|  |  |
| --- | --- |
| 3GPP TR 26.997 V0.2.0 (2024-05) | |
| Technical Report | |
| 3rd Generation Partnership Project;  Technical Specification Group Services and System Aspects;  Codec for Immersive Voice and Audio Services (IVAS);  Performance Characterization;  (Release 18) | |
|  | |
|  |  |
|  | |
| The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP. The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented. This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification. Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices. | |

|  |
| --- |
|  |
| ***3GPP***  Postal address  3GPP support office address  650 Route des Lucioles - Sophia Antipolis  Valbonne - FRANCE  Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16  Internet  http://www.3gpp.org |
| ***Copyright Notification***  No part may be reproduced except as authorized by written permission. The copyright and the foregoing restriction extend to reproduction in all media.  © 2024, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).  All rights reserved.  UMTS™ is a Trade Mark of ETSI registered for the benefit of its members  3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners LTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners  GSM® and the GSM logo are registered and owned by the GSM Association |

Contents

Foreword 7

1 Scope 8

2 References 8

3 Definitions of terms, symbols and abbreviations 9

3.1 Terms 9

3.2 Symbols 9

3.3 Abbreviations 9

4 General 10

4.1 Overview of the IVAS Codec Work Item 10

4.2 Presentation of the Following clauses 14

5 Terms of Reference 14

6 Selection Process 15

7 Introduction to the Testing of the IVAS codec 16

7.1 IVAS Selection Phase Testing 16

7.1.1 General methodology 16

7.1.1.1 Overview 16

7.1.1.2 P.800 16

7.1.1.3 BS.1534 16

7.1.2 Opinion Scales 17

7.1.2.1 ITU-T P.800 DCR 17

7.1.2.2 ITU-R BS.1534 18

7.1.3 Material 18

7.1.3.1 Overview 18

7.1.3.2 Speech Material for P.800 testing 19

7.1.3.3 Background Material 19

7.1.3.4 Music and Mixed Content Material for P.800 testing 19

7.1.3.5 Critical Generic Audio Items for BS.1534 testing 19

7.1.3.5.1 Steps of Critical Test Item Selection 19

7.1.3.5.2 Test Material 20

7.1.3.5.3 Training material 20

7.1.4 Listening Systems and Listening Environments 21

7.1.5 Experimental Procedure for P.800 experiments 21

7.1.6 Subjective Experiments 21

8 Important Notes about the Interpretation of Test Results 24

8.1 P.800 Testing 24

8.2 BS.1534 Testing 24

8.3 Analysis of Selection Phase Results 25

9 IVAS Performances 26

9.1 Mono 26

9.2 Stereo 26

9.2.1 Overview 26

9.2.2 Selection Experiment P800-1 (Stereo/Binaural, Clean Speech, Headphone Presentation) 27

9.2.3 Selection Experiment P800-2 (Stereo, Speech+Background, Headphone Presentation) 28

9.2.4 Selection Experiment P800-3 (Stereo, Mixed and Music, Headphone Presentation) 30

9.2.5 Selection Experiment BS1534-1a (Stereo, Generic Audio, 48 and 64 kbps, Headphone Presentation) 32

9.2.6 Selection Experiment BS1534-1b (Stereo, Generic Audio, 96 and 128 kbps, Headphone Presentation) 34

9.3 Scene-based Audio (SBA, Ambisonics) 35

9.3.1 Overview 35

9.3.2 Selection Experiment P800-4 (FOA, Clean Speech, Headphone Presentation) 35

9.3.3 Selection Experiment P800-5 (FOA, Speech+Background, Headphone Presentation) 37

9.3.4 Selection Experiment BS1534-4a (FOA, Generic Audio, 96, 128 and 160 kbps, Headphone Presentation) 39

9.3.5 Selection Experiment BS1534-4b (HOA2, Generic Audio, 160 and 192 kbps, Headphone Presentation) 41

9.3.6 Selection Experiment BS1534-5a (HOA3, Generic Audio, 192 and 256 kbps, Headphone Presentation) 42

9.3.7 Selection Experiment BS1534-5b (HOA3, Generic Audio, 384 and 512 kbps, 7.1+4 Loudspeaker Presentation) 44

9.4 Objects (Independent Streams with Metadata, ISM) 45

9.4.1 Overview 45

9.4.2 Selection Experiment P800-6 (1 Object, Clean speech, Headphone Presentation) 45

9.4.3 Selection Experiment P800-7 (2 Objects, Clean speech, Headphone Presentation) 47

9.4.4 Selection Experiment BS1534-6a (3 Objects, Generic Audio, 48, 64 and 96 kbps, Headphone Presentation) 49

9.4.5 Selection Experiment BS1534-6b (4 Objects, Generic Audio, 96, 128 and 256 kbps, Headphone Presentation) 51

9.5 Metadata-assisted Spatial Audio (MASA) 52

9.5.1 Overview 52

9.5.2 Selection Experiment P800-8 (MASA, Clean speech, Headphone Presentation) 52

9.5.3 Selection Experiment P800-9 (MASA, Speech+Background, Headphone Presentation) 54

9.5.4 Selection Experiment BS1534-7a (MASA, Generic Audio, 96 and 128 kbps, Headphone Presentation) 56

9.5.5 Selection Experiment BS1534-7b (MASA, Generic Audio, 192 and 256 kbps, Headphone Presentation) 58

9.6 Multi-Channel (MC) 59

9.6.1 Overview 59

9.6.2 Selection Experiment BS1534-2a (MC 5.1, Generic Audio, 64 and 96 kbps, 5.1 Loudspeaker Presentation) 59

9.6.3 Selection Experiment BS1534-2b (MC 5.1, Generic Audio, 128 and 160 kbps, 5.1 Loudspeaker Presentation) 61

9.6.4 Selection Experiment BS1534-3a (MC 7.1.4, Generic Audio, 128 and 160 kbps, 7.1+4 Loudspeaker Presentation) 62

9.6.5 Selection Experiment BS1534-3b (MC 7.1.4, Generic Audio, 384 and 512 kbps, 7.1+4 Loudspeaker Presentation) 64

9.7 Combined Objects and SBA (OSBA) 65

9.7.1 Overview 65

9.7.2 Optional additional information provided for codec selection 65

9.8 Combined Objects and MASA (OMASA) 67

9.8.1 Overview 67

9.8.2 Optional additional information provided for codec selection 67

9.9 Stereo operation with EVS compatible mono downmix stream 67

9.9.1 Overview 67

9.9.2 Optional additional information provided for codec selection 68

9.10 Rendering 69

9.11 Split Rendering 69

9.11.1 General 69

9.11.2 Overview 69

9.11.3 ISAR Selection Phase Requirements and Objectives 69

9.11.4 ISAR Selection Experiment ISAR\_BS1534-1 (HOA3, Generic Audio, 768 kbps Split Rendering Link, Headphone Presentation) 70

9.11.4.1 Overview 70

9.11.4.2 Test conditions 70

9.11.4.3 Result plots 70

9.11.4.4 Statistical analysis 72

9.11.4.5 Experimental conclusions 72

9.11.5 ISAR Selection Experiment ISAR\_BS1534-2 (MC 7.1+4, Generic Audio, 768 kbps Split Rendering Link, Headphone Presentation) 72

9.11.5.1 Overview 72

9.11.5.2 Test conditions 72

9.11.5.3 Result plots 72

9.11.5.4 Statistical analysis 74

9.11.5.5 Experimental conclusions 74

9.11.6 ISAR Selection Experiment ISAR\_BS1534-3 (4 Objects, Generic Audio, 768 kbps Split Rendering Link, Headphone Presentation) 74

9.11.6.1 Overview 74

9.11.6.2 Test conditions 74

9.11.6.3 Result plots 74

9.11.6.4 Statistical analysis 76

9.11.6.5 Experimental conclusions 76

9.11.7 ISAR Selection Experiment ISAR\_BS1534-4 (MASA 2TC, Generic Audio, 768 kbps Split Rendering Link, Headphone Presentation) 76

9.11.7.1 Overview 76

9.11.7.2 Test conditions 76

9.11.7.3 Result plots 76

9.11.7.4 Statistical analysis 78

9.11.7.5 Experimental conclusions 78

10 Objective Evaluations 78

10.1 Complexity and Delay Analysis 78

10.1.1 Complexity 78

10.1.2 Algorithmic Delay 78

Annex A: ToR Tests in Selection Phase 80

A.1 ToR Tests for Requirements 80

Annex B: Overall Characterization of the IVAS Codec 82

Annex C : IVAS Selection Phase Testplan 84

C.1 Experiment P800-1: Stereo/Binaural Clean Speech Test 84

C.1.1 Experiment setup 84

C.1.2 Content type categories and scene definitions 86

C.2 Experiment P800-2: Stereo Speech and Background Test 88

C.2.1 Experiment setup 88

C.2.2 Content type categories and scene definitions 89

C.3 Experiment P800-3: Stereo Mixed and Music Test 92

C.3.1 Experiment setup 92

C.4 Experiment P800-4: FOA Clean Speech Test 94

C.4.1 Experiment setup 94

C.4.2 Content type categories and scene definitions 95

C.5 Experiment P800-5: FOA Speech+background Test 97

C.5.1 Experiment setup 97

C.5.2 Content type categories and scene definitions 98

C.6 Experiment P800-6: 1-Object Clean Speech Test 100

C.6.1 Experiment setup 100

C.6.2 Content type categories and scene definitions 102

C.7 Experiment P800-7: 2-Objects Clean Speech Test 103

C.7.1 Experiment setup 103

C.7.2 Content type categories and scene definitions 106

C.8 Experiment P800-8: MASA Clean Speech Test 107

C.8.1 Experiment setup 107

C.8.2 Content type categories and scene definitions (Exp P800-8: Clean speech) 110

C.9 Experiment P800-9: MASA Speech+Background Test 111

C.9.1 Experiment setup 111

C.9.2 Content type categories and scene definitions (Exp P800-9: Speech+Background) 112

C.10 Experiment BS1534-1a: Stereo 114

C.11 Experiment BS1534-1b: Stereo 115

C.12 Experiment BS1534-2a: Multi-channel 5.1 117

C.13 Experiment BS1534-2b: Multi-channel 5.1 119

C.14 Experiment BS1534-3a: Multi-channel 7.1+4 120

C.15 Experiment BS1534-3b: Multi-channel 7.1+4 122

C.16 Experiment BS1534-4a: FOA 124

C.17 Experiment BS1534-4b: HOA2 126

C.18 Experiment BS1534-5a: HOA3 128

C.19 Experiment BS1534-5b: HOA3 129

C.20 Experiment BS1534-6a: Objects 130

C.21 Experiment BS1534-6b: Objects 131

C.22 Experiment BS1534-7a: MASA 133

C.23 Experiment BS1534-7b: MASA 134

Annex D: IVAS Characterization Phase Testplan 136

Annex E: IVAS Permanent Documents in 3GPP FTP-site 137

Annex F: Attachments 138

Annex G (informative): Change history 139

# Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (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 the present document, 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 document.

In the present document, modal verbs have the following meanings:

**shall** indicates a mandatory requirement to do something

**shall not** indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

**should** indicates a recommendation to do something

**should not** indicates a recommendation not to do something

**may** indicates permission to do something

**need not** indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

**can** indicates that something is possible

**cannot** indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

**will** indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**will not** indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**might** indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

**might not** indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

**is** (or any other verb in the indicative mood) indicates a statement of fact

**is not** (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

# 1 Scope

The present document provides information on the Immersive Voice and Audio Services (IVAS) codec Selection Phase which was run using the floating point (3GPP TS 26.258 and 3GPP TS 26.251). Experimental test results from the subjective quality testing are reported to illustrate the behaviour of the IVAS codec. Additional information is provided on implementation complexity of the IVAS codec and objective test results.

# 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 TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 26.250: "Codec for Immersive Voice and Audio Services (IVAS); General overview".

[3] void

[4] 3GPP TS 26.252: "Codec for Immersive Voice and Audio Services (IVAS); Test Sequences".

[5] 3GPP TS 26.253: "Codec for Immersive Voice and Audio Services (IVAS); Detailed algorithmic description incl. RTP payload format and SDP parameter definitions".

[6] 3GPP TS 26.254: "Codec for Immersive Voice and Audio Services (IVAS); Rendering".

[7] 3GPP TS 26.255: "Codec for Immersive Voice and Audio Services (IVAS); Error concealment of lost packets".

[8] 3GPP TS 26.256: "Codec for Immersive Voice and Audio Services (IVAS); Jitter buffer management".

[9] 3GPP TS 26.258: "Codec for Immersive Voice and Audio Services (IVAS); C code (floating-point)".

[10] IVAS-3 Permanent Document: "IVAS Performance Requirements".

[11] IVAS-4 Permanent Document: "IVAS Design Constraints".

[12] IVAS-5 Permanent Document: "IVAS Selection Rules".

[13] IVAS-6 Permanent Document: "IVAS Selection Deliverables".

[14] IVAS-7a Permanent Document: "Processing Plan for Selection Phase".

[15] IVAS-8a Permanent Document: "Test Plan for Selection Phase".

[16] EVS-7b Permanent Document: Processing Plan for Characterization Phase

[17] EVS-8b Permanent Document: "Test Plan for Characterization Phase".

[18] Recommendation ITU-T P.800: "Methods for subjective determination of transmission quality".

[19] Supplement ITU-T P.Suppl29: "ITU-T P.800 – Use Cases".

[20] Recommendation ITU-R BS.1534-3: "Method of the subjective assessment of intermediate quality level of audio systems".

[21] ITU-T Handbook of subjective testing practical procedures, 2011.

[22] Audio File Format Specifications: WAVE, https://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/WAVE.html.

[23] AFsp Package https://www-mmsp.ece.mcgill.ca/Documents/Downloads/AFsp/.

[24] IEEE Recommended Practice for Speech Quality Measurements, in IEEE Transactions on Audio and Electroacoustics, vol. 17, no. 3, pp. 225-246, September 1969, doi: 10.1109/TAU.1969.1162058.a.

[25] Recommendation ITU-R BS.2051-3 (05/2022): Advanced sound system for programme production.

[26] 3GPP TR 26.952: "Codec for Enhanced Voice Services (EVS); Performance characterization".

[27] 3GPP TS 26.445: "Codec for Enhanced Voice Services (EVS); Detailed Algorithmic Description".

[28] 3GPP TR 26.996: "Immersive Audio for Split Rendering Scenarios; Performance characterization".

[29] ISAR Permanent Document "Testing Aspects for Phase/Track 2/a".

[30] Recommendation ITU-T P.191 (03/2023): Software tools for speech and audio coding standardization.

[31] Recommendation ITU-R BS.1770-4 (10/2015): Algorithms to measure audio programme loudness and true-peak audio level.

# 3 Definitions of terms, symbols and abbreviations

## 3.1 Terms

For the purposes of the present document, the terms given in TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

## 3.2 Symbols

Void.

## 3.3 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in TR 21.905 [1].

CL Cross-check Laboratory

CuT Codec under Test

DCR Degradation Category Rating

DTX Discontinuous transmission

ESDRU Energy-based Spatial Distortion Reference Unit

EVS Enhanced Voice Services

FB Fullband

FEC Frame Erasure Concealment

FOA First-Order Ambisonics

GAL Global Analysis Laboratory

HL Host Laboratory

HOA2 Higher-Order Ambisonics, 2nd order

HOA3 Higher-Order Ambisonics, 3rd order

HRTF Head Related Transfer Function

ISM Independent Stream with Metadata

IVAS Immersive Voice and Audio Services

ISAR Immersive Audio for Split Rendering Scenarios

JBM Jitter Buffer Management

LFE Low Frequency Enhancement

LL Listening Laboratory

MASA Metadata-Assisted Spatial Audio

MC Multi-channel

MNRU Modulated Noise Reference Unit

NB Narrowband

OBA Object Based Audio

PLC Packet Loss Concealment

SBA Scene Based Audio

SDRU Spatial Distortion Reference Unit

SNR Signal-to-Noise Ratio

SWB Super Wideband

TC Transport Channels

WB Wideband

WMOPS Weighted Millions of Operations Per Second

# 4 General

## 4.1 Overview of the IVAS Codec Work Item

This clause provides an overview of the objectives before the actual work started, as a historical background. The standardized IVAS codec fulfilled all project objectives.

NOTE: The subsequent text in italics cites justification and objective from the IVAS Work Item Description. Since it is a direct citation, the tense of this text might generally refer to the future.

*The introduction of 4G/5G high-speed wireless access to telecommunications networks, combined with the availability of increasingly powerful hardware platforms, will enable advanced communications and multimedia services to be deployed more quickly and easily than ever before. Immersive services, applications and devices, as envisioned in 3GPP TR 22.891 are expected to provide an immersive user experience which, when compared to existing media services, will deliver a quantum leap in the quality of experience. Immersiveness will be a core service attribute of eXtended Reality (XR) ranging from VR over MR to AR, as described in TRs 26.918, 26.928 and 26.998. An immersive audio-visual experience implies, for the audio component, that a spatial sound impression is convincingly consistent with the presented visual scene and/or the real-world scene in the AR/MR case. In addition, the user should be able to move, within certain limits defined by the application, throughout the scene, and the audio component will adjust to reflect the user’s spatial orientation/position.*

*This WID will develop solutions specifically for conversational and non-conversational use cases where immersive content originates and is consumed in end-user devices.*

*The 3GPP Enhanced Voice Services (EVS) codec has delivered a highly significant improvement in user experience with the introduction of super-wideband (SWB) and full-band (FB) speech and audio coding, together with improved packet loss resiliency. For a truly immersive experience though, extended audio bandwidth is just one of the dimensions required and support beyond the mono and multi-mono currently offered by EVS is ideally required to immerse the user in a convincing virtual world in a resource-efficient manner. In addition, the currently specified stereo codecs in 3GPP, e.g., Enhanced aacPlus (eAAC+) and AMR-WB+ provide suitable quality and compression for stereo content in an adequate bit rate range, but lack the conversational features (e.g. sufficiently low latency) needed for conversational voice and teleconferencing. These coders also lack immersive audio support that is necessary for immersive services, including e.g., live streaming, XR and immersive teleconferencing.*

*The purpose of this work item is therefore to fill this technology gap and to address the increasing demand for rich and immersive multimedia services. In addition, teleconferencing applications over 4G/5G will benefit from this next generation codec used as an improved conversational coder supporting multi-stream coding (e.g., channel, object and scene-based audio). Use cases for this next generation codec include, but are not limited to, conversational voice, multi-stream teleconferencing, VR conversational and user generated live and non-live content streaming, AR/MR. The approach proposed is to build upon the EVS codec with the goal of developing a single codec with attractive features and performance (e.g. excellent audio quality, low delay, spatial audio coding support, appropriate range of bit rates, high-quality error resiliency, practical implementation complexity). In the scope of 3GPP the predominant audio rendering instrument is envisaged to be headphones but configurations with e.g. tablet speaker playback may also be of relevance.*

*The overall objective of this work item is to develop a single general-purpose audio codec for immersive 4G and 5G services and applications including the XR use cases envisioned in 3GPP TRs 26.918 and 26.928 and possibly relying on devices described in 26.998. The following objectives should be achieved with the work item:*

- *The solution is expected to meet the terms of reference (design constraints, performance requirements) developed as part of this WI.*

- *The solution is expected to handle encoding/decoding/rendering of speech, music and generic sound.*

*- It is expected to support encoding of channel-based audio (e.g. mono, stereo or 5.1) and scene-based audio (e.g., higher-order ambisonics) inputs including spatial information about the sound field and sound sources. The solution is expected to provide support for diegetic and non-diegetic input.*

*- It is expected to provide a decoder for the encoded format and a renderer with sufficiently low motion to sound latency.*

*- The solution is expected to operate with low latency to enable conversational services over 4G/5G.*

*- The solution is expected to support high error robustness under various transmission conditions from clean channels to channels with packet loss and delay jitter and to be optimized for 4G/5G.*

*- The solution is expected to provide support for a range of service capabilities, e.g., from mono to stereo to fully immersive audio encoding/decoding/rendering.*

*- The solution is expected to be implementable on a wide range of UEs and other end-user devices to address various needs in terms of balancing user experience and implementation complexity / cost.*

*- The solution is expected to provide support for immersive Real-time Communication services including MTSI and potentially streaming services offered by the 5G system through the definition of a new immersive audio media component. Support for MTSI services is also accomplished by the provision of bit-exact EVS operation as part of the solution.*

*The developments under this work item should lead to a set of new specifications defining among others textual description of the IVAS codec. Following 3GPP practice, fixed-point and floating-point C code and associated test vectors should also be part of this set of specifications. RTP payload format, SDP parameter definitions, jitter buffer management, rendering and packet loss concealment should be specified as part of the set of the codec specifications.*

The codec for Immersive Voice and Audio Services is part of a framework comprising besides encoder and decoder, renderer and a number of auxiliary functions associated with the support of stereo and immersive audio formats.

The IVAS codec is an extension of the 3GPP Enhanced Voice Services (EVS) codec; it provides full and bit exact EVS codec functionality for mono speech/audio signal input.

On top of that the IVAS codec is optimized for encoding and decoding of stereo and immersive audio formats, using tools such as Single Channel Element (SCE) coding, Channel Pair Element (CPE) coding and multi-channel coding by means of the Multi-channel Coding Tool (MCT). The stereo modes comprise a hybrid time-domain/DFT-domain/MDCT-domain coding scheme including inter channel alignment (ICA). Immersive audio formats comprise multi-channel audio (5.1, 5.1.2, 5.1.4, 7.1, 7.1.4 setups), scene-based audio (Ambisonics up to order 3), metadata-assisted spatial audio (MASA), and object-based audio (Independent Stream with Metadata (ISM) up to 4 ISMs). In addition, the following combined immersive audio formats are supported: object-based audio with scene-based audio (OSBA, up to 4 ISMs with Ambisonics) and object-based audio with metadata-assisted spatial audio (OMASA, up to 4 ISMs with MASA).

The codec features VAD/DTX/CNG for rate efficient stereo and immersive conversational voice transmissions, an error concealment mechanism to combat the effects of transmission errors and lost packets. Jitter buffer management is also provided.

The IVAS codec operates on 20 ms audio frames. It is capable of switching its bit rate upon command instantly at (active) frame boundaries.

In addition, split rendering functionality was added by the 3GPP work item on “Immersive Audio for Split Rendering Scenarios” (ISAR) [33].

NOTE: The subsequent text in italics cites justification and objective from the ISAR Work Item Description [33]. Since it is a direct citation, the tense of this text might generally refer to the future.

*Work currently carried out under the MeCAR and related work items assumes a common XR Baseline Client architecture. An essential characteristic is that a functional split is envisioned between a Presentation Engine comprising a set of composite renderers that are controlled by a Scene Manager and an XR Runtime performing a set of functions that interface with a platform to perform commonly required operations, e.g. post-rendering, prior to final output. The relevant interface between Presentation engine and end device may be a 5G physical interface between, e.g., between a smartphone or 5G EDGE and a lightweight device (AR glasses) like those considered in 5G EDGe-Dependent AR (EDGAR) and 5G Wireless Tethered AR UEs as described in 3GPP TR 26.998 or those considered under FS\_SmarTAR.*

*The functional split assumed in split renderer architectures is a result of stringent implementation and operational requirements applicable for rendering of XR media on XR devices. For head-tracked immersive audio, the need to rely on a split renderer architecture, may depend on various factors among which the round-trip latency between the renderer in the presentation engine and the lightweight device is a decisive parameter. There are scenarios where this latency may be substantial which may prefer a split rendering approach with pose correction in the end device for binaural audio in a similar way as for video unless decoding and head-tracked binaural audio rendering on the lightweight device does not exceed its strict complexity constraints.* *In other scenarios, that latency may be sufficiently low, in which case the head-tracked binaural rendering can exclusively be done in the presentation engine. It is notable that the transmission over the interface may generally be bit rate constrained and dependent on the specific physical interface.*

*Binaural audio rendering comprises of signal processing functionalities that may include:*

*- Binauralization of audio input based on head rotation (3DoF),*

*- Binauralization of audio input based on listener position and head rotation (6DoF),*

*- Room acoustics synthesis.*

*Audio input to be rendered may be a combination of diegetic immersive (3D audio) and non-diegetic sounds. The diegetic immersive sounds need to be binauralized using the up-to-date head rotation data. The head rotation data is typically originating from the head-tracker available from the lightweight end device. The room acoustic synthesis can be performed using room impulse response data or parametric representation thereof, typically supplied to the Presentation Engine.*

*Depending on constraints and design preferences of the lightweight device (AR glasses, earbuds, etc.) and the properties of the interface between Presentation Engine and end device, solutions are needed that among more are compliant with TRs 26.928 and 26.998. The solutions shall address given interface characteristics and not impose any new requirements for them.*

*Another aspect is the currently ongoing standardization of the EVS Codec Extension for Immersive Voice and Audio Services (IVAS) codec. While low complex rendering for lightweight devices is not a specific design objective, the IVAS codec work item should ideally provide solutions that would enable using IVAS services over head-tracked lightweight clients meeting relevant requirements.*

*Bearing in mind the evolution of the AR/XR technologies, it would be desirable to design low complex solutions for head-tracked binaural audio rendering on lightweight devices that under certain limitations are agnostic in a sense that the pre-renderer component in the presentation engine could be connected with any immersive binaural audio framework through suitable APIs.*

*The solutions to be specified are intended to add to the number of rendering options to enable immersive audio services on a broad range of devices, including light-weight AR glasses or earbuds. The pre-rendering part of the solutions is expected to become non-mandatory but shall fulfill the relevant requirements set out under this work item. It should interface through a fully specified intermediate bitstream with a fully specified split rendering decoder. For end device implementations claiming support of a specific solution, a fully compliant implementation of at least the split rendering decoder shall be required. Other end device implementations not claiming support of a specific solution remain at the discretion of the implementor.*

*The overall objective of this work item is to develop solutions for immersive binaural audio on head-tracked devices that are compatible with the envisaged split architectures (MeCAR, 26.998). The solutions should consider low-complex and lightweight devices and demonstrate operational benefits over solutions with full decoding and rendering in the end device. The following objectives should be achieved with the work item:*

*- Provide format specification for intermediate representation(s).*

*- Provide functional requirements for (pre-)renderer operations to be carried out by Presentation Engine.*

*- Define suitable APIs.*

*- Provide encoder, bitstream and decoder specification for intermediate representations including audio with and without post-rendering control metadata.*

*- Provide a specification for decoded intermediate representations to provide binaural audio output with and without head-tracker input and post-rendering control metadata.*

*- Consider potential solutions offered by the IVAS work item, and specify the necessary interfaces.*

*The work item shall in a first phase identify and agree relevant requirements to be documented in a TR. This shall cover:*

*- Design constraints related to complexity and memory as well as constraints related to relevant interfaces between presentation engine and end device such as bit rate, latency, down- and upstream traffic characteristics.*

*- Design constraints related to functional capability requirements such as rendering of non-diegetic sounds, 3DoF rendering of diegetic immersive sounds, 6DoF rendering of diegetic immersive sounds, including simultaneous rendering of different sound categories, and room acoustics synthesis.*

*- Performance requirements.*

*The solution(s) are characterized for the range of relevant interface characteristics between presentation engine and lightweight device. The case where the immersive audio is decoded and rendered within the end device should be considered as a reference.*

*The requirements will be documented in a first technical report. The developments under this work item shall lead to a new specification defining among others textual descriptions of the involved renderers and codec (incl. frame loss concealment) of the intermediate representation(s). The performance of the developed solutions in relation to the requirements will be documented in a second technical report. Solutions meeting the ISAR split rendering requirement may be added to the set of IVAS codec specifications (by means of CRs) if they are found suitable for IVAS. The developed solutions should also be referenced in the MeCAR specification.*

*Specific split rendering solutions for IVAS should comprise a non-mandatory default split rendering encoder for the specified internal and stand-alone IVAS renderers. In addition, for a given specific solution there should be specified interfaces offering the possibility either to connect a given (proprietary) renderer for IVAS to the intermediate representation encoder or to use proprietary pre-renderers/intermediate encoders to produce compliant intermediate bitstreams.* *Such proprietary solutions shall be compliant with the relevant requirements documented in the first technical report. ISAR end device implementations for IVAS claiming support of a specific solution shall be required to have at least a fully compliant split rendering decoder. Other decoder/post-renderer implementations not claiming support of a specific ISAR solution for IVAS remain at the discretion of the implementor.*

A special feature of the IVAS renderer is that it supports split operation with pre-rendering and transcoding to a head-trackable intermediate representation that can be transmitted to a post-rendering end-device. This enables moving a large part of the processing load and memory requirements for IVAS decoding and rendering to a (more) capable node/UE while offloading the final rendering end-device. For the split rendering bitstreams, bit-rates ranging from 256 kbps to 768 kbps are supported.

## 4.2 Presentation of the Following clauses

Clause 5 outlines the Terms of reference for the IVAS project. In clause 6, the selection process in 3GPP is presented. An overview of selection [and characterization tests] can be found in clause 7. The subjective tests provide statistical data which are subject to variations; important notes about interpretation of results are described in clause 8.

The actual subjective test results are presented in clause 9, and objective evaluations are presented in clause 10.

# 5 Terms of Reference

3GPP sets the codec Terms of Reference as Design Constraints and Performance Requirements.

The design constraints specified in the IVAS-4 Permanent Document [11] set the framework for the IVAS codec in terms of capability and resource usage. As such they list functionalities that are divided into mandatory, recommended and optional features to be provided by IVAS codec candidates. In the final standard, all modes have an equal status and they together form the IVAS codec.

Codec features were defined as follows:

The IVAS codec is an extension of the 3GPP Enhanced Voice Services (EVS) codec [2]. It provides full and bit exact EVS codec functionality for mono speech/audio signal input. It further provides:

- Encoding and decoding of stereo and immersive audio formats such as multi-channel audio, scene-based audio (Ambisonics), metadata-assisted spatial audio (MASA), object-based audio (ISM), and their combination.

- VAD/DTX/CNG for rate efficient stereo and immersive conversational voice transmissions

- Error concealment mechanisms to combat the effects of transmission errors and lost packets. Jitter buffer management is also provided.

- The IVAS codec operates on 20 ms audio frames. In addition, rendering is possible with 5 ms granularity.

- Support for bit rate switching upon command.

- Stereo and immersive audio coding at the following discrete bit rates [kbps]: 13.2, 16.4, 24.4, 32, 48, 64, 80, 128, 160, 192, 256, 384, and 512, with supported bit rate ranges listed in Table ‎4.2‑1.

- Support for WB, SWB and FB audio, with the supported bitrate range listed in Table ‎4.2‑2.

Table ‎4.2‑1: Ranges of supported bitrates for stereo and immersive coding of the IVAS codec

|  |  |
| --- | --- |
| Input audio format | Range of supported bitrates [kbps] |
| Stereo | 13.2 – 256 |
| Scene-based audio (SBA) | 13.2 – 512 |
| Metadata assisted spatial audio (MASA) | 13.2 – 512 |
| Object-based audio (ISM) | 13.2 – 512 |
| Multi-channel audio (MC) | 13.2 – 512 |
| Combined ISM and MASA (OMASA) | 13.2 – 512 |
| Combined ISM and SBA (OSBA) | 13.2 – 512 |

Note, that for the object-based audio format (ISM) the range of supported bitrates varies based on the number of objects as follows: 13.2 kbps – 128 kbps for 1 ISM, 16.4 kbps – 256 kbps for 2 ISMs, 24.4 kbps – 384 kbps for 3 ISMs, 24.4 kbps – 512 kbps for 4 ISMs.

Table ‎4.2‑2: Supported audio bandwidth per input audio format and bitrate

|  |  |  |  |
| --- | --- | --- | --- |
| Input audio format | Bitrates supporting WB  [kbps] | Bitrates supporting SWB  [kbps] | Bitrates supporting FB  [kbps] |
| Stereo | 13.2 – 256 | 13.2 – 256 | 32 – 256 |
| Scene-based audio (SBA) | 13.2 – 512 | 13.2 – 512 | 32 – 512 |
| Metadata assisted spatial audio (MASA) | 13.2 – 512 | 13.2 – 512 | 32 – 512 |
| Object-based audio (ISM), 1 object | 13.2 – 128 | 13.2 – 128 | 16.4 – 128 |
| Object-based audio (ISM), 2 objects | 13.2 – 256 | 24.4 – 256 | 32 – 256 |
| Object-based audio (ISM), 3 objects | 24.4 – 384 | 24.4 – 384 | 48 – 384 |
| Object-based audio (ISM), 4 objects | 24.4 – 512 | 24.4 – 512 | 64 – 512 |
| Multi-channel audio (MC) | 13.2 – 512 | 13.2 – 512 | 32 – 512 |
| Combined ISM and MASA (OMASA) | 13.2 – 512 | 13.2 – 512 | 32 – 512 |
| Combined ISM and SBA (OSBA) | 13.2 – 512 | 13.2 – 512 | 32 – 512 |

The IVAS-4 Permanent Document [11] also sets constraints on maximum algorithmic delay (40 ms); frame length (20 ms); maximum computational complexity (10\*EVS); memory limits (10\*EVS); and limit of the output gain.

A full description of the performance requirements can be found in IVAS-3 Permanent Document: Performance Requirements [10].

# 6 Selection Process

3GPP runs codec selection as a rigorous process, outlined below.

Codec selection in 3GPP follows pre-defined procedures. Proponents are obliged to provide certain information about their candidate to facilitate selection, and strict rules are set prior to selection to provide guidance on selecting the candidate to be standardized. Verification serves the purpose of cross-check and provision of additional (technical) information.

Selection Deliverables are specified in IVAS-6 Permanent Document [13].

Proponents were required to provide the following information about their candidate for selection (named selection deliverables):

- Candidate Codec Executable (floating-point)

- Candidate Codec Source Code (floating-point)

- High level technical description of the candidate algorithm

- Report covering the compliance to Design Constraints

- Draft overview specification

- IPR declaration

- Funding payment (proponents paid for selection testing)

- Legal framework to cover use model parameters, unprocessed audio test material, processed audio test material, and test results

- Optional additional information incl. demo material

Selection rules are specified in IVAS-5 Permanent Document [13].

The strict 3GPP selection process involved the following rules (which were agreed before selection) to determine the candidate to be standardized:

- Provision of full set of selection phase deliverables

- Compliance with design constraints

- Codec performance analysed based on test conditions, see IVAS-8a [15].

In the 3GPP SA4#125 meeting the selection deliverables and selection test results were reviewed and based on this information, 3GPP SA4 selected the IVAS codec candidate jointly developed by the IVAS Public Collaboration as the 3GPP IVAS standard. Subsequently the SA#101 plenary meeting approved the IVAS codec selection and the set of IVAS specifications [4], [5], [6], [7], [8], and [9].

# 7 Introduction to the Testing of the IVAS codec

## 7.1 IVAS Selection Phase Testing

### 7.1.1 General methodology

#### 7.1.1.1 Overview

The following test methodologies were used in the IVAS Selection test: P.800 [18] under consideration of P.Suppl29 [19] was used in experiments designed to evaluate the Immersive conversation use case scenarios, and BS.1534 [20] was used in experiments designed to evaluate the Generic immersive audio use case scenarios. The high-level configuration of experiments for both methodologies is outlined below.

#### 7.1.1.2 P.800

- Test duration should not exceed 2 hours per listening panel. Typical value of voting period was used for estimation of test durations, but actual voting period is not specified.

- Randomizations constructed under randomized blocks experimental design described in [21].

- 6 categories for each test. Categories are defined for each experiment separately.

- 6 samples/category (1 for each listening panel) plus 1 sample/category for preliminaries.

- 30 naïve listeners, 6 listening panels (5 listeners per panel), each panel with an independent randomization

- 180 votes for each condition.

- Total number of conditions for each experiment: 36

- Number of trials: number of test conditions x 6 talkers/categories = 216 trials.

#### 7.1.1.3 BS.1534

- Number of items per experiment: 12

- 14 experienced listeners

- Maximum total number of conditions: 8

- Number of anchor conditions: 2

- Direct

- 7 kHz low-pass anchor

NOTE: the exact number of anchors may vary depending on actual experiment.

### 7.1.2 Opinion Scales

#### 7.1.2.1 ITU-T P.800 DCR

For ITU-T P.800 DCR tests, the opinion scale as defined in Table ‎7.1‑1 is used.

Table ‎7.1‑1: Opinion scale for ITU-T P.800 DCR test

|  |  |
| --- | --- |
| Impairment | Scale |
| No impairment | 5 |
| Small impairment | 4 |
| Moderate impairment | 3 |
| Large impairment | 2 |
| Very large impairment | 1 |

The following instructions (translated properly to the listening labs language and be given to the listeners) are used. The instructions given to the listeners shall be provided for information in the listening lab report.

|  |
| --- |
| **INSTRUCTIONS TO NAÏVE LISTENERS FOR P.800 DCR TEST**  In this experiment you will be evaluating systems that might be used for future immersive telecommunication services using spatial audio. Spatial audio means that you can locate various sound sources around yourself. For example, a first talker may appear to talk from the left-hand side and a second talker from the right-hand side, a talker can be moving, etc.  In each trial, you will hear a *reference* audio sample followed by a *test* sample. The *test* sample has the same content as the *reference* sample, but it is possibly impaired after it has passed through a telecommunication system.  Your task is to evaluate the overall impairment of the second sample compared to the first sample, comprising both degradations in the sound quality (e.g., due to additional noise, roughness, clicks or other distortions), and/or differences in the spatial representation (e.g., sound source location, distance, spatial width, movement, etc.).  You should listen carefully to both samples within a trial. When they have finished, select the category that best describes your overall impression about the amount of any impairment you can perceive in the second sample relative to the first sample:  5 - No impairment  4 - Small impairment  3 - Moderate impairment  2 - Large impairment  1 - Very large impairment  Note that the level of impairments present in different *test* samples is expected to span the complete range of the rating scale during the experiment.  Please do not discuss your opinions with other listeners participating in the experiment. If you have any questions, please ask the test administrator. |

**Voting screen** (after playback of sample pair)

|  |
| --- |
| Please rate the OVERALL IMPAIRMENT of the second sample compared to the first sample:  5 - No impairment  4 - Small impairment  3 - Moderate impairment  2 - Large impairment  1 - Very large impairment |

#### 7.1.2.2 ITU-R BS.1534

For BS.1534 tests, the continuous quality scale (QCS) according to [20] is used.



Figure ‎7.1‑1: BS.1534 Continuous Quality Scale

### 7.1.3 Material

#### 7.1.3.1 Overview

All audio material was sampled at 48 kHz with Full Band (FB) content. The audio material was delivered to the HL as 16-bit little endian WAVE format files [22] following the naming convention provided in the IVAS Processing Plan (IVAS-7a). For multi-track audio, the audio tracks are ordered according to Table 5 of IVAS Processing Plan (IVAS-7a). Additionally, it should be verified that the audio material can be processed with the AFsp package tools [23].

The following categories of audio content were used in IVAS Selection Test using P.800:

- Clean speech: Except for experiment P800-6 (1 object), each sample contains two (or more) different talkers in conversation scenario. The talkers transition from one to another as in natural conversation, possibly with partial overlap.

- Speech with background: the details about the environment are specified in IVAS-8a Annex E.

- Music and Mixed content – categories specified in IVAS-8a clause 4.5.3.

The following category of audio content was be used in IVAS Selection Test using BS.1534:

- Generic audio – critical generic audio items including speech with and/or without background, music, mixed.

#### 7.1.3.2 Speech Material for P.800 testing

Except for the experiment P800-3 (Music and mixed content stereo experiment), P.800 test experiments have used artificially created immersive audio. LLs provided clean speech mono audio samples and music and mixed content stereo samples. SA4 would provide scene descriptions and scripts to create the immersive audio.

The recording SNR should be in accordance with P.800 at least 40 dB but preferably 50 dB or higher, the leading and trailing inactivity portions should be shorter than 20 ms. The reverberation time RT60 should be in accordance to P.800 less than 500 ms, preferably below 200 ms. The length of the sentences should typically correspond to the length of traditional Harvard sentences [24].

#### 7.1.3.3 Background Material

- Immersive conversation use case scenario (P.800 testing): A mix-based approach using separate background recordings was used. The minimum lengths of noise files shall be 80 s. The following guideline is applied to the noise types used.

- Car noise is intended to test the performance of the codec under steady state background noise and should be recorded in a moving car. A constant speed between 80 km/h (50 mph) and 110 km/h (70 mph) is recommended. The make and model of the car should be reasonably common in the country of the recording. Typically, the windows of the car should be closed, and the radio turned off.

- Office noise is intended to represent a typical office environment. This noise type should also contain typical office sounds, such as keyboard noise, computer fans, telephones ringing, printers, air conditioner, etc.

