS

To process an *Input8* file through the Opus codec at XXX bit/s, use:

```
opus_demo.exe restricted-lowdelay 8000 1 XXX -cbr -bandwidth NB
[-no_delay_comp] [-epf patternFile] -framesize 10 -complexity 0 Input8 Output8
```

To process an *Input16* file through the Opus codec at XXX bit/s, use:

```
opus_demo.exe restricted-lowdelay 16000 1 XXX -cbr -bandwidth WB
[-no_delay_comp] [-epf patternFile] -framesize 10 -complexity 0 Input16 Output16
```

To process an *Input32* file through the Opus codec at XXX bit/s, apply rate-change from 32kHz to 48kHz to generate *Input48* and use:

```
opus_demo.exe restricted-lowdelay 48000 1 XXX -cbr -bandwidth FB
[-no_delay_comp] [-epf patternFile] -framesize 10 -complexity 0 Input48 Output48
```

And resample *Output48* to *Output32*

Where *patternFile* indicates an error pattern file in 16 bit format triggering concealment.

NOTE 1: patternFile is not in G.192 format.

NOTE 2: Opus encoder and decoder provide delay compensated output files except if used for tandem conditions, where the option "-no_delay_comp" is to be enabled. The delay compensation is separated into half delay portion for encoder and half for decoder to align the bit stream comparable to all other codecs for best PLC testing.

B.4.4.3.3.8 LC3plus

Convert the *Input8*, *Input16*, *Input24*, *Input32*, *Input48 to Input.wav*.

To process an *Input.wav* file through the LC3plus codec at XXX bit/s, use:

```
LC3plus.exe -v -E -formatG192 -cfgG192 CONFIG.txt -d D -frame_ms F [-epmode EP]
Input.wav bitstream
LC3plus.exe -v -E -formatG192 -cfgG192 CONFIG.txt -d D [-epmode EP] bitstream
Output.wav
```

Where *F* indicated the used frame duration in ms, i.e. 10, 5 or 2.5. *EP* indicates the used error protection class on encoder side, i.e. 1 to 4, and enabled epmode on decoder side.

Convert the *Output.wav* to *Output8, Output16, Output32, Output24*.

NOTE: LC3plus encoder and decoder provide delay compensated output files except if used for tandem conditions, where the option "-d 0" is enabled. The delay compensation is separated into half delay portion for encoder and half for decoder to align the bit stream comparable to all other codecs for best PLC testing, indicated by "-d 2".

B.4.4.4 Error Insertion Device (EID)

B.4.4.4.1 Introduction

For the conditions where random frame erasures are desired, frame erasure patterns are applied to the bitstream using tools from Recommendation ITU-T G.191 [6].

B.4.4.4.2 Frame error tool

For the G.711 A/μ-law, the following processing shall be used:

```
eid-int –ep g192 –factor 2 ep.g192 ep10.g192
g711iplc.exe ep10.g192 input.8k ouput.8k
```

where:

`ep.g192` is the error pattern file assuming 20 ms frames and `ep10.g192` is the error pattern file for the g711iplc tool where all entries are doubled to take the 10 ms frame grid of the g711iplc tool into account

`input.8k` is the G.711 decoded output file

`output.8k` is the G.711 decoded output file with packet loss concealment

For the G.726, the following processing shall be used:

```
G726_applyBE input epBER.g192 output
G726_applyFL input epPLC.g192 output
```

where:

`epBER/PLC.g192` is the error pattern file assuming 20 ms frames.

For all other reference codecs and LC3plus, the following processing shall be used:

`eid-xor.exe –vbr –ber g192bsin epBER.g192 g192bsout` (for bit errors)

`eid-xor.exe –vbr –fer g192bsin epPLC.g192 g192bsout` (for packet loss)

where:

`g192bsin` is the input bit stream

`epBER.g192` is the error pattern file indicating bit errors `epPLC.g192` is the error pattern file indicating frame losses

`g192bsout` is the output bit stream

B.4.4.4.3 Pattern generation

B.4.4.4.3.1 Random pattern

The error patterns used are generated using the gen-patt tool as follows:

```
gen-patt.exe -tailstat -fer -g192 -gamma 0 -rate XXX -tol 0.001      -reset -n
LENGTH -start 501 ep.g192
```

where `XXX` is the required erasure rate, e.g. 0.03 for 3 % and 0.06 for 6 % FER.