- Street noise is intended to represent a typical street environment. It should contain unsteady traffic noise for example recorded at traffic lights where cars stop, human noise such as steps. It should not contain speech, but baby cries are allowed.

- Generic immersive audio use case scenario (BS.1534 testing): Primarily, full recordings of complete immersive scenes including background will be used. A mix-based approach might be used in addition.

#### 7.1.3.4 Music and Mixed Content Material for P.800 testing

Music and mixed content samples shall contain meaningful contents and the duration of each sample shall be approximately 8 s and at least 7 s. The following categories were used:

- Classical music

- Modern instrumental music

- Modern vocal music

- Radio Jingle

- Movie Trailer

- Advertisement

This means that LL shall provide 7 samples per category, 6 for evaluation and 1 for preliminaries.

#### 7.1.3.5 Critical Generic Audio Items for BS.1534 testing

##### 7.1.3.5.1 Steps of Critical Test Item Selection

The following steps were performed:

- Call for test material according to the generic audio signal categories described below.

- Material Collection entity collects candidate material submitted in response to the call and selects a number of critical items to be used in the Selection test.

- Material Collection entity selects a limited set of training items to be used in a training phase.

##### 7.1.3.5.2 Test Material

First, a call was sent out for test material according to a number of generic audio signal categories as specified below. All 3GPP members were invited to submit test material to the Material Collection entity. The submitting organization assigned the items to the below-mentioned audio signal categories. Then, the Material Collection entity identified 12 critical items per experiment, plus four items for training, which are representative for assumed typical IVAS application scenarios.

Generic audio signal categories:

Stereo – generic stereo audio signals with a focus on music categories:

- Pop, with and/or without vocals

- Classic, with and/or without vocals

- Single instruments

- a capella vocals, solo and/or choir

- Mixed speech and music

- Speech with and/or without background noise

Multi-Channel (5.1 and 7.1.4) – generic channel-based audio signals from produced content:

- Music including concerts with live audience

- Film soundtracks with and/or without speech dialogue

- Effects (e,g, nature, city/transport sounds)

Scene-Based Audio / MASA – generic immersive audio signals in the form of complex scenes, captured and/or produced content which may or may not include speech:

- Nature sounds (e.g. forest, water, wind)

- City sounds (e.g. traffic, bus, train)

- Music including concerts with live audience

- Babble-like sound (e.g. market, restaurant, conference)

- Event/Sport-like sound

- Conferencing scene with and/or without background noise/music

Object-Based Audio - Realistic immersive audio signals, e.g.:

- Scenarios comprising voice, music, background objects.

- Conversational scenarios of several talkers with or without background, with or without partial overtalk of no more than two talkers. Talkers may be moving around the scene at natural pace. However, it is not expected that all talkers are active all the time, with unnaturally rapid displacements.

The length in time of the items is 10s at a maximum.

The Material Collection entity is reported to SA4 a list indicating the number of proposed items per submitting organization.

##### 7.1.3.5.3 Training material

Limited material was used in the training phase in which the subjects familiarize with the testing methodology and environment.

The training was conducted with four sound items. These items were identified by the Material Collection entity and shall not be re-used in the blind grading phase. The training phase shall be executed as a separate short BS.1534 session.

### 7.1.4 Listening Systems and Listening Environments

The IVAS Selection Test has used the following listening systems:

- Stereo headphones for binaural listening, e.g.:

- Beyerdynamic DT 770 Pro for P.800 experiments

- Sennheiser HD 650 for BS1534 experiments

- Loudspeaker listening system – 7.1+4 loudspeaker setup [25].

### 7.1.5 Experimental Procedure for P.800 experiments

Initially the experimenter provided a written copy of the experiment instructions to the listeners. When the listeners have had acknowledged that they understand the instructions, they were presented with a practice session to rate the preliminary conditions. After the practice session was completed, the experimenter asked if there are any questions. Only questions about the rating procedures or the meaning of the instructions should be answered. Any technical questions on matters such as the experimental methodology or details of the types of distortions they are rating must not be answered.

### 7.1.6 Subjective Experiments

The purpose of the 23 experiments (Experiments P800-1 – P800-9, and BS1534-1a – BS1534-7b) is to evaluate the performances of the IVAS codec candidate algorithm with respect to the performance requirements and objectives defined in (IVAS-3).

The details provided in this clause and in corresponding Annexes of IVAS-8a are those that are specific to each particular experiment. Generic information can be found in IVAS-8a clause 4. LLs were asked to use the information in IVAS-8a clause 4 in conjunction with the information given in IVAS-8a clause 5 and its Annexes.

Table ‎7.1‑2 shows a high-level overview of P.800 experiments. Table ‎7.1‑3 shows a high-level overview of BS.1534 experiments. Finally, Table ‎7.1‑4 shows allocation of experiments to LLs and languages proposed by LLs for each P.800 experiment.

Detail conditions for each subjective experiment are defined in IVAS-8a Annex E for P.800 experiments and in IVAS-8a Annex F for BS.1534 experiments.

Table ‎7.1‑2: High-level overview of P.800 experiments

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Exp | Input format | Source material | Listening environment | Bitrates kbps | FER/jitter | DTX |
| P800-1 | Stereo | Clean speech | Headphones | ≤ 48 | ≤ 5% | Y |
| P800-2 | Stereo | Speech+Background | Headphones | ≤ 64 | ≤ 5% | Y |
| P800-3 | Stereo | Mixed & Music | Headphones | ≤ 64 | ≤ 5% | Y |
| P800-4 | FOA | Clean speech | Headphones | ≤ 96 | ≤ 5% | N |
| P800-5 | FOA | Speech+Background | Headphones | ≤ 96 | ≤ 5% | Y |
| P800-6 | 1 Object | Clean speech | Headphones | ≤ 64 | ≤ 5% | Y |
| P800-7 | 2 Objects | Clean speech | Headphones | ≤ 64 | ≤ 5% | Y |
| P800-8 | MASA | Clean speech | Headphones | ≤ 80 | ≤ 5% | N |
| P800-9 | MASA | Speech+Background | Headphones | ≤ 80 | ≤ 5% | Y |

Table ‎7.1‑3: High-level overview of BS.1534 experiments

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Exp | Input format | Source material | Listening environment | Bitrates kbps |
| BS1534-1a | Stereo | Generic Audio | Headphones | 48, 64 |
| BS1534-1b | Stereo | Generic Audio | Headphones | 96, 128 |
| BS1534-2a | 5.1 | Generic Audio | 5.1 | 64, 96 |
| BS1534-2b | 5.1 | Generic Audio | 5.1 | 128, 160 |
| BS1534-3a | 7.1.4 | Generic Audio | 7.1 + 4 | 128, 160 |
| BS1534-3b | 7.1.4 | Generic Audio | 7.1 + 4 | 384, 512 |
| BS1534-4a | FOA | Generic Audio | Headphones | 96, 128, 160 |
| BS1534-4b | HOA2 | Generic Audio | Headphones | 160, 192 |
| BS1534-5a | HOA3 | Generic Audio | Headphones | 192, 256 |
| BS1534-5b | HOA3 | Generic Audio | 7.1 + 4 | 384, 512 |
| BS1534-6a | Objects | Generic Audio | Headphones | 48, 64, 96 |
| BS1534-6b | Objects | Generic Audio | Headphones | 96, 128, 192 |
| BS1534-7a | MASA | Generic Audio | Headphones | 96, 128 |
| BS1534-7b | MASA | Generic Audio | Headphones | 192, 256 |

Notes:

- Stereo may include binauralized samples (without head tracking).

- For inputs 7.1+4, FOA, HOA2, HOA3, Objects & MASA vertical dimension is assumed in the samples.

- DTX on/off is assumed within the same experiment, where DTX on is used for relevant conditions.

- All experiments except for stereo P.800 experiments are assumed Full Band experiments, i.e., the direct reference condition is always FB. P.800 stereo experiments are SWB experiments.

Minimum requirements for P.800 experiments: 6 talkers (3 male + 3 female) per experiment, 14 single sentences per talker.

Table ‎7.1‑4 shows allocation of LLs so that each experiment is conducted twice, each time by a different LL. For P.800 experiments, each experiment is run twice with different languages.

Table ‎7.1‑4: Allocation of experiments to LLs and P.800 languages

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Exp  Source material  Listening environment | | | Languages | | | |
| Force Technology | Head Acoustics/ IKS | MQ University | Mesaqin |
| Lab a | Lab b | Lab c | Lab d |
| P800-1 | Clean speech | Headphones | JAP |  |  | FR |
| P800-2 | Speech+Background | Headphones |  | GER |  | MAN |
| P800-3 | Mixed & Music | Headphones | DAN |  |  | MAN |
| P800-4 | Clean speech | Headphones | JAP |  | ENG |  |
| P800-5 | Speech+Background | Headphones | DAN | GER |  |  |
| P800-6 | Clean speech | Headphones | JAP |  | ENG |  |
| P800-7 | Clean speech | Headphones | DAN |  |  | MAN |
| P800-8 | Clean speech | Headphones | DAN | GER |  |  |
| P800-9 | Speech+Background | Headphones | JAP |  |  | FR |
| BS1534-1a | Generic Audio | Headphones | x |  |  | x |
| BS1534-1b | Generic Audio | Headphones |  | x |  | x |
| BS1534-2a | Generic Audio | 5.1 | x | x |  |  |
| BS1534-2b | Generic Audio | 5.1 | x | x |  |  |
| BS1534-3a | Generic Audio | 7.1 + 4 | x | x |  |  |
| BS1534-3b | Generic Audio | 7.1 + 4 | x | x |  |  |
| BS1534-4a | Generic Audio | Headphones | x |  |  | x |
| BS1534-4b | Generic Audio | Headphones |  | x |  | x |
| BS1534-5a | Generic Audio | Headphones | x |  |  | x |
| BS1534-5b | Generic Audio | 7.1+4 | x | x |  |  |
| BS1534-6a | Generic Audio | Headphones |  | x |  | x |
| BS1534-6b | Generic Audio | Headphones |  | x |  | x |
| BS1534-7a | Generic Audio | Headphones |  | x |  | x |
| BS1534-7b | Generic Audio | Headphones |  | x |  | x |
| Total |  | | | | | |

where

* JAP = Japanese
* FR = French
* GER = German
* MAN = Mandarin
* DAN = Danish
* ENG = English

# 8 Important Notes about the Interpretation of Test Results

## 8.1 P.800 Testing

Mean Opinion Scores can only be representative of the test conditions in which they were recorded (speech/music material, processing, listening conditions, language, and cultural background of the listening subject). Listening tests performed with other conditions than those used in the testing could lead to a different set of MOS results. On the other hand, the relative performances of different codecs under test is considered more reliable and less impacted by cultural difference between listening subjects than absolute MOS values. When looking at the relative differences of the codecs in the same test, it should be noted that a difference of typically 0.15 - 0.2 MOS between two test results would not usually be found statistically significant; appropriate statistical significance tests such as Student's T-test should be used to get an accurate figure of statistically significant difference between conditions within an experiment.

The subjective testing is conducted using limited amount of source material in order to keep the size of the experiment within reasonable limits. Sometimes this can cause some irregularities to the test results. Also, the performance of the tested codecs is not always known when designing the test, thus balancing the test conditions may not always be perfect. This may result in imperfect utilisation of the ranking scale and difficulties to discriminate the codecs with quality very close to each other.

Furthermore, in a number of experiments both clean and erroneous channel conditions were presented in the same experiment. It can be expected that the separation of the different clean channel conditions is less in those experiments compared to experiments where only clean channel conditions are presented. During the setup of the listening experiments SA4 experts made every effort to minimize effects like scale saturation and alike. However, the large number of conditions to be tested and the limited number of experiments that could be conducted made certain compromises unavoidable.

The resolution of the testing is limited. The listeners only use a scale from 1 to 5 to rank the different codecs. However, during the tests presented in the present document, we are characterising a large number of different IVAS modes, most of which are very high quality codecs and this may cause sometimes a "saturation" effect in the test, i.e. the listeners cannot discriminate the different codecs because of the limited range and scale.

Taking into account the comments presented above, the reader is advised to exercise some precautions when looking and comparing the individual scores of the tests. Usually, looking at the whole picture and overall trends in the test in question may give better interpretation of the performance of the codecs. This precaution should be especially taken into account when looking at the experiments conducted using impaired channels which may present rather big variability of results over the limited amount of tested conditions.

Throughout the present document, test results are presented in the following graphical forms.

- **Line-graphs** - Summary results are presented in line-graphs which compare the Reference codecs and the IVAS codec for various test parameters, (e.g. Bit-rate, Frame Error Rate, DTX on/off, etc.). The line-graphs only include conditions from within a test, i.e., no comparisons are made across tests. The graphs include 95% confidence intervals.

## 8.2 BS.1534 Testing

In contrast to P.800, the BS.1534 (MUSHRA) test methodology uses the continuous quality score (CQS). The CQS consists of identical graphical scales which are divided into five equal intervals with the adjectives “Excellent”, “Good”, “Fair”, “Poor” and “Bad” as also shown in Figure ‎7.1‑1 from top to bottom. The subject records his/her assessment of the quality in a suitable form, typically with the use of sliders on an electronic display. The assessor (subject) is asked to rate the quality of all stimuli, according to the five-interval CQS. The assessments for each test condition are converted linearly to normalized scores in the range 0 to 100, where 0 corresponds to the bottom of the scale (bad quality).

The MUSHRA method has the advantage of displaying many stimuli at the same time so that the assessor is able to carry out any comparison between them directly. In order to achieve reliable results, experienced assessors are chosen to carry out the listening tests. In combination with suitable test material (which was selected by an expert panel), both also contributes to a high resolution of the tests.

The size of a listening panel is not solely a consideration of the desired resolution but a large number of experienced subjects participating contributes to the generalization of the test results. Although 14 experienced subjects participated in the assessment per test for the IVAS selection tests, it should be understood that the results are in principle only valid for precisely that group of experienced listeners actually involved in the actual test. Taking this account, the reader is advised to exercise some precautions when looking and comparing the individual scores of the tests. Usually, looking at the whole picture and overall trends in the test in question may give better interpretation of the performance of the codecs. In order to illustrate the general trends, also graphs combining results from both (independent) participating labs are given.

Throughout the present document, test results are presented in the following graphical forms.

- **Point-graphs,** **Bar-graphs** - Summary results are presented in point-graphs/bar-graphs which display mean value and 95% confidence intervals. The graphs include the IVAS codec as well as the Reference codecs, anchors and hidden reference. The point-graphs only include conditions from within a test, i.e., no comparisons are made across tests.

As stated already for the P.800 tests, appropriate statistical significance tests such as Student's T-test should be used to get an accurate figure of statistically significant difference between conditions within an experiment.

## 8.3 Analysis of Selection Phase Results

The Selection phase GAL report provides the following analysis of test results:

For both test designs (P800 and BS1534), the ToR requirements in IVAS-8a [16] are defined in the same way. All codec-under-test (CuT) conditions have to perform either better than (BT) or not worse than (NWT) certain reference conditions. Due to the uncertainty of the auditory data, it is necessary to consider not only the averaged per-condition results, but also the variance/standard deviation and the number of votes. For this reason, Student's t-test for independent groups is evaluated on each result of the test and the corresponding reference condition. The t-statistic (referred as "T-Stat" in the following) is calculated without the assumption that the variances of test and reference are equal (Welch's t-test).

T-Stat is compared to t(M), the single-sided 95% confidence level of the inverse t-distribution, which depends on the degree of freedom M, i.e., the number of votes per condition. The check result R is then determined according to equation 1. Note that R might also show that the CuT is worse than (WT) the reference.

(1)

Based on clause 3.3.6.2 of IVAS-8a, it is indicated which and how many requirements have to be met for each condition in each experiment. The tabular data shown in the following clauses utilizes status flags for this purpose:

- FAIL: requirement was not met (cell in tabular data marked in red)

- PASS: requirement was met (cell in tabular data not marked)

- EXCEED: requirement was met and is significantly better (cell in tabular data marked in blue).

Table ‎8.3‑1 illustrates how the status is determined for each combination of requirement and check result.

Table ‎8.3‑1: Status based on check result

|  |  |  |
| --- | --- | --- |
| Requirement | Check result | Status |
| NWT | WT | FAIL |
| NWT | NWT | PASS |
| NWT | BT | EXCEED |
| BT | WT | FAIL |
| BT | NWT | FAIL |
| BT | BT | PASS |

In some cases, a ToR requirement is defined as the logical disjunction (OR) of two separate checks, which lead to two status values. The resulting status with respect to the requirement is determined depending on the involved criterion types as defined in Table ‎8.3‑2.

Table ‎8.3‑2: Status disjunction per criteria

|  |  |  |  |
| --- | --- | --- | --- |
| Status #1 | Status #2 | | Result status |
| (NWT check) | (NWT check) | (BT check) |
| EXCEED | EXCEED |  | EXCEED |
| EXCEED | PASS |  | PASS |
| EXCEED | FAIL |  | PASS |
| PASS | PASS |  | PASS |
| PASS | FAIL |  | PASS |
| FAIL | FAIL |  | FAIL |
| EXCEED |  | PASS | EXCEED |
| EXCEED |  | FAIL | PASS |
| PASS |  | PASS | PASS |
| PASS |  | FAIL | PASS |
| FAIL |  | PASS | PASS |
| FAIL |  | FAIL | FAIL |

NOTE: Each ToR requirement includes at least one NWT check, but never a disjunction of two BT checks. Thus, the cases listed in Table ‎8.3‑2 cover all possible cases of the ToR evaluation.

# 9 IVAS Performances

## 9.1 Mono

The IVAS codec is an extension of the 3GPP Enhanced Voice Services (EVS) codec; it provides full and bit exact EVS codec functionality for mono speech/audio signal input. The performance of the EVS codec is documented in 3GPP TR 26.952 [25].

## 9.2 Stereo

### 9.2.1 Overview

In Selection phase, five experiments have been conducted to evaluate the performance of the IVAS codec with stereo content. While the experiments P800-1, P800-2 and P800-3 were conducted as P.800 DCR tests, the experiments BS1534-1a and BS1534-1b were conducted as BS.1534 tests. All experiments were conducted using headphone presentation.

- Selection Experiment P800-1: Stereo clean speech under clean and impaired channel conditions, headphone presentation

- Selection Experiment P800-2: Stereo speech + background under clean channel conditions, DTX off and on, headphone presentation

- Selection Experiment P800-3: Stereo mixed and music under clean and impaired channel conditions, headphone presentation

- Selection Experiment BS1534-1a: Stereo generic audio, 48 and 64 kbps, headphone presentation

- Selection Experiment BS1534-1b: Stereo generic audio, 96 and 128 kbps, headphone presentation

### 9.2.2 Selection Experiment P800-1 (Stereo/Binaural, Clean Speech, Headphone Presentation)

Selection Experiment P800-1 evaluates IVAS for Stereo/Binaural clean speech under clean and impaired channel conditions with headphone presentation. See Annex C.1 for details.

The complete statistical evaluation of the requirement ToR tests for experiment P800-1 is given in the following table. The evaluation is done separately for the data from the two listening laboratories.

Table ‎9.2‑1: Statistical overview on the results of P800-1

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | DTX | FER | Req. | MOS | Std. | Cond. | Bitrate | MOS | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | | | |
| a | c25 | 1 | 13.2 | off |  | NWT | 3.93 | 1.08 | c10 | 2x8 | 3.45 | 1.06 | 4.27 | BT | EXCEED |
| 2 | 13.2 | off |  | BT | 3.93 | 1.08 | c09 | 2x7.2 | 3.31 | 1.15 | 5.33 | BT | PASS |
| c26 | 1 | 16.4 | off |  | NWT | 4.14 | 1.07 | c11 | 2x9.6 | 3.83 | 1.02 | 2.87 | BT | EXCEED |
| 2 | 16.4 | off |  | BT | 4.14 | 1.07 | c10 | 2x8 | 3.45 | 1.06 | 6.18 | BT | PASS |
| c27 | 1 | 24.4 | off |  | NWT | 4.41 | 0.9 | c12 | 2x13.2 | 4.38 | 0.76 | 0.32 | NWT | PASS |
| 2 | 24.4 | off |  | BT | 4.41 | 0.9 | c11 | 2x9.6 | 3.83 | 1.02 | 5.69 | BT | PASS |
| c28 | 1 | 32 | off |  | NWT | 4.61 | 0.66 | c13 | 2x16.4 | 4.57 | 0.72 | 0.61 | NWT | PASS |
| 2 | 32 | off |  | BT | 4.61 | 0.66 | c12 | 2x13.2 | 4.38 | 0.76 | 3.13 | BT | PASS |
| c29 | 1 | 48 | off |  | NWT | 4.72 | 0.68 | c15 | 2x32 | 4.67 | 0.62 | 0.73 | NWT | PASS |
| 2 | 48 | off |  | BT | 4.72 | 0.68 | c14 | 2x24.4 | 4.72 | 0.56 | 0.08 | NWT | FAIL |
| c30 | 1 | 13.2 | off | 5% | NWT | 3.33 | 1.22 | c17 | 2x8 | 2.88 | 1.14 | 3.58 | BT | EXCEED |
| 2 | 13.2 | off | 5% | BT | 3.33 | 1.22 | c16 | 2x7.2 | 2.83 | 1.17 | 3.97 | BT | PASS |
| c31 | 1 | 16.4 | off | 5% | NWT | 3.52 | 1.11 | c18 | 2x9.6 | 3.11 | 1.21 | 3.36 | BT | EXCEED |
| 2 | 16.4 | off | 5% | BT | 3.52 | 1.11 | c17 | 2x8 | 2.88 | 1.14 | 5.4 | BT | PASS |
| c32 | 1 | 24.4 | off | 5% | NWT | 3.74 | 0.99 | c19 | 2x13.2 | 3.62 | 1.12 | 1.09 | NWT | PASS |
| 2 | 24.4 | off | 5% | BT | 3.74 | 0.99 | c18 | 2x9.6 | 3.11 | 1.21 | 5.38 | BT | PASS |
| c33 | 1 | 32 | off | 5% | NWT | 3.9 | 1.09 | c20 | 2x16.4 | 3.64 | 1.07 | 2.3 | BT | EXCEED |
| 2 | 32 | off | 5% | BT | 3.9 | 1.09 | c19 | 2x13.2 | 3.62 | 1.12 | 2.43 | BT | PASS |
| c34 | 1 | 48 | off | 5% | NWT | 4.15 | 0.86 | c22 | 2x32 | 3.83 | 1.05 | 3.19 | BT | EXCEED |
| 2 | 48 | off | 5% | BT | 4.15 | 0.86 | c21 | 2x24.4 | 4.09 | 1.04 | 0.56 | NWT | FAIL |
| c35 | 1 | 24.4 | on |  | NWT | 4.4 | 0.84 | c23 | 2x13.2 | 4.23 | 0.85 | 1.87 | BT | EXCEED |
| c36 | 1 | 13.2 | on | 5% | NWT | 3.27 | 1.19 | c24 | 2x8 | 2.77 | 1.13 | 4.14 | BT | EXCEED |
| d | c25 | 1 | 13.2 | off |  | NWT | 3.5 | 0.99 | c10 | 2x8 | 2.07 | 0.84 | 14.75 | BT | EXCEED |
| 2 | 13.2 | off |  | BT | 3.5 | 0.99 | c09 | 2x7.2 | 2.1 | 0.92 | 13.9 | BT | PASS |
| c26 | 1 | 16.4 | off |  | NWT | 3.82 | 0.88 | c11 | 2x9.6 | 3.22 | 1.07 | 5.78 | BT | EXCEED |
| 2 | 16.4 | off |  | BT | 3.82 | 0.88 | c10 | 2x8 | 2.07 | 0.84 | 19.32 | BT | PASS |
| c27 | 1 | 24.4 | off |  | NWT | 4.21 | 0.86 | c12 | 2x13.2 | 3.82 | 0.97 | 4.09 | BT | EXCEED |
| 2 | 24.4 | off |  | BT | 4.21 | 0.86 | c11 | 2x9.6 | 3.22 | 1.07 | 9.67 | BT | PASS |
| c28 | 1 | 32 | off |  | NWT | 4.31 | 0.69 | c13 | 2x16.4 | 4.11 | 0.8 | 2.54 | BT | EXCEED |
| 2 | 32 | off |  | BT | 4.31 | 0.69 | c12 | 2x13.2 | 3.82 | 0.97 | 5.52 | BT | PASS |
| c29 | 1 | 48 | off |  | NWT | 4.56 | 0.63 | c15 | 2x32 | 4.32 | 0.77 | 3.24 | BT | EXCEED |
| 2 | 48 | off |  | BT | 4.56 | 0.63 | c14 | 2x24.4 | 4.38 | 0.68 | 2.51 | BT | PASS |
| c30 | 1 | 13.2 | off | 5% | NWT | 2.92 | 1.11 | c17 | 2x8 | 1.76 | 0.72 | 11.74 | BT | EXCEED |
| 2 | 13.2 | off | 5% | BT | 2.92 | 1.11 | c16 | 2x7.2 | 1.69 | 0.69 | 12.55 | BT | PASS |
| c31 | 1 | 16.4 | off | 5% | NWT | 3.23 | 1 | c18 | 2x9.6 | 2.64 | 1.06 | 5.48 | BT | EXCEED |
| 2 | 16.4 | off | 5% | BT | 3.23 | 1 | c17 | 2x8 | 1.76 | 0.72 | 16.05 | BT | PASS |
| c32 | 1 | 24.4 | off | 5% | NWT | 3.39 | 1.01 | c19 | 2x13.2 | 3.12 | 0.97 | 2.55 | BT | EXCEED |
| 2 | 24.4 | off | 5% | BT | 3.39 | 1.01 | c18 | 2x9.6 | 2.64 | 1.06 | 6.89 | BT | PASS |
| c33 | 1 | 32 | off | 5% | NWT | 3.6 | 0.96 | c20 | 2x16.4 | 3.29 | 1 | 2.97 | BT | EXCEED |
| 2 | 32 | off | 5% | BT | 3.6 | 0.96 | c19 | 2x13.2 | 3.12 | 0.97 | 4.69 | BT | PASS |
| c34 | 1 | 48 | off | 5% | NWT | 3.76 | 1.03 | c22 | 2x32 | 3.61 | 0.9 | 1.52 | NWT | PASS |
| 2 | 48 | off | 5% | BT | 3.76 | 1.03 | c21 | 2x24.4 | 3.76 | 0.87 | 0.05 | NWT | FAIL |
| c35 | 1 | 24.4 | on |  | NWT | 4.14 | 0.81 | c23 | 2x13.2 | 3.86 | 0.95 | 2.99 | BT | EXCEED |
| c36 | 1 | 13.2 | on | 5% | NWT | 2.97 | 1.03 | c24 | 2x8 | 1.82 | 0.74 | 12.22 | BT | EXCEED |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.2‑2: Summary of the results of P800-1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | DTX | FER | ToR | Status |
| a | c25 | 13.2 | off |  | NWT c10 OR BT c09 | EXCEED |
| c26 | 16.4 | off |  | NWT c11 OR BT c10 | EXCEED |
| c27 | 24.4 | off |  | NWT c12 OR BT c11 | PASS |
| c28 | 32 | off |  | NWT c13 OR BT c12 | PASS |
| c29 | 48 | off |  | NWT c15 OR BT c14 | PASS |
| c30 | 13.2 | off | 5% | NWT c17 OR BT c16 | EXCEED |
| c31 | 16.4 | off | 5% | NWT c18 OR BT c17 | EXCEED |
| c32 | 24.4 | off | 5% | NWT c19 OR BT c18 | PASS |
| c33 | 32 | off | 5% | NWT c20 OR BT c19 | EXCEED |
| c34 | 48 | off | 5% | NWT c22 OR BT c21 | PASS |
| c35 | 24.4 | on |  | NWT c23 | EXCEED |
| c36 | 13.2 | on | 5% | NWT c24 | EXCEED |
| d | c25 | 13.2 | off |  | NWT c10 OR BT c09 | EXCEED |
| c26 | 16.4 | off |  | NWT c11 OR BT c10 | EXCEED |
| c27 | 24.4 | off |  | NWT c12 OR BT c11 | EXCEED |
| c28 | 32 | off |  | NWT c13 OR BT c12 | EXCEED |
| c29 | 48 | off |  | NWT c15 OR BT c14 | EXCEED |
| c30 | 13.2 | off | 5% | NWT c17 OR BT c16 | EXCEED |
| c31 | 16.4 | off | 5% | NWT c18 OR BT c17 | EXCEED |
| c32 | 24.4 | off | 5% | NWT c19 OR BT c18 | EXCEED |
| c33 | 32 | off | 5% | NWT c20 OR BT c19 | EXCEED |
| c34 | 48 | off | 5% | NWT c22 OR BT c21 | PASS |
| c35 | 24.4 | on |  | NWT c23 | EXCEED |
| c36 | 13.2 | on | 5% | NWT c24 | EXCEED |

The following diagrams show the results for a range of conditions from experiment P800-1 as rate-distortion curves. The first two diagrams only show results for clean channel conditions, i.e. conditions c09 – c15 for EVS conditions and c25 – c29 for IVAS conditions. The second two diagrams show results for conditions with 5% simulated frame loss, i.e. conditions c16 – c22 for EVS conditions and c30 – c34 for IVAS conditions.



Figure ‎9.2‑1: P800-1 (stereo/binaural, clean speech, headphone presentation) rate-distortion curves for clean and impaired channel conditions, DTX off

### 9.2.3 Selection Experiment P800-2 (Stereo, Speech+Background, Headphone Presentation)

Selection Experiment P800-2 evaluates IVAS for Stereo speech + background under clean channel conditions with DTX off and on using headphone presentation. See, Annex C.2 for details.

The complete statistical evaluation of the requirement ToR tests for experiment P800-2 is given in the following table. The evaluation is done separately for the data from the two listening laboratories.

Table ‎9.2‑3: Statistical overview on the results of P800-2

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | DTX | Req. | MOS | Std. | Cond. | Bitrate | MOS | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | | |
| b | c26 | 1 | 13.2 | off | NWT | 3.28 | 0.95 | c12 | 2x8 | 2.45 | 1.12 | 7.62 | BT | EXCEED |
| 2 | 13.2 | off | BT | 3.28 | 0.95 | c11 | 2x7.2 | 2.57 | 1.11 | 6.52 | BT | PASS |
| c27 | 1 | 16.4 | off | NWT | 3.54 | 0.95 | c13 | 2x9.6 | 3.43 | 0.98 | 1.15 | NWT | PASS |
| 2 | 16.4 | off | BT | 3.54 | 0.95 | c12 | 2x8 | 2.45 | 1.12 | 10.01 | BT | PASS |
| c28 | 1 | 24.4 | off | NWT | 4.04 | 0.85 | c14 | 2x13.2 | 3.81 | 0.89 | 2.54 | BT | EXCEED |
| 2 | 24.4 | off | BT | 4.04 | 0.85 | c13 | 2x9.6 | 3.43 | 0.98 | 6.39 | BT | PASS |
| c29 | 1 | 32 | off | NWT | 4.21 | 0.75 | c15 | 2x16.4 | 4.19 | 0.77 | 0.21 | NWT | PASS |
| 2 | 32 | off | BT | 4.21 | 0.75 | c14 | 2x13.2 | 3.81 | 0.89 | 4.62 | BT | PASS |
| c30 | 1 | 48 | off | NWT | 4.71 | 0.5 | c17 | 2x32 | 4.41 | 0.73 | 4.54 | BT | EXCEED |
| 2 | 48 | off | BT | 4.71 | 0.5 | c16 | 2x24.4 | 4.54 | 0.6 | 2.95 | BT | PASS |
| c31 | 1 | 64 | off | NWT | 4.72 | 0.48 | c18 | 2x48 | 4.82 | 0.47 | -1.89 | WT | FAIL |
| 2 | 64 | off | BT | 4.72 | 0.48 | c17 | 2x32 | 4.41 | 0.73 | 4.76 | BT | PASS |
| c32 | 1 | 13.2 | on | NWT | 3.33 | 1.02 | c20 | 2x8 | 2.48 | 1.06 | 7.78 | BT | EXCEED |
| 2 | 13.2 | on | BT | 3.33 | 1.02 | c19 | 2x7.2 | 2.39 | 1.02 | 8.72 | BT | PASS |
| c33 | 1 | 16.4 | on | NWT | 3.72 | 0.88 | c21 | 2x9.6 | 3.34 | 0.99 | 3.84 | BT | EXCEED |
| 2 | 16.4 | on | BT | 3.72 | 0.88 | c20 | 2x8 | 2.48 | 1.06 | 12.07 | BT | PASS |
| c34 | 1 | 24.4 | on | NWT | 4 | 0.72 | c22 | 2x13.2 | 3.66 | 1 | 3.75 | BT | EXCEED |
| 2 | 24.4 | on | BT | 4 | 0.72 | c21 | 2x9.6 | 3.34 | 0.99 | 7.27 | BT | PASS |
| c35 | 1 | 32 | on | NWT | 4.16 | 0.78 | c23 | 2x16.4 | 4.1 | 0.83 | 0.66 | NWT | PASS |
| 2 | 32 | on | BT | 4.16 | 0.78 | c22 | 2x13.2 | 3.66 | 1 | 5.31 | BT | PASS |
| c36 | 1 | 48 | on | NWT | 4.52 | 0.59 | c25 | 2x32 | 4.28 | 0.73 | 3.47 | BT | EXCEED |
| 2 | 48 | on | BT | 4.52 | 0.59 | c24 | 2x24.4 | 4.44 | 0.67 | 1.24 | NWT | FAIL |
| d | c26 | 1 | 13.2 | off | NWT | 3.93 | 1.02 | c12 | 2x8 | 2.55 | 1.1 | 12.38 | BT | EXCEED |
| 2 | 13.2 | off | BT | 3.93 | 1.02 | c11 | 2x7.2 | 2.71 | 1.16 | 10.63 | BT | PASS |
| c27 | 1 | 16.4 | off | NWT | 4.21 | 0.88 | c13 | 2x9.6 | 3.67 | 1.14 | 5.06 | BT | EXCEED |
| 2 | 16.4 | off | BT | 4.21 | 0.88 | c12 | 2x8 | 2.55 | 1.1 | 15.84 | BT | PASS |
| c28 | 1 | 24.4 | off | NWT | 4.43 | 0.78 | c14 | 2x13.2 | 4.12 | 0.97 | 3.41 | BT | EXCEED |
| 2 | 24.4 | off | BT | 4.43 | 0.78 | c13 | 2x9.6 | 3.67 | 1.14 | 7.46 | BT | PASS |
| c29 | 1 | 32 | off | NWT | 4.59 | 0.67 | c15 | 2x16.4 | 4.41 | 0.78 | 2.4 | BT | EXCEED |
| 2 | 32 | off | BT | 4.59 | 0.67 | c14 | 2x13.2 | 4.12 | 0.97 | 5.44 | BT | PASS |
| c30 | 1 | 48 | off | NWT | 4.73 | 0.54 | c17 | 2x32 | 4.56 | 0.7 | 2.69 | BT | EXCEED |
| 2 | 48 | off | BT | 4.73 | 0.54 | c16 | 2x24.4 | 4.63 | 0.62 | 1.65 | NWT | FAIL |
| c31 | 1 | 64 | off | NWT | 4.71 | 0.52 | c18 | 2x48 | 4.77 | 0.52 | -1.11 | NWT | PASS |
| 2 | 64 | off | BT | 4.71 | 0.52 | c17 | 2x32 | 4.56 | 0.7 | 2.37 | BT | PASS |
| c32 | 1 | 13.2 | on | NWT | 3.91 | 1.02 | c20 | 2x8 | 2.67 | 1.13 | 10.94 | BT | EXCEED |
| 2 | 13.2 | on | BT | 3.91 | 1.02 | c19 | 2x7.2 | 2.62 | 1.11 | 11.51 | BT | PASS |
| c33 | 1 | 16.4 | on | NWT | 4.27 | 0.84 | c21 | 2x9.6 | 3.68 | 1.18 | 5.45 | BT | EXCEED |
| 2 | 16.4 | on | BT | 4.27 | 0.84 | c20 | 2x8 | 2.67 | 1.13 | 15.23 | BT | PASS |
| c34 | 1 | 24.4 | on | NWT | 4.4 | 0.74 | c22 | 2x13.2 | 4.24 | 0.88 | 1.82 | BT | EXCEED |
| 2 | 24.4 | on | BT | 4.4 | 0.74 | c21 | 2x9.6 | 3.68 | 1.18 | 6.92 | BT | PASS |
| c35 | 1 | 32 | on | NWT | 4.54 | 0.71 | c23 | 2x16.4 | 4.37 | 0.8 | 2.16 | BT | EXCEED |
| 2 | 32 | on | BT | 4.54 | 0.71 | c22 | 2x13.2 | 4.24 | 0.88 | 3.49 | BT | PASS |
| c36 | 1 | 48 | on | NWT | 4.62 | 0.59 | c25 | 2x32 | 4.51 | 0.77 | 1.54 | NWT | PASS |
| 2 | 48 | on | BT | 4.62 | 0.59 | c24 | 2x24.4 | 4.57 | 0.68 | 0.74 | NWT | FAIL |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.2‑4: Summary of the results of P800-2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | DTX | ToR | Status |
| b | c26 | 13.2 | off | NWT c12 OR BT c11 | EXCEED |
| c27 | 16.4 | off | NWT c13 OR BT c12 | PASS |
| c28 | 24.4 | off | NWT c14 OR BT c13 | EXCEED |
| c29 | 32 | off | NWT c15 OR BT c14 | PASS |
| c30 | 48 | off | NWT c17 OR BT c16 | EXCEED |
| c31 | 64 | off | NWT c18 OR BT c17 | PASS |
| c32 | 13.2 | on | NWT c20 OR BT c19 | EXCEED |
| c33 | 16.4 | on | NWT c21 OR BT c20 | EXCEED |
| c34 | 24.4 | on | NWT c22 OR BT c21 | EXCEED |
| c35 | 32 | on | NWT c23 OR BT c22 | PASS |
| c36 | 48 | on | NWT c25 OR BT c24 | PASS |
| d | c26 | 13.2 | off | NWT c12 OR BT c11 | EXCEED |
| c27 | 16.4 | off | NWT c13 OR BT c12 | EXCEED |
| c28 | 24.4 | off | NWT c14 OR BT c13 | EXCEED |
| c29 | 32 | off | NWT c15 OR BT c14 | EXCEED |
| c30 | 48 | off | NWT c17 OR BT c16 | PASS |
| c31 | 64 | off | NWT c18 OR BT c17 | PASS |
| c32 | 13.2 | on | NWT c20 OR BT c19 | EXCEED |
| c33 | 16.4 | on | NWT c21 OR BT c20 | EXCEED |
| c34 | 24.4 | on | NWT c22 OR BT c21 | EXCEED |
| c35 | 32 | on | NWT c23 OR BT c22 | EXCEED |
| c36 | 48 | on | NWT c25 OR BT c24 | PASS |

The following diagrams show the results for a range of conditions from experiment P800-2 as rate-distortion curves. The first two diagrams only show results for active coding conditions (DTX off), i.e. conditions c11 – c18 for EVS conditions and c26 – c31 for IVAS conditions. The second two diagrams show results for DTX conditions (DTX on), i.e. conditions c19 – c25 for EVS conditions and c32 – c36 for IVAS conditions.

Figure ‎9.2‑2: P800-2 (stereo, speech + background, headphone presentation) rate distortion curves for conditions with DTX off and on

### 9.2.4 Selection Experiment P800-3 (Stereo, Mixed and Music, Headphone Presentation)

Selection Experiment P800-3 evaluates IVAS for Stereo mixed and music under clean and impaired channel conditions using headphone presentation. See Annex C.3 for details.

The complete statistical evaluation of the requirement ToR tests for experiment P800-3 is given in the following table. The evaluation is done separately for the data from the two listening laboratories.

Table ‎9.2‑5: Statistical overview on the results of P800-3

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | | | EVS Reference | | | | | Evaluation | | |
| Value | Bitrate | DTX | FER | Req. | MOS | Std. | Cond. | Bitrate | DTX | MOS | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | | | | |
| a | c24 | 1 | 13.2 | off |  | NWT | 3.26 | 1.16 | c10 | 2x8 |  | 1.95 | 1 | 11.43 | BT | EXCEED |
| 2 | 13.2 | off |  | BT | 3.26 | 1.16 | c09 | 2x7.2 |  | 1.89 | 1.02 | 11.83 | BT | PASS |
| c25 | 1 | 16.4 | off |  | NWT | 3.79 | 0.93 | c11 | 2x9.6 |  | 2.96 | 1.25 | 7.18 | BT | EXCEED |
| 2 | 16.4 | off |  | BT | 3.79 | 0.93 | c10 | 2x8 |  | 1.95 | 1 | 18 | BT | PASS |
| c26 | 1 | 24.4 | off |  | NWT | 4.13 | 0.88 | c12 | 2x13.2 |  | 3.41 | 1.17 | 6.62 | BT | EXCEED |
| 2 | 24.4 | off |  | BT | 4.13 | 0.88 | c11 | 2x9.6 |  | 2.96 | 1.25 | 10.32 | BT | PASS |
| c27 | 1 | 32 | off |  | NWT | 4.24 | 0.76 | c13 | 2x16.4 |  | 3.69 | 1.09 | 5.54 | BT | EXCEED |
| 2 | 32 | off |  | BT | 4.24 | 0.76 | c12 | 2x13.2 |  | 3.41 | 1.17 | 8.06 | BT | PASS |
| c28 | 1 | 48 | off |  | NWT | 4.44 | 0.71 | c15 | 2x32 |  | 4.14 | 0.92 | 3.42 | BT | EXCEED |
| 2 | 48 | off |  | BT | 4.44 | 0.71 | c14 | 2x24.4 |  | 4.12 | 0.87 | 3.86 | BT | PASS |
| c29 | 1 | 64 | off |  | NWT | 4.47 | 0.65 | c16 | 2x48 |  | 4.55 | 0.64 | -1.22 | NWT | PASS |
| 2 | 64 | off |  | BT | 4.47 | 0.65 | c15 | 2x32 |  | 4.14 | 0.92 | 3.87 | BT | PASS |
| c30 | 1 | 13.2 | off | 5% | NWT | 2.74 | 1.07 | c18 | 2x8 |  | 1.66 | 0.94 | 10.28 | BT | EXCEED |
| 2 | 13.2 | off | 5% | BT | 2.74 | 1.07 | c17 | 2x7.2 |  | 1.6 | 0.84 | 11.32 | BT | PASS |
| c31 | 1 | 16.4 | off | 5% | NWT | 3.26 | 1.14 | c19 | 2x9.6 |  | 2.71 | 1.18 | 4.5 | BT | EXCEED |
| 2 | 16.4 | off | 5% | BT | 3.26 | 1.14 | c18 | 2x8 |  | 1.66 | 0.94 | 14.6 | BT | PASS |
| c32 | 1 | 24.4 | off | 5% | NWT | 3.42 | 1.13 | c20 | 2x13.2 |  | 2.78 | 1.16 | 5.34 | BT | EXCEED |
| 2 | 24.4 | off | 5% | BT | 3.42 | 1.13 | c19 | 2x9.6 |  | 2.71 | 1.18 | 5.85 | BT | PASS |
| c33 | 1 | 32 | off | 5% | NWT | 3.44 | 1.11 | c21 | 2x16.4 |  | 2.98 | 1.1 | 3.96 | BT | EXCEED |
| 2 | 32 | off | 5% | BT | 3.44 | 1.11 | c20 | 2x13.2 |  | 2.78 | 1.16 | 5.56 | BT | PASS |
| c34 | 1 | 48 | off | 5% | NWT | 3.91 | 0.94 | c23 | 2x32 |  | 3.41 | 1.12 | 4.6 | BT | EXCEED |
| 2 | 48 | off | 5% | BT | 3.91 | 0.94 | c22 | 2x24.4 |  | 3.33 | 1.12 | 5.3 | BT | PASS |
| c35 | 1 | 24.4 | on |  | NWT | 4.17 | 0.87 | c12 | 2x13.2 | off | 3.41 | 1.17 | 7 | BT | EXCEED |
| 2 | 24.4 | on |  | BT | 4.17 | 0.87 | c11 | 2x9.6 | off | 2.96 | 1.25 | 10.71 | BT | PASS |
| c36 | 1 | 13.2 | on | 5% | NWT | 2.76 | 1.16 | c18 | 2x8 | off | 1.66 | 0.94 | 9.98 | BT | EXCEED |
| 2 | 13.2 | on | 5% | BT | 2.76 | 1.16 | c17 | 2x7.2 | off | 1.6 | 0.84 | 10.92 | BT | PASS |
| d | c24 | 1 | 13.2 | off |  | NWT | 3.46 | 1.05 | c10 | 2x8 |  | 2.18 | 0.98 | 11.97 | BT | EXCEED |
| 2 | 13.2 | off |  | BT | 3.46 | 1.05 | c09 | 2x7.2 |  | 2.04 | 0.95 | 13.44 | BT | PASS |
| c25 | 1 | 16.4 | off |  | NWT | 4.06 | 0.85 | c11 | 2x9.6 |  | 3.29 | 1.2 | 6.97 | BT | EXCEED |
| 2 | 16.4 | off |  | BT | 4.06 | 0.85 | c10 | 2x8 |  | 2.18 | 0.98 | 19.43 | BT | PASS |
| c26 | 1 | 24.4 | off |  | NWT | 4.29 | 0.72 | c12 | 2x13.2 |  | 3.64 | 1.04 | 6.89 | BT | EXCEED |
| 2 | 24.4 | off |  | BT | 4.29 | 0.72 | c11 | 2x9.6 |  | 3.29 | 1.2 | 9.51 | BT | PASS |
| c27 | 1 | 32 | off |  | NWT | 4.42 | 0.67 | c13 | 2x16.4 |  | 3.97 | 0.89 | 5.41 | BT | EXCEED |
| 2 | 32 | off |  | BT | 4.42 | 0.67 | c12 | 2x13.2 |  | 3.64 | 1.04 | 8.5 | BT | PASS |
| c28 | 1 | 48 | off |  | NWT | 4.58 | 0.56 | c15 | 2x32 |  | 4.33 | 0.79 | 3.46 | BT | EXCEED |
| 2 | 48 | off |  | BT | 4.58 | 0.56 | c14 | 2x24.4 |  | 4.43 | 0.67 | 2.39 | BT | PASS |
| c29 | 1 | 64 | off |  | NWT | 4.63 | 0.57 | c16 | 2x48 |  | 4.61 | 0.53 | 0.46 | NWT | PASS |
| 2 | 64 | off |  | BT | 4.63 | 0.57 | c15 | 2x32 |  | 4.33 | 0.79 | 4.13 | BT | PASS |
| c30 | 1 | 13.2 | off | 5% | NWT | 3.12 | 1.17 | c18 | 2x8 |  | 1.88 | 0.99 | 10.85 | BT | EXCEED |
| 2 | 13.2 | off | 5% | BT | 3.12 | 1.17 | c17 | 2x7.2 |  | 1.83 | 1 | 11.29 | BT | PASS |
| c31 | 1 | 16.4 | off | 5% | NWT | 3.48 | 1.06 | c19 | 2x9.6 |  | 2.79 | 1.25 | 5.64 | BT | EXCEED |
| 2 | 16.4 | off | 5% | BT | 3.48 | 1.06 | c18 | 2x8 |  | 1.88 | 0.99 | 14.7 | BT | PASS |
| c32 | 1 | 24.4 | off | 5% | NWT | 3.9 | 0.92 | c20 | 2x13.2 |  | 3.21 | 1.16 | 6.3 | BT | EXCEED |
| 2 | 24.4 | off | 5% | BT | 3.9 | 0.92 | c19 | 2x9.6 |  | 2.79 | 1.25 | 9.64 | BT | PASS |
| c33 | 1 | 32 | off | 5% | NWT | 4.03 | 0.93 | c21 | 2x16.4 |  | 3.52 | 1.21 | 4.55 | BT | EXCEED |
| 2 | 32 | off | 5% | BT | 4.03 | 0.93 | c20 | 2x13.2 |  | 3.21 | 1.16 | 7.47 | BT | PASS |
| c34 | 1 | 48 | off | 5% | NWT | 4.14 | 0.84 | c23 | 2x32 |  | 3.76 | 1.09 | 3.74 | BT | EXCEED |
| 2 | 48 | off | 5% | BT | 4.14 | 0.84 | c22 | 2x24.4 |  | 3.83 | 1.04 | 3.08 | BT | PASS |
| c35 | 1 | 24.4 | on |  | NWT | 4.38 | 0.79 | c12 | 2x13.2 | off | 3.64 | 1.04 | 7.58 | BT | EXCEED |
| 2 | 24.4 | on |  | BT | 4.38 | 0.79 | c11 | 2x9.6 | off | 3.29 | 1.2 | 10.09 | BT | PASS |
| c36 | 1 | 13.2 | on | 5% | NWT | 3.09 | 1.15 | c18 | 2x8 | off | 1.88 | 0.99 | 10.65 | BT | EXCEED |
| 2 | 13.2 | on | 5% | BT | 3.09 | 1.15 | c17 | 2x7.2 | off | 1.83 | 1 | 11.09 | BT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.2‑6: Summary of the results of P800-3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | DTX | FER | ToR | Status |
| a | c24 | 13.2 | off |  | NWT c10 OR BT c09 | EXCEED |
| c25 | 16.4 | off |  | NWT c11 OR BT c10 | EXCEED |
| c26 | 24.4 | off |  | NWT c12 OR BT c11 | EXCEED |
| c27 | 32 | off |  | NWT c13 OR BT c12 | EXCEED |
| c28 | 48 | off |  | NWT c15 OR BT c14 | EXCEED |
| c29 | 64 | off |  | NWT c16 OR BT c15 | PASS |
| c30 | 13.2 | off | 5% | NWT c18 OR BT c17 | EXCEED |
| c31 | 16.4 | off | 5% | NWT c19 OR BT c18 | EXCEED |
| c32 | 24.4 | off | 5% | NWT c20 OR BT c19 | EXCEED |
| c33 | 32 | off | 5% | NWT c21 OR BT c20 | EXCEED |
| c34 | 48 | off | 5% | NWT c23 OR BT c22 | EXCEED |
| c35 | 24.4 | on |  | NWT c12 OR BT c11 | EXCEED |
| c36 | 13.2 | on | 5% | NWT c18 OR BT c17 | EXCEED |
| d | c24 | 13.2 | off |  | NWT c10 OR BT c09 | EXCEED |
| c25 | 16.4 | off |  | NWT c11 OR BT c10 | EXCEED |
| c26 | 24.4 | off |  | NWT c12 OR BT c11 | EXCEED |
| c27 | 32 | off |  | NWT c13 OR BT c12 | EXCEED |
| c28 | 48 | off |  | NWT c15 OR BT c14 | EXCEED |
| c29 | 64 | off |  | NWT c16 OR BT c15 | PASS |
| c30 | 13.2 | off | 5% | NWT c18 OR BT c17 | EXCEED |
| c31 | 16.4 | off | 5% | NWT c19 OR BT c18 | EXCEED |
| c32 | 24.4 | off | 5% | NWT c20 OR BT c19 | EXCEED |
| c33 | 32 | off | 5% | NWT c21 OR BT c20 | EXCEED |
| c34 | 48 | off | 5% | NWT c23 OR BT c22 | EXCEED |
| c35 | 24.4 | on |  | NWT c12 OR BT c11 | EXCEED |
| c36 | 13.2 | on | 5% | NWT c18 OR BT c17 | EXCEED |

The following diagrams show the results for a range of conditions from experiment P800-3 as rate-distortion curves. The first two diagrams only show results for clean channel conditions, i.e. conditions c09 – c16 for EVS conditions and c24 – c29 for IVAS conditions. The second two diagrams show results for conditions with 5% simulated frame loss, i.e. conditions c17 – c24 for EVS conditions and c30 – c34 for IVAS conditions.

Figure ‎9.2‑3: P800-3 (stereo, mixed & music) rate distortion curves for clean and impaired channel conditions

### 9.2.5 Selection Experiment BS1534-1a (Stereo, Generic Audio, 48 and 64 kbps, Headphone Presentation)

Selection Experiment BS1534-1a evaluates IVAS for Stereo generic audio at 48 and 64 kbps using headphone presentation. See Annex C.10 for details.

The averaged results per condition for experiment BS1534-1a are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS conditions with increasing bitrate (c03 – c05) and IVAS conditions with increasing bitrate (c06 – c07).

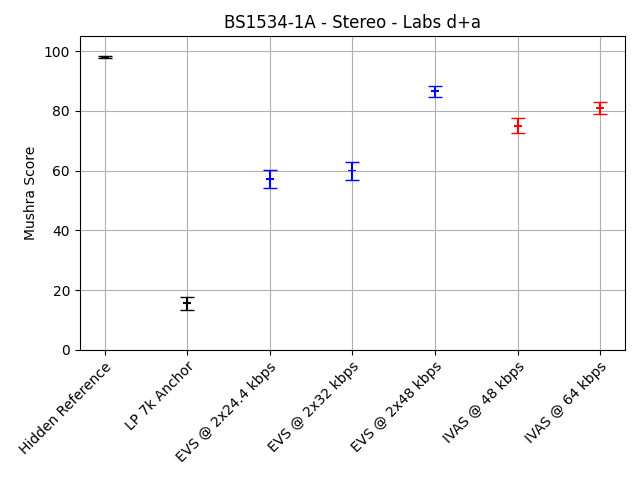
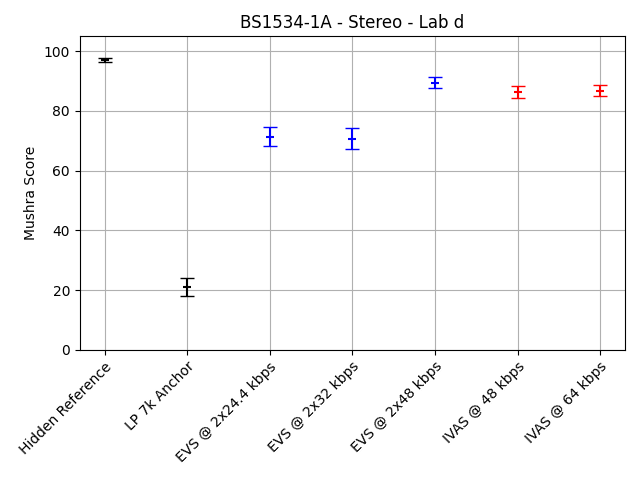
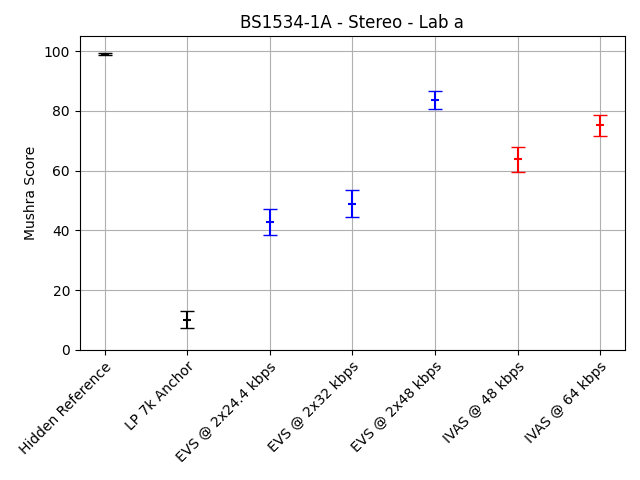


Figure ‎9.2‑4: BS1534-1a (Stereo, generic audio, 48 and 64 kbps, headphone presentation) MUSHRA plots for labs a and d, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-1a is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.2‑7: Statistical overview on the results of BS1534-1a

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| a | c06 | 1 | 48 | NWT | 63.8 | 27.3 | c04 | 2x32 | 48.9 | 29.2 | 4.81 | BT | EXCEED |
| 2 | 48 | BT | 63.8 | 27.3 | c03 | 2x24.4 | 42.9 | 28.1 | 6.92 | BT | PASS |
| c07 | 1 | 64 | NWT | 75.2 | 22.8 | c05 | 2x48 | 83.6 | 19.4 | -3.64 | WT | FAIL |
| 2 | 64 | BT | 75.2 | 22.8 | c04 | 2x32 | 48.9 | 29.2 | 9.18 | BT | PASS |
| d | c06 | 1 | 48 | NWT | 86.5 | 13.3 | c04 | 2x32 | 70.7 | 23.2 | 7.66 | BT | EXCEED |
| 2 | 48 | BT | 86.5 | 13.3 | c03 | 2x24.4 | 71.4 | 21.4 | 7.76 | BT | PASS |
| c07 | 1 | 64 | NWT | 86.8 | 11.6 | c05 | 2x48 | 89.5 | 11.8 | -2.1 | WT | FAIL |
| 2 | 64 | BT | 86.8 | 11.6 | c04 | 2x32 | 70.7 | 23.2 | 8.08 | BT | PASS |
| d+a | c06 | 1 | 48 | NWT | 75.1 | 24.3 | c04 | 2x32 | 59.8 | 28.5 | 7.5 | BT | EXCEED |
| 2 | 48 | BT | 75.1 | 24.3 | c03 | 2x24.4 | 57.1 | 28.7 | 8.77 | BT | PASS |
| c07 | 1 | 64 | NWT | 81 | 19 | c05 | 2x48 | 86.6 | 16.3 | -4.06 | WT | FAIL |
| 2 | 64 | BT | 81 | 19 | c04 | 2x32 | 59.8 | 28.5 | 11.36 | BT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.2‑8: Summary of the results of BS1534-1a

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| a | c06 | 48 | NWT c04 OR BT c03 | EXCEED |
| c07 | 64 | NWT c05 OR BT c04 | PASS |
| d | c06 | 48 | NWT c04 OR BT c03 | EXCEED |
| c07 | 64 | NWT c05 OR BT c04 | PASS |
| d+a | c06 | 48 | NWT c04 OR BT c03 | EXCEED |
| c07 | 64 | NWT c05 OR BT c04 | PASS |

### 9.2.6 Selection Experiment BS1534-1b (Stereo, Generic Audio, 96 and 128 kbps, Headphone Presentation)

Selection Experiment BS1534-1b evaluates IVAS for Stereo generic audio at 96 and 128 kbps using headphone presentation. Se IVAS-8a, Annex F.2 for details.

The averaged results per condition for experiment BS1534-1b are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS conditions with increasing bitrate (c03 – c05) and IVAS conditions with increasing bitrate (c06 – c07).

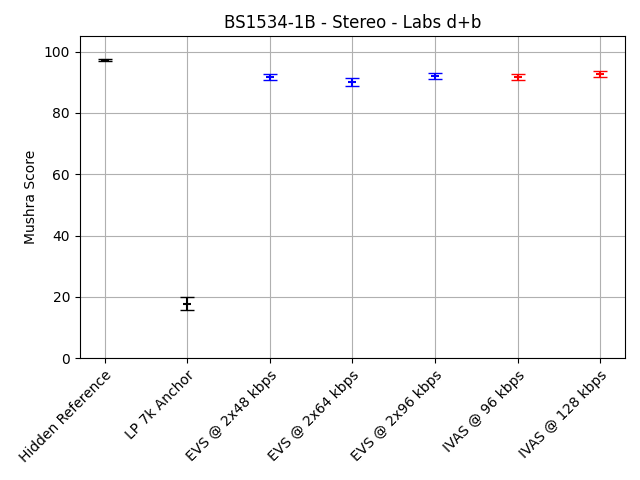
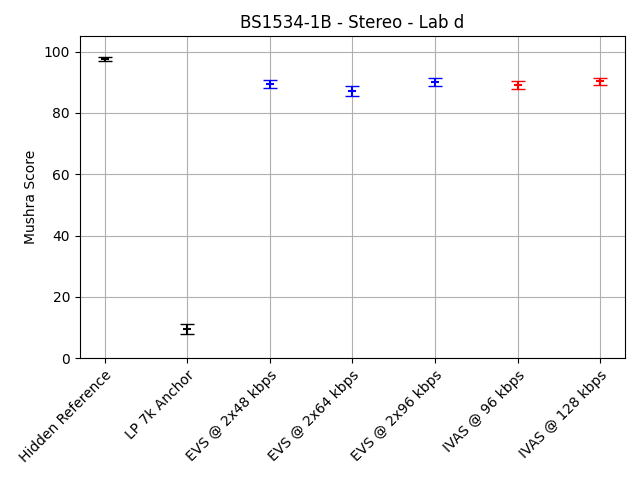
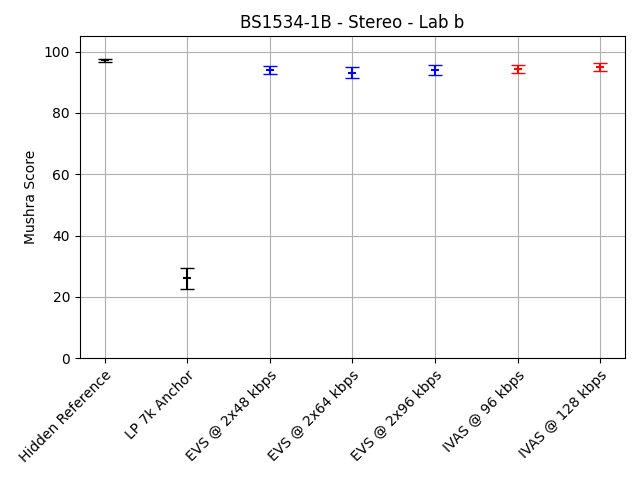


Figure ‎9.2‑5: BS1534-1b (Stereo, generic audio, 96 and 128 kbps, headphone presentation) MUSHRA plots for labs b and d, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-1b is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.2‑9: Statistical overview on the results of BS1534-1b

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| b | c06 | 1 | 96 | NWT | 94.4 | 9 | c04 | 2x64 | 93.1 | 11.9 | 1.13 | NWT | PASS |
| 2 | 96 | BT | 94.4 | 9 | c03 | 2x48 | 94 | 9.5 | 0.37 | NWT | FAIL |
| c07 | 1 | 128 | NWT | 95 | 8.5 | c05 | 2x96 | 93.9 | 10.9 | 0.98 | NWT | PASS |
| 2 | 128 | BT | 95 | 8.5 | c04 | 2x64 | 93.1 | 11.9 | 1.68 | BT | PASS |
| d | c06 | 1 | 96 | NWT | 89.2 | 8.9 | c04 | 2x64 | 87.2 | 10.5 | 1.8 | BT | EXCEED |
| 2 | 96 | BT | 89.2 | 8.9 | c03 | 2x48 | 89.4 | 8 | -0.3 | NWT | FAIL |
| c07 | 1 | 128 | NWT | 90.4 | 7.5 | c05 | 2x96 | 90.2 | 8.7 | 0.26 | NWT | PASS |
| 2 | 128 | BT | 90.4 | 7.5 | c04 | 2x64 | 87.2 | 10.5 | 3.16 | BT | PASS |
| d+b | c06 | 1 | 96 | NWT | 91.8 | 9.3 | c04 | 2x64 | 90.2 | 11.6 | 1.98 | BT | EXCEED |
| 2 | 96 | BT | 91.8 | 9.3 | c03 | 2x48 | 91.7 | 9 | 0.06 | NWT | FAIL |
| c07 | 1 | 128 | NWT | 92.7 | 8.3 | c05 | 2x96 | 92.1 | 10 | 0.9 | NWT | PASS |
| 2 | 128 | BT | 92.7 | 8.3 | c04 | 2x64 | 90.2 | 11.6 | 3.24 | BT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.2‑10: Summary of the results of BS1534-1b

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| b | c06 | 96 | NWT c04 OR BT c03 | PASS |
| c07 | 128 | NWT c05 OR BT c04 | PASS |
| d | c06 | 96 | NWT c04 OR BT c03 | PASS |
| c07 | 128 | NWT c05 OR BT c04 | PASS |
| d+b | c06 | 96 | NWT c04 OR BT c03 | PASS |
| c07 | 128 | NWT c05 OR BT c04 | PASS |

## 9.3 Scene-based Audio (SBA, Ambisonics)

### 9.3.1 Overview

In Selection phase, six experiments have been conducted to evaluate the performance of the IVAS codec with Ambisonics content. While the experiments P800-4 and P800-5 were conducted as P.800 DCR tests on FOA content, the experiments BS1534-4a, BS1534-4b, BS1534-5a and BS1534-5b were conducted as BS.1534 tests for different Ambisonics orders. The experiments P800-4, P800-5, BS1534-4a, BS1534-4b, BS1534-5a were conducted using headphone presentation, whereas experiment BS1534-5b was conducted using 7.1+4 loudspeaker presentation.

- Selection Experiment P800-4: Clean speech, FOA under clean and impaired channel conditions, headphone presentation

- Selection Experiment P800-5: Speech + background, FOA under clean channel conditions, DTX off and on, headphone presentation

- Selection Experiment BS1534-4a: Generic audio, FOA input, HOA3 output, 96 kbps, 128 kbps and 160 kbps, headphone presentation

- Selection Experiment BS1534-4b: Generic audio, HOA2 input, HOA3 output, 160 and 192 kbps, headphone presentation

- Selection Experiment BS1534-5a: Generic audio, HOA3 input, HOA3 output, 192 and 256 kbps, headphone presentation

- Selection Experiment BS1534-5b: Generic audio, HOA3 input, HOA3 output, 384 and 512 kbps, 7.1+4 loudspeaker presentation

### 9.3.2 Selection Experiment P800-4 (FOA, Clean Speech, Headphone Presentation)

Selection Experiment P800-4 evaluates IVAS for FOA clean speech under clean and impaired channel conditions using headphone presentation. See Annex C.4 for details.

The complete statistical evaluation of the requirement ToR tests for experiment P800-4 is given in the following table. The evaluation is done separately for the data from the two listening laboratories.

Table ‎9.3‑1: Statistical overview on the results of P800-4

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | DTX | FER | Req. | MOS | Std. | Cond. | Bitrate | MOS | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | | | |
| a | c24 | 1 | 16.4 | off |  | NWT | 3.13 | 1.06 | c09 | 3x7.2 | 2.47 | 1.14 | 5.66 | BT | EXCEED |
| c25 | 1 | 24.4 | off |  | NWT | 3.49 | 0.95 | c10 | 4x7.2 | 2.54 | 1.07 | 8.85 | BT | EXCEED |
| c26 | 1 | 32 | off |  | NWT | 3.74 | 0.98 | c12 | 4x9.6 | 3.4 | 0.98 | 3.33 | BT | EXCEED |
| 2 | 32 | off |  | BT | 3.74 | 0.98 | c11 | 4x8 | 2.73 | 1.11 | 9.24 | BT | PASS |
| c27 | 1 | 48 | off |  | NWT | 4.31 | 0.76 | c14 | 4x16.4 | 4.34 | 0.77 | -0.41 | NWT | PASS |
| 2 | 48 | off |  | BT | 4.31 | 0.76 | c13 | 4x13.2 | 4.1 | 0.88 | 2.43 | BT | PASS |
| c28 | 1 | 64 | off |  | NWT | 4.61 | 0.59 | c15 | 4x24.4 | 4.52 | 0.66 | 1.35 | NWT | PASS |
| 2 | 64 | off |  | BT | 4.61 | 0.59 | c14 | 4x16.4 | 4.34 | 0.77 | 3.68 | BT | PASS |
| c29 | 1 | 80 | off |  | NWT | 4.66 | 0.51 | c15 | 4x24.4 | 4.52 | 0.66 | 2.17 | BT | EXCEED |
| 2 | 80 | off |  | BT | 4.66 | 0.51 | c14 | 4x16.4 | 4.34 | 0.77 | 4.52 | BT | PASS |
| c30 | 1 | 96 | off |  | NWT | 4.72 | 0.47 | c16 | 4x32 | 4.52 | 0.64 | 3.46 | BT | EXCEED |
| 2 | 96 | off |  | BT | 4.72 | 0.47 | c15 | 4x24.4 | 4.52 | 0.66 | 3.32 | BT | PASS |
| c31 | 1 | 24.4 | off | 5% | NWT | 3.07 | 1.08 | c17 | 4x7.2 | 2.27 | 1.05 | 7.07 | BT | EXCEED |
| c32 | 1 | 32 | off | 5% | NWT | 3.31 | 0.97 | c19 | 4x9.6 | 2.93 | 1.09 | 3.53 | BT | EXCEED |
| 2 | 32 | off | 5% | BT | 3.31 | 0.97 | c18 | 4x8 | 2.38 | 1.14 | 8.34 | BT | PASS |
| c33 | 1 | 48 | off | 5% | NWT | 3.81 | 0.95 | c21 | 4x16.4 | 3.74 | 0.88 | 0.64 | NWT | PASS |
| 2 | 48 | off | 5% | BT | 3.81 | 0.95 | c20 | 4x13.2 | 3.52 | 0.99 | 2.83 | BT | PASS |
| c34 | 1 | 64 | off | 5% | NWT | 4.1 | 0.96 | c22 | 4x24.4 | 3.96 | 0.87 | 1.44 | NWT | PASS |
| 2 | 64 | off | 5% | BT | 4.1 | 0.96 | c21 | 4x16.4 | 3.74 | 0.88 | 3.66 | BT | PASS |
| c35 | 1 | 80 | off | 5% | NWT | 4.15 | 0.91 | c22 | 4x24.4 | 3.96 | 0.87 | 2.01 | BT | EXCEED |
| 2 | 80 | off | 5% | BT | 4.15 | 0.91 | c21 | 4x16.4 | 3.74 | 0.88 | 4.3 | BT | PASS |
| c36 | 1 | 96 | off | 5% | NWT | 4.02 | 0.92 | c23 | 4x32 | 3.85 | 0.87 | 1.83 | BT | EXCEED |
| 2 | 96 | off | 5% | BT | 4.02 | 0.92 | c22 | 4x24.4 | 3.96 | 0.87 | 0.65 | NWT | FAIL |
| c | c24 | 1 | 16.4 | off |  | NWT | 3.53 | 0.92 | c09 | 3x7.2 | 2.44 | 1 | 10.82 | BT | EXCEED |
| c25 | 1 | 24.4 | off |  | NWT | 3.68 | 0.86 | c10 | 4x7.2 | 2.44 | 0.89 | 13.45 | BT | EXCEED |
| c26 | 1 | 32 | off |  | NWT | 3.77 | 0.83 | c12 | 4x9.6 | 3.51 | 0.89 | 2.88 | BT | EXCEED |
| 2 | 32 | off |  | BT | 3.77 | 0.83 | c11 | 4x8 | 2.59 | 0.96 | 12.44 | BT | PASS |
| c27 | 1 | 48 | off |  | NWT | 4.39 | 0.63 | c14 | 4x16.4 | 4.29 | 0.67 | 1.46 | NWT | PASS |
| 2 | 48 | off |  | BT | 4.39 | 0.63 | c13 | 4x13.2 | 4.1 | 0.73 | 4.01 | BT | PASS |
| c28 | 1 | 64 | off |  | NWT | 4.61 | 0.54 | c15 | 4x24.4 | 4.5 | 0.62 | 1.81 | BT | EXCEED |
| 2 | 64 | off |  | BT | 4.61 | 0.54 | c14 | 4x16.4 | 4.29 | 0.67 | 5 | BT | PASS |
| c29 | 1 | 80 | off |  | NWT | 4.64 | 0.57 | c15 | 4x24.4 | 4.5 | 0.62 | 2.22 | BT | EXCEED |
| 2 | 80 | off |  | BT | 4.64 | 0.57 | c14 | 4x16.4 | 4.29 | 0.67 | 5.34 | BT | PASS |
| c30 | 1 | 96 | off |  | NWT | 4.66 | 0.51 | c16 | 4x32 | 4.46 | 0.62 | 3.35 | BT | EXCEED |
| 2 | 96 | off |  | BT | 4.66 | 0.51 | c15 | 4x24.4 | 4.5 | 0.62 | 2.69 | BT | PASS |
| c31 | 1 | 24.4 | off | 5% | NWT | 3.14 | 0.99 | c17 | 4x7.2 | 2.18 | 0.96 | 9.34 | BT | EXCEED |
| c32 | 1 | 32 | off | 5% | NWT | 3.51 | 0.89 | c19 | 4x9.6 | 3.08 | 1.02 | 4.19 | BT | EXCEED |
| 2 | 32 | off | 5% | BT | 3.51 | 0.89 | c18 | 4x8 | 2.26 | 0.92 | 13.1 | BT | PASS |
| c33 | 1 | 48 | off | 5% | NWT | 3.94 | 0.88 | c21 | 4x16.4 | 3.63 | 0.92 | 3.32 | BT | EXCEED |
| 2 | 48 | off | 5% | BT | 3.94 | 0.88 | c20 | 4x13.2 | 3.49 | 0.97 | 4.65 | BT | PASS |
| c34 | 1 | 64 | off | 5% | NWT | 4.13 | 0.76 | c22 | 4x24.4 | 4.09 | 0.75 | 0.49 | NWT | PASS |
| 2 | 64 | off | 5% | BT | 4.13 | 0.76 | c21 | 4x16.4 | 3.63 | 0.92 | 5.61 | BT | PASS |
| c35 | 1 | 80 | off | 5% | NWT | 4.18 | 0.7 | c22 | 4x24.4 | 4.09 | 0.75 | 1.23 | NWT | PASS |
| 2 | 80 | off | 5% | BT | 4.18 | 0.7 | c21 | 4x16.4 | 3.63 | 0.92 | 6.44 | BT | PASS |
| c36 | 1 | 96 | off | 5% | NWT | 4.14 | 0.8 | c23 | 4x32 | 3.81 | 0.76 | 4.05 | BT | EXCEED |
| 2 | 96 | off | 5% | BT | 4.14 | 0.8 | c22 | 4x24.4 | 4.09 | 0.75 | 0.61 | NWT | FAIL |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.3‑2: Summary of the results of P800-4

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | DTX | FER | ToR | Status |
| a | c24 | 16.4 | off |  | NWT c09 | EXCEED |
| c25 | 24.4 | off |  | NWT c10 | EXCEED |
| c26 | 32 | off |  | NWT c12 or BT c11 | EXCEED |
| c27 | 48 | off |  | NWT c14 or BT c13 | PASS |
| c28 | 64 | off |  | NWT c15 or BT c14 | PASS |
| c29 | 80 | off |  | NWT c15 or BT c14 | EXCEED |
| c30 | 96 | off |  | NWT c16 or BT c15 | EXCEED |
| c31 | 24.4 | off | 5% | NWT c17 | EXCEED |
| c32 | 32 | off | 5% | NWT c19 or BT c18 | EXCEED |
| c33 | 48 | off | 5% | NWT c21 or BT c20 | PASS |
| c34 | 64 | off | 5% | NWT c22 or BT c21 | PASS |
| c35 | 80 | off | 5% | NWT c22 or BT c21 | EXCEED |
| c36 | 96 | off | 5% | NWT c23 or BT c22 | PASS |
| c | c24 | 16.4 | off |  | NWT c09 | EXCEED |
| c25 | 24.4 | off |  | NWT c10 | EXCEED |
| c26 | 32 | off |  | NWT c12 or BT c11 | EXCEED |
| c27 | 48 | off |  | NWT c14 or BT c13 | PASS |
| c28 | 64 | off |  | NWT c15 or BT c14 | EXCEED |
| c29 | 80 | off |  | NWT c15 or BT c14 | EXCEED |
| c30 | 96 | off |  | NWT c16 or BT c15 | EXCEED |
| c31 | 24.4 | off | 5% | NWT c17 | EXCEED |
| c32 | 32 | off | 5% | NWT c19 or BT c18 | EXCEED |
| c33 | 48 | off | 5% | NWT c21 or BT c20 | EXCEED |
| c34 | 64 | off | 5% | NWT c22 or BT c21 | PASS |
| c35 | 80 | off | 5% | NWT c22 or BT c21 | PASS |
| c36 | 96 | off | 5% | NWT c23 or BT c22 | PASS |

The following diagrams show the results for a range of conditions from experiment P800-4 as rate-distortion curves. The first two diagrams only show results for clean channel conditions, i.e. conditions c1 – c16 for EVS conditions and c24 – c30 for IVAS conditions. The second two diagrams show results for conditions with 5% simulated frame loss, i.e. conditions c17 – c23 for EVS conditions and c31 – c36 for IVAS conditions.

Figure ‎9.3‑1: P800-4 (Ambisonics FOA input/output, clean speech, headphone presentation) rate distortion curves for clean and impaired channel conditions

### 9.3.3 Selection Experiment P800-5 (FOA, Speech+Background, Headphone Presentation)

Selection Experiment P800-5 evaluates IVAS for FOA speech + background under clean channel conditions, DTX off and on using headphone presentation. See Annex C.5 for details.

The complete statistical evaluation of the requirement ToR tests for experiment P800-5 is given in the following table. The evaluation is done separately for the data from the two listening laboratories.