Different error patterns should be generated for each experiment. `LENGTH` is the number of 20 ms or 10 ms frames indicating the length of the whole input file (preamble and concatenated speech). Patterns are generated at highest erasure rate and lower erasures rates are generated by randomly changing bad frames to good frames.

B.4.4.4.3.2 Bursty pattern (PLP)

A random pattern is generated for the erasure rate 1,43 % and 20 ms frame size with the file name PLP_1_20_g192. To generate:

- PLP_2_20_g192: Each bad frame is followed by one additional bad frames
- PLP_4_20_g192: Each bad frame is followed by three additional bad frames
- PLP_7_20_g192: Each bad frame is followed by six additional bad frames

B.4.4.4.3.3 Converting pattern

To adapt to different frame sizes of codecs, e.g. from 20 ms to 10 ms, use:

```
eid-int.exe -ep g192 -factor 2 PLP_2_20_g192 PLP_2_10_g192
```

B.4.4.4.3.4 DECT RSSI pattern

The patterns are generating by randomly selecting frame out of measured DECT error profiles given a certain Receive Side Signal Indication (RSSI). The generated pattern for bit errors is denoted as "epSS_PPBER.192" and the pattern for frame loss is denoted as "epSS_PPPLC.192" where SS indicates the RSSI 80, 56, 48, 40 and PP is the payload length in bytes.

B.4.4.4.4 Switching profile generation

To generate EP mode switching profiles for LC3plus conditions, use:

```
gen-rate-profile.exe –layers B1,B2,…,Bx SWF 10 B1 LENGTH SEEDx
```

where `B1,B2,…,Bx` are all EP modes rates starting from the lowest one (`B1`) up to the highest one (`Bx`) multiplied by 100, i.e. 100, 200, 300, 400 for EP mode 1,2,3,4.

B.4.5 Binaries used in characterization phase

B.4.5.1 Introduction

All binaries are compiled and tested under Win32 platforms. The following clauses document the origin and the compilation of the binaries.

B.4.5.2 Tools

Table B.7: Recommendation ITU-T G.191

Source	Recommendation ITU-T G.191 [6] S4-120344 "Filter masks for EVS testing"
URL	http://ftp.3gpp.org/tsg_sa/WG4_CODEC/TSGS4_68/Docs/S4-120344.zip
Version / Release	G.191 [6]
Description	Software tools for speech and audio coding standardization
Comments	G.191 filter tool patched with S4-120344 to enable support for HP50 and SHQ filter
Executables	`oper, astrip, concat, sv56demo, filter, scaldemo, actlev, eid-xor, gen-patt, mnrudemo, gen-rate-profile, eid-int`
Status	Available

Table B.8: AMR-NB Error Insertion

Source	Orange: S4-120998
URL	http://ftp.3gpp.org/tsg_sa/WG4_CODEC/TSGS4_70/Docs/S4-120998.zip
Version / Release	-
Description	Error insertion device for AMR-NB bit streams
Comments	
Executables	`eid-amr`
Status	Available

Table B.9: WAV - PCM Converter

Source	McGill University, Telecommunications & Signal Processing Laboratory, Audio File Programs and Routines
URL	http://www-mmsp.ece.mcgill.ca/Documents/Downloads/AFsp/AFsp-v9r0.tar.gz
Version / Release	Release v9r0; Software version of CopyAudio: v6r0 2003-05-08
Description	CopyAudio tool from the AFsp
Comments	
Executables	`CopyAudio.exe`
Status	Available

Table B.10: SWB MNRU

Source	NTT: AHEVS-165
URL	http://ftp.3gpp.org/tsg_sa/WG4_CODEC/Ad-hoc_EVS/Docs/AHEVS-165.zip
Version / Release	-
Description	Recommendation ITU-T P.50 [17] MNRU
Comments	
Executables	`p50mnru.exe`
Status	Available