Table ‎9.3‑3: Statistical overview on the results of P800-5

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | DTX | Req. | MOS | Std. | Cond. | Bitrate | MOS | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | | |
| a | c24 | 1 | 16.4 | off | NWT | 3.26 | 0.96 | c09 | 3x7.2 | 2.93 | 1.04 | 3.06 | BT | EXCEED |
| c25 | 1 | 24.4 | off | NWT | 3.68 | 1 | c10 | 4x7.2 | 3.04 | 1.08 | 5.82 | BT | EXCEED |
| c26 | 1 | 32 | off | NWT | 3.84 | 0.96 | c12 | 4x9.6 | 3.74 | 0.88 | 1.08 | NWT | PASS |
| 2 | 32 | off | BT | 3.84 | 0.96 | c11 | 4x8 | 3.15 | 1.03 | 6.6 | BT | PASS |
| c27 | 1 | 48 | off | NWT | 4.45 | 0.66 | c14 | 4x16.4 | 4.27 | 0.68 | 2.58 | BT | EXCEED |
| 2 | 48 | off | BT | 4.45 | 0.66 | c13 | 4x13.2 | 4 | 0.85 | 5.62 | BT | PASS |
| c28 | 1 | 64 | off | NWT | 4.54 | 0.59 | c15 | 4x24.4 | 4.47 | 0.71 | 1.12 | NWT | PASS |
| 2 | 64 | off | BT | 4.54 | 0.59 | c14 | 4x16.4 | 4.27 | 0.68 | 4.12 | BT | PASS |
| c29 | 1 | 80 | off | NWT | 4.56 | 0.65 | c15 | 4x24.4 | 4.47 | 0.71 | 1.31 | NWT | PASS |
| 2 | 80 | off | BT | 4.56 | 0.65 | c14 | 4x16.4 | 4.27 | 0.68 | 4.18 | BT | PASS |
| c30 | 1 | 96 | off | NWT | 4.54 | 0.65 | c16 | 4x32 | 4.38 | 0.68 | 2.38 | BT | EXCEED |
| 2 | 96 | off | BT | 4.54 | 0.65 | c15 | 4x24.4 | 4.47 | 0.71 | 1.07 | NWT | FAIL |
| c31 | 1 | 16.4 | on | NWT | 3.26 | 1 | c17 | 3x7.2 | 2.78 | 1.09 | 4.28 | BT | EXCEED |
| c32 | 1 | 24.4 | on | NWT | 3.49 | 0.98 | c18 | 4x7.2 | 2.96 | 1.01 | 5.09 | BT | EXCEED |
| c33 | 1 | 32 | on | NWT | 3.8 | 0.88 | c20 | 4x9.6 | 3.64 | 0.9 | 1.72 | BT | EXCEED |
| 2 | 32 | on | BT | 3.8 | 0.88 | c19 | 4x8 | 2.97 | 0.98 | 8.42 | BT | PASS |
| c34 | 1 | 48 | on | NWT | 4.26 | 0.7 | c22 | 4x16.4 | 4.13 | 0.66 | 1.79 | BT | EXCEED |
| 2 | 48 | on | BT | 4.26 | 0.7 | c21 | 4x13.2 | 3.87 | 0.88 | 4.69 | BT | PASS |
| c35 | 1 | 64 | on | NWT | 4.25 | 0.76 | c23 | 4x24.4 | 4.2 | 0.86 | 0.58 | NWT | PASS |
| 2 | 64 | on | BT | 4.25 | 0.76 | c22 | 4x16.4 | 4.13 | 0.66 | 1.56 | NWT | FAIL |
| c36 | 1 | 80 | on | NWT | 4.3 | 0.75 | c23 | 4x24.4 | 4.2 | 0.86 | 1.18 | NWT | PASS |
| 2 | 80 | on | BT | 4.3 | 0.75 | c22 | 4x16.4 | 4.13 | 0.66 | 2.26 | BT | PASS |
| b | c24 | 1 | 16.4 | off | NWT | 3.53 | 0.97 | c09 | 3x7.2 | 2.82 | 0.99 | 6.91 | BT | EXCEED |
| c25 | 1 | 24.4 | off | NWT | 3.78 | 1.01 | c10 | 4x7.2 | 2.93 | 1.04 | 7.88 | BT | EXCEED |
| c26 | 1 | 32 | off | NWT | 3.96 | 0.86 | c12 | 4x9.6 | 3.81 | 0.99 | 1.53 | NWT | PASS |
| 2 | 32 | off | BT | 3.96 | 0.86 | c11 | 4x8 | 2.87 | 1.05 | 10.75 | BT | PASS |
| c27 | 1 | 48 | off | NWT | 4.61 | 0.57 | c14 | 4x16.4 | 4.45 | 0.64 | 2.44 | BT | EXCEED |
| 2 | 48 | off | BT | 4.61 | 0.57 | c13 | 4x13.2 | 4.16 | 0.8 | 6.15 | BT | PASS |
| c28 | 1 | 64 | off | NWT | 4.77 | 0.46 | c15 | 4x24.4 | 4.63 | 0.55 | 2.6 | BT | EXCEED |
| 2 | 64 | off | BT | 4.77 | 0.46 | c14 | 4x16.4 | 4.45 | 0.64 | 5.4 | BT | PASS |
| c29 | 1 | 80 | off | NWT | 4.71 | 0.53 | c15 | 4x24.4 | 4.63 | 0.55 | 1.45 | NWT | PASS |
| 2 | 80 | off | BT | 4.71 | 0.53 | c14 | 4x16.4 | 4.45 | 0.64 | 4.21 | BT | PASS |
| c30 | 1 | 96 | off | NWT | 4.75 | 0.45 | c16 | 4x32 | 4.6 | 0.59 | 2.71 | BT | EXCEED |
| 2 | 96 | off | BT | 4.75 | 0.45 | c15 | 4x24.4 | 4.63 | 0.55 | 2.31 | BT | PASS |
| c31 | 1 | 16.4 | on | NWT | 3.52 | 1.04 | c17 | 3x7.2 | 2.72 | 0.99 | 7.5 | BT | EXCEED |
| c32 | 1 | 24.4 | on | NWT | 3.76 | 0.94 | c18 | 4x7.2 | 2.73 | 1.02 | 9.97 | BT | EXCEED |
| c33 | 1 | 32 | on | NWT | 3.79 | 0.99 | c20 | 4x9.6 | 3.83 | 0.88 | -0.34 | NWT | PASS |
| 2 | 32 | on | BT | 3.79 | 0.99 | c19 | 4x8 | 2.88 | 1.06 | 8.49 | BT | PASS |
| c34 | 1 | 48 | on | NWT | 4.52 | 0.58 | c22 | 4x16.4 | 4.33 | 0.72 | 2.73 | BT | EXCEED |
| 2 | 48 | on | BT | 4.52 | 0.58 | c21 | 4x13.2 | 4.02 | 0.84 | 6.65 | BT | PASS |
| c35 | 1 | 64 | on | NWT | 4.63 | 0.53 | c23 | 4x24.4 | 4.51 | 0.68 | 1.9 | BT | EXCEED |
| 2 | 64 | on | BT | 4.63 | 0.53 | c22 | 4x16.4 | 4.33 | 0.72 | 4.41 | BT | PASS |
| c36 | 1 | 80 | on | NWT | 4.73 | 0.47 | c23 | 4x24.4 | 4.51 | 0.68 | 3.69 | BT | EXCEED |
| 2 | 80 | on | BT | 4.73 | 0.47 | c22 | 4x16.4 | 4.33 | 0.72 | 6.22 | BT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.3‑4: Summary of the results of P800-5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | DTX | ToR | Status |
| a | c24 | 16.4 | Off | NWT c09 | EXCEED |
| c25 | 24.4 | Off | NWT c10 | EXCEED |
| c26 | 32 | Off | NWT c12 or BT c11 | PASS |
| c27 | 48 | Off | NWT c14 or BT c13 | EXCEED |
| c28 | 64 | Off | NWT c15 or BT c14 | PASS |
| c29 | 80 | off | NWT c15 or BT c14 | PASS |
| c30 | 96 | off | NWT c16 or BT c15 | PASS |
| c31 | 16.4 | on | NWT c17 | EXCEED |
| c32 | 24.4 | on | NWT c18 | EXCEED |
| c33 | 32 | on | NWT c20 or BT c19 | EXCEED |
| c34 | 48 | on | NWT c22 or BT c21 | EXCEED |
| c35 | 64 | on | NWT c23 or BT c22 | PASS |
| c36 | 80 | on | NWT c23 or BT c22 | PASS |
| b | c24 | 16.4 | off | NWT c09 | EXCEED |
| c25 | 24.4 | off | NWT c10 | EXCEED |
| c26 | 32 | off | NWT c12 or BT c11 | PASS |
| c27 | 48 | off | NWT c14 or BT c13 | EXCEED |
| c28 | 64 | off | NWT c15 or BT c14 | EXCEED |
| c29 | 80 | off | NWT c15 or BT c14 | PASS |
| c30 | 96 | off | NWT c16 or BT c15 | EXCEED |
| c31 | 16.4 | on | NWT c17 | EXCEED |
| c32 | 24.4 | on | NWT c18 | EXCEED |
| c33 | 32 | on | NWT c20 or BT c19 | PASS |
| c34 | 48 | on | NWT c22 or BT c21 | EXCEED |
| c35 | 64 | on | NWT c23 or BT c22 | EXCEED |
| c36 | 80 | on | NWT c23 or BT c22 | EXCEED |

The following diagrams show the results for a range of conditions from experiment P800-5 as rate-distortion curves. The first two diagrams only show results for active coding conditions (DTX off), i.e. conditions c10 – c16 for EVS conditions and c24 – c30 for IVAS conditions. The second two diagrams show results for DTX conditions (DTX on), i.e. conditions c18 – c23 for EVS conditions and c31 – c36 for IVAS conditions.

Figure ‎9.3‑2: P800-5 (Ambisonics FOA input/output, speech + background, headphone presentation) rate distortion curves for conditions with DTX off and on

### 9.3.4 Selection Experiment BS1534-4a (FOA, Generic Audio, 96, 128 and 160 kbps, Headphone Presentation)

Selection Experiment BS1534-4a evaluates IVAS for Ambisonics generic audio for FOA input and HOA3 output at 96 kbps, 128 kbps and 160 kbps using headphone presentation. See , Annex C.16 for details.

The averaged results per condition for experiment BS1534-4a are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS conditions with increasing bitrate (c03 – c05) and IVAS conditions with increasing bitrate (c06 – c08).

Ein Bild, das Text, Reihe, Diagramm, Zahl enthält.

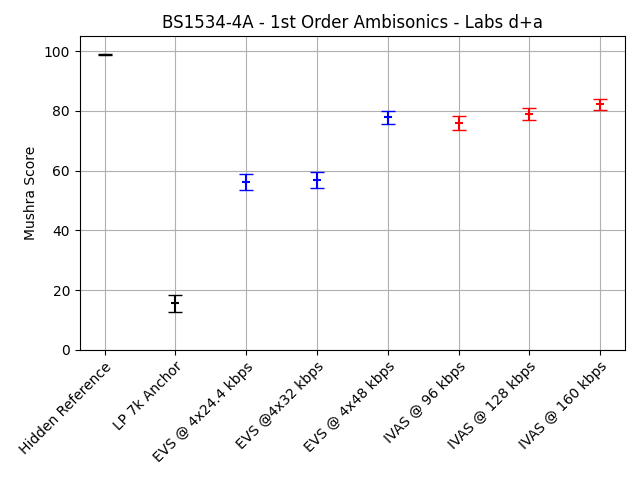
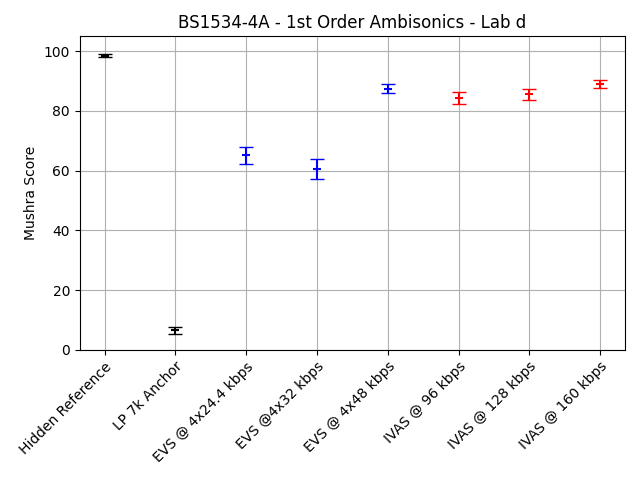
Automatisch generierte Beschreibung

Figure ‎9.3‑3: BS1534-4a (Ambisonics FOA input and HOA3 output, generic Audio, 96, 128 and 160 kbps, headphone presentation) MUSHRA plots for labs a and d, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-4a is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.3‑5: Statistical overview on the results of BS1534-4a

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| a | c06 | 1 | 96 | NWT | 67.7 | 25 | c04 | 4x32 | 53.4 | 27.2 | 5 | BT | EXCEED |
| 2 | 96 | BT | 67.7 | 25 | c03 | 4x24.4 | 47.3 | 28.7 | 6.93 | BT | PASS |
| c07 | 1 | 128 | NWT | 72.3 | 23.6 | c05 | 4x48 | 68.4 | 23 | 1.53 | NWT | PASS |
| 2 | 128 | BT | 72.3 | 23.6 | c04 | 4x32 | 53.4 | 27.2 | 6.78 | BT | PASS |
| c08 | 1 | 160 | NWT | 75.4 | 21 | c05 | 4x48 | 68.4 | 23 | 2.93 | BT | EXCEED |
| 2 | 160 | BT | 75.4 | 21 | c04 | 4x32 | 53.4 | 27.2 | 8.28 | BT | PASS |
| d | c06 | 1 | 96 | NWT | 84.3 | 13.4 | c04 | 4x32 | 60.5 | 22.6 | 11.74 | BT | EXCEED |
| 2 | 96 | BT | 84.3 | 13.4 | c03 | 4x24.4 | 65.1 | 19 | 10.68 | BT | PASS |
| c07 | 1 | 128 | NWT | 85.7 | 12.3 | c05 | 4x48 | 87.5 | 10.3 | -1.48 | NWT | PASS |
| 2 | 128 | BT | 85.7 | 12.3 | c04 | 4x32 | 60.5 | 22.6 | 12.67 | BT | PASS |
| c08 | 1 | 160 | NWT | 89.2 | 8.7 | c05 | 4x48 | 87.5 | 10.3 | 1.6 | NWT | PASS |
| 2 | 160 | BT | 89.2 | 8.7 | c04 | 4x32 | 60.5 | 22.6 | 15.33 | BT | PASS |
| d+a | c06 | 1 | 96 | NWT | 76 | 21.7 | c04 | 4x32 | 57 | 25.2 | 10.48 | BT | EXCEED |
| 2 | 96 | BT | 76 | 21.7 | c03 | 4x24.4 | 56.2 | 25.9 | 10.72 | BT | PASS |
| c07 | 1 | 128 | NWT | 79 | 19.9 | c05 | 4x48 | 77.9 | 20.2 | 0.67 | NWT | PASS |
| 2 | 128 | BT | 79 | 19.9 | c04 | 4x32 | 57 | 25.2 | 12.53 | BT | PASS |
| c08 | 1 | 160 | NWT | 82.3 | 17.5 | c05 | 4x48 | 77.9 | 20.2 | 2.99 | BT | EXCEED |
| 2 | 160 | BT | 82.3 | 17.5 | c04 | 4x32 | 57 | 25.2 | 15.12 | BT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.3‑6: Summary of the results of BS1534-4a

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| a | c06 | 96 | NWT c04 OR BT c03 | EXCEED |
| c07 | 128 | NWT c05 OR BT c04 | PASS |
| c08 | 160 | NWT c05 OR BT c04 | EXCEED |
| d | c06 | 96 | NWT c04 OR BT c03 | EXCEED |
| c07 | 128 | NWT c05 OR BT c04 | PASS |
| c08 | 160 | NWT c05 OR BT c04 | PASS |
| d+a | c06 | 96 | NWT c04 OR BT c03 | EXCEED |
| c07 | 128 | NWT c05 OR BT c04 | PASS |
| c08 | 160 | NWT c05 OR BT c04 | EXCEED |

### 9.3.5 Selection Experiment BS1534-4b (HOA2, Generic Audio, 160 and 192 kbps, Headphone Presentation)

Selection Experiment BS1534-4b evaluates IVAS for Ambisonics generic audio for HOA2 input and HOA3 output at 160 and 192 kbps using headphone presentation. See Annex C.17 for details.

The averaged results per condition for experiment BS1534-4b are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS conditions with increasing bitrate (c03 – c05) and IVAS conditions with increasing bitrate (c06 – c07).

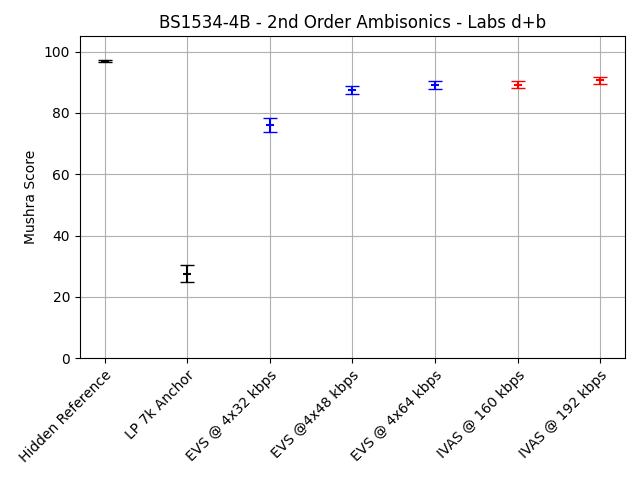
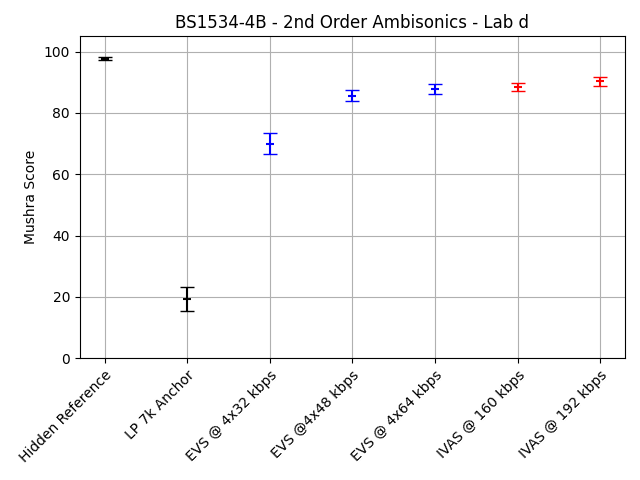
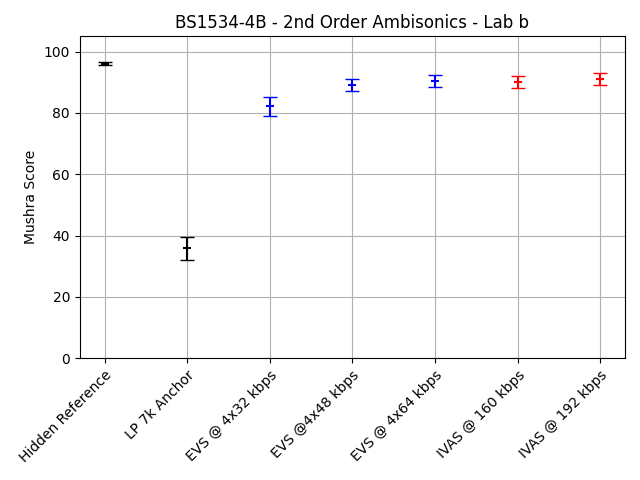


Figure ‎9.3‑4: BS1534-4b (Ambisonics HOA2 input and HOA3 output, generic Audio, 160 and 192 kbps, headphone presentation) MUSHRA plots for labs b and d, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-4b is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.3‑7: Statistical overview on the results of BS1534-4b

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| b | c06 | 1 | 160 | NWT | 90 | 13.1 | c04 | 4x48 | 89.1 | 13.3 | 0.63 | NWT | PASS |
| 2 | 160 | BT | 90 | 13.1 | c03 | 4x32 | 82.2 | 20.2 | 4.21 | BT | PASS |
| c07 | 1 | 192 | NWT | 91 | 13.1 | c05 | 4x64 | 90.4 | 13.3 | 0.4 | NWT | PASS |
| 2 | 192 | BT | 91 | 13.1 | c04 | 4x48 | 89.1 | 13.3 | 1.29 | NWT | FAIL |
| d | c06 | 1 | 160 | NWT | 88.4 | 8.3 | c04 | 4x48 | 85.7 | 11.2 | 2.54 | BT | EXCEED |
| 2 | 160 | BT | 88.4 | 8.3 | c03 | 4x32 | 70 | 21.6 | 10.29 | BT | PASS |
| c07 | 1 | 192 | NWT | 90.3 | 8.9 | c05 | 4x64 | 87.8 | 10.1 | 2.4 | BT | EXCEED |
| 2 | 192 | BT | 90.3 | 8.9 | c04 | 4x48 | 85.7 | 11.2 | 4.16 | BT | PASS |
| d+b | c06 | 1 | 160 | NWT | 89.2 | 11 | c04 | 4x48 | 87.4 | 12.4 | 2.01 | BT | EXCEED |
| 2 | 160 | BT | 89.2 | 11 | c03 | 4x32 | 76.1 | 21.8 | 9.85 | BT | PASS |
| c07 | 1 | 192 | NWT | 90.6 | 11.2 | c05 | 4x64 | 89.1 | 11.9 | 1.72 | BT | EXCEED |
| 2 | 192 | BT | 90.6 | 11.2 | c04 | 4x48 | 87.4 | 12.4 | 3.54 | BT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

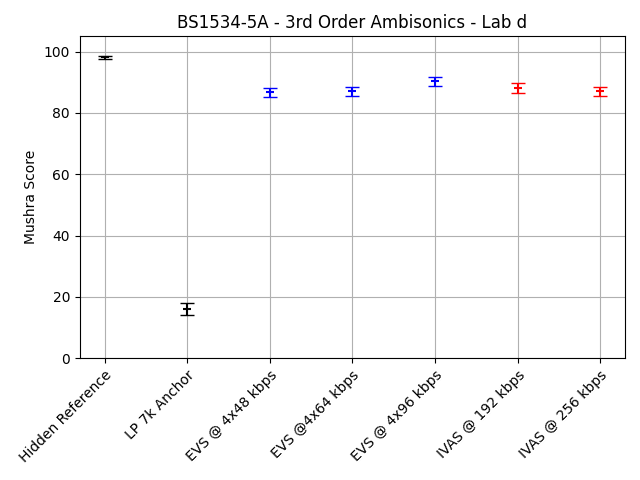
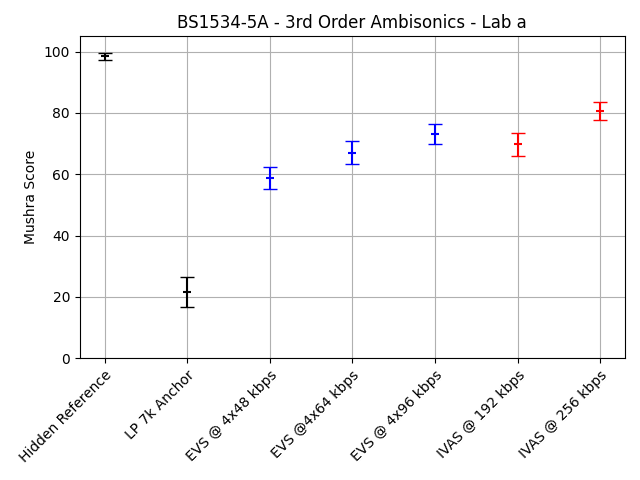
Table ‎9.3‑8: Summary of the results of BS1534-4b

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| b | c06 | 160 | NWT c04 OR BT c03 | PASS |
| c07 | 192 | NWT c05 OR BT c04 | PASS |
| d | c06 | 160 | NWT c04 OR BT c03 | EXCEED |
| c07 | 192 | NWT c05 OR BT c04 | EXCEED |
| d+b | c06 | 160 | NWT c04 OR BT c03 | EXCEED |
| c07 | 192 | NWT c05 OR BT c04 | EXCEED |

### 9.3.6 Selection Experiment BS1534-5a (HOA3, Generic Audio, 192 and 256 kbps, Headphone Presentation)

Selection Experiment BS1534-5a evaluates IVAS for Ambisonics generic audio, HOA3 input and HOA3 output at 192 and 256 kbps using headphone presentation. See Annex C.18 for details.

The averaged results per condition for experiment BS1534-5a are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS conditions with increasing bitrate (c03 – c05) and IVAS conditions with increasing bitrate (c06 – c07).

Ein Bild, das Text, Reihe, Diagramm, Zahl enthält.

Automatisch generierte Beschreibung

Figure ‎9.3‑5: BS1534-5a (Ambisonics, HOA3 input/output generic audio, 192 and 256 kbps, headphone presentation) MUSHRA plots for labs a and d, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-5a is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.3‑9: Statistical overview on the results of BS1534-5a

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| a | c06 | 1 | 192 | NWT | 69.8 | 24.4 | c04 | 4x64 | 67.1 | 24.5 | 1.03 | NWT | PASS |
| 2 | 192 | BT | 69.8 | 24.4 | c03 | 4x48 | 58.8 | 23.6 | 4.22 | BT | PASS |
| c07 | 1 | 256 | NWT | 80.6 | 20.1 | c05 | 4x96 | 73 | 21.6 | 3.34 | BT | EXCEED |
| 2 | 256 | BT | 80.6 | 20.1 | c04 | 4x64 | 67.1 | 24.5 | 5.54 | BT | PASS |
| d | c06 | 1 | 192 | NWT | 88.1 | 10.4 | c04 | 4x64 | 87.1 | 10 | 0.86 | NWT | PASS |
| 2 | 192 | BT | 88.1 | 10.4 | c03 | 4x48 | 86.7 | 10.1 | 1.22 | NWT | FAIL |
| c07 | 1 | 256 | NWT | 87 | 9.7 | c05 | 4x96 | 90.4 | 10 | -3.13 | WT | FAIL |
| 2 | 256 | BT | 87 | 9.7 | c04 | 4x64 | 87.1 | 10 | -0.08 | NWT | FAIL |
| d+a | c06 | 1 | 192 | NWT | 78.9 | 20.8 | c04 | 4x64 | 77.1 | 21.2 | 1.14 | NWT | PASS |
| 2 | 192 | BT | 78.9 | 20.8 | c03 | 4x48 | 72.7 | 22.9 | 3.67 | BT | PASS |
| c07 | 1 | 256 | NWT | 83.8 | 16.1 | c05 | 4x96 | 81.7 | 18.9 | 1.57 | NWT | PASS |
| 2 | 256 | BT | 83.8 | 16.1 | c04 | 4x64 | 77.1 | 21.2 | 4.64 | BT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.3‑10: Summary of the results of BS1534-5a

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| a | c06 | 192 | NWT c04 OR BT c03 | PASS |
| c07 | 256 | NWT c05 OR BT c04 | EXCEED |
| d | c06 | 192 | NWT c04 OR BT c03 | PASS |
| c07 | 256 | NWT c05 OR BT c04 | FAIL |
| d+a | c06 | 192 | NWT c04 OR BT c03 | PASS |
| c07 | 256 | NWT c05 OR BT c04 | PASS |

### 9.3.7 Selection Experiment BS1534-5b (HOA3, Generic Audio, 384 and 512 kbps, 7.1+4 Loudspeaker Presentation)

Selection Experiment BS1534-5b evaluates IVAS for Ambisonics generic audio, HOA3 input and HOA3 output at 384 and 512 kbps using 7.1+4 loudspeaker presentation. See Annex C.19 for details.

The averaged results per condition for experiment BS1534-5b are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS conditions with increasing bitrate (c03 – c04) and IVAS conditions with increasing bitrate (c05 – c06).

Ein Bild, das Text, Reihe, Screenshot, Diagramm enthält.

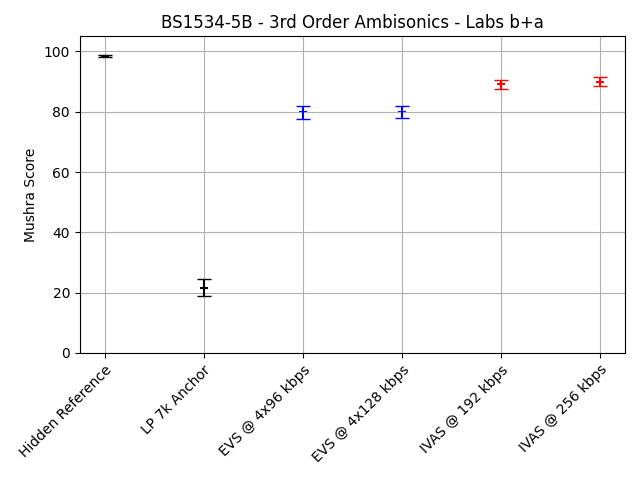
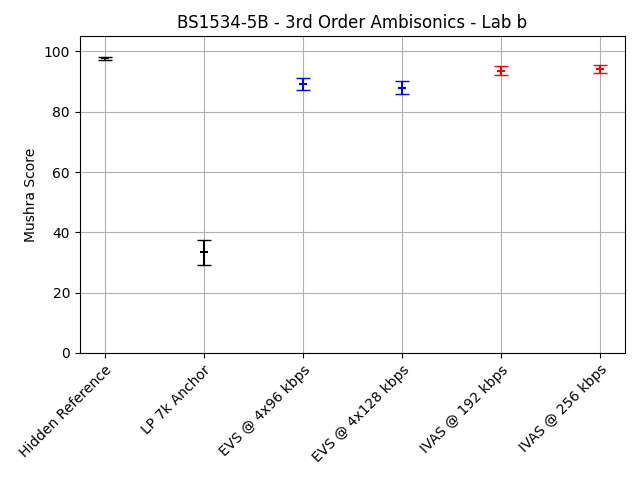
Automatisch generierte Beschreibung

Figure ‎9.3‑6: BS1534-5b (Ambisonics, HOA3 input/output, generic audio, 384 and 512 kbps, 7.1+4 loudspeaker presentation) MUSHRA plots for labs a and b, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-5b is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.3‑11: Statistical overview on the results of BS1534-5b

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| a | c05 | 1 | 384 | NWT | 84.5 | 16.8 | c04 | 4x128 | 71.6 | 18.1 | 6.78 | BT | EXCEED |
| 2 | 384 | BT | 84.5 | 16.8 | c03 | 4x96 | 70.4 | 19.3 | 7.15 | BT | PASS |
| c06 | 1 | 512 | NWT | 85.8 | 16.1 | c04 | 4x128 | 71.6 | 18.1 | 7.62 | BT | EXCEED |
| b | c05 | 1 | 384 | NWT | 93.6 | 9.5 | c04 | 4x128 | 88 | 14.4 | 4.24 | BT | EXCEED |
| 2 | 384 | BT | 93.6 | 9.5 | c03 | 4x96 | 89.1 | 13.3 | 3.55 | BT | PASS |
| c06 | 1 | 512 | NWT | 94.2 | 8.5 | c04 | 4x128 | 88 | 14.4 | 4.8 | BT | EXCEED |
| b+a | c05 | 1 | 384 | NWT | 89.1 | 14.4 | c04 | 4x128 | 79.8 | 18.3 | 7.32 | BT | EXCEED |
| 2 | 384 | BT | 89.1 | 14.4 | c03 | 4x96 | 79.8 | 19 | 7.15 | BT | PASS |
| c06 | 1 | 512 | NWT | 90 | 13.5 | c04 | 4x128 | 79.8 | 18.3 | 8.23 | BT | EXCEED |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.3‑12: Summary of the results of BS1534-5b

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| a | c05 | 384 | NWT c04 OR BT c03 | EXCEED |
| c06 | 512 | NWT c04 | EXCEED |
| b | c05 | 384 | NWT c04 OR BT c03 | EXCEED |
| c06 | 512 | NWT c04 | EXCEED |
| b+a | c05 | 384 | NWT c04 OR BT c03 | EXCEED |
| c06 | 512 | NWT c04 | EXCEED |

## 9.4 Objects (Independent Streams with Metadata, ISM)

### 9.4.1 Overview

In Selection phase, four experiments have been conducted to evaluate the performance of the IVAS codec with object content. While the experiments P800-6 and P800-7 were conducted as P.800 DCR tests on 1 or 2 object content, the experiments BS1534-6a and BS1534-6b were conducted as BS.1534 test on 3 or 4 object content. All experiments were conducted using headphone presentation.

- Selection Experiment P800-6: Clean speech, 1 object under clean and impaired channel conditions, DTX off and on, headphone presentation

- Selection Experiment P800-7: Clean speech, 2 objects under clean and impaired channel conditions, DTX off and on, headphone presentation

- Selection Experiment BS1534-6a: Generic audio, 3 objects, 48, 64 and 96 kbps, headphone presentation

- Selection Experiment BS1534-6b: Generic audio, 4 objects, 96, 128 and 256 kbps, headphone presentation

### 9.4.2 Selection Experiment P800-6 (1 Object, Clean speech, Headphone Presentation)

Selection Experiment P800-6 evaluates IVAS for 1 object clean speech, under clean and impaired channel conditions, DTX off and on using headphone presentation. See Annex C.6 for details.

The complete statistical evaluation of the requirement ToR tests for experiment P800-6 is given in the following table. The evaluation is done separately for the data from the two listening laboratories.

Table ‎9.4‑1: Statistical overview on the results of P800-6

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | | | EVS Reference | | | Evaluation | | |
| Value | Bitrate | DTX | FER | Req. | MOS | Std. | Cond. | MOS | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | | |
| a | c24 | 1 | 13.2 | Off |  | NWT | 4.49 | 0.77 | c11 | 4.46 | 0.75 | 0.48 | NWT | PASS |
| c25 | 1 | 16.4 | Off |  | NWT | 4.44 | 0.73 | c12 | 4.53 | 0.72 | -1.1 | NWT | PASS |
| c26 | 1 | 24.4 | Off |  | NWT | 4.58 | 0.67 | c13 | 4.61 | 0.6 | -0.49 | NWT | PASS |
| c27 | 1 | 32 | Off |  | NWT | 4.62 | 0.68 | c14 | 4.59 | 0.61 | 0.48 | NWT | PASS |
| c28 | 1 | 48 | Off |  | NWT | 4.58 | 0.65 | c15 | 4.63 | 0.62 | -0.67 | NWT | PASS |
| c29 | 1 | 64 | Off |  | NWT | 4.57 | 0.6 | c16 | 4.7 | 0.56 | -2.18 | WT | FAIL |
| c30 | 1 | 13.2 | Off | 5% | NWT | 3.67 | 1.14 | c17 | 3.56 | 1.15 | 0.96 | NWT | PASS |
| c31 | 1 | 16.4 | Off | 5% | NWT | 3.49 | 1.13 | c18 | 3.74 | 1.07 | -2.15 | WT | FAIL |
| c32 | 1 | 24.4 | Off | 5% | NWT | 3.82 | 1.07 | c19 | 3.77 | 1.06 | 0.49 | NWT | PASS |
| c33 | 1 | 32 | Off | 5% | NWT | 3.8 | 1.08 | c20 | 3.85 | 1.04 | -0.45 | NWT | PASS |
| c34 | 1 | 13.2 | On |  | NWT | 4.4 | 0.86 | c21 | 4.44 | 0.73 | -0.52 | NWT | PASS |
| c35 | 1 | 16.4 | On |  | NWT | 4.45 | 0.74 | c22 | 4.51 | 0.69 | -0.81 | NWT | PASS |
| c36 | 1 | 24.4 | On |  | NWT | 4.54 | 0.63 | c23 | 4.66 | 0.56 | -1.94 | WT | FAIL |
| c | c24 | 1 | 13.2 | Off |  | NWT | 4.23 | 0.71 | c11 | 4.14 | 0.76 | 1.09 | NWT | PASS |
| c25 | 1 | 16.4 | Off |  | NWT | 4.34 | 0.69 | c12 | 4.44 | 0.62 | -1.45 | NWT | PASS |
| c26 | 1 | 24.4 | Off |  | NWT | 4.49 | 0.66 | c13 | 4.52 | 0.6 | -0.42 | NWT | PASS |
| c27 | 1 | 32 | Off |  | NWT | 4.52 | 0.59 | c14 | 4.51 | 0.55 | 0.1 | NWT | PASS |
| c28 | 1 | 48 | Off |  | NWT | 4.48 | 0.64 | c15 | 4.59 | 0.6 | -1.69 | WT | FAIL |
| c29 | 1 | 64 | Off |  | NWT | 4.57 | 0.59 | c16 | 4.58 | 0.61 | -0.17 | NWT | PASS |
| c30 | 1 | 13.2 | Off | 5% | NWT | 3.4 | 0.82 | c17 | 3.43 | 0.89 | -0.37 | NWT | PASS |
| c31 | 1 | 16.4 | Off | 5% | NWT | 3.44 | 0.89 | c18 | 3.48 | 0.96 | -0.35 | NWT | PASS |
| c32 | 1 | 24.4 | Off | 5% | NWT | 3.58 | 0.89 | c19 | 3.68 | 0.89 | -1.07 | NWT | PASS |
| c33 | 1 | 32 | Off | 5% | NWT | 3.65 | 0.88 | c20 | 3.55 | 0.88 | 1.08 | NWT | PASS |
| c34 | 1 | 13.2 | On |  | NWT | 4.26 | 0.69 | c21 | 4.29 | 0.68 | -0.52 | NWT | PASS |
| c35 | 1 | 16.4 | On |  | NWT | 4.46 | 0.63 | c22 | 4.31 | 0.7 | 2.21 | BT | EXCEED |
| c36 | 1 | 24.4 | On |  | NWT | 4.56 | 0.58 | c23 | 4.6 | 0.61 | -0.7 | NWT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.4‑2: Summary of the results of P800-6

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | DTX | FER | ToR | Status |
| a | c24 | 13.2 | Off |  | NWT C11 | PASS |
| c25 | 16.4 | Off |  | NWT C12 | PASS |
| c26 | 24.4 | Off |  | NWT C13 | PASS |
| c27 | 32 | Off |  | NWT C14 | PASS |
| c28 | 48 | Off |  | NWT C15 | PASS |
| c29 | 64 | Off |  | NWT C16 | FAIL |
| c30 | 13.2 | Off | 5% | NWT C17 | PASS |
| c31 | 16.4 | Off | 5% | NWT C18 | FAIL |
| c32 | 24.4 | Off | 5% | NWT C19 | PASS |
| c33 | 32 | Off | 5% | NWT C20 | PASS |
| c34 | 13.2 | On |  | NWT C21 | PASS |
| c35 | 16.4 | On |  | NWT C22 | PASS |
| c36 | 24.4 | On |  | NWT C23 | FAIL |
| c | c24 | 13.2 | Off |  | NWT C11 | PASS |
| c25 | 16.4 | Off |  | NWT C12 | PASS |
| c26 | 24.4 | Off |  | NWT C13 | PASS |
| c27 | 32 | Off |  | NWT C14 | PASS |
| c28 | 48 | Off |  | NWT C15 | FAIL |
| c29 | 64 | Off |  | NWT C16 | PASS |
| c30 | 13.2 | Off | 5% | NWT C17 | PASS |
| c31 | 16.4 | Off | 5% | NWT C18 | PASS |
| c32 | 24.4 | Off | 5% | NWT C19 | PASS |
| c33 | 32 | Off | 5% | NWT C20 | PASS |
| c34 | 13.2 | On |  | NWT C21 | PASS |
| c35 | 16.4 | On |  | NWT C22 | EXCEED |
| c36 | 24.4 | On |  | NWT C23 | PASS |

The following diagrams show the results for a range of conditions from experiment P800-6 as rate-distortion curves. The first two diagrams only show results for clean channel conditions, i.e. conditions c11 – c16 for EVS conditions and c24 – c29 for IVAS conditions. The second two diagrams show results for conditions with 5% simulated frame loss, i.e. conditions c17 – c21 for EVS conditions and c30 – c34 for IVAS conditions.

Figure ‎9.4‑1: P800-6 (1 object, clean speech, headphone presentation) rate distortion curves for clean and impaired channel conditions

### 9.4.3 Selection Experiment P800-7 (2 Objects, Clean speech, Headphone Presentation)

Selection Experiment P800-7 evaluates IVAS for 2 objects, clean speech, under clean and impaired channel conditions, DTX off and on using headphone presentation. See Annex C.7 for details.

The complete statistical evaluation of the requirement ToR tests for experiment P800-7 is given in the following table. The evaluation is done separately for the data from the two listening laboratories.

Table ‎9.4‑3: Statistical overview on the results of P800-7

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | DTX | FER | Req. | MOS | Std. | Cond. | Bitrate | MOS | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | | | |
| a | c24 | 1 | 16.4 | Off |  | NWT | 3.08 | 1.14 | c11 | 2 x 8.0 | 2.63 | 1.06 | 3.87 | BT | EXCEED |
| c25 | 1 | 24.4 | Off |  | NWT | 4.42 | 0.68 | c12 | 2 x 13.2 | 4.41 | 0.73 | 0.15 | NWT | PASS |
| c26 | 1 | 32 | Off |  | NWT | 4.59 | 0.59 | c13 | 2 x 16.4 | 4.51 | 0.72 | 1.2 | NWT | PASS |
| c27 | 1 | 48 | Off |  | NWT | 4.57 | 0.58 | c14 | 2 x 24.4 | 4.58 | 0.66 | -0.17 | NWT | PASS |
| c28 | 1 | 64 | Off |  | NWT | 4.55 | 0.66 | c15 | 2 x 32 | 4.5 | 0.74 | 0.67 | NWT | PASS |
| c29 | 1 | 16.4 | Off | 5% | NWT | 2.79 | 1.16 | c16 | 2 x 8.0 | 2.46 | 1.09 | 2.76 | BT | EXCEED |
| c30 | 1 | 24.4 | Off | 5% | NWT | 3.92 | 1.04 | c17 | 2 x 13.2 | 3.87 | 0.91 | 0.54 | NWT | PASS |
| c31 | 1 | 32 | Off | 5% | NWT | 4.05 | 0.84 | c18 | 2 x 16.4 | 3.94 | 0.91 | 1.2 | NWT | PASS |
| c32 | 1 | 48 | Off | 5% | NWT | 4.07 | 0.83 | c19 | 2 x 24.4 | 4.07 | 0.89 | 0.06 | NWT | PASS |
| c33 | 1 | 16.4 | On |  | NWT | 3.03 | 1.13 | c20 | 2 x 8.0 | 2.76 | 1.12 | 2.3 | BT | EXCEED |
| c34 | 1 | 24.4 | On |  | NWT | 4.45 | 0.69 | c21 | 2 x 13.2 | 4.36 | 0.74 | 1.25 | NWT | PASS |
| c35 | 1 | 32 | On |  | NWT | 4.58 | 0.67 | c22 | 2 x 16.4 | 4.47 | 0.7 | 1.47 | NWT | PASS |
| c36 | 1 | 48 | On |  | NWT | 4.54 | 0.65 | c23 | 2 x 24.4 | 4.54 | 0.6 | 0.08 | NWT | PASS |
| d | c24 | 1 | 16.4 | Off |  | NWT | 3.96 | 0.89 | c11 | 2 x 8.0 | 3.19 | 0.96 | 7.79 | BT | EXCEED |
| c25 | 1 | 24.4 | Off |  | NWT | 4.66 | 0.54 | c12 | 2 x 13.2 | 4.64 | 0.55 | 0.38 | NWT | PASS |
| c26 | 1 | 32 | Off |  | NWT | 4.7 | 0.52 | c13 | 2 x 16.4 | 4.63 | 0.59 | 1.23 | NWT | PASS |
| c27 | 1 | 48 | Off |  | NWT | 4.7 | 0.51 | c14 | 2 x 24.4 | 4.76 | 0.44 | -1.12 | NWT | PASS |
| c28 | 1 | 64 | Off |  | NWT | 4.74 | 0.49 | c15 | 2 x 32 | 4.71 | 0.49 | 0.54 | NWT | PASS |
| c29 | 1 | 16.4 | Off | 5% | NWT | 3.47 | 0.97 | c16 | 2 x 8.0 | 2.89 | 1.1 | 5.34 | BT | EXCEED |
| c30 | 1 | 24.4 | Off | 5% | NWT | 4.26 | 0.81 | c17 | 2 x 13.2 | 4.23 | 0.83 | 0.27 | NWT | PASS |
| c31 | 1 | 32 | Off | 5% | NWT | 4.32 | 0.75 | c18 | 2 x 16.4 | 4.33 | 0.79 | -0.14 | NWT | PASS |
| c32 | 1 | 48 | Off | 5% | NWT | 4.32 | 0.78 | c19 | 2 x 24.4 | 4.43 | 0.71 | -1.41 | NWT | PASS |
| c33 | 1 | 16.4 | On |  | NWT | 3.86 | 0.91 | c20 | 2 x 8.0 | 3.32 | 0.91 | 5.56 | BT | EXCEED |
| c34 | 1 | 24.4 | On |  | NWT | 4.63 | 0.6 | c21 | 2 x 13.2 | 4.64 | 0.53 | -0.1 | NWT | PASS |
| c35 | 1 | 32 | On |  | NWT | 4.64 | 0.57 | c22 | 2 x 16.4 | 4.62 | 0.57 | 0.36 | NWT | PASS |
| c36 | 1 | 48 | On |  | NWT | 4.66 | 0.54 | c23 | 2 x 24.4 | 4.73 | 0.47 | -1.35 | NWT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.4‑4: Summary of the results of P800-7

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | DTX | FER | ToR | Status |
| a | c24 | 16.4 | Off |  | NWT C11 | EXCEED |
| c25 | 24.4 | Off |  | NWT C12 | PASS |
| c26 | 32 | Off |  | NWT C13 | PASS |
| c27 | 48 | Off |  | NWT C14 | PASS |
| c28 | 64 | Off |  | NWT C15 | PASS |
| c29 | 16.4 | Off | 5% | NWT C16 | EXCEED |
| c30 | 24.4 | Off | 5% | NWT C17 | PASS |
| c31 | 32 | Off | 5% | NWT C18 | PASS |
| c32 | 48 | Off | 5% | NWT C19 | PASS |
| c33 | 16.4 | On |  | NWT C20 | EXCEED |
| c34 | 24.4 | On |  | NWT C21 | PASS |
| c35 | 32 | On |  | NWT C22 | PASS |
| c36 | 48 | On |  | NWT C23 | PASS |
| d | c24 | 16.4 | Off |  | NWT C11 | EXCEED |
| c25 | 24.4 | Off |  | NWT C12 | PASS |
| c26 | 32 | Off |  | NWT C13 | PASS |
| c27 | 48 | Off |  | NWT C14 | PASS |
| c28 | 64 | Off |  | NWT C15 | PASS |
| c29 | 16.4 | Off | 5% | NWT C16 | EXCEED |
| c30 | 24.4 | Off | 5% | NWT C17 | PASS |
| c31 | 32 | Off | 5% | NWT C18 | PASS |
| c32 | 48 | Off | 5% | NWT C19 | PASS |
| c33 | 16.4 | On |  | NWT C20 | EXCEED |
| c34 | 24.4 | On |  | NWT C21 | PASS |
| c35 | 32 | On |  | NWT C22 | PASS |
| c36 | 48 | On |  | NWT C23 | PASS |

The following diagrams show the results for a range of conditions from experiment P800-7 as rate-distortion curves. The first two diagrams only show results for clean channel conditions, i.e. conditions c11 – c15 for EVS conditions and c24 – c29 for IVAS conditions. The second two diagrams show results for conditions with 5% simulated frame loss, i.e. conditions c16 – c19 for EVS conditions and c29 – c32 for IVAS conditions.

Figure ‎9.4‑2: P800-7 (2 objects, clean speech, headphone presentation) rate distortion curves for clean and impaired channel conditions

### 9.4.4 Selection Experiment BS1534-6a (3 Objects, Generic Audio, 48, 64 and 96 kbps, Headphone Presentation)

Selection Experiment BS1534-6a evaluates IVAS for 3 objects, generic audio at 48, 64 and 96 kbps using headphone presentation. See Annex C.20 for details.

The averaged results per condition for experiment BS1534-6a are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS conditions with increasing bitrate (c03 – c05) and IVAS conditions with increasing bitrate (c06 – c08).

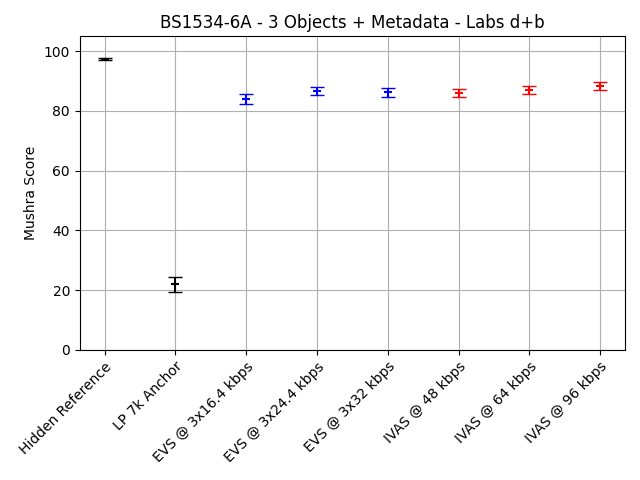
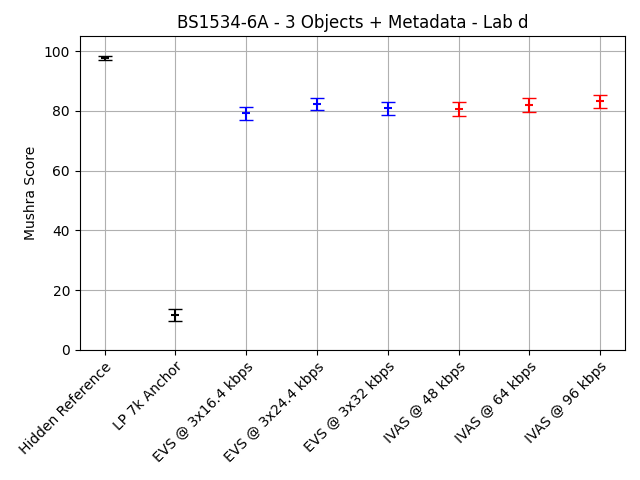
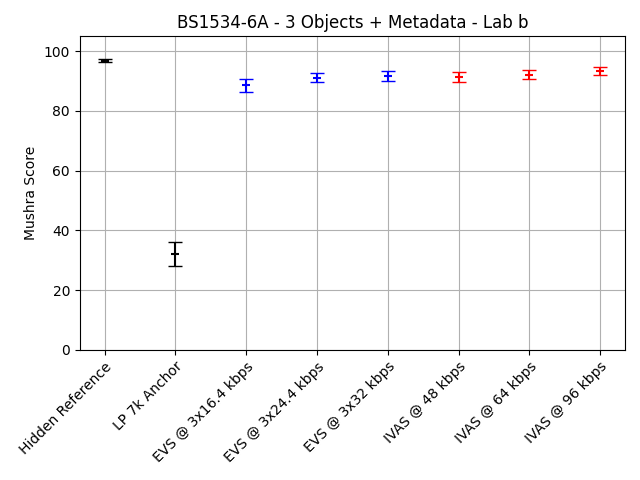


Figure ‎9.4‑3: BS1534-6a (3 objects, generic audio, 48, 64 and 96 kbps, headphone presentation) MUSHRA plots for labs b and d, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-6a is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.4‑5: Statistical overview on the results of BS1534-6a

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| b | c06 | 1 | 48 | NWT | 91.4 | 10.2 | c03 | 3x16.4 | 88.6 | 14.6 | 1.97 | BT | EXCEED |
| c07 | 1 | 64 | NWT | 92.2 | 9.2 | c04 | 3x24.4 | 91.2 | 10.8 | 0.91 | NWT | PASS |
| c08 | 1 | 96 | NWT | 93.4 | 8.4 | c05 | 3x32 | 91.7 | 11.1 | 1.59 | NWT | PASS |
| d | c06 | 1 | 48 | NWT | 80.7 | 15.2 | c03 | 3x16.4 | 79.2 | 13.9 | 0.94 | NWT | PASS |
| c07 | 1 | 64 | NWT | 82 | 15.1 | c04 | 3x24.4 | 82.4 | 13.7 | -0.26 | NWT | PASS |
| c08 | 1 | 96 | NWT | 83.2 | 14.4 | c05 | 3x32 | 80.8 | 14.3 | 1.5 | NWT | PASS |
| d+b | c06 | 1 | 48 | NWT | 86 | 14 | c03 | 3x16.4 | 83.9 | 15 | 1.88 | BT | EXCEED |
| c07 | 1 | 64 | NWT | 87.1 | 13.5 | c04 | 3x24.4 | 86.8 | 13.1 | 0.28 | NWT | PASS |
| c08 | 1 | 96 | NWT | 88.3 | 12.9 | c05 | 3x32 | 86.2 | 13.9 | 1.97 | BT | EXCEED |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.4‑6: Summary of the results of BS1534-6a

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| b | c06 | 48 | NWT c03 | EXCEED |
| c07 | 64 | NWT c04 | PASS |
| c08 | 96 | NWT c05 | PASS |
| d | c06 | 48 | NWT c03 | PASS |
| c07 | 64 | NWT c04 | PASS |
| c08 | 96 | NWT c05 | PASS |
| d+b | c06 | 48 | NWT c03 | EXCEED |
| c07 | 64 | NWT c04 | PASS |
| c08 | 96 | NWT c05 | EXCEED |

### 9.4.5 Selection Experiment BS1534-6b (4 Objects, Generic Audio, 96, 128 and 256 kbps, Headphone Presentation)

Selection Experiment BS1534-6b evaluates IVAS for 4 objects, generic audio, at 96, 128 and 256 kbps using headphone presentation. See Annex C.21. for details.

The averaged results per condition for experiment BS1534-6b are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS conditions with increasing bitrate (c03 – c05) and IVAS conditions with increasing bitrate (c06 – c08).

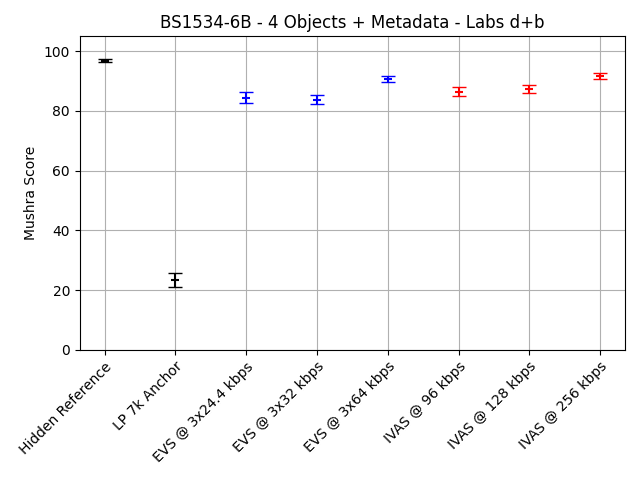
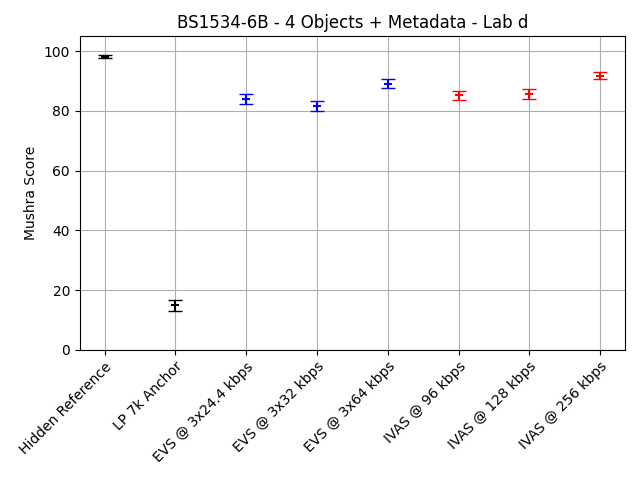
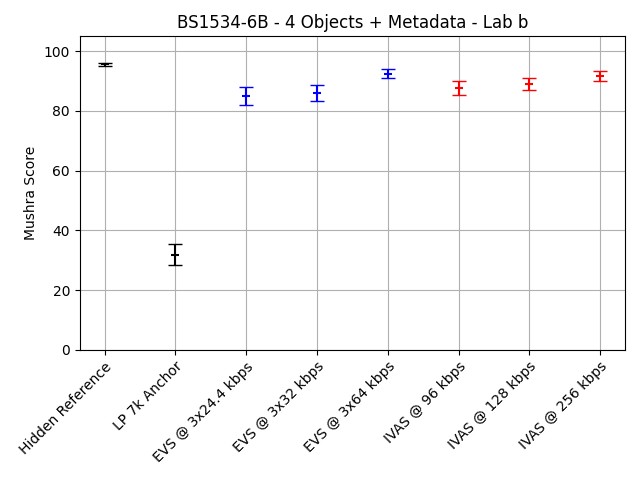


Figure ‎9.4‑4: BS1534-6b (4 objects, generic audio, 96, 128 and 192 kbps, headphone presentation) MUSHRA plots for labs b and d, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-6b is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.4‑7: Statistical overview on the results of BS1534-6b

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| b | c06 | 1 | 96 | NWT | 87.7 | 15.7 | c03 | 4x24.4 | 84.9 | 19.4 | 1.44 | NWT | PASS |
| c07 | 1 | 128 | NWT | 89 | 14 | c04 | 4x32 | 86 | 17.8 | 1.72 | BT | EXCEED |
| c08 | 1 | 256 | NWT | 91.7 | 11 | c05 | 4x64 | 92.4 | 9.9 | -0.64 | NWT | PASS |
| d | c06 | 1 | 96 | NWT | 85.2 | 10.7 | c03 | 4x24.4 | 84.1 | 11.3 | 0.97 | NWT | PASS |
| c07 | 1 | 128 | NWT | 85.7 | 10.4 | c04 | 4x32 | 81.6 | 11 | 3.51 | BT | EXCEED |
| c08 | 1 | 256 | NWT | 91.8 | 8.1 | c05 | 4x64 | 89.1 | 9.3 | 2.82 | BT | EXCEED |
| d+b | c06 | 1 | 96 | NWT | 86.5 | 13.5 | c03 | 4x24.4 | 84.5 | 15.9 | 1.73 | BT | EXCEED |
| c07 | 1 | 128 | NWT | 87.4 | 12.4 | c04 | 4x32 | 83.8 | 14.9 | 3.35 | BT | EXCEED |
| c08 | 1 | 256 | NWT | 91.7 | 9.6 | c05 | 4x64 | 90.8 | 9.7 | 1.31 | NWT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.4‑8: Summary of the results of BS1534-6b

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| b | c06 | 96 | NWT c03 | PASS |
| c07 | 128 | NWT c04 | EXCEED |
| c08 | 256 | NWT c05 | PASS |
| d | c06 | 96 | NWT c03 | PASS |
| c07 | 128 | NWT c04 | EXCEED |
| c08 | 256 | NWT c05 | EXCEED |
| d+b | c06 | 96 | NWT c03 | EXCEED |
| c07 | 128 | NWT c04 | EXCEED |
| c08 | 256 | NWT c05 | PASS |

## 9.5 Metadata-assisted Spatial Audio (MASA)

### 9.5.1 Overview

In Selection phase, four experiments have been conducted to evaluate the performance of the IVAS codec wit MASA content. While the experiments P800-8 and P800-9 were conducted as P.800 DCR tests, the experiments BS1534-7a and BS1534-7b were conducted as BS.1534 tests. All experiments were conducted using headphone presentation.

- Selection Experiment P800-8: Clean speech, MASA under clean and impaired channel conditions, headphone presentation

- Selection Experiment P800-9: Speech + background, MASA under clean channel conditions, DTX off and on, headphone presentation

- Selection Experiment BS1534-7a: Generic audio, MASA, 96 and 128 kbps, headphone presentation

- Selection Experiment BS1534-7b: Generic audio, MASA, 192 and 256 kbps, headphone presentation

### 9.5.2 Selection Experiment P800-8 (MASA, Clean speech, Headphone Presentation)

Selection Experiment P800-8 evaluates IVAS for MASA clean speech under clean and impaired channel conditions using headphone presentation. See Annex C.8 for details.

The complete statistical evaluation of the requirement ToR tests for experiment P800-8 is given in the following table. The evaluation is done separately for the data from the two listening laboratories.

Table ‎9.5‑1: Statistical overview on the results of P800-8

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | DTX | FER | Req. | MOS | Std. | Cond. | Bitrate | MOS | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | | | |
| a | c25 | 1 | 13.2 | Off |  | NWT | 3.61 | 0.91 | c09 | 3x7.2 | 2.91 | 1.06 | 6.7 | BT | EXCEED |
| 2 | 13.2 | Off |  | NWT | 3.61 | 0.91 | c15 | 2x7.2 | 2.93 | 1.24 | 5.91 | BT | EXCEED |
| c26 | 1 | 16.4 | Off |  | NWT | 3.91 | 0.83 | c10 | 4x7.2 | 2.9 | 1.09 | 9.89 | BT | EXCEED |
| 2 | 16.4 | Off |  | NWT | 3.91 | 0.83 | c15 | 2x7.2 | 2.93 | 1.24 | 8.78 | BT | EXCEED |
| c27 | 1 | 24.4 | Off |  | NWT | 4.08 | 0.78 | c11 | 4x8 | 2.96 | 1.08 | 11.28 | BT | EXCEED |
| 2 | 24.4 | Off |  | NWT | 4.08 | 0.78 | c16 | 2x8 | 3.12 | 1.23 | 8.94 | BT | EXCEED |
| c28 | 1 | 32 | Off |  | NWT | 4.28 | 0.77 | c12 | 4x9.6 | 3.62 | 1 | 6.96 | BT | EXCEED |
| 2 | 32 | Off |  | NWT | 4.28 | 0.77 | c17 | 2x9.6 | 3.67 | 1.1 | 6.09 | BT | EXCEED |
| c29 | 1 | 48 | Off |  | NWT | 4.46 | 0.67 | c13 | 4x16.4 | 4.28 | 0.76 | 2.36 | BT | EXCEED |
| 2 | 48 | Off |  | NWT | 4.46 | 0.67 | c18 | 2x16.4 | 4.45 | 0.68 | 0.08 | NWT | PASS |
| c30 | 1 | 64 | Off |  | NWT | 4.59 | 0.6 | c14 | 4x24.4 | 4.48 | 0.65 | 1.76 | BT | EXCEED |
| 2 | 64 | Off |  | NWT | 4.59 | 0.6 | c19 | 2x24.4 | 4.67 | 0.57 | -1.26 | NWT | PASS |
| c31 | 1 | 80 | Off |  | NWT | 4.68 | 0.5 | c14 | 4x24.4 | 4.48 | 0.65 | 3.27 | BT | EXCEED |
| 2 | 80 | Off |  | NWT | 4.68 | 0.5 | c19 | 2x24.4 | 4.67 | 0.57 | 0.11 | NWT | PASS |
| c32 | 1 | 13.2 | Off | 5% | NWT | 3.36 | 0.98 | c20 | 3x7.2 | 2.57 | 1.04 | 7.38 | BT | EXCEED |
| c33 | 1 | 16.4 | Off | 5% | NWT | 3.68 | 1 | c21 | 4x7.2 | 2.51 | 1.16 | 10.22 | BT | EXCEED |
| c34 | 1 | 24.4 | Off | 5% | NWT | 3.83 | 0.92 | c22 | 4x8 | 2.73 | 1.07 | 10.5 | BT | EXCEED |
| c35 | 1 | 48 | Off | 5% | NWT | 4.22 | 0.8 | c23 | 4x16.4 | 4 | 0.9 | 2.48 | BT | EXCEED |
| c36 | 1 | 64 | Off | 5% | NWT | 4.37 | 0.76 | c24 | 4x24.4 | 4.29 | 0.74 | 0.99 | NWT | PASS |
| b | c25 | 1 | 13.2 | Off |  | NWT | 3.36 | 1.03 | c09 | 3x7.2 | 2.37 | 0.98 | 9.32 | BT | EXCEED |
| 2 | 13.2 | Off |  | NWT | 3.36 | 1.03 | c15 | 2x7.2 | 2.41 | 0.94 | 9.11 | BT | EXCEED |
| c26 | 1 | 16.4 | Off |  | NWT | 3.78 | 0.92 | c10 | 4x7.2 | 2.36 | 1 | 14.09 | BT | EXCEED |
| 2 | 16.4 | Off |  | NWT | 3.78 | 0.92 | c15 | 2x7.2 | 2.41 | 0.94 | 13.97 | BT | EXCEED |
| c27 | 1 | 24.4 | Off |  | NWT | 3.98 | 0.89 | c11 | 4x8 | 2.43 | 0.91 | 16.43 | BT | EXCEED |
| 2 | 24.4 | Off |  | NWT | 3.98 | 0.89 | c16 | 2x8 | 2.57 | 1.05 | 13.84 | BT | EXCEED |
| c28 | 1 | 32 | Off |  | NWT | 4.04 | 0.92 | c12 | 4x9.6 | 3.54 | 1 | 4.94 | BT | EXCEED |
| 2 | 32 | Off |  | NWT | 4.04 | 0.92 | c17 | 2x9.6 | 3.62 | 1.16 | 3.78 | BT | EXCEED |
| c29 | 1 | 48 | Off |  | NWT | 4.46 | 0.74 | c13 | 4x16.4 | 4.13 | 0.92 | 3.73 | BT | EXCEED |
| 2 | 48 | Off |  | NWT | 4.46 | 0.74 | c18 | 2x16.4 | 4.46 | 0.68 | 0.07 | NWT | PASS |
| c30 | 1 | 64 | Off |  | NWT | 4.75 | 0.48 | c14 | 4x24.4 | 4.42 | 0.72 | 5.17 | BT | EXCEED |
| 2 | 64 | Off |  | NWT | 4.75 | 0.48 | c19 | 2x24.4 | 4.58 | 0.66 | 2.74 | BT | EXCEED |
| c31 | 1 | 80 | Off |  | NWT | 4.74 | 0.51 | c14 | 4x24.4 | 4.42 | 0.72 | 5 | BT | EXCEED |
| 2 | 80 | Off |  | NWT | 4.74 | 0.51 | c19 | 2x24.4 | 4.58 | 0.66 | 2.6 | BT | EXCEED |
| c32 | 1 | 13.2 | Off | 5% | NWT | 3.09 | 1.07 | c20 | 3x7.2 | 2.05 | 0.86 | 10.16 | BT | EXCEED |
| c33 | 1 | 16.4 | Off | 5% | NWT | 3.27 | 1.04 | c21 | 4x7.2 | 2.11 | 0.9 | 11.3 | BT | EXCEED |
| c34 | 1 | 24.4 | Off | 5% | NWT | 3.51 | 1.01 | c22 | 4x8 | 2.26 | 0.88 | 12.51 | BT | EXCEED |
| c35 | 1 | 48 | Off | 5% | NWT | 4.23 | 0.86 | c23 | 4x16.4 | 3.9 | 0.94 | 3.51 | BT | EXCEED |
| c36 | 1 | 64 | Off | 5% | NWT | 4.34 | 0.76 | c24 | 4x24.4 | 4.14 | 0.89 | 2.36 | BT | EXCEED |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.5‑2: Summary of the results of P800-8

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | DTX | FER | ToR | Status |
| a | c25 | 13.2 | Off |  | NWT c09 OR NWT c15 | EXCEED |
| c26 | 16.4 | Off |  | NWT c10 OR NWT c15 | EXCEED |
| c27 | 24.4 | Off |  | NWT c11 OR NWT c16 | EXCEED |
| c28 | 32 | Off |  | NWT c12 OR NWT c17 | EXCEED |
| c29 | 48 | Off |  | NWT c13 OR NWT c18 | PASS |
| c30 | 64 | Off |  | NWT c14 OR NWT c19 | PASS |
| c31 | 80 | Off |  | NWT c14 OR NWT c19 | PASS |
| c32 | 13.2 | Off | 5% | NWT c20 | EXCEED |
| c33 | 16.4 | Off | 5% | NWT c21 | EXCEED |
| c34 | 24.4 | Off | 5% | NWT c22 | EXCEED |
| c35 | 48 | Off | 5% | NWT c23 | EXCEED |
| c36 | 64 | Off | 5% | NWT c24 | PASS |
| b | c25 | 13.2 | Off |  | NWT c09 OR NWT c15 | EXCEED |
| c26 | 16.4 | Off |  | NWT c10 OR NWT c15 | EXCEED |
| c27 | 24.4 | Off |  | NWT c11 OR NWT c16 | EXCEED |
| c28 | 32 | Off |  | NWT c12 OR NWT c17 | EXCEED |
| c29 | 48 | Off |  | NWT c13 OR NWT c18 | PASS |
| c30 | 64 | Off |  | NWT c14 OR NWT c19 | EXCEED |
| c31 | 80 | Off |  | NWT c14 OR NWT c19 | EXCEED |
| c32 | 13.2 | Off | 5% | NWT c20 | EXCEED |
| c33 | 16.4 | Off | 5% | NWT c21 | EXCEED |
| c34 | 24.4 | Off | 5% | NWT c22 | EXCEED |
| c35 | 48 | Off | 5% | NWT c23 | EXCEED |
| c36 | 64 | Off | 5% | NWT c24 | EXCEED |

The following diagrams show the results for a range of conditions from experiment P800-8 as rate-distortion curves. The first two diagrams only show results for clean channel conditions, i.e. conditions c10 – c14 for EVS conditions and c25 – c31 for IVAS conditions. The second two diagrams show results for conditions with 5% simulated frame loss, i.e. conditions c21 – c24 for EVS conditions and c32 – c36 for IVAS conditions.

Figure ‎9.5‑1: P800-8 (MASA, clean speech, headphone presentation) rate distortion curves for clean and impaired channel conditions

### 9.5.3 Selection Experiment P800-9 (MASA, Speech+Background, Headphone Presentation)

Selection Experiment P800-9 evaluates IVAS for MASA fors peech + background conditions under clean channel conditions, DTX off and on using headphone presentation. See Annex C.9 for details.

The complete statistical evaluation of the requirement ToR tests for experiment P800-9 is given in the following table. The evaluation is done separately for the data from the two listening laboratories.

Table ‎9.5‑3: Statistical overview on the results of P800-9

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | DTX | Req. | MOS | Std. | Cond. | Bitrate | MOS | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | | |
| a | c24 | 1 | 13.2 | off | NWT | 3.46 | 1.1 | c09 | 3x7.2 | 3.44 | 1.1 | 0.15 | NWT | PASS |
| c25 | 1 | 16.4 | off | NWT | 3.69 | 0.97 | c10 | 4x7.2 | 3.47 | 1.11 | 2.07 | BT | EXCEED |
| c26 | 1 | 24.4 | off | NWT | 4.23 | 0.86 | c11 | 4x8 | 3.68 | 1.01 | 5.5 | BT | EXCEED |
| c27 | 1 | 32 | off | NWT | 4.31 | 0.79 | c12 | 4x9.6 | 4.04 | 0.9 | 3.04 | BT | EXCEED |
| c28 | 1 | 48 | off | NWT | 4.39 | 0.83 | c13 | 4x16.4 | 4.38 | 0.75 | 0.07 | NWT | PASS |
| c29 | 1 | 64 | off | NWT | 4.44 | 0.82 | c14 | 4x24.4 | 4.49 | 0.71 | -0.62 | NWT | PASS |
| c30 | 1 | 80 | off | NWT | 4.44 | 0.78 | c14 | 4x24.4 | 4.49 | 0.71 | -0.63 | NWT | PASS |
| c31 | 1 | 13.2 | on | NWT | 3.37 | 1.12 | c18 | 3x7.2 | 3.33 | 1.09 | 0.33 | NWT | PASS |
| 2 | 13.2 | on | NWT | 3.37 | 1.12 | c15 | 2x7.2 | 3.35 | 1.06 | 0.15 | NWT | PASS |
| c32 | 1 | 16.4 | on | NWT | 3.8 | 0.95 | c19 | 4x7.2 | 3.33 | 1.09 | 4.36 | BT | EXCEED |
| 2 | 16.4 | on | NWT | 3.8 | 0.95 | c15 | 2x7.2 | 3.35 | 1.06 | 4.22 | BT | EXCEED |
| c33 | 1 | 24.4 | on | NWT | 4.09 | 0.91 | c20 | 4x8 | 3.33 | 1.05 | 7.33 | BT | EXCEED |
| c34 | 1 | 32 | on | NWT | 4.04 | 0.96 | c21 | 4x9.6 | 4.06 | 0.92 | -0.17 | NWT | PASS |
| 2 | 32 | on | NWT | 4.04 | 0.96 | c16 | 2x9.6 | 4.16 | 0.91 | -1.19 | NWT | PASS |
| c35 | 1 | 48 | on | NWT | 4.29 | 0.8 | c22 | 4x16.4 | 4.31 | 0.77 | -0.21 | NWT | PASS |
| 2 | 48 | on | NWT | 4.29 | 0.8 | c17 | 2x16.4 | 4.33 | 0.78 | -0.53 | NWT | PASS |
| c36 | 1 | 64 | on | NWT | 4.4 | 0.79 | c23 | 4x24.4 | 4.34 | 0.8 | 0.67 | NWT | PASS |
| d | c24 | 1 | 13.2 | off | NWT | 3.44 | 0.89 | c09 | 3x7.2 | 2.49 | 0.8 | 10.57 | BT | EXCEED |
| c25 | 1 | 16.4 | off | NWT | 3.78 | 0.92 | c10 | 4x7.2 | 2.66 | 0.88 | 11.79 | BT | EXCEED |
| c26 | 1 | 24.4 | off | NWT | 3.71 | 0.86 | c11 | 4x8 | 2.67 | 0.87 | 11.35 | BT | EXCEED |
| c27 | 1 | 32 | off | NWT | 3.84 | 0.76 | c12 | 4x9.6 | 3.73 | 0.78 | 1.36 | NWT | PASS |
| c28 | 1 | 48 | off | NWT | 4.28 | 0.64 | c13 | 4x16.4 | 4.15 | 0.78 | 1.71 | BT | EXCEED |
| c29 | 1 | 64 | off | NWT | 4.24 | 0.71 | c14 | 4x24.4 | 4.32 | 0.74 | -1.02 | NWT | PASS |
| c30 | 1 | 80 | off | NWT | 4.29 | 0.67 | c14 | 4x24.4 | 4.32 | 0.74 | -0.38 | NWT | PASS |
| c31 | 1 | 13.2 | on | NWT | 3.25 | 0.88 | c18 | 3x7.2 | 2.52 | 0.8 | 8.24 | BT | EXCEED |
| 2 | 13.2 | on | NWT | 3.25 | 0.88 | c15 | 2x7.2 | 2.46 | 0.88 | 8.49 | BT | EXCEED |
| c32 | 1 | 16.4 | on | NWT | 3.58 | 0.86 | c19 | 4x7.2 | 2.57 | 0.86 | 11.13 | BT | EXCEED |
| 2 | 16.4 | on | NWT | 3.58 | 0.86 | c15 | 2x7.2 | 2.46 | 0.88 | 12.25 | BT | EXCEED |
| c33 | 1 | 24.4 | on | NWT | 3.65 | 0.84 | c20 | 4x8 | 2.61 | 0.87 | 11.63 | BT | EXCEED |
| c34 | 1 | 32 | on | NWT | 3.86 | 0.74 | c21 | 4x9.6 | 3.73 | 0.91 | 1.46 | NWT | PASS |
| 2 | 32 | on | NWT | 3.86 | 0.74 | c16 | 2x9.6 | 3.83 | 0.83 | 0.34 | NWT | PASS |
| c35 | 1 | 48 | on | NWT | 4.06 | 0.78 | c22 | 4x16.4 | 3.89 | 0.91 | 1.88 | BT | EXCEED |
| 2 | 48 | on | NWT | 4.06 | 0.78 | c17 | 2x16.4 | 3.96 | 0.81 | 1.14 | NWT | PASS |
| c36 | 1 | 64 | on | NWT | 4.22 | 0.73 | c23 | 4x24.4 | 4.09 | 0.79 | 1.6 | NWT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.5‑4: Summary of the results of P800-9

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | DTX | ToR | Status |
| a | c24 | 13.2 | off | NWT c09 | PASS |
| c25 | 16.4 | off | NWT c10 | EXCEED |
| c26 | 24.4 | off | NWT c11 | EXCEED |
| c27 | 32 | off | NWT c12 | EXCEED |
| c28 | 48 | off | NWT c13 | PASS |
| c29 | 64 | off | NWT c14 | PASS |
| c30 | 80 | off | NWT c14 | PASS |
| c31 | 13.2 | on | NWT c18 or NWT c15 | PASS |
| c32 | 16.4 | on | NWT c19 or NWT c15 | EXCEED |
| c33 | 24.4 | on | NWT c20 | EXCEED |
| c34 | 32 | on | NWT c21 or NWT c16 | PASS |
| c35 | 48 | on | NWT c22 or NWT c17 | PASS |
| c36 | 64 | on | NWT c23 | PASS |
| d | c24 | 13.2 | off | NWT c09 | EXCEED |
| c25 | 16.4 | off | NWT c10 | EXCEED |
| c26 | 24.4 | off | NWT c11 | EXCEED |
| c27 | 32 | off | NWT c12 | PASS |
| c28 | 48 | off | NWT c13 | EXCEED |
| c29 | 64 | off | NWT c14 | PASS |
| c30 | 80 | off | NWT c14 | PASS |
| c31 | 13.2 | on | NWT c18 or NWT c15 | EXCEED |
| c32 | 16.4 | on | NWT c19 or NWT c15 | EXCEED |
| c33 | 24.4 | on | NWT c20 | EXCEED |
| c34 | 32 | on | NWT c21 or NWT c16 | PASS |
| c35 | 48 | on | NWT c22 or NWT c17 | PASS |
| c36 | 64 | on | NWT c23 | PASS |

The following diagrams show the results for a range of conditions from experiment P800-9 as rate-distortion curves. The first two diagrams only show results for active coding conditions (DTX off), i.e. conditions c10 – c14 for EVS conditions and c24 – c30 for IVAS conditions. The second two diagrams show results for DTX conditions (DTX on), i.e. conditions c19 – c23 for EVS conditions and c31 – c36 for IVAS conditions.

Figure ‎9.5‑2: Figure 16: P800-9 (MASA, speech + background, headphone presentation) rate distortion curves for conditions with DTX off and on

### 9.5.4 Selection Experiment BS1534-7a (MASA, Generic Audio, 96 and 128 kbps, Headphone Presentation)

Selection Experiment BS1534-7a evaluates IVAS for MASA, generic audio, at 96 and 128 kbps using headphone presentation. See Annex C.22 for details.

The averaged results per condition for experiment BS1534-7a are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS multi-mono FOA conditions with increasing bitrate (c03 – c04), EVS dual-mono MASA conditions with increasing bitrate (c05 – c06) and IVAS conditions with increasing bitrate (c07 – c08).

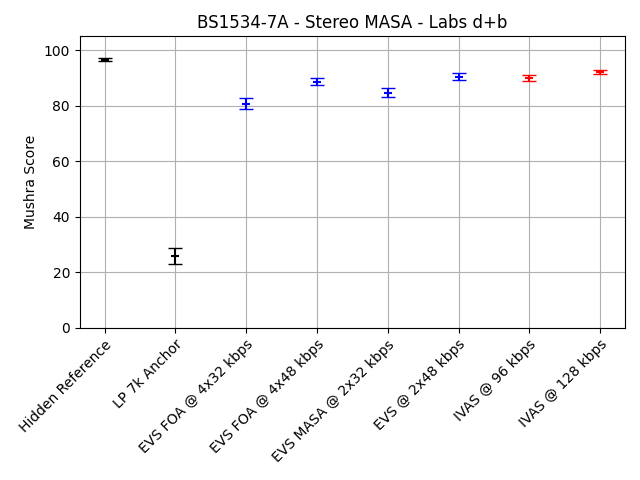
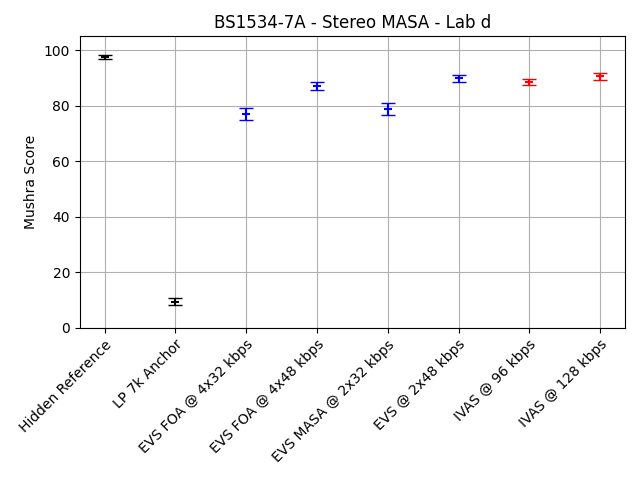
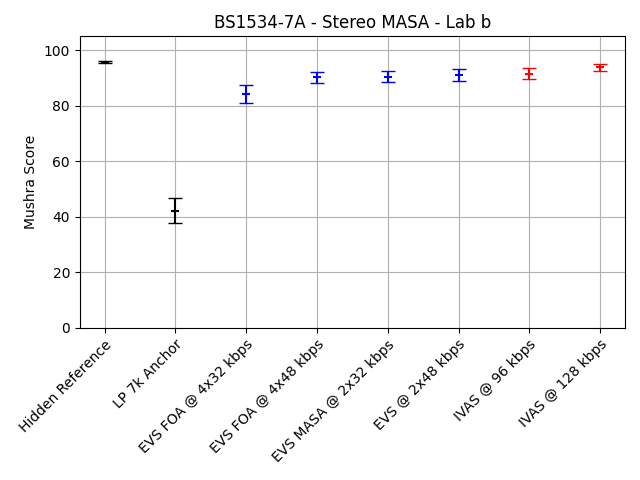


Figure ‎9.5‑3: BS1534-7a (MASA, generic audio, 96, 128 and 192 kbps, headphone presentation) MUSHRA plots for labs b and d, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-7a is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.5‑5: Statistical overview on the results of BS1534-7a

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| b | c07 | 1 | 96 | NWT | 91.6 | 11.9 | c03 | 4x32 | 84.1 | 20.9 | 4.01 | BT | EXCEED |
| 2 | 96 | NWT | 91.6 | 11.9 | c05 | 2x32, unq. MD | 90.5 | 12.6 | 0.85 | NWT | PASS |
| c08 | 1 | 128 | NWT | 93.9 | 7.7 | c04 | 4x48 | 90.2 | 13.1 | 3.14 | BT | EXCEED |
| 2 | 128 | NWT | 93.9 | 7.7 | c06 | 2x48, unq. MD | 91.2 | 14.1 | 2.14 | BT | EXCEED |
| d | c07 | 1 | 96 | NWT | 88.5 | 7.9 | c03 | 4x32 | 77.2 | 14.3 | 9.05 | BT | EXCEED |
| 2 | 96 | NWT | 88.5 | 7.9 | c05 | 2x32, unq. MD | 78.8 | 15 | 7.45 | BT | EXCEED |
| c08 | 1 | 128 | NWT | 90.6 | 7.5 | c04 | 4x48 | 87 | 8.9 | 3.91 | BT | EXCEED |
| 2 | 128 | NWT | 90.6 | 7.5 | c06 | 2x48, unq. MD | 89.8 | 8.8 | 0.82 | NWT | PASS |
| d+b | c07 | 1 | 96 | NWT | 90.1 | 10.2 | c03 | 4x32 | 80.7 | 18.2 | 8.27 | BT | EXCEED |
| 2 | 96 | NWT | 90.1 | 10.2 | c05 | 2x32, unq. MD | 84.6 | 15 | 5.49 | BT | EXCEED |
| c08 | 1 | 128 | NWT | 92.2 | 7.7 | c04 | 4x48 | 88.6 | 11.3 | 4.81 | BT | EXCEED |
| 2 | 128 | NWT | 92.2 | 7.7 | c06 | 2x48, unq. MD | 90.5 | 11.8 | 2.2 | BT | EXCEED |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.5‑6: Summary of the results of BS1534-7a

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| b | c07 | 96 | NWT c03 OR NWT c05 | PASS |
| c08 | 128 | NWT c04 OR NWT c06 | EXCEED |
| d | c07 | 96 | NWT c03 OR NWT c05 | EXCEED |
| c08 | 128 | NWT c04 OR NWT c06 | PASS |
| d+b | c07 | 96 | NWT c03 OR NWT c05 | EXCEED |
| c08 | 128 | NWT c04 OR NWT c06 | EXCEED |

### 9.5.5 Selection Experiment BS1534-7b (MASA, Generic Audio, 192 and 256 kbps, Headphone Presentation)

Selection Experiment BS1534-7b evaluates IVAS for MASA generic audio, at 192 and 256 kbps using headphone presentation. See Annex C.23 for details.

The averaged results per condition for experiment BS1534-7b are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS multi-mono FOA conditions with increasing bitrate (c03 – c04), EVS dual-mono MASA conditions with increasing bitrate (c05 – c06) and IVAS conditions with increasing bitrate (c07 – c08).

Ein Bild, das Text, Screenshot, Reihe, Diagramm enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Reihe, Diagramm, Screenshot enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Reihe, Screenshot, Diagramm enthält.