Table B.11: Randomization Tool

Source	S4-121078
URL	http://ftp.3gpp.org/tsg_sa/WG4_CODEC/TSGS4_70/Docs/S4-121078.zip
Version / Release	-
Description	Tool for providing all randomizations depending on a master seed
Comments	
Executables	`random.exe`
Status	Available

B.4.5.3 Codecs

Table B.12: AMR-NB

Source	3GPP TS 26.073: ANSI-C code for the Adaptive Multi Rate (AMR) speech codec
URL	-
Version / Release	ETSI TS 126 073 [9]
Description	AMR narrow band fix point encoder and decoder software
Comments	Compiled with VAD version 1 and 2
Executables	`amr_cod_vad1.exe, amr_cod_vad2.exe, amr_dec.exe`
Status	Available

Table B.13: G.711

Source	Recommendation ITU-T G.191 [6]
URL	-
Version / Release	Version 3.3 of 02.Feb. 2010
Description	G.711 codec
Comments	
Executables	`g711demo.exe, g711iplc.exe`
Status	Available

Table B.14: G.726

Source	Recommendation ITU-T G.191 [6]
URL	-
Version / Release	Version 1.4 of 03. Feb.2010
Description	G.726 codec
Comments	
Executables	`g726demo.exe`
Status	Available

Table B.15: AMR-WB

Source	3GPP TS 26.173: ANSI-C code for the Adaptive Multi-Rate - Wideband (AMR-WB) speech codec
URL	-
Version / Release	ETSI TS 126 173 [8]
Description	AMR-WB fixed point encoder and decoder software
Comments	
Executables	`amrwb_cod.exe, amrwb_dec.exe`
Status	Available

Table B.16: G.722

Source	Recommendation ITU-T G.191 [6]
URL	-
Version / Release	COPYRIGHT CNET LANNION A TSS/CMC Date 24/Aug/90 COPYRIGHT Ericsson AB. Date 22/May/06 COPYRIGHT France Telecom R&D Date 23/Aug/06
Description	G.722 encoder and decoder software
Comments	
Executables	`encg722.exe, decg722.exe`
Status	Available

Table B.17: EVS

Source	3GPP 26.441
URL	-
Version / Release	ETSI TS 126 441 [18]
Description	EVS fix-point encoder and decoder software
Comments	
Executables	`EVS_cod.exe, EVS_dec.exe`
Status	Available

Table B.18: OPUS

Source	Xiph.org
URL	https://ftp.osuosl.org/pub/xiph/releases/opus/opus-1.3.tar.gz
Version / Release	1.3
Description	Opus codec software compiled in fix-point arithmetic
Comments	Option to change delay alignment; added EPF support
Executables	`opus_demo_1_3_0`
Status	Available

Table B.19: LC3plus

Source	ETSI
URL	-
Version / Release	ETSI TS 103 634 [5]
Description	LC3plus codec software compiled in fix-point arithmetic
Comments	
Executables	`LC3plus`
Status	Available

Annex C (normative): Results of the subjective characterization test for the ETSI LC3plus codec

C.1 Introduction

This annex presents a report on subjective testing performed. MESAQIN.com performed the experiment according to the procedures and test plan specified in the main part of the present document. No deviations from or exceptions to the listening test procedures and specifications described in the test plan have been observed or experienced.

The raw voting data along with their statistical processing (MOS calculation and conditional STD and CI95) are contained in subfolder "Raw voting data along with their statistical processing" in archive ts_103624v010201p0.zip which accompanies the present document.

C.2 Summary of tests conducted and results obtained

C.2.1 Experiments

MESAQIN.com performed 8 different experiments (i.e. 7 experiments in English, one experiment in Mandarin) within the project. The experiment name, methodology (as defined in Recommendation ITU-T P.800 [1]) and language for each experiment are listed in Table C.1.