Automatisch generierte Beschreibung

Figure ‎9.5‑4: BS1534-7b (MASA, generic audio, 192 and 256 kbps, headphone presentation) MUSHRA plots for labs b and d, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-7b is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.5‑7: Statistical overview on the results of BS1534-7b

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| b | c07 | 1 | 192 | NWT | 94.5 | 6.5 | c03 | 4x64 | 93.9 | 8 | 0.71 | NWT | PASS |
| 2 | 192 | NWT | 94.5 | 6.5 | c05 | 2x64, unq. MD | 94.2 | 6.6 | 0.37 | NWT | PASS |
| c08 | 1 | 256 | NWT | 94.2 | 6.1 | c04 | 4x96 | 94.5 | 6 | -0.41 | NWT | PASS |
| 2 | 256 | NWT | 94.2 | 6.1 | c06 | 2x96, unq. MD | 94.5 | 6.3 | -0.37 | NWT | PASS |
| d | c07 | 1 | 192 | NWT | 90.6 | 8.4 | c03 | 4x64 | 87.2 | 10.5 | 3.26 | BT | EXCEED |
| 2 | 192 | NWT | 90.6 | 8.4 | c05 | 2x64, unq. MD | 89.4 | 8.8 | 1.33 | NWT | PASS |
| c08 | 1 | 256 | NWT | 90.4 | 8.6 | c04 | 4x96 | 88.8 | 9.3 | 1.64 | NWT | PASS |
| 2 | 256 | NWT | 90.4 | 8.6 | c06 | 2x96, unq. MD | 90.7 | 9.1 | -0.23 | NWT | PASS |
| d+b | c07 | 1 | 192 | NWT | 92.5 | 7.7 | c03 | 4x64 | 90.6 | 9.9 | 2.87 | BT | EXCEED |
| 2 | 192 | NWT | 92.5 | 7.7 | c05 | 2x64, unq. MD | 91.8 | 8.1 | 1.23 | NWT | PASS |
| c08 | 1 | 256 | NWT | 92.3 | 7.7 | c04 | 4x96 | 91.7 | 8.3 | 1.07 | NWT | PASS |
| 2 | 256 | NWT | 92.3 | 7.7 | c06 | 2x96, unq. MD | 92.6 | 8.1 | -0.39 | NWT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.5‑8: Summary of the results of BS1534-7b

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| b | c07 | 192 | NWT c03 OR NWT c05 | PASS |
| c08 | 256 | NWT c04 OR NWT c06 | PASS |
| d | c07 | 192 | NWT c03 OR NWT c05 | PASS |
| c08 | 256 | NWT c04 OR NWT c06 | PASS |
| d+b | c07 | 192 | NWT c03 OR NWT c05 | PASS |
| c08 | 256 | NWT c04 OR NWT c06 | PASS |

## 9.6 Multi-Channel (MC)

### 9.6.1 Overview

In Selection phase, four experiments have been conducted to evaluate the performance of the IVAS codec with multi-channel content. All experiments were conducted as BS.1534 tests. While experiments BS1534-2a and BS1534-2b were conducted using 5.1 loudspeaker presentation, experiments BS1534-3a and BS1534-3b were conducted using 7.1+4 loudspeaker presentation.

- Selection Experiment BS1534-2a: Generic audio, multi-channel 5.1, 64 and 96 kbps, loudspeaker 5.1 presentation

- Selection Experiment BS1534-2b: Generic audio, multi-channel 5.1, 128 and 160 kbps, loudspeaker 5.1 presentation

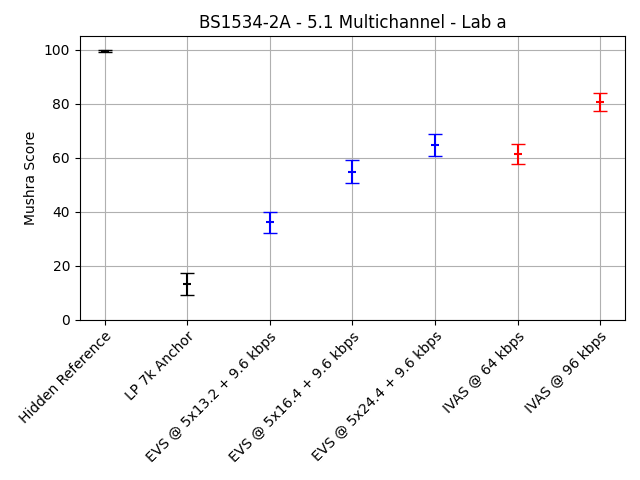
- Selection Experiment BS1534-3a: Generic audio, multi-channel 7.1.4, 128 and 160 kbps, loudspeaker 7.1+4 presentation

- Selection Experiment BS1534-3b: Generic audio, multi-channel 7.1.4, 384 and 512 kbps, loudspeaker 7.1+4 presentation

### 9.6.2 Selection Experiment BS1534-2a (MC 5.1, Generic Audio, 64 and 96 kbps, 5.1 Loudspeaker Presentation)

Selection Experiment BS1534-2a evaluates IVAS for multi-channel 5.1, generic audio at 64 and 96 kbps using loudspeaker 5.1 presentation. See Annex C.123 for details.

The averaged results per condition for experiment BS1534-2a are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS conditions with increasing bitrate (c03 – c05) and IVAS conditions with increasing bitrate (c06 – c07).

Ein Bild, das Text, Reihe, Diagramm, Zahl enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Reihe, Diagramm, Zahl enthält.

Automatisch generierte Beschreibung

Figure ‎9.6‑2: BS1534-2a (MC 5.1, generic audio, 64 and 96 kbps, 5.1 loudspeaker presentation) MUSHRA plots for labs a and b, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-2a is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.6‑1: Statistical overview on the results of BS1534-2a

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| a | c06 | 1 | 64 | NWT | 61.4 | 24.2 | c04 | 5x16.4 | 54.8 | 28.3 | 2.29 | BT | EXCEED |
| 2 | 64 | BT | 61.4 | 24.2 | c03 | 5x13.2 | 36.1 | 25.5 | 9.35 | BT | PASS |
| c07 | 1 | 96 | NWT | 80.6 | 21.9 | c05 | 5x24.4 | 64.7 | 26.7 | 5.95 | BT | EXCEED |
| 2 | 96 | BT | 80.6 | 21.9 | c04 | 5x16.4 | 54.8 | 28.3 | 9.32 | BT | PASS |
| b | c06 | 1 | 64 | NWT | 85.6 | 19 | c04 | 5x16.4 | 76.6 | 24.8 | 3.71 | BT | EXCEED |
| 2 | 64 | BT | 85.6 | 19 | c03 | 5x13.2 | 65.2 | 28.9 | 7.63 | BT | PASS |
| c07 | 1 | 96 | NWT | 88.9 | 15.5 | c05 | 5x24.4 | 86.1 | 18.6 | 1.51 | NWT | PASS |
| 2 | 96 | BT | 88.9 | 15.5 | c04 | 5x16.4 | 76.6 | 24.8 | 5.45 | BT | PASS |
| b+a | c06 | 1 | 64 | NWT | 73.5 | 24.9 | c04 | 5x16.4 | 65.7 | 28.7 | 3.75 | BT | EXCEED |
| 2 | 64 | BT | 73.5 | 24.9 | c03 | 5x13.2 | 50.6 | 30.9 | 10.57 | BT | PASS |
| c07 | 1 | 96 | NWT | 84.7 | 19.4 | c05 | 5x24.4 | 75.4 | 25.3 | 5.36 | BT | EXCEED |
| 2 | 96 | BT | 84.7 | 19.4 | c04 | 5x16.4 | 65.7 | 28.7 | 10.05 | BT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

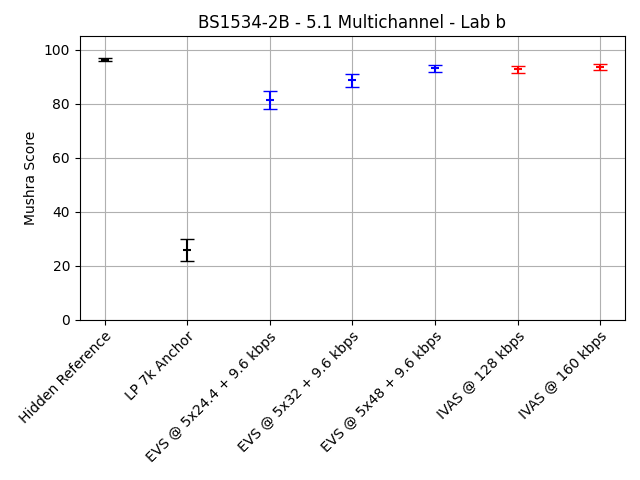
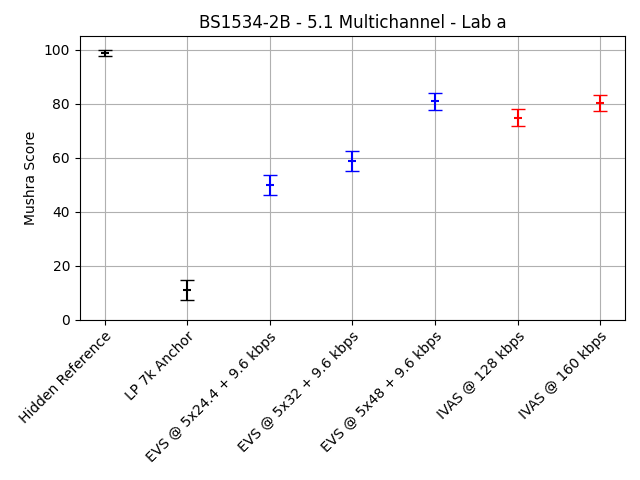
Table ‎9.6‑2: Summary of the results of BS1534-2a

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| a | c06 | 64 | NWT c04 OR BT c03 | EXCEED |
| c07 | 96 | NWT c05 OR BT c04 | EXCEED |
| b | c06 | 64 | NWT c04 OR BT c03 | EXCEED |
| c07 | 96 | NWT c05 OR BT c04 | PASS |
| b+a | c06 | 64 | NWT c04 OR BT c03 | EXCEED |
| c07 | 96 | NWT c05 OR BT c04 | EXCEED |

### 9.6.3 Selection Experiment BS1534-2b (MC 5.1, Generic Audio, 128 and 160 kbps, 5.1 Loudspeaker Presentation)

Selection Experiment BS1534-2b evaluates IVAS for multi-channel 5.1, generic audio at 128 and 160 kbps using loudspeaker 5.1 presentation. See Annex C.13 for details.

The averaged results per condition for experiment BS1534-2b are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS conditions with increasing bitrate (c03 – c05) and IVAS conditions with increasing bitrate (c06 – c07).

Ein Bild, das Text, Reihe, Diagramm, Zahl enthält.

Automatisch generierte Beschreibung

Figure ‎9.6‑3: BS1534-2b (MC 5.1, generic audio, 128 and 160 kbps, 5.1 loudspeaker presentation) MUSHRA plots for labs b and d, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-2b is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.6‑3: Statistical overview on the results of BS1534-2b

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| a | c06 | 1 | 128 | NWT | 74.9 | 21.4 | c04 | 5x32 | 58.9 | 24.6 | 6.36 | BT | EXCEED |
| 2 | 128 | BT | 74.9 | 21.4 | c03 | 5x24.4 | 49.9 | 24.1 | 10.07 | BT | PASS |
| c07 | 1 | 160 | NWT | 80.2 | 18.8 | c05 | 5x48 | 81 | 20.2 | -0.34 | NWT | PASS |
| 2 | 160 | BT | 80.2 | 18.8 | c04 | 5x32 | 58.9 | 24.6 | 8.93 | BT | PASS |
| b | c06 | 1 | 128 | NWT | 92.8 | 8.2 | c04 | 5x32 | 88.8 | 15.7 | 2.96 | BT | EXCEED |
| 2 | 128 | BT | 92.8 | 8.2 | c03 | 5x24.4 | 81.4 | 22.3 | 6.23 | BT | PASS |
| c07 | 1 | 160 | NWT | 93.6 | 7.4 | c05 | 5x48 | 93.1 | 8.6 | 0.56 | NWT | PASS |
| 2 | 160 | BT | 93.6 | 7.4 | c04 | 5x32 | 88.8 | 15.7 | 3.59 | BT | PASS |
| b+a | c06 | 1 | 128 | NWT | 83.9 | 18.5 | c04 | 5x32 | 73.8 | 25.5 | 5.84 | BT | EXCEED |
| 2 | 128 | BT | 83.9 | 18.5 | c03 | 5x24.4 | 65.6 | 28 | 9.94 | BT | PASS |
| c07 | 1 | 160 | NWT | 86.9 | 15.7 | c05 | 5x48 | 87 | 16.7 | -0.09 | NWT | PASS |
| 2 | 160 | BT | 86.9 | 15.7 | c04 | 5x32 | 73.8 | 25.5 | 8.01 | BT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

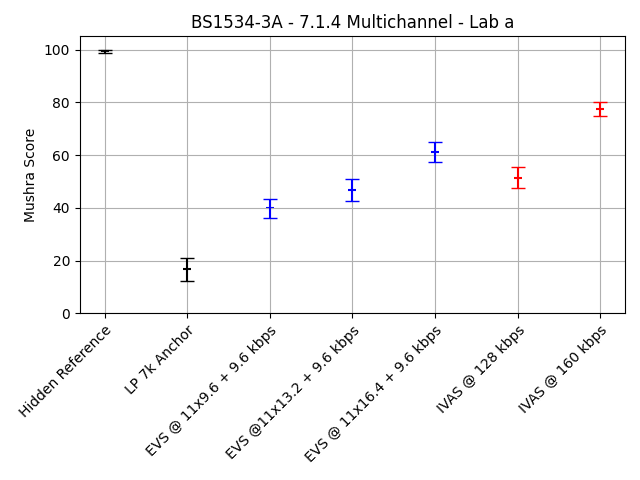
Table ‎9.6‑4: Summary of the results of BS1534-2b

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| a | c06 | 128 | NWT c04 OR BT c03 | EXCEED |
| c07 | 160 | NWT c05 OR BT c04 | PASS |
| b | c06 | 128 | NWT c04 OR BT c03 | EXCEED |
| c07 | 160 | NWT c05 OR BT c04 | PASS |
| b+a | c06 | 128 | NWT c04 OR BT c03 | EXCEED |
| c07 | 160 | NWT c05 OR BT c04 | PASS |

### 9.6.4 Selection Experiment BS1534-3a (MC 7.1.4, Generic Audio, 128 and 160 kbps, 7.1+4 Loudspeaker Presentation)

Selection Experiment BS1534-3a evaluates IVAS for multi-channel 7.1.4, generic audio at 128 and  kbps using loudspeaker 7.1+4 presentation. See Annex C.15 for details.

The averaged results per condition for experiment BS1534-3a are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS conditions with increasing bitrate (c03 – c05) and IVAS conditions with increasing bitrate (c06 – c07).

Ein Bild, das Text, Reihe, Diagramm, Zahl enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Reihe, Diagramm, Zahl enthält.

Automatisch generierte Beschreibung

Figure ‎9.6‑4: BS1534-3a (MC 7.1.4, generic audio, 128 and 160 kbps, 7.1+4 loudspeaker presentation) MUSHRA plots for labs a and b, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-3a is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.6‑5: Statistical overview on the results of BS1534-3a

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| a | c06 | 1 | 128 | NWT | 51.5 | 26.3 | c04 | 11x13.2 | 46.8 | 26.7 | 1.61 | NWT | PASS |
| 2 | 128 | BT | 51.5 | 26.3 | c03 | 11x9.6 | 39.8 | 23.9 | 4.25 | BT | PASS |
| c07 | 1 | 160 | NWT | 77.5 | 17.7 | c05 | 11x16.4 | 61.1 | 24.6 | 7.01 | BT | EXCEED |
| 2 | 160 | BT | 77.5 | 17.7 | c04 | 11x13.2 | 46.8 | 26.7 | 12.42 | BT | PASS |
| b | c06 | 1 | 128 | NWT | 85.2 | 19.8 | c04 | 11x13.2 | 72.2 | 27.3 | 4.99 | BT | EXCEED |
| 2 | 128 | BT | 85.2 | 19.8 | c03 | 11x9.6 | 64.9 | 28.2 | 7.64 | BT | PASS |
| c07 | 1 | 160 | NWT | 90.6 | 12.3 | c05 | 11x16.4 | 84.7 | 18.9 | 3.41 | BT | EXCEED |
| 2 | 160 | BT | 90.6 | 12.3 | c04 | 11x13.2 | 72.2 | 27.3 | 7.95 | BT | PASS |
| b+a | c06 | 1 | 128 | NWT | 68.4 | 28.8 | c04 | 11x13.2 | 59.5 | 29.8 | 3.91 | BT | EXCEED |
| 2 | 128 | BT | 68.4 | 28.8 | c03 | 11x9.6 | 52.4 | 29 | 7.18 | BT | PASS |
| c07 | 1 | 160 | NWT | 84.1 | 16.6 | c05 | 11x16.4 | 72.9 | 24.9 | 6.84 | BT | EXCEED |
| 2 | 160 | BT | 84.1 | 16.6 | c04 | 11x13.2 | 59.5 | 29.8 | 13.18 | BT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.6‑6: Summary of the results of BS1534-3a

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| a | c06 | 128 | NWT c04 OR BT c03 | PASS |
| c07 | 160 | NWT c05 OR BT c04 | EXCEED |
| b | c06 | 128 | NWT c04 OR BT c03 | EXCEED |
| c07 | 160 | NWT c05 OR BT c04 | EXCEED |
| b+a | c06 | 128 | NWT c04 OR BT c03 | EXCEED |
| c07 | 160 | NWT c05 OR BT c04 | EXCEED |

### 9.6.5 Selection Experiment BS1534-3b (MC 7.1.4, Generic Audio, 384 and 512 kbps, 7.1+4 Loudspeaker Presentation)

Selection Experiment BS1534-3b evaluates IVAS for multi-channel 7.1.4, generic audio at, 384 and 512 kbps, loudspeaker 7.1+4 presentation. See Annex C.15 for details.

The averaged results per condition for experiment BS1534-3b are depicted in the following figures. The three figures show the individual results for the two labs and the results for a joint evaluation, respectively. The conditions are shown grouped by Hidden Reference (c01), LP 7k anchor (c02), EVS conditions with increasing bitrate (c03 – c05) and IVAS conditions with increasing bitrate (c06 – c07).

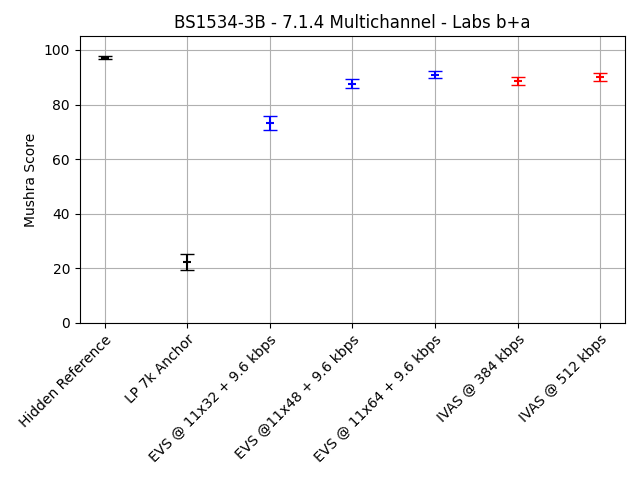
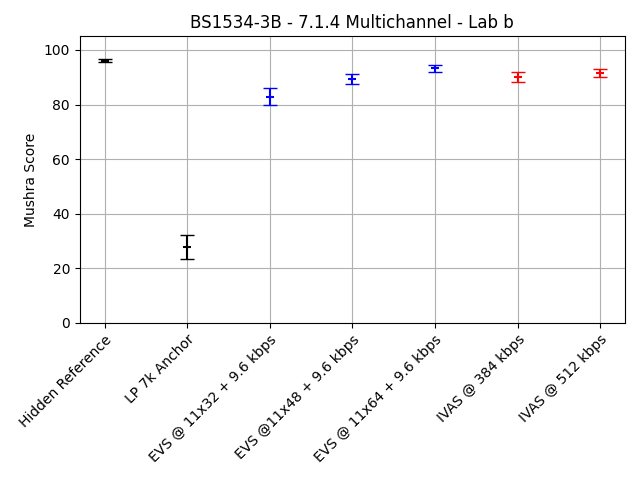
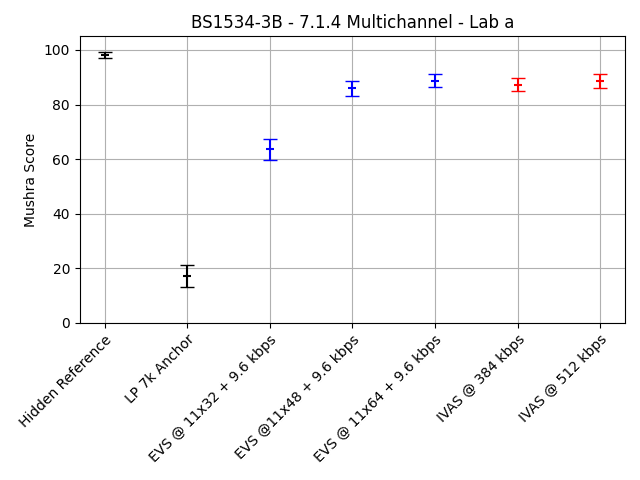


Figure ‎9.6‑5: BS1534-3b (MC 7.1.4, generic audio, 384 and 512 kbps, 7.1+4 loudspeaker presentation) MUSHRA plots for labs a and b, both labs combined

The complete statistical evaluation of the requirement ToR tests for experiment BS1534-3b is given in the following table. The evaluation is done separately for the data from the two listening laboratories and for a combination of the two data sets.

Table ‎9.6‑7: Statistical overview on the results of BS1534-3b

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Type | CuT | | | | EVS Reference | | | | Evaluation | | |
| Value | Bitrate | Req. | Score | Std. | Cond. | Bitrate | Score | Std. | T-Stat | Result | State |
| Lab | Cond. | ToR# |  | | | | | | | | | | |
| a | c06 | 1 | 384 | NWT | 87.3 | 16 | c04 | 11x48 | 85.9 | 17.6 | 0.72 | NWT | PASS |
| 2 | 384 | BT | 87.3 | 16 | c03 | 11x32 | 63.6 | 25 | 10.32 | BT | PASS |
| c07 | 1 | 512 | NWT | 88.6 | 15.8 | c05 | 11x64 | 88.7 | 16 | -0.04 | NWT | PASS |
| 2 | 512 | BT | 88.6 | 15.8 | c04 | 11x48 | 85.9 | 17.6 | 1.48 | NWT | FAIL |
| b | c06 | 1 | 384 | NWT | 90.1 | 11.5 | c04 | 11x48 | 89.3 | 12.4 | 0.6 | NWT | PASS |
| 2 | 384 | BT | 90.1 | 11.5 | c03 | 11x32 | 82.9 | 20.1 | 4.04 | BT | PASS |
| c07 | 1 | 512 | NWT | 91.5 | 10.1 | c05 | 11x64 | 93.3 | 8 | -1.76 | WT | FAIL |
| 2 | 512 | BT | 91.5 | 10.1 | c04 | 11x48 | 89.3 | 12.4 | 1.8 | BT | PASS |
| b+a | c06 | 1 | 384 | NWT | 88.7 | 14 | c04 | 11x48 | 87.6 | 15.3 | 0.94 | NWT | PASS |
| 2 | 384 | BT | 88.7 | 14 | c03 | 11x32 | 73.3 | 24.6 | 9.99 | BT | PASS |
| c07 | 1 | 512 | NWT | 90.1 | 13.3 | c05 | 11x64 | 91 | 12.8 | -0.9 | NWT | PASS |
| 2 | 512 | BT | 90.1 | 13.3 | c04 | 11x48 | 87.6 | 15.3 | 2.23 | BT | PASS |

The following table provides a summary of the results. For this summary, the requirements that are defined as a disjunction of two separate checks have been combined into an overall status for this requirement as described before.

Table ‎9.6‑8: Summary of the results of BS1534-3b

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lab | Cond. | Bitrate | ToR | Status |
| a | c06 | 384 | NWT c04 OR BT c03 | PASS |
| c07 | 512 | NWT c05 OR BT c04 | PASS |
| b | c06 | 384 | NWT c04 OR BT c03 | PASS |
| c07 | 512 | NWT c05 OR BT c04 | PASS |
| b+a | c06 | 384 | NWT c04 OR BT c03 | PASS |
| c07 | 512 | NWT c05 OR BT c04 | PASS |

## 9.7 Combined Objects and SBA (OSBA)

### 9.7.1 Overview

Provided as optional additional information for codec selection, two experiments have been conducted to evaluate the performance of the IVAS codec with the combination of combined objects and SBA content. All experiments were conducted as BS.1534 tests using headphone presentation.

- Optional additional information experiment 1: Generic audio, combined objects and SBA, 256 kbps, headphone presentation

- Optional additional information experiment 2: Generic audio, combined objects and SBA, 512 kbps, headphone presentation

### 9.7.2 Optional additional information provided for codec selection

The following quality assessments has been provided as optional additional information for codec selection:

IVAS provides a combined format mode to code ISMs and an SBA scene in a single instance of the codec. At low bitrates, the objects are pre-rendered into the SBA scene, which is then coded and reproduced as usual. At high bitrates, a high-quality mode exists, which allocates extra downmix channels for the objects and codes their metadata separately.

This high-quality mode can be shown to exhibit a strong quality benefit over both the separate coding of the two formats by separate instances of IVAS and pre-rendering. This is demonstrated by the results of a MUSHRA listening test plotted in Figure ‎9.7‑1. The high-quality mode (osba\_256kpbs) scores significantly higher than the separate-coding (ism\_128kbps\_sba\_128kbps) and the pre-rendering (prerendered\_256kbps).

Ein Bild, das Reihe, Diagramm, Text, Zahl enthält.

Automatisch generierte Beschreibung

Figure ‎9.7‑1: BS.1534 comparison between combined vs. Separate operation of ISM + SBA at 256 kbps

Further benefit can be observed at 512 kbps by the results of a separate MUSHRA listening test (6 expert listeners), shown in Figure ‎9.7‑2. The high-quality mode (osba\_512) scores significantly higher than the separate-coding (base\_2x256kbps) and the pre-rendering (prerend\_512), bringing the average quality into the excellent range.

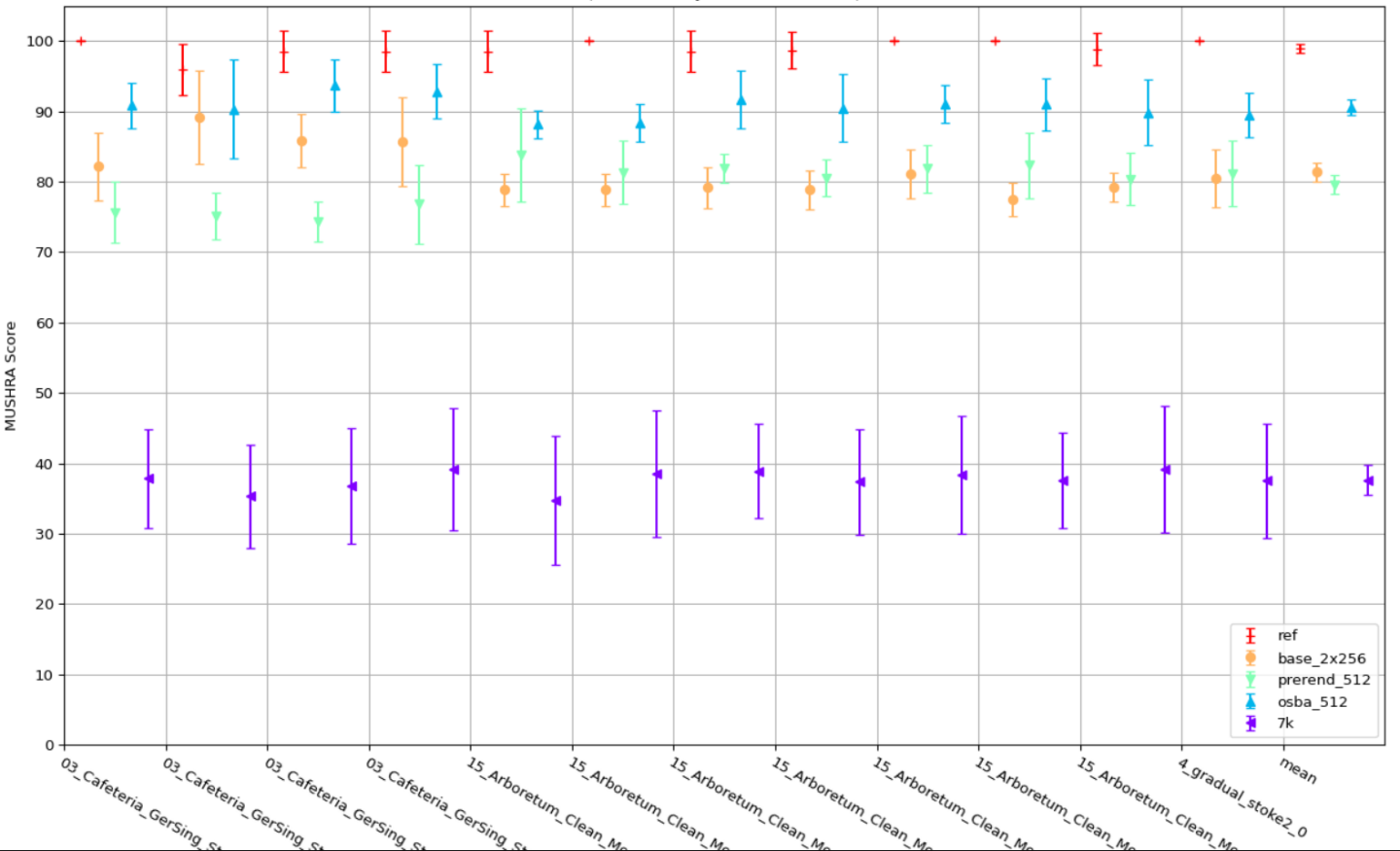


Figure ‎9.7‑2: BS.1534 comparison between combined vs. separate operation of ISM + SBA at 512 kbps

## 9.8 Combined Objects and MASA (OMASA)

### 9.8.1 Overview

Provided as optional additional information provided for codec selection, one experiments has been conducted to evaluate the performance of the IVAS codec with the combination of combined objects and MASA content. The experiment was conducted as BS.1534 test using headphone presentation.

- Optional additional information experiment 3: Generic audio, combined objects and MASA, 48, 64, 80 and 128 kbps, headphone presentation

### 9.8.2 Optional additional information provided for codec selection

The following quality assessment has been provided as optional additional information for codec selection:

Combined ISM (objects) and spatial capture with MASA provide an efficient way to, e.g., capture and transmit the environment and talker(s) at the same time. Combined coding of ISMs and MASA provides both increased coding efficiency and reduced complexity and memory usage compared to separate codec instances. Combined ISM and MASA operation is possible with 1 to 4 ISMs at all bitrates from 13.2 to 512 kbps. At the lowest bitrates ISMs and MASA are pre-renderer into MASA format and with increasing bitrate objects are parametrically presented, most important object separately coded, and finally all ISMs separately coded with the accompanying metadata in addition to the spatial MASA signal. The combined format coding automatically distributes allowed bit budget optimally between ISMs, MASA, and spatial metadata.

When comparing to separate coding of ISMs and MASA, combined coding (labeled ‘oMASA’ in Figure ‎9.8‑1) provides significant quality improvement especially at lower bitrates.

Ein Bild, das Text, Screenshot, Reihe, Zahl enthält.

Automatisch generierte Beschreibung

Figure ‎9.8‑1: BS.1534 comparison between combined vs. separate operation of ISM and MASA at various bitrates

## 9.9 Stereo operation with EVS compatible mono downmix stream

### 9.9.1 Overview

Provided as optional additional information provided for codec selection, one experiments has been conducted to evaluate the performance of the IVAS codec for downmixing stereo content to an EVS compatible mono downmix. The experiment was conducted as P.800 DCR test using headphone presentation.

- Optional additional information experiment 4: Speech and music, stereo input, mono output, 13.2 and 24.4 kbps, headphone presentation

### 9.9.2 Optional additional information provided for codec selection

The following quality assessment has been provided as optional additional information for codec selection:

For dynamic downmix tool for interoperable EVS the following in house listening test has been conducted. The results are shown in Figure ‎9.9‑1.

- P.800 DCR 48 kHz stereo input

- Direct, MNRU, SDRU

- CuT1: average (static downmix) + EVS

- CuT2: trunk (dynamic downmix) + EVS

- CuT3: tandem: mono output of IVAS stereo + EVS

- Bit rate

- 13.2 kbps, 24.4 kbps, downmix only

- Input items:

- Japanese speech with some music 60 items.

- Including non-overlap, overlap talkers

- Including 6 types of speaker and microphone position

- Listeners: 24 (naïve)

According to the results, dynamic downmix could provide better or not worse than static downmix and tandem coding. Note that tandem coding consists of EVS encoding of the mono output of IVAS stereo coding with additional delay.

Ein Bild, das Text, Screenshot, Diagramm, Reihe enthält.

Automatisch generierte Beschreibung

Figure ‎9.9‑1: MOS scores of P.800 DCR test for downmix tools and tandem coding

## 9.10 Rendering

NOTE: Testing of rendering aspects and the inclusion of the corresponding results is scheduled for the characterization phase of IVAS.

## 9.11 Split Rendering

### 9.11.1 General

An IVAS specific split rendering solution was added to IVAS as part of the Workitem “Immersive Audio for Split Rendering Scenarios” (ISAR). A detailed performance characterization of ISAR can be found in 3GPP TR 26.996 (“Immersive Audio for Split Rendering Scenarios; Performance characterization”) [28]. In order for this document to contain a full characterization of all IVAS immersive audio features, the test results of the IVAS specific solutionare replicated here. For details on test setup, please see 3GPP TR 26.996.

### 9.11.2 Overview

In the ISAR Selection phase, four experiments have been conducted to evaluate the performance of the IVAS specific ISAR solution. Table ‎9.11‑1shows a high-level overview of the experiments. Each experiment was carried out twice (in experiments a and b), once by the solution proponent and once by a cross-checker (XC).

Table ‎9.11‑1: High-level overview of ISAR selection experiments

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Exp | Input format | Source material | Listening  environment | Bitrates  (kbps) | Listening Lab |
| ISAR\_BS1534-1a ISAR\_BS1534-1b | SBA (HOA3) | Generic Audio | Headphones | IVAS: 512,  CuT: 768 | Dolby Qualcomm (XC) |
| ISAR\_BS1534-2a ISAR\_BS1534-2b | Multi-channel 7.1+4 | Generic Audio | Headphones | IVAS: 512,  CuT: 768 | Fraunhofer Ittiam (XC) |
| ISAR\_BS1534-3a ISAR\_BS1534-3b | Objects (ISM-4) | Generic Audio | Headphones | IVAS: 512,  CuT: 768 | Fraunhofer Nokia (XC) |
| ISAR\_BS1534-4a ISAR\_BS1534-4b | MASA (2 TC) | Generic Audio | Headphones | IVAS: 512,  CuT: 768 | Dolby Bytedance (XC) |

### 9.11.3 ISAR Selection Phase Requirements and Objectives

All experiments check the same requirements defined in TR 26.865 [2], namely that the QoE of the ISAR split rendering system is no worse than the 0-DOF native transcoding reference system using the same operation point of the native coding system (IVAS coding at 512 kbps) and best possible operation point for transcoding (IVAS stereo at 256 kbps). The 4 experiments evaluate the requirement for the 4 different main head-trackable IVAS coding formats, i.e., SBA (HOA3), MC 7.1.4, ISM-4 and MASA.

The objective is that QoE provided by split rendering solution should be as close as possible to quality of native coding reference system using same operation point. There is no statistical test to verify if this objective is met. However, a statement will be made based on the observed test scores how close the quality of the tested ISAR split rendering solution for the given immersive audio input format is to the quality of the native coding reference system.

Conclusion of all 8 experiments testing the requirement that the ISAR split rendering solution for IVAS shall be no worse than the 0-DOF transcoding reference system is that this requirement is met across all tested immersive input audio formats. It can generally be observed that the achievable quality is even clearly better whereby a quality level in the ‘excellent’ range close to the quality of the native IVAS coding reference system is achieved whereas the 0-DOF transcoding alternative offers substantially lower quality.

### 9.11.4 ISAR Selection Experiment ISAR\_BS1534-1 (HOA3, Generic Audio, 768 kbps Split Rendering Link, Headphone Presentation)

#### 9.11.4.1 Overview

Selection Experiment ISAR\_BS1534-1 evaluates Split Rendering for IVAS HOA3 audio at 768 kbps. See [29] Annex B.1 for details.

#### 9.11.4.2 Test conditions

The test conditions in all experiments are as follows:

- **c01 (Ref):** Native coding reference system operated at 512 kbps with IVAS decoding and head-tracked rendering at end-device

- **c02 (LP7 anchor):** 7 kHz low-pass filtered native coding reference system

- **c03 (0-DOF):** 0-DOF native transcoding reference system with 512 kbps coded IVAS content, decoded and binaurally rendered to (outdated) pose available at pre-renderer, subsequently transcoded to IVAS stereo at 256 kbps

- **c04 (CuT):** Codec under Test, i.e., ISAR split rendering system operated at the required bit rate of 768 kbps for the intermediate immersive audio representation

#### 9.11.4.3 Result plots

Provided below are the result plots and tables with statistical analysis result for the two ISAR\_BS1534-1 experiments.

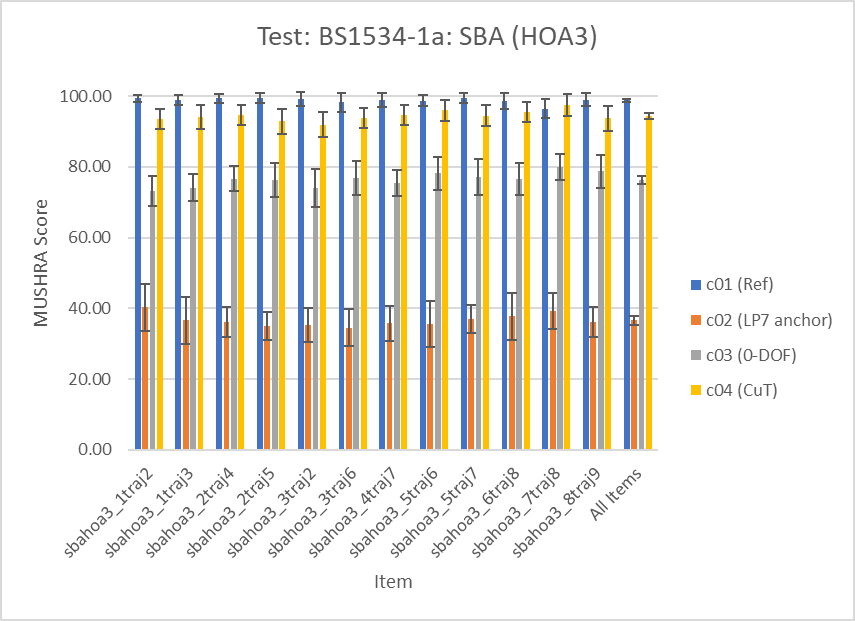


Figure ‎9.11‑1: Results of ISAR\_BS1534-1a test for SBA input audio

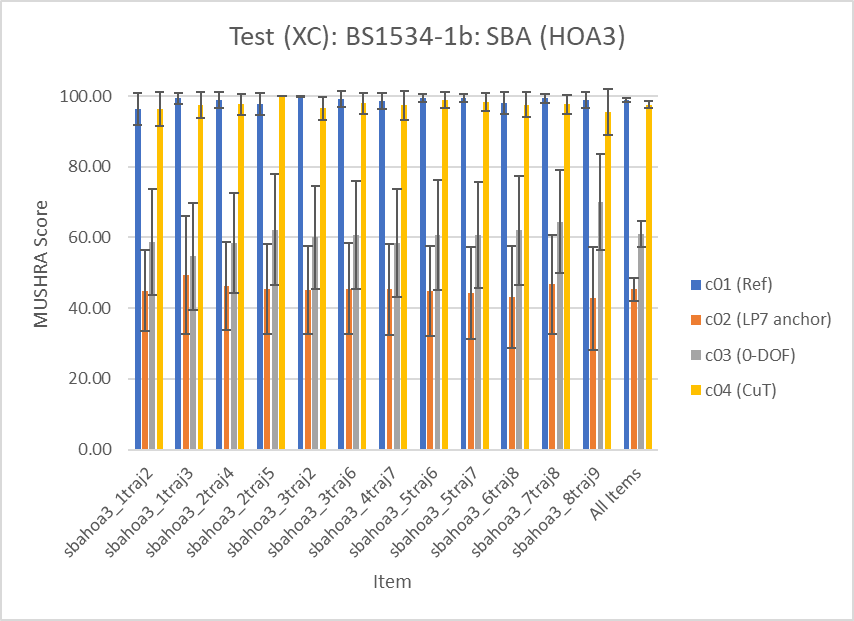


Figure ‎9.11‑2: Results of ISAR\_BS1534-1b test for SBA input audio

#### 9.11.4.4 Statistical analysis

Table ‎9.11‑2: Result of statistical analysis of ISAR\_BS1534-1a test checking CuT NWT 0-DOF Reference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean Diff. (c03 - c04) | Stdev Diff. | SEMD | T | Prob. | ToR |
| -17.9917 | 5.7941 | 0.5289 | -34.0156 | 1.0000 | Pass |

Table ‎9.11‑3: Result of statistical analysis of ISAR\_BS1534-1b test checking CuT NWT 0-DOF Reference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean Diff. (c03 - c04) | Stdev Diff. | SEMD | T | Prob. | ToR |
| -36.6583 | 20.3336 | 1.8562 | -19.7492 | 1.0000 | Pass |

#### 9.11.4.5 Experimental conclusions

Conclusion of both experiments is that the ISAR split rendering solution for SBA input meets the requirement to be no worse than the 0-DOF transcoding reference system. The experiments indicate that the achievable quality is even clearly better whereby a quality level in the ‘excellent’ range is achieved compared to the 0-DOF transcoding reference providing quality in the ‘good’ range. The objective to provide a quality level as close as possible to the native coding reference system is met in the sense that the quality score of the split rendering system is in the high ‘excellent’ range which indicates only very minor audible differences.