Table C.1: Allocation of experiments and languages

Exp.	Type	Methodology	Language
1	NB clean	ACR	EN
2	NB error	ACR	EN
3	WB clean	ACR	EN
4	WB error	ACR	EN
5	SWB clean	DCR	EN
6	SWB error	ACR	EN
7a	FB_MB_fullscale	ACR	EN
7b	FB_MB_fullscale	ACR	MAN

C.2.2 Speech material

MESAQIN.com used processed speech material delivered by Fraunhofer Institute in 48-kHz sampled with 16-bit resolution uncompressed wav files. The processed speech material for all tests conformed to restrictions indicated in "Practical procedures for subjective testing", ITU-T Handbook, 2011 [14]. The list of conditions is available in clause 6 of the present document. No sample name blinding or anonymization has been applied as it was principally not needed.

C.2.3 Listening Environment

The tests were performed in an acoustically treated critical listening room in MESAQIN.com Prague laboratories that conforms to the requirements of Recommendation ITU-T P.800 [1] in full. Its background noise during the tests was below than 30 dB SPL(A) with no peaks in audible acoustic frequency range and its reverberation time (60 dB) is 185 ms.

Listening equipment conformed to specified requirements (Recommendation ITU-T P.800 [1] and clause 5 of the present document) in full. Diffuse-field equalized Sennheiser headphones HD-650 have been used for all experiments. All headphones used have been calibrated and verified before and after performed experiments as required by [1] and [14]. A professional digital voting device has been used to collect the votes.

C.2.4 Test instructions and Presentation order

The instructions for subjects in each of the languages tested by the MESAQIN.com are available in clause B.3. The instructions to subjects were available in written during the entire test sessions. During the training session of each test, the instructions were verbally explained and if needed briefly discussed by a dedicated expert person that was able to answer questions from the subjects in their native language.

Test duration never exceeded 1,5 hours per listening panel. Test duration comprised of up to 50 % of actual listening time and test overhead including administration, initial briefing, preliminaries, and breaks.

The playout loudness calibration used: 73 dB SPL equals to -26 dBov.

For each speech sample, the MOS-LQS following ACR methodology for Experiments 1 to 4 and 6 to 7 have been assessed. In Experiment 5, the MOS-LQS following DCR methodology has been assessed.

C.2.5 Test Subjects

For each listening test, only native speakers of the tested language were used. Only normal-hearing subjects have been used for the experiments. Their hearing normality has been verified by subject self-assessment questionnaire during the subject hiring phase and verified prior the session by expert test. Each experiment required 24 listeners (6 panels of 4 listeners each). The age and gender information for the set of subjects used in each tested language tested by the MESAQIN.com are provided in clause C.2.6

C.2.6 Age and gender information

Table C.2 provides age and gender information across all subjects within each language tested by MESAQIN.com.

Table C.2

Language	#females/#males	Mean Age (years)	Age StdDev (years)
English	1,00	33,27	11,44
Chinese	1,00	30,67	10,45

C.2.7 Raw data delivery and statistical analysis

All raw voting data have been delivered to the xls data delivery files prodided in the electronic attachment. They contain Means, Standard Deviations and CI95 of MOS-LQ for every condition within each test. Also, as required by in clause 7.1 of the present document, each condition in clauses 6.1 to 6.4 of the present document, labelled as requirement has been compared to the corresponding CuT condition using a Student's Dependent Groups t-test (single-sided at 95 % confidence level) on the subjective scores. This comparison has proven that the CuT meets or exceeds the requirements in all examined cases. Additionally, the same rule has been applied to the performance objectives, whereas the comparison between CuT and performance objective have only informative character. Also here the results shown that CuT meets or exceeds the performance objectives in all examined cases. The detailed results including all statistical evaluations are available in the digital attachment, separately for each experiment.

C.2.8 Discussion of any problems encountered during testing and the solution used to address the problem

During the tests no problems have been encountered.

Annex D (informative): Bibliography

- ETSI TS 126 071: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; Mandatory speech CODEC speech processing functions; AMR speech Codec; General description (3GPP TS 26.071)".

History

Document history		
V1.1.1	November 2019	Publication
V1.2.1	October 2021	Publication