### 9.11.5 ISAR Selection Experiment ISAR\_BS1534-2 (MC 7.1+4, Generic Audio, 768 kbps Split Rendering Link, Headphone Presentation)

#### 9.11.5.1 Overview

Selection Experiment ISAR\_BS1534-2 evaluates Split Rendering for IVAS 7.1+4 audio at 768 kbps. See [29] Annex B.2 for details.

#### 9.11.5.2 Test conditions

The test conditions in all experiments are as follows:

- **c01 (Ref):** Native coding reference system operated at 512 kbps with IVAS decoding and head-tracked rendering at end-device

- **c02 (LP7 anchor):** 7 kHz low-pass filtered native coding reference system

- **c03 (0-DOF):** 0-DOF native transcoding reference system with 512 kbps coded IVAS content, decoded and binaurally rendered to (outdated) pose available at pre-renderer, subsequently transcoded to IVAS stereo at 256 kbps

- **c04 (CuT):** Codec under Test, i.e., ISAR split rendering system operated at the required bit rate of 768 kbps for the intermediate immersive audio representation

#### 9.11.5.3 Result plots

Provided below are the result plots and tables with statistical analysis result for the two ISAR\_BS1534-2 experiments.

Ein Bild, das Text, Screenshot, parallel, Reihe enthält.

Automatisch generierte Beschreibung

Figure ‎9.11‑3: Results of ISAR\_BS1534-2a test for MC 7.1.4 input audio

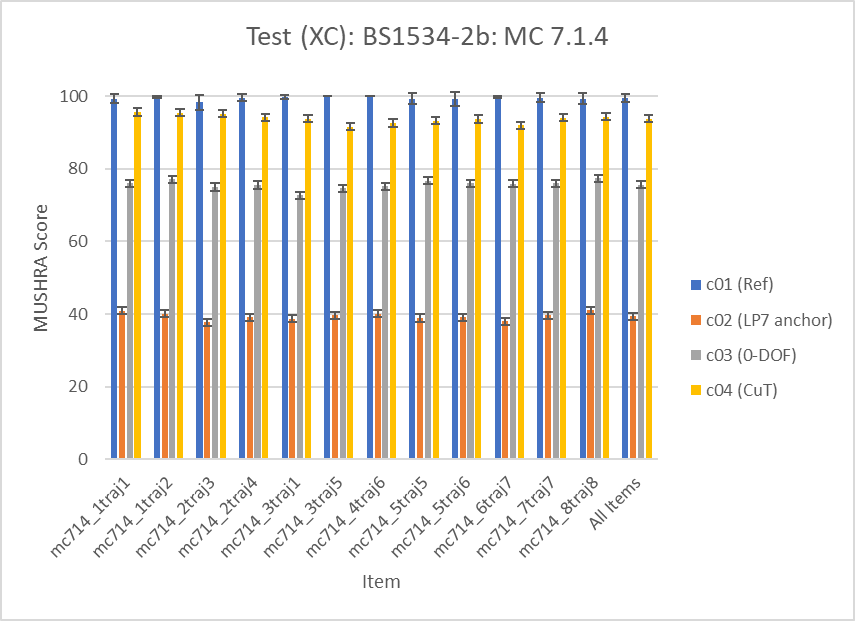


Figure ‎9.11‑4: Results of ISAR\_BS1534-2b test for MC 7.1.4 input audio

#### 9.11.5.4 Statistical analysis

Table ‎9.11‑4: Result of statistical analysis of ISAR\_BS1534-2a test checking CuT NWT 0-DOF Reference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean Diff. (c03 - c04) | Stdev Diff. | SEMD | t | Prob. | ToR |
| -31.9583 | 15.1321 | 1.3814 | -23.1353 | 1.0000 | Pass |

Table ‎9.11‑5: Result of statistical analysis of ISAR\_BS1534-2b test checking CuT NWT 0-DOF Reference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean Diff. (c03 - c04) | Stdev Diff. | SEMD | t | Prob. | ToR |
| -18.1667 | 6.0881 | 0.5558 | -32.6879 | 1.0000 | Pass |

#### 9.11.5.5 Experimental conclusions

Conclusion of both experiments is that the ISAR split rendering solution for Multi-Channel 7.1+4 input meets the requirement to be no worse than the 0-DOF transcoding reference system. The experiments indicate that the achievable quality is even clearly better whereby a quality level in the ‘excellent’ range is achieved compared to the 0-DOF transcoding reference providing quality in the ‘good’ range. The objective to provide a quality level as close as possible to the native coding reference system is met in the sense that the quality score of the split rendering system is in the high ‘excellent’ range, which indicates only very minor audible differences.

### 9.11.6 ISAR Selection Experiment ISAR\_BS1534-3 (4 Objects, Generic Audio, 768 kbps Split Rendering Link, Headphone Presentation)

#### 9.11.6.1 Overview

Selection Experiment ISAR\_BS1534-3 evaluates Split Rendering for 4 object audio (ISM4) at 768 kbps. See [29] Annex B.3 for details.

#### 9.11.6.2 Test conditions

The test conditions in all experiments are as follows:

- **c01 (Ref):** Native coding reference system operated at 512 kbps with IVAS decoding and head-tracked rendering at end-device

- **c02 (LP7 anchor):** 7 kHz low-pass filtered native coding reference system

- **c03 (0-DOF):** 0-DOF native transcoding reference system with 512 kbps coded IVAS content, decoded and binaurally rendered to (outdated) pose available at pre-renderer, subsequently transcoded to IVAS stereo at 256 kbps

- **c04 (CuT):** Codec under Test, i.e., ISAR split rendering system operated at the required bit rate of 768 kbps for the intermediate immersive audio representation

#### 9.11.6.3 Result plots

Provided below are the result plots and tables with statistical analysis result for the two ISAR\_BS1534-3 experiments.

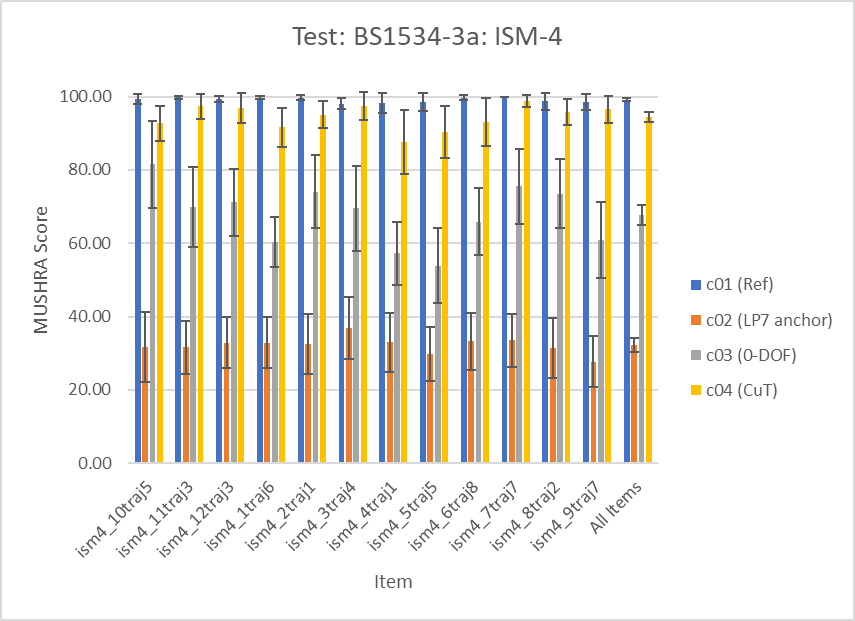


Figure ‎9.11‑5: Results of ISAR\_BS1534-3a test for ISM-4 input audio

Ein Bild, das Text, Screenshot, parallel, Reihe enthält.

Automatisch generierte Beschreibung

Figure ‎9.11‑6: Results of ISAR\_BS1534-3b test for ISM-4 input audio

#### 9.11.6.4 Statistical analysis

Table ‎9.11‑6: Result of statistical analysis of ISAR\_BS1534-3a test checking CuT NWT 0-DOF Reference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean Diff. (c03 - c04) | Stdev Diff. | SEMD | T | Prob. | ToR |
| -26.6500 | 14.7310 | 1.3448 | -19.8178 | 1.0000 | Pass |

Table ‎9.11‑7: Result of statistical analysis of ISAR\_BS1534-3b test checking CuT NWT 0-DOF Reference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean Diff. (c03 - c04) | Stdev Diff. | SEMD | t | Prob. | ToR |
| -32.4083 | 22.0761 | 2.0153 | -16.0814 | 1.0000 | Pass |

#### 9.11.6.5 Experimental conclusions

Conclusion of both experiments is that the ISAR split rendering solution for ISM-4 input meets the requirement to be no worse than the 0-DOF transcoding reference system. The experiments indicate that the achievable quality is even clearly better whereby a quality level in the ‘excellent’ range is achieved compared to the 0-DOF transcoding reference providing quality in the ‘good’ range. The objective to provide a quality level as close as possible to the native coding reference system is met in the sense that the quality score of the split rendering system is in the high ‘excellent’ range, which indicates only very minor audible differences.

### 9.11.7 ISAR Selection Experiment ISAR\_BS1534-4 (MASA 2TC, Generic Audio, 768 kbps Split Rendering Link, Headphone Presentation)

#### 9.11.7.1 Overview

Selection Experiment ISAR\_BS1534-4 evaluates Split Rendering for IVAS MASA audio using 2 transport channels (TC) at 768 kbps. See [29] Annex B.4 for details.

#### 9.11.7.2 Test conditions

The test conditions in all experiments are as follows:

- **c01 (Ref):** Native coding reference system operated at 512 kbps with IVAS decoding and head-tracked rendering at end-device

- **c02 (LP7 anchor):** 7 kHz low-pass filtered native coding reference system

- **c03 (0-DOF):** 0-DOF native transcoding reference system with 512 kbps coded IVAS content, decoded and binaurally rendered to (outdated) pose available at pre-renderer, subsequently transcoded to IVAS stereo at 256 kbps

- **c04 (CuT):** Codec under Test, i.e., ISAR split rendering system operated at the required bit rate of 768 kbps for the intermediate immersive audio representation

#### 9.11.7.3 Result plots

Provided below are the result plots and tables with statistical analysis result for the two ISAR\_BS1534-4 experiments.

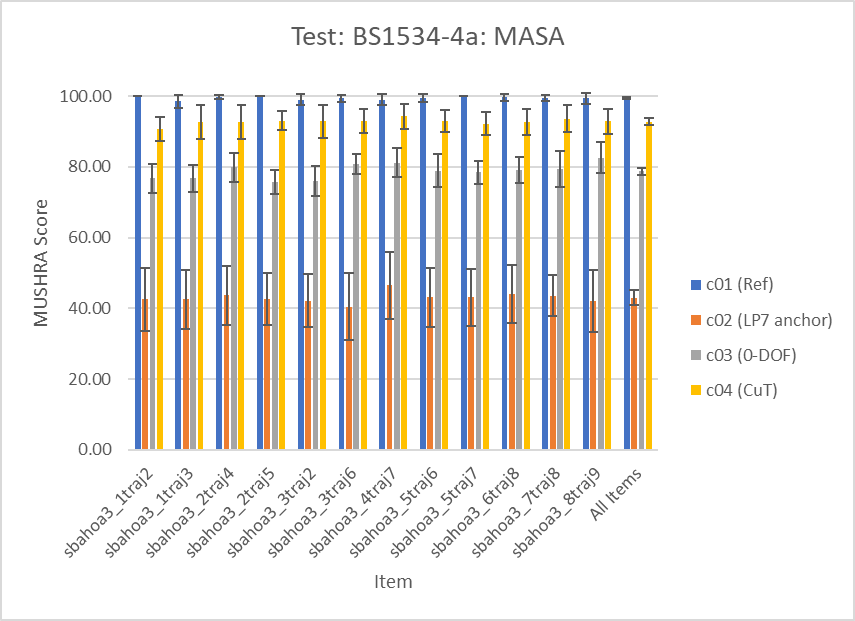


Figure ‎9.11‑7: Results of ISAR\_BS1534-4a test for MASA input audio

Ein Bild, das Text, Screenshot, parallel, Reihe enthält.

Automatisch generierte Beschreibung

Figure ‎9.11‑8: Results of ISAR\_BS1534-4b test for MASA input audio

#### 9.11.7.4 Statistical analysis

Table ‎9.11‑8: Result of statistical analysis of ISAR\_BS1534-4a test checking CuT NWT 0-DOF Reference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean Diff. (c03 - c04) | Stdev Diff. | SEMD | t | Prob. | ToR |
| -14.0167 | 5.0493 | 0.4609 | -30.4091 | 1.0000 | Pass |

Table ‎9.11‑9: Result of statistical analysis of ISAR\_BS1534-4b test checking CuT NWT 0-DOF Reference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean Diff. (c03 - c04) | Stdev Diff. | SEMD | T | Prob. | ToR |
| -13.7000 | 22.1500 | 2.0220 | -6.7754 | 1.0000 | Pass |

#### 9.11.7.5 Experimental conclusions

Conclusion of both experiments is that the ISAR split rendering solution for MASA input meets the requirement to be no worse than the 0-DOF transcoding reference system. The experiments indicate that the achievable quality is even clearly better whereby a quality level in the ‘excellent’ range is achieved compared to the 0-DOF transcoding reference providing quality in the ‘good’ range. The objective to provide a quality level as close as possible to the native coding reference system is met in the sense that the quality score of the split rendering system is in the high ‘excellent’ range, which indicates only very minor audible differences.

# 10 Objective Evaluations

## 10.1 Complexity and Delay Analysis

### 10.1.1 Complexity

NOTE: A complexity evaluation is expected to be included after completion of the IVAS fixed-point code.

### 10.1.2 Algorithmic Delay

The input signals (audio, or audio and metadata) are processed using 20 ms frames. The codec algorithmic delay depends on the input/output audio formats as described in Table ‎10.1‑1.

Table ‎10.1‑1: IVAS algorithmic delay for different input/output format combinations (rounded to integer milliseconds; in case multiple values are provided they depend on the bitrate)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Decoder output format | | | | | | |
|  |  | Mono | Stereo | Multi-Channel | Binaural audio | Scene-based audio | Object-based audio | Metadata-assisted spatial audio |
| **Encoder input format** | Mono | 32 |  |  |  |  |  |  |
| Stereo | 32 | 32 | 32 |  |  |  |  |
| Binaural | 32 |  |  | 32 |  |  |  |
| Multi-channel | 32 | 32 | 32 / 37 | 32 / 37 | 32 / 37 |  |  |
| Scene-based audio | 33 | 33 | 38 | 38 | 38 |  |  |
| Object-based audio | 32 | 32 | 32 / 37 | 32 / 37 | 32 / 37 | 32 / 37 |  |
| Metadata-assisted spatial audio | 32 / 37 | 37 | 37 | 37 | 37 |  | 32 (NOTE |
| OSBA | 33 | 33 | 38 | 38 | 38 |  |  |
| OMASA | 32 / 37 | 37 | 37 | 37 | 37 |  |  |
| NOTE: Metadata-assisted spatial audio (MASA) decoder output allows also for mono or stereo decoder output at 32 ms algorithmic delay by stripping the metadata file | | | | | | | | |

The algorithmic delay related to the core-coder coding in IVAS is 32 ms similarly as in EVS though its splitting between the encoder and the decoder is slightly different. It consists of 8.75 ms for the encoder look-ahead and 3.25 ms for the decoder delay related to the time-domain BWE and resampling in the DFT domain.

Further, the IVAS delay consists of 5 ms delay related to the rendering to the related output configuration, thus making the overall delay of 32 ms in some set-ups and 37 ms in other set-ups.

Finally, in SBA format, an additional encoder delay of 1 ms is present and it is related to the filter-bank analyses prior to the encoding.

It is also noted that the delay figures exclude any HRIR/BRIR induced delay.

The codec delay for mono (EVS) operation and stereo downmix operation in EVS compatible operation is 32 ms as described in clause 4.3 in [27].

Annex A:  
ToR Tests in Selection Phase

# A.1 ToR Tests for Requirements

Table ‎A.1-1 summarizes the results for the Requirements ToR tests over the 23 selection experiments as analyzed in the GAL report. Each row of the table shows results of ToR tests for a single Experiment - results for Test 1 on the left and for Test 2 on the right. For each Experiment, the table shows the test name, #Requirement ToR's followed by the Test 1 listening lab (LL), #ToRs passed, #ToRs failed, Test 2 LL, #ToRs passed, #ToRs failed and finally the Percent of ToRs passed across both Tests within the Experiment.

On the far right side of the table, "Percent ToRs Passed" values are shown for each of the four Groups of Experiments: 100% for Stereo, 100% for Multichannel, 98.6% for SBA, 93.8% for Objects, and 100% for MASA. Finally, at the bottom of the table, "Percent ToRs Passed" for the entire Selection Phase is reported - 5 of 288 ToRs were failed for a Percent Passed value of 98.3%.

Table A.1‑1: Summary ToR Test Results for Requirements

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Test | #Req | Test 1 | | | Test 2 | | | Passed | Group |
| LL | #Pass | #Fail | LL | #Pass | #Fail |
| P800-1 | 12 | a | 12 | 0 | d | 12 | 0 | 100% | Stereo  100% |
| P800-2 | 11 | b | 11 | 0 | d | 11 | 0 | 100% |
| P800-3 | 13 | a | 13 | 0 | d | 13 | 0 | 100% |
| BS1534-1a | 2 | a | 2 | 0 | d | 2 | 0 | 100% |
| BS1534-1b | 2 | b | 2 | 0 | d | 2 | 0 | 100% |
| BS1534-2a | 2 | a | 2 | 0 | b | 2 | 0 | 100% | Multichan.  100% |
| BS2534-2b | 2 | a | 2 | 0 | b | 2 | 0 | 100% |
| BS1534-3a | 2 | a | 2 | 0 | b | 2 | 0 | 100% |
| BS1534-3b | 2 | a | 2 | 0 | b | 2 | 0 | 100% |
| P800-4 | 13 | a | 13 | 0 | c | 13 | 0 | 100% | SBA  98.6% |
| P800-5 | 13 | a | 13 | 0 | b | 13 | 0 | 100% |
| BS1534-4a | 3 | a | 3 | 0 | d | 3 | 0 | 100% |
| BS1534-4b | 2 | b | 2 | 0 | d | 2 | 0 | 100% |
| BS1534-5a | 2 | a | 2 | 0 | d | 1 | 1 | 75.0% |
| BS1534-5b | 2 | a | 2 | 0 | b | 2 | 0 | 100% |
| P800-6 | 13 | a | 10 | 3 | c | 12 | 1 | 84.6% | Objects  93.8% |
| P800-7 | 13 | a | 13 | 0 | d | 13 | 0 | 100% |
| BS1534-6a | 3 | b | 3 | 0 | d | 3 | 0 | 100% |
| BS1534-6b | 3 | b | 3 | 0 | d | 3 | 0 | 100% |
| P800-8 | 12 | a | 12 | 0 | b | 12 | 0 | 100% | MASA  100% |
| P800-9 | 13 | a | 13 | 0 | d | 13 | 0 | 100% |
| BS1534-7a | 2 | b | 2 | 0 | d | 2 | 0 | 100% |
| BS1534-7b | 2 | b | 2 | 0 | d | 2 | 0 | 100% |
| Total | 144 |  | 141 | 3 |  | 142 | 2 | 98.3% |  |

The tests were carried out by the listening labs according to Table ‎A.1-2:

Table A.1‑2: List of listening labs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| LL identifier | Listening Lab Company | P800 tests | BS1534 tests (headphone) | BS1534 tests (loudspeaker) |
| a | FORCE Technology | 8 | 3 | 5 |
| b | HEAD acoustics GmbH / IKS (RWTH Aachen) | 3 | 6 | 5 |
| c | Macquarie University | 2 | 0 | 0 |
| d | Mesaqin.com | 5 | 9 | 0 |

Across all P.800 experiments, there are 226 requirement ToR tests (i.e. 113 requirements per labs). The CuT failed four of them, it passed 94 times and exceeded the requirement 128 times. The four failed tests are all in P800-6.

Table A.1‑3: Overview of all failed tests in P.800 experiments

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Exp. | Lab | Cond. | Bitrate | DTX | FER | ToR | Status |
| P800-6 | a | c29 | 64 | Off |  | NWT C16 | FAIL |
| c31 | 16.4 | Off | 5% | NWT C18 | FAIL |
| c36 | 24.4 | On |  | NWT C23 | FAIL |
| c | c28 | 48 | Off |  | NWT C15 | FAIL |

Across all BS.1534 experiments, there are 93 requirement ToR tests (i.e. 31 requirements per lab + a combined analysis for both labs). It should be noted that the GAL report includes for BS.1534 experiments also an analysis of the combination of the data sets from both labs. The CuT failed one of them, it passed 48 times and exceeded the requirement 44 times. The failed test is in BS1534-5A.

Table A.1‑3: Overview of all failed tests in BS.1534 experiments

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Exp. | Lab | Cond. | Bitrate | ToR | Status |
| BS1534-5A | d | c07 | 256 | NWT c05 OR BT c04 | FAIL |

In total, the CuT failed on four out of 226 requirement ToR tests for the P.800 experiments, it passed 94 times and exceeded the requirement 128 times. In the BS.1534 tests, the CuT failed on one of the 93 requirement ToR tests, it passed 48 times and exceeded the requirement 44 times. There is no instance in any experiment where the CuT failed a requirement ToR test for the data sets of both listening laboratories. Additionally, the CuT did not fail any requirement ToR test for the joint analyses of the BS.1534 data sets.

Annex B:   
Overall Characterization of the IVAS Codec

IVAS is the next generation codec in 3GPP. It is an extension of the 3GPP Enhanced Voice Services (EVS) codec offering:

- Complete bit-exact EVS codec functionality for mono speech/audio signal input

- Support of stereo and binaural audio

- Support of audio formats beyond stereo which include multi-channel audio (5.1, 5.1.2, 5.1.4, 7.1, 7.1.4), scene-based audio (Ambisonics up to 3rd order), metadata-assisted spatial audio (MASA), and object-based audio.

- Support of combined immersive audio formats: object-based audio with scene-based audio (OSBA) and object-based audio with metadata-assisted spatial audio (OMASA)

- VAD/DTX/CNG for rate efficient stereo and immersive conversational voice transmissions

- Error concealment mechanisms to combat the effects of transmission errors and lost packets

- Jitter buffer management

- Binaural rendering functionality for headphone playback including head-tracking and scene orientation control, and loudspeaker rendering functionality for loudspeaker playback

- Binaural split rendering functionality with pre-rendering and transcoding to a head-trackable intermediate representation that can be transmitted to a post-rendering end-device for headphone playback.

The codec is optimized for services over 5G mobile networks and implementations on 5G devices with:

- Operation on 20 ms audio frames

- Multi-rate/multi-mode operation at the following discrete bit rates [kbps]: 13.2, 16.4, 24.4, 32, 48, 64, 80, 96, 128, 160, 192, 256, 384, and 512

- Ability to switch bitrate upon command

- Support of sampling frequencies of 8 kHz (only EVS interoperable coding), 16 kHz, 32 kHz and 48 kHz (fullband audio content)

- Low algorithmic delay (≤38 ms)

- Complexity and memory footprint within design constraint limits defining three levels, suitable for different device types and application scenarios

IVAS is the first 3GPP communication codec which natively supports stereo and immersive audio. This enables completely new use-cases beyond traditional telephony: In a stereo or immersive telephony use-case, for example, a participant can capture and convey an immersive scene to a remote participant, e.g., to share the full immersive experience of an event. For spatial conferencing applications, the flexibility of the IVAS codec will provide multiple options for:

- Ad-hoc conferencing calls with the transmission of the physical immersive scene picked up by a UE, e.g., placed on a table. Rendering of the immersive scene makes it easier to distinguish the talkers’ voice, clearly separated from ambient sounds, leading to more natural and effort-less conferencing.

- More complex scenarios with multiple participants, transmitted as individual streams and spatially rendered on the receiving UE to match the video scene, for example.

- Scenarios where an intermediate call server combines multiple participants into an immersive scene.

Moreover, the flexibility of the IVAS codec with respect to supported audio formats, bitrates, rendering options and suitability for different device types might enable completely new usage scenarios for communication. In addition, the IVAS codec will support content distribution use-cases including streaming of stereo/immersive content and advanced VR/AR applications.

3GPP's rigorous and transparent standardization process involved the definition of demanding terms of reference (ToRs). During the selection phase, the IVAS codec was tested against in total 144 ToRs in 46 subjective tests performed in six languages by four independents test labs. During these tests, the IVAS codec has met the requirements in 98.3% of the cases with no systematic fails in both labs.

Annex C :  
IVAS Selection Phase Testplan

# C.1 Experiment P800-1: Stereo/Binaural Clean Speech Test

## C.1.1 Experiment setup

Tables C.1-1 to C.1-3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table C.1-1: Conditions for Experiment P800-1

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT |
| Bitrates | 13.2, 16.4, 24.4, 32, 48 |
| DTX | DTX on and off |
| Input level | -26 LKFS [31] |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0%, 5% |
|  |  |
| **Codec references** |  |
| Codec references  Bitrates | EVS dual mono  2x7.2, 2x8.0, 2x9.6, 2x13.2, 2x16.4, 2x24.4, 2x32 kbps |
| Input level  DTX | -26 LKFS [31]  DTX on and off |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0%, 5% |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS [31] |
| P.50 MNRU  ESDRU | Q=16, 20, 24, 28 dB  *α* = 0.1, 0.4, 0.7 |
| Input frequency mask | HP50 |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Model-based relying on convolution of raw mono clean speech sentences with Room Impulse Responses respective to various talker positions relative to a capture point as described in the ITU-T Reverberation Tool [30] |
| Audio sampling frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to SWB |
| Kind of samples | Sentence pair uttered by different talkers and genders (3 male and 3 female) |
| Number of categories | 6 Different environments and talker interactions |
| Number of samples | 6 + 1 (preliminaries) samples per category |
| Listening Level | 73 dB SPL |
| Listeners | Naïve listeners |
| Randomizations | 6 panels of 5 listeners |
| Rating Scale | Following clause 7.1.2.1 |
| Listening System | Headphones, in accordance with clause 7.1.4 |
| Listening Environment | No room noise |
| Languages | Japanese, French |

Table C.1-2: Preliminaries for Experiment P800-1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | EVS | 2x13.2 | 5% |
| 2 | c02 |  | MNRU Q=28 dB | - | - |
| 3 | c06 |  | ESDRU = 0.7 | - | - |
| 4 | c13 |  | EVS | 2x16.4 | No errors |
| 5 | c08 |  | ESDRU = 0.1 | - | - |
| 6 | c16 |  | EVS | 2x7.2 | 5% |
| 7 | c04 |  | MNRU Q=20 dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c15 |  | EVS | 2x32 | No errors |
| 10 | c07 |  | ESDRU = 0.4 | - | - |
| 11 | c05 |  | MNRU Q=16 dB | - | - |
| 12 | c21 |  | EVS | 2x24.4 | 5% |

Table C.1-3: Test conditions for Experiment P800-1,  
clean speech under clean and impaired channel conditions

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **FER/Profile** | **ToR** |
| c01 | Reference | - | - | - |  |
| c02 | MNRU Q=28 dB | - | - | - |  |
| c03 | MNRU Q=24 dB | - | - | - |  |
| c04 | MNRU Q=20 dB | - | - | - |  |
| c05 | MNRU Q=16 dB | - | - | - |  |
| c06 | ESDRU | - | - | - |  |
| c07 | ESDRU | - | - | - |  |
| c08 | ESDRU | - | - | - |  |
| c09 | EVS | 2x7.2 | off | No errors |  |
| c10 | EVS | 2x8 | off | No errors |  |
| c11 | EVS | 2x9.6 | off | No errors |  |
| c12 | EVS | 2x13.2 | off | No errors |  |
| c13 | EVS | 2x16.4 | off | No errors |  |
| c14 | EVS | 2x24.4 | off | No errors |  |
| c15 | EVS | 2x32 | off | No errors |  |
| c16 | EVS | 2x7.2 | off | 5% |  |
| c17 | EVS | 2x8 | off | 5% |  |
| c18 | EVS | 2x9.6 | off | 5% |  |
| c19 | EVS | 2x13.2 | off | 5% |  |
| c20 | EVS | 2x16.4 | off | 5% |  |
| c21 | EVS | 2x24.4 | off | 5% |  |
| c22 | EVS | 2x32 | off | 5% |  |
| c23 | EVS | 2x13.2 | on | No errors |  |
| c24 | EVS | 2x8 | on | 5% |  |
| c25 | CuT | 13.2 | off | No errors | NWT c10 OR BT c09 |
| c26 | CuT | 16.4 | off | No errors | NWT c11 OR BT c10 |
| c27 | CuT | 24.4 | off | No errors | NWT c12 OR BT c11 |
| c28 | CuT | 32 | off | No errors | NWT c13 OR BT c12 |
| c29 | CuT | 48 | off | No errors | NWT c15 OR BT c14 |
| c30 | CuT | 13.2 | off | 5% | NWT c17 OR BT c16 |
| c31 | CuT | 16.4 | off | 5% | NWT c18 OR BT c17 |
| c32 | CuT | 24.4 | off | 5% | NWT c19 OR BT c18 |
| c33 | CuT | 32 | off | 5% | NWT c20 OR BT c19 |
| c34 | CuT | 48 | off | 5% | NWT c22 OR BT c21 |
| c35 | CuT | 24.4 | on | No errors | NWT c23 |
| c36 | CuT | 13.2 | on | 5% | NWT c24 |

## C.1.2 Content type categories and scene definitions

Table C.1-4: Content type categories and scene definitions

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Category | Room | Reverb | Microphone Setup | ***Background*** | SNR  [dB] | Overtalk [s]  See NOTE 1 | Bandwidth | Talker positions See NOTE 2 | Talker selection by panel |
| cat 1 | small | anechoic | A-B (100cm) | Low level idle noise | 45 | 1 | Max available up to SWB | 1-7  5-3  2-6  4-1  3-4  7-2 | P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3 |
| cat 2 | large | anechoic | A-B (150cm) | Low level idle noise | 45 | -1 | max available up to SWB | 5-11  1-6  3-7  5-8  9-7  10-9 | P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2 |
| cat 3 | small | anechoic | M-S | Low level idle noise | 45 | -1 | max available up to SWB | 1-7  5-3  2-6  4-1  3-4  7-2 | P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1 |
| cat 4 | small | echoic | A-B (100cm) | Low level idle noise | 45 | 1 | max available up to SWB | 1-7  5-3  2-6  4-1  3-4  7-2 | P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3 |
| cat 5 | large | echoic | A-B (150cm) | Low level idle noise | 45 | 1 | max available up to SWB | 2-8  9-4  6-10  11-8  10-12  12-1 | P1: f3m3 P2: m1f1 P3: f2m2 P4: m3f3 P5: f1m1 P6: m2f2 |
| cat 6 | small | echoic | Binaural | Low level idle noise | 45 | -1 | max available up to SWB | 1-7  5-3  2-6  4-1  3-4  7-2 | P1: m2f2 P2: f3m3 P3: m1f1 P4: f2m2 P5: m3f3 P6: f1m1 |
| NOTE 1: Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.  NOTE 2: The talker positions are part of the scene definition of the different categories. They correspond to the talker positions as depicted in Figures 14.3 and 14.5 of [30] for the large and the small room, respectively. | | | | | | | | | |

# C.2 Experiment P800-2: Stereo Speech and Background Test

## C.2.1 Experiment setup

Tables C.2-1 to C.2-3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table C.2-1: Conditions for Experiment P800-2

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64 |
| DTX | DTX on and off |
| Input level | -26 LKFS [31] |
| Input frequency mask | HP50 |
| Noise | 15dB |
| Error Conditions | No errors |
|  |  |
| **Codec references** |  |
| Codec references  Bitrates | EVS dual mono  2x7.2, 2x8.0, 2x9.6, 2x13.2, 2x16.4, 2x24.4, 2x32, 2x448 kbps |
| Input level  DTX | -26 LKFS [31]  DTX off, DTX on |
| Input frequency mask | HP50 |
| Noise | 15dB |
| Error Conditions | No errors |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS [31] |
| P.50 MNRU  ESDRU | Q=12, 17, 22, 27, 32 dB  *α* = 0.1, 0.3, 0.5, 0.7 |
| Input frequency mask | HP50 |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Model-based relying on convolution of raw mono clean speech sentences with Room Impulse Responses respective to various talker positions relative to a capture point as described in the ITU-T Reverberation Tool [30] and impulse responses provided by MC. |
| Audio sampling frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to SWB |
| Kind of samples | Sentence pair uttered by different talkers and genders (3 male and 3 female) |
| Number of categories | 6 Different environments with different background types (car, office, street) and talker interactions |
| Number of samples | 6 + 1 (preliminaries) samples per category |
| Listening Level | 73 dB SPL |
| Listeners | Naïve listeners |
| Randomizations | 6 panels of 5 listeners |
| Rating Scale | Following clause 7.1.2.1 |
| Listening System | Headphones, in accordance with clause 7.1.4 |
| Listening Environment | No room noise |
| Languages | German, Mandarin |

Table C.2-2: Preliminaries for Experiment P800-2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **DTX** |
| 1 | c22 |  | EVS | 2x13.2 | on |
| 2 | c02 |  | MNRU Q=32 dB | - | - |
| 3 | c07 |  | ESDRU = 0.7 | - | - |
| 4 | c15 |  | EVS | 2x16.4 | off |
| 5 | c10 |  | ESDRU = 0.1 | - | - |
| 6 | c19 |  | EVS | 2x7.2 | on |
| 7 | c04 |  | MNRU Q=22 dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c17 |  | EVS | 2x32 | off |
| 10 | c08 |  | ESDRU = 0.5 | - | - |
| 11 | c05 |  | MNRU Q=17 dB | - | - |
| 12 | c24 |  | EVS | 2x24.4 | on |

Table C.2-3: Test conditions for Experiment P800-2,  
clean speech with background noise under clean channel conditions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **ToR** |
| c01 | Reference | - | - |  |
| c02 | MNRU Q=32 dB | - | - |  |
| c03 | MNRU Q=27 dB | - | - |  |
| c04 | MNRU Q=22 dB | - | - |  |
| c05 | MNRU Q=17 dB | - | - |  |
| c06 | MNRU Q=12 dB |  |  |  |
| c07 | ESDRU *α* = 0.7 | - | - |  |
| c08 | ESDRU *α* = 0.5 | - | - |  |
| c09 | ESDRU *α* = 0.3 |  |  |  |
| c10 | ESDRU *α* = 0.1 | - | - |  |
| c11 | EVS | 2x7.2 | off |  |
| c12 | EVS | 2x8 | off |  |
| c13 | EVS | 2x9.6 | off |  |
| c14 | EVS | 2x13.2 | off |  |
| c15 | EVS | 2x16.4 | off |  |
| c16 | EVS | 2x24.4 | off |  |
| c17 | EVS | 2x32 | off |  |
| c18 | EVS | 2x48 | off |  |
| c19 | EVS | 2x7.2 | on |  |
| c20 | EVS | 2x8 | on |  |
| c21 | EVS | 2x9.6 | on |  |
| c22 | EVS | 2x13.2 | on |  |
| c23 | EVS | 2x16.4 | on |  |
| c24 | EVS | 2x24.4 | on |  |
| c25 | EVS | 2x32 | on |  |
| c26 | CuT | 13.2 | off | NWT c12 OR BT c11 |
| c27 | CuT | 16.4 | off | NWT c13 OR BT c12 |
| c28 | CuT | 24.4 | off | NWT c14 OR BT c13 |
| c29 | CuT | 32 | off | NWT c15 OR BT c14 |
| c30 | CuT | 48 | off | NWT c17 OR BT c16 |
| c31 | CuT | 64 | off | NWT c18 OR BT c17 |
| c32 | CuT | 13.2 | on | NWT c20 OR BT c19 |
| c33 | CuT | 16.4 | on | NWT c21 OR BT c20 |
| c34 | CuT | 24.4 | on | NWT c22 OR BT c21 |
| c35 | CuT | 32 | on | NWT c23 OR BT c22 |
| c36 | CuT | 48 | on | NWT c25 OR BT c24 |

## C.2.2 Content type categories and scene definitions

Table C.2-4: Content type categories and scene definitions

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Category | Room | Reverb | Microphone Setup | Background | SNR  In dB | Overtalk [s]see NOTE 1 | Bandwidth | Talker positions see NOTE 2 | Talker selection by panel |
| cat 1 | car | car | A-B Cardioid pair20 cm | car | 15 | -1 | Max available up to SWB | Driver-Passenger  BackRight-Driver  Driver-BackCenter  BackLeft-Driver  BackRight-BackLeft  BackCenter-BackRight | P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3 |
| cat 2 | - | - | A-B (20cm) | street | 15 | -1 | max available up to SWB | Regularly spaced around the microphone | P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2 |
| cat 3 | small | echoic | M-S | office | 15 | -1 | max available up to SWB | 1-7  5-3  2-6  4-1  3-4  7-2 | P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1 |
| cat 4 | small | echoic | A-B (100cm) | office | 15 | -1 | max available up to SWB | 1-7  5-3  2-6  4-1  3-4  7-2 | P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3 |
| cat 5 | large | echoic | A-B (150cm) | office | 15 | -1 | max available up to SWB | 2-8  9-4  6-10  11-8  10-12  12-1 | P1: f3m3 P2: m1f1 P3: f2m2 P4: m3f3 P5: f1m1 P6: m2f2 |
| cat 6 | small | echoic | Binaural | office | 15 | -1 | max available up to SWB | 1-7  5-3  2-6  4-1  3-4  7-2 | P1: m2f2 P2: f3m3 P3: m1f1 P4: f2m2 P5: m3f3 P6: f1m1 |
| NOTE 1: Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.  NOTE 2: The talker positions are part of the scene definition of the different categories. For office noise, they correspond to the talker positions as depicted in Figures 14.3 and 14.5 of [30] for the large and the small room. | | | | | | | | | |

# C.3 Experiment P800-3: Stereo Mixed and Music Test

## C.3.1 Experiment setup

Tables C.3-1 to C.3-3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table C.3-1: Conditions for Experiment P800-3

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64 |
| DTX | DTX on and off |
| Input level | -26 LKFS [31] |
| Input frequency mask | 20KBP |
| Noise | No noise |
| Error Conditions | 0%, 5% |
|  |  |
| **Codec references** |  |
| Codec references | EVS dual mono |
| Bitrates | 2x7.2, 2x8.0, 2x9.6, 2x13.2, 2x16.4, 2x24.4, 2x32, 2x48 kbps |
| Input level  DTX | -26 LKFS [31]  DTX off |
| Input frequency mask | 20KBP |
| Noise | No noise |
| Error Conditions | 0%, 5% |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS [31] |
| P.50 MNRU  ESDRU | Q = 12, 17, 22, 27 dB  = 0.1, 0.4, 0.7 |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Categories | Following clause 7.1.3.4 |
| Audio sampling frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Number of categories | 3 mixed content and 3 music |
| Number of samples | 6 + 1 (preliminaries) samples per category |
| Listening Level | 73 dB SPL |
| Listeners | Naïve listeners |
| Randomizations | 6 panels of 5 listeners |
| Rating Scale | Following clause 7.1.2.1 |
| Listening System | Headphones, in accordance with clause 7.1.4 |
| Listening Environment | No room noise |
| Languages | Danish, Mandarin |

Table C.3-2: Preliminaries for Experiment P800-3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | C20 |  | EVS | 2x13.2 | 5% |
| 2 | c02 |  | MNRU Q=28 dB | - | - |
| 3 | c06 |  | ESDRU = 0.7 | - | - |
| 4 | c13 |  | EVS | 2x16.4 | No errors |
| 5 | c08 |  | ESDRU = 0.1 | - | - |
| 6 | c17 |  | EVS | 2x7.2 | 5% |
| 7 | c04 |  | MNRU Q=23 dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c15 |  | EVS | 2x32 | No errors |
| 10 | c07 |  | ESDRU = 0.4 | - | - |
| 11 | c05 |  | MNRU Q=12 dB | - | - |
| 12 | c22 |  | EVS | 2x24.4 | 5% |

Table C.3-3: Test conditions for Experiment P800-3,  
mixed contents and music under clean and impaired channel conditions

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **FER/Profile** | **ToR** |
| c01 | Reference | - | - | - |  |
| c02 | MNRU Q=27 dB | - | - | - |  |
| c03 | MNRU Q=22 dB | - | - | - |  |
| c04 | MNRU Q=17 dB | - | - | - |  |
| c05 | MNRU Q=12 dB | - | - | - |  |
| c06 | ESDRU | - | - | - |  |
| c07 | ESDRU | - | - | - |  |
| c08 | ESDRU 1 | - | - | - |  |
| c09 | EVS | 2x7.2 | off | No errors |  |
| c10 | EVS | 2x8 | off | No errors |  |
| c11 | EVS | 2x9.6 | off | No errors |  |
| c12 | EVS | 2x13.2 | off | No errors |  |
| c13 | EVS | 2x16.4 | off | No errors |  |
| c14 | EVS | 2x24.4 | off | No errors |  |
| c15 | EVS | 2x32 | off | No errors |  |
| c16 | EVS | 2x48 | off | No errors |  |
| c17 | EVS | 2x7.2 | off | 5% |  |
| c18 | EVS | 2x8 | off | 5% |  |
| c19 | EVS | 2x9.6 | off | 5% |  |
| c20 | EVS | 2x13.2 | off | 5% |  |
| c21 | EVS | 2x16.4 | off | 5% |  |
| c22 | EVS | 2x24.4 | off | 5% |  |
| c23 | EVS | 2x32 | off | 5% |  |
| c24 | CuT | 13.2 | off | No errors | NWT c10 OR BT c09 |
| c25 | CuT | 16.4 | off | No errors | NWT c11 OR BT c10 |
| c26 | CuT | 24.4 | off | No errors | NWT c12 OR BT c11 |
| c27 | CuT | 32 | off | No errors | NWT c13 OR BT c12 |
| c28 | CuT | 48 | off | No errors | NWT c15 OR BT c14 |
| c29 | CuT | 64 | off | No errors | NWT c16 OR BT c15 |
| c30 | CuT | 13.2 | off | 5% | NWT c18 OR BT c17 |
| c31 | CuT | 16.4 | off | 5% | NWT c19 OR BT c18 |
| c32 | CuT | 24.4 | off | 5% | NWT c20 OR BT c19 |
| c33 | CuT | 32 | off | 5% | NWT c21 OR BT c20 |
| c34 | CuT | 48 | off | 5% | NWT c23 OR BT c22 |
| c35 | CuT | 24.4 | on | No errors | NWT c12 OR BT c11(1 |
| c36 | CuT | 13.2 | on | 5% | NWT c18 OR BT c17(1 |

(1These references are not requirements, they are just for information.

# C.4 Experiment P800-4: FOA Clean Speech Test

## C.4.1 Experiment setup

Tables C.4-1 to C.4-3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table C.4-1: Conditions for Experiment P800-4 (Clean Speech)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT |
| Bitrates | IVAS candidate operated at 16.4, 24.4, 32, 48, 64, 80 and 96 kbps |
| DTX | DTX off |
| Input level | -26 LKFS [31] |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0%, 5% |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS |
| Bitrates | 3x7.2, 4x7.2, 4x8, 4x9.6, 4x13.2, 4x16.4, 4x24.4, 4x32 kbps |
| Input level  DTX | -26 LKFS [31]  DTX off |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0%, 5% |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS [31] |
| P.50 MNRU  ESDRU | Q= 20, 24, 28, 32 dB  α = 0.4, 0.6, 0.8 |
| Input frequency mask | HP50 |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Model-based relying on convolution of raw mono clean speech sentences convolved with (FOA) Spatial Room Impulse Responses respective to various talker positions relative to a capture point and spatial (FOA) ambient noise mixing. |
| Binaural renderer | FOA to binaural (external) rendering |
| Audio sampling frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pair uttered by different talkers and genders (3 male and 3 female) |
| Number of categories | 6 Different environments and talker interactions |
| Number of samples | 6 + 1 (preliminaries) samples per category |
| Listening Level | 73 dB SPL |
| Listeners | Naïve listeners |
| Randomizations | 6 panels of 5 listeners |
| Rating Scale | Following clause 7.1.2.1 |
| Listening System | Headphones, in accordance with clause 7.1.4 |
| Listening Environment | No room noise |
| Languages | Japanese, English |

Table C.4-2: Preliminaries for Experiment P800-4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c20 |  | EVS | 4x13.2 | 5% |
| 2 | c02 |  | MNRU Q=32 dB | - | - |
| 3 | c06 |  | ESDRU = 0.8 | - | - |
| 4 | c14 |  | EVS | 4x16.4 | No errors |
| 5 | c08 |  | ESDRU = 0.55 | - | - |
| 6 | c17 |  | EVS | 4x7.2 | 5% |
| 7 | c04 |  | MNRU Q=24 dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c16 |  | EVS | 4x32 | No errors |
| 10 | c07 |  | ESDRU = 0.675 | - | - |
| 11 | c05 |  | MNRU Q=20 dB | - | - |
| 12 | c22 |  | EVS | 4x24.4 | 5% |

Table C.4-3: Test conditions for Experiment P800-4,  
clean speech under clean and impaired channel conditions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **FER/Profile** | **ToR** |
| c01 | Reference | - | - |  |
| c02 | MNRU Q=32 dB | - | - |  |
| c03 | MNRU Q=28 dB | - | - |  |
| c04 | MNRU Q=24 dB | - | - |  |
| c05 | MNRU Q=20 dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU | - | - |  |
| c09 | EVS | 3x7.2 | No errors |  |
| c10 | EVS | 4x7.2 | No errors |  |
| c11 | EVS | 4x8 | No errors |  |
| c12 | EVS | 4x9.6 | No errors |  |
| c13 | EVS | 4x13.2 | No errors |  |
| c14 | EVS | 4x16.4 | No errors |  |
| c15 | EVS | 4x24.4 | No errors |  |
| c16 | EVS | 4x32 | No errors |  |
| c17 | EVS | 4x7.2 | 5% |  |
| c18 | EVS | 4x8 | 5% |  |
| c19 | EVS | 4x9.6 | 5% |  |
| c20 | EVS | 4x13.2 | 5% |  |
| c21 | EVS | 4x16.4 | 5% |  |
| c22 | EVS | 4x24.4 | 5% |  |
| c23 | EVS | 4x32 | 5% |  |
| c24 | CuT | 16.4 | No errors | NWT c09 |
| c25 | CuT | 24.4 | No errors | NWT c10 |
| c26 | CuT | 32 | No errors | NWT c12 or BT c11 |
| c27 | CuT | 48 | No errors | NWT c14 or BT c13 |
| c28 | CuT | 64 | No errors | NWT c15 or BT c14 |
| c29 | CuT | 80 | No errors | NWT c15 or BT c14 |
| c30 | CuT | 96 | No errors | NWT c16 or BT c15 |
| c31 | CuT | 24.4 | 5% | NWT c17 |
| c32 | CuT | 32 | 5% | NWT c19 or BT c18 |
| c33 | CuT | 48 | 5% | NWT c21 or BT c20 |
| c34 | CuT | 64 | 5% | NWT c22 or BT c21 |
| c35 | CuT | 80 | 5% | NWT c22 or BT c21 |
| c36 | CuT | 96 | 5% | NWT c23 or BT c22 |

## C.4.2 Content type categories and scene definitions

Table C.4-4: Content type categories and scene definitions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Category** | **Environment** | **Background** | **SNR [dB]** | **Overtalk [s] see NOTE 1** | **Bandwidth** | **Talker positions** | **Talker selection by panel see NOTE 2** |
| cat 1 | room\_1\_FOA | room\_1\_cleanbg\_FOA | 45 | 1 | Max |  | P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3 |
| cat 2 | room\_4\_FOA | room\_[1/4]\_cleanbg\_FOA | 45 | -1 | Max |  | P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2 |
| cat 3 | room\_2\_FOA | room\_2\_cleanbg\_FOA | 45 | 1 | Max |  | P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1 |
| cat 4 | room\_5\_FOA | room\_[2/5]\_cleanbg\_FOA | 45 | -1 | Max |  | P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3 |
| cat 5 | room\_3\_FOA | room\_3\_cleanbg\_FOA | 45 | 1 | Max |  | P1: f3m3 P2: m1f1 P3: f2m2 P4: m3f3 P5: f1m1 P6: m2f2 |
| cat 6 | room\_6\_FOA | room\_[3/6]\_cleanbg\_FOA | 45 | -1 | Max |  | P1: m2f2 P2: f3m3 P3: m1f1 P4: f2m2 P5: m3f3 P6: f1m1 |
| NOTE 1 Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.  NOTE 2 All sentences by the 6 talkers shall be unique. | | | | | | | |

# C.5 Experiment P800-5: FOA Speech+background Test

## C.5.1 Experiment setup

Tables C.5-1 to C.5-3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table C.5-1: Conditions for Experiment P800-5

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT |
| bitrates | 16.4, 24.4, 32, 48, 64, 80, 96 kbps |
| DTX | DTX off, DTX on |
| Input level | -26 LKFS [31] |
| Input frequency mask | HP50 |
| Noise | 10, 15 dB |
| Error Conditions | No errors |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS |
| Bitrates | 3x7.2, 4x7.2, 4x8, 4x9.6, 4x13.2, 4x16.4, 4x24.4, 4x32 kbps |
| Input level  DTX | -26 LKFS [31]  DTX off, DTX on |
| Input frequency mask | HP50 |
| Noise | 10, 15dB |
| Error Conditions | No errors |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS [31] |
| P.50 MNRU  ESDRU | Q= 17, 21, 25, 29 dB  *α* = 0.4, 0.6, 0.8 |
| Input frequency mask | HP50 |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Model-based relying on convolution of raw mono clean speech sentences convolved with (FOA) Spatial Room Impulse Responses respective to various talker positions relative to a capture point and spatial (FOA) ambient noise mixing |
| Binaural renderer | FOA to binaural (external) rendering |
| Audio sampling frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pair uttered by different talkers and genders (3 male and 3 female) |
| Number of categories | 6 Different environments with different background types (car, office, street) and talker interactions |
| Number of samples | 6 + 1 (preliminaries) samples per category |
| Listening Level | 73 dB SPL |
| Listeners | Naïve listeners |
| Randomizations | 6 panels of 5 listeners |
| Rating Scale | Following clause 7.1.2.1 |
| Listening System | Headphones, in accordance with clause 7.1.4 |
| Listening Environment | No room noise |
| Languages | Danish, German |

Table C.5-2: Preliminaries for Experiment P800-5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **DTX** |
| 1 | c21 |  | EVS | 4x13.2 | on |
| 2 | c02 |  | MNRU Q=29 dB | - | - |
| 3 | c06 |  | ESDRU = 0.8 | - | - |
| 4 | c14 |  | EVS | 4x16.4 | off |
| 5 | c08 |  | ESDRU = 0.4 | - | - |
| 6 | c18 |  | EVS | 4x7.2 | on |
| 7 | c04 |  | MNRU Q=21 dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c16 |  | EVS | 4x32 | off |
| 10 | c07 |  | ESDRU = 0.6 | - | - |
| 11 | c05 |  | MNRU Q=17 dB | - | - |
| 12 | c23 |  | EVS | 4x24.4 | on |

Table C.5-3: Test conditions for Experiment P800-5,  
clean speech with background noise under clean channel conditions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **ToR** |
| c01 | Reference | - | - |  |
| c02 | MNRU Q=29 dB | - | - |  |
| c03 | MNRU Q=25 dB | - | - |  |
| c04 | MNRU Q=21 dB | - | - |  |
| c05 | MNRU Q=17 dB | - | - |  |
| c06 | ESDRU 8 | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU4 | - | - |  |
| c09 | EVS | 3x7.2 | off |  |
| c10 | EVS | 4x7.2 | off |  |
| c11 | EVS | 4x8 | off |  |
| c12 | EVS | 4x9.6 | off |  |
| c13 | EVS | 4x13.2 | off |  |
| c14 | EVS | 4x16.4 | off |  |
| c15 | EVS | 4x24.4 | off |  |
| c16 | EVS | 4x32 | off |  |
| c17 | EVS | 3x7.2 | on |  |
| c18 | EVS | 4x7.2 | on |  |
| c19 | EVS | 4x8 | on |  |
| c20 | EVS | 4x9.6 | on |  |
| c21 | EVS | 4x13.2 | on |  |
| c22 | EVS | 4x16.4 | on |  |
| c23 | EVS | 4x24.4 | on |  |
|  |  |  |  |  |
| c24 | CuT | 16.4 | off | NWT c09 |
| c25 | CuT | 24.4 | off | NWT c10 |
| c26 | CuT | 32 | off | NWT c12 or BT c11 |
| c27 | CuT | 48 | off | NWT c14 or BT c13 |
| c28 | CuT | 64 | off | NWT c15 or BT c14 |
| c29 | CuT | 80 | off | NWT c15 or BT c14 |
| c30 | CuT | 96 | off | NWT c16 or BT c15 |
| c31 | CuT | 16.4 | on | NWT c17 |
| c32 | CuT | 24.4 | on | NWT c18 |
| c33 | CuT | 32 | On | NWT c20 or BT c19 |
| c34 | CuT | 48 | On | NWT c22 or BT c21 |
| c35 | CuT | 64 | On | NWT c23 or BT c22 |
| c36 | CuT | 80 | On | NWT c23 or BT c22 |

## C.5.2 Content type categories and scene definitions

Table C.5-4: Content type categories and scene definitions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Category** | **Environment** | **Background** | **SNR[dB]** | **Overtalk [s] see NOTE 1** | **Bandwidth** | **Talker position** | **Talker selection by panel see NOTE 2** |
| cat 1 | car\_1\_FOA | car\_1\_bg\_FOA | 10 | -1 | Max |  | P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3 |
| cat 2 | car\_[1/2]\_FOA | car\_[1/2]\_bg\_FOA | 10 | -1 | Max |  | P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2 |
| cat 3 | out\_1\_FOA | street\_1\_bg\_FOA | 15 | -1 | Max |  | P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1 |
| cat 4 | out\_[1/2]\_FOA | [park\_1\_bg\_FOA / nature\_1\_bg\_FOA / event\_1\_bg\_FOA / street\_[1/2]\_bg\_FOA] | 15 | -1 | Max |  | P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3 |
| cat 5 | room\_[X]\_FOA | office\_1\_bg\_FOA | 15 | -1 | Max |  | P1: f3m3 P2: m1f1 P3: f2m2 P4: m3f3 P5: f1m1 P6: m2f2 |
| cat 6 | room\_[X]\_FOA | [cafeteria\_1\_bg\_FOA / mall\_1\_bg\_FOA/ office[1/2]\_bg\_FOA] | 15 | -1 | Max |  | P1: m2f2 P2: f3m3 P3: m1f1 P4: f2m2 P5: m3f3 P6: f1m1 |
| NOTE 1 Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.  NOTE 2 All sentences by the 6 talkers shall be unique. | | | | | | | |

# C.6 Experiment P800-6: 1-Object Clean Speech Test

## C.6.1 Experiment setup

Tables C.6-1 to C.6-3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

The test Categories correspond to different talkers.

Table E.6.1: Conditions for Experiment P800-6

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64 kbps |
| DTX | DTX ON and OFF |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0%, 5% FERs |
|  |  |
| **Codec references** |  |
| Codec references | EVS |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64 kbps |
| DTX | DTX ON and OFF |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0%, 5% FERs |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS [31] |
| MNRU | Q= 15, 23, 31, 39, 47 dB |
| ESDRU | 0.1, 0.3, 0.5, 0.7 |
| Input frequency mask | HP50 |
|  |  |
| **Common Conditions** |  |
| Audio sampling frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Number of talkers (categories) | 3 male and 3 female |
| Number of speech samples | 6 + 1 (preliminaries) sentence pairs per talker |
| Listening Level | 73 dB SPL |
| Listeners | Naïve Listeners |
| Randomizations (panels) | 6 panels of 5 listeners |
| Rating Scale | Following clause 7.1.2.1 |
| Listening System | Headphones, in accordance with clause 7.1.4 |
| Languages | Japanese, English |

Table C.6-2: Preliminaries for Experiment P800-6

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial** | **Label** | **Sample** | **Condition** | **Bitrate [kb/s]** | **DTX** | **FER [%]** |
| **1** | C19 |  | EVS | 24.4 | Off | 5 |
| **2** | C04 |  | MNRU 31 | - | - | - |
| **3** | C21 |  | EVS | 13.2 | On | 0 |
| **4** | C09 |  | ESDRU 0.5 | - | - | - |
| **5** | C14 |  | EVS | 32 | Off | 0 |
| **6** | C02 |  | MNRU 15 | - | - | - |
| **7** | C12 |  | EVS | 16.4 | Off | 0 |
| **8** | C01 |  | Direct | - | - | - |
| **9** | C17 |  | EVS | 13.2 | Off | 5 |
| **10** | C06 |  | MNRU 47 | - | - | - |
| **11** | C23 |  | EVS | 24.4 | On | 0 |
| **12** | C07 |  | ESDRU 0.1 | - | - | - |

Table C.6-3 Test conditions for Experiment P800-6

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Label | **Condition** | **Bitrate [kb/s]** | **DTX** | **FER [%]** | **Ref condition** | **ToR** |
| **C01** | Direct |  | - | - | - | - |
| **C02** | MNRU 15 |  | - | - | - | - |
| **C03** | MNRU 23 |  | - | - | - | - |
| **C04** | MNRU 31 |  | - | - | - | - |
| **C05** | MNRU 39 |  | - | - | - | - |
| **C06** | MNRU 47 |  | - | - | - | - |
| **C07** | ESDRU 0.1 |  | - | - | - | - |
| **C08** | ESDRU 0.3 |  | - | - | - | - |
| **C09** | ESDRU 0.5 |  | - | - | - | - |
| **C10** | ESDRU 0.7 |  | - | - | - | - |
| **C11** | EVS | 13.2 | Off | 0 | - | - |
| **C12** | EVS | 16.4 | Off | 0 | - | - |
| **C13** | EVS | 24.4 | Off | 0 | - | - |
| **C14** | EVS | 32 | Off | 0 | - | - |
| **C15** | EVS | 48 | Off | 0 | - | - |
| **C16** | EVS | 64 | Off | 0 |  |  |
| **C17** | EVS | 13.2 | Off | 5 | - | - |
| **C18** | EVS | 16.4 | Off | 5 | - | - |
| **C19** | EVS | 24.4 | Off | 5 | - | - |
| **C20** | EVS | 32 | Off | 5 |  |  |
| **C21** | EVS | 13.2 | On | 0 | - | - |
| **C22** | EVS | 16.4 | On | 0 | - | - |
| **C23** | EVS | 24.4 | On | 0 | - | - |
| **C24** | CuT | 13.2 | Off | 0 | C11 | NWT |
| **C25** | CuT | 16.4 | Off | 0 | C12 | NWT |
| **C26** | CuT | 24.4 | Off | 0 | C13 | NWT |
| **C27** | CuT | 32 | Off | 0 | C14 | NWT |
| **C28** | CuT | 48 | Off | 0 | C15 | NWT |
| **C29** | CuT | 64 | Off | 0 | C16 | NWT |
| **C30** | CuT | 13.2 | Off | 5 | C17 | NWT |
| **C31** | CuT | 16.4 | Off | 5 | C18 | NWT |
| **C32** | CuT | 24.4 | Off | 5 | C19 | NWT |
| **C33** | CuT | 32 | Off | 5 | C20 | NWT |
| **C34** | CuT | 13.2 | On | 0 | C21 | NWT |
| **C35** | CuT | 16.4 | On | 0 | C22 | NWT |
| **C36** | CuT | 24.4 | On | 0 | C23 | NWT |

## C.6.2 Content type categories and scene definitions

A leading and trailing silence is present for each sample, in accordance with IVAS-7a. The metadata corresponds to the whole duration of the samples. This means that for moving objects, only a part of the trajectory corresponds to active speech. The following scenes are used:

a) Talker sitting at a table (elevation 0°), at different azimuths.

b) Standing talker (elevation 35°), at different azimuths.

c) Smaller talker (child) walking around a table in the positive sense (counterclockwise), elevation 0°. Azimuth varies continuously for the sentence pair.

d) Adult talker walking around a table in the negative sense (clockwise), elevation 35°. Azimuth varies continuously for the sentence pair.

e) Elevation displacement: Elevation varies continuously for the sentence pair. Azimuth is constant for a sentence pair, but different for each sentence pair.

f) Azimuth and elevation displacement: Azimuth and elevation vary continuously.

Each of the sentences uttered by a certain talker is encoded using different scene. To balance the test, in addition to listeners of each panel listening to all talkers (Categories), all scenes are also covered in each panel. Allocation of scenes to each panel is given in the Table below:

Table C.6-4: Allocation of scenes for each to talker (category cat 1-cat 6) and listening panel (P1-P6)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Category** | **Talker initial elevation** | **Elevation change** see NOTE 1 | **Talker initial azimuth** | **Azimuth change** see NOTE 2 | **Panel** |
| **cat 1:**  M1 | 0°  35°  0°  35°  -90°  35° | static  static  static  static  0.3°/ frame  -0.2°/ frame | 0°  180°  120°  180°  120°  0° | static  static  1°/ frame  -1°/ frame  static  0.5°/ frame | P1  P2  P3  P4  P5  P6 |
| **cat 2:**  F1 | 35°  0°  35°  -90°  35°  0° | static  static  static  0.3°/ frame  -0.2°/ frame  static | 120°  60°  120°  60°  300°  300° | static  1°/ frame  -1°/ frame  static  0.5°/ frame  static | P1  P2  P3  P4  P5  P6 |
| **cat 3:**  M2 | 0°  35°  -90°  35°  0°  35° | static  static  0.3°/ frame  -0.2°/ frame  static  static | 0°  60°  0°  240°  240°  60° | 1°/ frame  -1°/ frame  static  0.5°/ frame  static  static | P1  P2  P3  P4  P5  P6 |
| **cat 4:**  F2 | 35°  -90°  35°  0°  35°  0° | static  0.3°/ frame  -0.2°/ frame  static  static  static | 0°  300°  180°  180°  0°  300° | -1°/ frame  static  0.5°/ frame  static  static  1°/ frame | P1  P2  P3  P4  P5  P6 |
| **cat 5:**  M3 | -90°  35°  0°  35°  0°  35° | 0.3°/ frame  -0.2°/ frame  static  static  static  static | 240°  120°  120°  300°  240°  300° | static  0.5°/ frame  static  static  1°/ frame  -1°/ frame | P1  P2  P3  P4  P5  P6 |
| **cat 6:**  F3 | 35°  0°  35°  0°  35°  -90° | -0.2°/ frame  static  static  static  static  0.3°/ frame | 60°  60°  240°  180°  240°  180° | 0.5°/ frame  static  static  1°/ frame  -1°/ frame  static | P1  P2  P3  P4  P5  P6 |
| NOTE 1 The positive sense for elevation is from bottom up  NOTE 2 The positive sense for azimuth is counterclockwise | | | | | |

# C.7 Experiment P800-7: 2-Objects Clean Speech Test

## C.7.1 Experiment setup

Tables C.7-1 to C.7-3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

The test Categories correspond to different talker pairs.

Table C.7-1: Conditions for Experiment P800-7

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT |
| Bitrates | 16.4, 24.4, 32, 48, 64 kbps |
| DTX | DTX ON and OFF |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0%, 5% FERs |
|  |  |
| **Codec references** |  |
| Codec references | EVS |
| Bitrates | 2x8.0, 2x13.2, 2x16.4, 2x24.4, 2x32 kbps |
| DTX | DTX ON and OFF |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0%, 5% FERs |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS [31] |
| MNRU | Q= 15, 23, 31, 39, 47 dB |
| ESDRU | 0.1, 0.3, 0.5, 0.7 |
| Input frequency mask | HP50 |
|  |  |
| **Common Conditions** |  |
| Audio sampling frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Number of talkers (categories) | 3 male and 3 female |
| Number of speech samples | 6 + 1 (preliminaries) sentence pairs per talker. |
| Listening Level | 73 dB SPL |
| Listeners | Naïve Listeners |
| Randomizations (panels) | 6 panels of 5 listeners |
| Rating Scale | Following clause 7.1.2.1 |
| Listening System | Headphones, in accordance with clause 7.1.4 |
| Languages | Danish, Mandarin |

Table C.7-2: Preliminaries for Experiment P800-7

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial** | **Label** | **Sample** | **Condition** | **Bitrate [kb/s]** | **DTX** | **FER [%]** |
| **1** | C18 |  | EVS | 2 x 16.4 | Off | 5 |
| **2** | C04 |  | MNRU 31 | - | - | - |
| **3** | C20 |  | EVS | 2 x 8.0 | On | 0 |
| **4** | C09 |  | ESDRU 0.5 | - | - | - |
| **5** | C14 |  | EVS | 2 x 24.4 | Off | 0 |
| **6** | C02 |  | MNRU 15 | - | - | - |
| **7** | C12 |  | EVS | 2 x 13.2 | Off | 0 |
| **8** | C01 |  | Direct | - | - | - |
| **9** | C16 |  | EVS | 2 x 8.0 | Off | 5 |
| **10** | C06 |  | MNRU 47 | - | - | - |
| **11** | C22 |  | EVS | 2 x 16.4 | On | 0 |
| **12** | C07 |  | ESDRU 0.1 | - | - | - |

Table C.7-3 Test conditions for Experiment P800-7:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kb/s]** | **DTX** | **FER [%]** | **Ref condition** | **ToR** |
| **C01** | Direct | - | - | - | - | - |
| **C02** | MNRU 15 | - | - | - | - | - |
| **C03** | MNRU 23 | - | - | - | - | - |
| **C04** | MNRU 31 | - | - | - | - | - |
| **C05** | MNRU 39 | - | - | - | - | - |
| **C06** | MNRU 47 | - | - | - | - | - |
| **C07** | ESDRU 0.1 | - | - | - | - | - |
| **C08** | ESDRU 0.3 | - | - | - | - | - |
| **C09** | ESDRU 0.5 | - | - | - | - | - |
| **C10** | ESDRU 0.7 | - | - | - | - | - |
| **C11** | EVS | 2 x 8.0 | Off | 0 | - | - |
| **C12** | EVS | 2 x 13.2 | Off | 0 | - | - |
| **C13** | EVS | 2 x 16.4 | Off | 0 | - | - |
| **C14** | EVS | 2 x 24.4 | Off | 0 | - | - |
| **C15** | EVS | 2 x 32 | Off | 0 | - | - |
| **C16** | EVS | 2 x 8.0 | Off | 5 | - | - |
| **C17** | EVS | 2 x 13.2 | Off | 5 | - | - |
| **C18** | EVS | 2 x 16.4 | Off | 5 | - | - |
| **C19** | EVS | 2 x 24.4 | Off | 5 | - | - |
| **C20** | EVS | 2 x 8.0 | On | 0 | - | - |
| **C21** | EVS | 2 x 13.2 | On | 0 | - | - |
| **C22** | EVS | 2 x 16.4 | On | 0 | - | - |
| **C23** | EVS | 2 x 24.4 | On | 0 |  |  |
| **C24** | CuT | 16.4 | Off | 0 | C11 | NWT |
| **C25** | CuT | 24.4 | Off | 0 | C12 | NWT |
| **C26** | CuT | 32 | Off | 0 | C13 | NWT |
| **C27** | CuT | 48 | Off | 0 | C14 | NWT |
| **C28** | CuT | 64 | Off | 0 | C15 | NWT |
| **C29** | CuT | 16.4 | Off | 5 | C16 | NWT |
| **C30** | CuT | 24.4 | Off | 5 | C17 | NWT |
| **C31** | CuT | 32 | Off | 5 | C18 | NWT |
| **C32** | CuT | 48 | Off | 5 | C19 | NWT |
| **C33** | CuT | 16.4 | On | 0 | C20 | NWT |
| **C34** | CuT | 24.4 | On | 0 | C21 | NWT |
| **C35** | CuT | 32 | On | 0 | C22 | NWT |
| **C36** | CuT | 48 | On | 0 | C23 | NWT |

## C.7.2 Content type categories and scene definitions

The listening database consists of artificially created spatial audio samples from monophonic clean speech recordings where always 1 female and 1 male talker are combined in conversation-like scenarios following the Scene descriptions below.

A leading and trailing silence is present for each artificially created spatial audio sample, in accordance with IVAS-7a. The metadata corresponds to the whole duration of the sample. This means that for moving objects, only a part of the trajectory corresponds to active speech.

In one half of the samples, the 2nd talker’s utterance follows the 1st talker’s utterance simulating natural conversation. The gap between the utterances is set to 1 s. In the other half of the samples, the situation is similar, but the utterances partially overlap. The targeted overlap is also 1 s. Non-overlapping sentence pairs are used for Scenes a., c., and e. as described below. Overlapping sentence pairs are used for Scenes b., d., and f. The following scenes are used:

a) Two talkers sitting at a table (elevation 0°), at different azimuths. To increase positional variation, both the absolute azimuths and the difference of the azimuths of both talkers vary for each sentence pair.

b) Two standing talkers (elevation 35°), at different azimuths. To increase positional variation, both the absolute azimuths and the difference of the azimuths of both talkers vary for each sentence pair.

c) One talker sitting at a table (elevation 0°), second talker standing beside the table (elevation 45°). Non-overlapping utterances.

d) One talker sitting at a table (elevation 0°), second talker walking around the table (elevation 45°). The azimuth of the 2nd talker varies continually, positive sense is counter clockwise.

e) Two talkers walking side-by-side around the table (elevation 45°). The azimuth is the same for both talkers and varies continually.

f) Two talkers walking around the table in opposite directions (elevation 30°), starting at the same position. Azimuths of both talkers vary continually.

The following table lists the test Categories corresponding to different talker pairs. Each of the sentence pairs uttered by a certain talker pair is associated to different scenes. To balance the test, listeners of each panel listen to all talker pairs, and all scenes are covered in each panel.

Table C.7-4: Allocation of scenes for each talker pair (category cat 1 – cat 6) and listening panel (P1-P6)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Category** | **Overtalk**  **[s]** Note 1 | **1st talker elevation** | **2nd talker elevation** | **1st talker initial azimuth** | **1st talker azimuth change** see NOTE 2 | **2nd talker initial azimuth** | **2nd talker azimuth change** see NOTE 2 | **Panel** |
| **cat 1:**  M1 + F1 | -1  1  -1  1  -1  1 | 0°  35°  0°  0°  45°  30° | 0°  35°  45°  45°  45°  30° | 0°  10°  20°  200°  340°  120° | static  static  static  static  -1°/ frame  1°/ frame | 50°  110°  170°  30°  340°  120° | static  static  static  -1°/ frame  -1°/ frame  -1°/ frame | P1  P2  P3  P4  P5  P6 |
| **cat 2:**  M2 + F2 | 1  -1  1  -1  1  -1 | 35°  0°  0°  45°  30°  0° | 35°  45°  45°  45°  30°  0° | 20°  30°  250°  290°  180°  10° | static  static  static  -1°/ frame  1°/ frame  static | 170°  230°  340°  290°  180°  110° | static  static  -1°/ frame  -1°/ frame  -1°/ frame  static | P1  P2  P3  P4  P5  P6 |
| **cat 3:**  M3 + F3 | -1  1  -1  1  -1  1 | 0°  0°  45°  30°  0°  35° | 45°  45°  45°  30°  0°  35° | 40°  300°  180°  240°  20°  30° | static  static  1°/ frame  1°/ frame  static  static | 290°  290°  180°  240°  170°  230° | static  -1°/ frame  1°/ frame  1°/ frame  static  static | P1  P2  P3  P4  P5  P6 |
| **cat 4:**  F1 + M2 | 1  -1  1  -1  1  -1 | 0°  45°  30°  0°  35°  0° | 45°  45°  30°  0°  35°  45° | 50°  130°  300°  30°  40°  50° | static  1°/ frame  1°/ frame  static  static  static | 180°  130°  300°  230°  290°  350° | 1°/ frame  1°/ frame  -1°/ frame  static  static  static | P1  P2  P3  P4  P5  P6 |
| **cat 5:**  F2 + M3 | -1  1  -1  1  -1  1 | 45°  30°  0°  35°  0°  0° | 45°  30°  0°  35°  45°  45° | 80°  0°  40°  50°  0°  100° | 1°/ frame  1°/ frame  static  static  static  static | 80°  0°  290°  350°  50°  130° | 1°/ frame  -1°/ frame  static  static  static  1°/ frame | P1  P2  P3  P4  P5  P6 |
| **cat 6:**  F3 + M1 | 1  -1  1  -1  1  -1 | 30°  0°  35°  0°  0°  45° | 30°  0°  35°  45°  45°  45° | 60°  50°  0°  10°  150°  30° | 1°/ frame  static  static  static  static  -1°/ frame | 60°  350°  50°  110°  80°  30° | -1°/ frame  static  static  static  1°/ frame  -1°/ frame | P1  P2  P3  P4  P5  P6 |
| NOTE 1 Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.  NOTE 2 The positive sense for azimuth is counterclockwise | | | | | | | | |

C.8 Experiment P800-8: MASA Clean Speech Test

## C.8.1 Experiment setup

Stereo-MASA inputs are tested in IVAS selection phase. The stereo-MASA inputs can have 1-direction spatial metadata (based, e.g., on MASA analysis of original FOA content) or 2-direction spatial metadata (based, e.g., on MASA analysis of original HOA2 content).

Tables C.8-1 to C.8-3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table C.8-1: Conditions for Experiment P800-8 (Clean Speech)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64, 80 kbps |
| DTX | DTX off |
| Input level | -26 LKFS [31] |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0%, 5% |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS, Dual-mono EVS with unquantized metadata |
| Bitrates | 3x7.2, 4x7.2, 4x8, 4x9.6, 4x16.4, 4x24.4 kbps |
|  | 2x7.2, 2x8, 2x9.6, 2x16.4, 2x24.4 kbps |
| Input level  DTX | -26 LKFS [31]  DTX off |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0%, 5% |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS [31] |
| P.50 MNRU  ESDRU | Q = 20, 24, 28, 32 dB  *α* = 0.4, 0.6, 0.8 |
| Input frequency mask | HP50 |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Model-based generation according to convolution of raw mono clean speech sentences with FOA Spatial Room Impulse Responses corresponding to the talker positions relative to a capture point and spatial FOA background. MASA format generation from FOA according to MASA analysis |
| Binaural renderer | IVAS MASA C Reference Software binaural rendering |
| Audio sampling frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pair uttered by different talkers and genders (3 male and 3 female) |
| Number of categories | 6 Different environments and talker interactions |
| Number of samples | 6 + 1 (preliminaries) samples per category |
| Listening Level | 73 dB SPL |
| Listeners | Naïve listeners |
| Randomizations | 6 panels of 5 listeners |
| Rating Scale | Following clause 7.1.2.1 |
| Listening System | Headphones, in accordance with clause 7.1.4 |
| Listening Environment | No room noise |
| Languages | Danish, German |

Table C.8-2: Preliminaries for Experiment P800-8

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c23 | EVS | 4x9.6 | 5% |
| 2 | c02 | MNRU Q=32 dB | - | - |
| 3 | c06 | ESDRU = 0.8 | - | - |
| 4 | c13 | EVS | 4x16.4 | No errors |
| 5 | c08 | ESDRU = 0.4 | - | - |
| 6 | c21 | EVS | 4x7.2 | 5% |
| 7 | c04 | MNRU Q=24 dB | - | - |
| 8 | c01 | Reference | - | - |
| 9 | c14 | EVS | 4x24.4 | No errors |
| 10 | c07 | ESDRU = 0.6 | - | - |
| 11 | c05 | MNRU Q=20 dB | - | - |
| 12 | c22 | EVS | 4x8 | 5% |

Table C.8-3: Test conditions for Experiment P800-8,  
clean speech under clean and impaired channel conditions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **FER/Profile** | **ToR** |
| c01 | Reference | - | - |  |
| c02 | MNRU Q=32 dB | - | - |  |
| c03 | MNRU Q=28 dB | - | - |  |
| c04 | MNRU Q=24 dB | - | - |  |
| c05 | MNRU Q=20 dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU | - | - |  |
| c09 | EVS | 3x7.2 | No errors |  |
| c10 | EVS | 4x7.2 | No errors |  |
| c11 | EVS | 4x8 | No errors |  |
| c12 | EVS | 4x9.6 | No errors |  |
| c13 | EVS | 4x16.4 | No errors |  |
| c14 | EVS | 4x24.4 | No errors |  |
| c15 | EVS(+unquantized metadata) | 2x7.2 | No errors |  |
| c16 | EVS(+unquantized metadata) | 2x8 | No errors |  |
| c17 | EVS(+unquantized metadata) | 2x9.6 | No errors |  |
| c18 | EVS(+unquantized metadata) | 2x16.4 | No errors |  |
| c19 | EVS(+unquantized metadata) | 2x24.4 | No errors |  |
| c20 | EVS | 3x7.2 | 5% |  |
| c21 | EVS | 4x7.2 | 5% |  |
| c22 | EVS | 4x8 | 5% |  |
| c23 | EVS | 4x16.4 | 5% |  |
| c24 | EVS | 4x24.4 | 5% |  |
| c25 | CuT | 13.2 | No errors | NWT c09 OR NWT c15 |
| c26 | CuT | 16.4 | No errors | NWT c10 OR NWT c15 |
| c27 | CuT | 24.4 | No errors | NWT c11 OR NWT c16 |
| c28 | CuT | 32 | No errors | NWT c12 OR NWT c17 |
| c29 | CuT | 48 | No errors | NWT c13 OR NWT c18 |
| c30 | CuT | 64 | No errors | NWT c14 OR NWT c19 |
| c31 | CuT | 80 | No errors | NWT c14 OR NWT c19 |
| c32 | CuT | 13.2 | 5% | NWT c20 |
| c33 | CuT | 16.4 | 5% | NWT c21 |
| c34 | CuT | 24.4 | 5% | NWT c22 |
| c35 | CuT | 48 | 5% | NWT c23 |
| C36 | CuT | 64 | 5% | NWT c24 |

## C.8.2 Content type categories and scene definitions (Exp P800-8: Clean speech)

Table C.8-4 Content type categories and scene definitions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Category** | **Environment** | **Background** | **SNR [dB]** | **Overtalk [s]see NOTE 1** | **Bandwidth** | **Talker positions** | **Talker selection by panel see NOTE 2** |
| cat 1 | room\_1\_MASA | room\_1\_cleanbg\_MASA | 45 | 1 | Max |  | P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3 |
| cat 2 | room\_4\_MASA | room\_4\_cleanbg\_MASA | 45 | -1 | Max |  | P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2 |
| cat 3 | room\_2\_MASA | room\_2\_cleanbg\_MASA | 45 | 1 | Max |  | P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1 |
| cat 4 | room\_5\_MASA | room\_5\_cleanbg\_MASA | 45 | -1 | Max |  | P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3 |
| cat 5 | room\_3\_MASA | room\_3\_cleanbg\_MASA | 45 | 1 | Max |  | P1: f3m3 P2: m1f1 P3: f2m2 P4: m3f3 P5: f1m1 P6: m2f2 |
| cat 6 | room\_6\_MASA | room\_6\_cleanbg\_MASA | 45 | -1 | Max |  | P1: m2f2 P2: f3m3 P3: m1f1 P4: f2m2 P5: m3f3 P6: f1m1 |
| NOTE 1 Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.  NOTE 2 All sentences by the 6 talkers shall be unique. | | | | | | | |

# C.9 Experiment P800-9: MASA Speech+Background Test

## C.9.1 Experiment setup

Stereo-MASA inputs are tested in IVAS selection phase. The stereo-MASA inputs can have 1-direction spatial metadata (based, e.g., on MASA analysis of original FOA content) or 2-direction spatial metadata (based, e.g., on MASA analysis of original HOA2 content).

Tables C.9-1 to C.9-3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table C.9-1: Conditions for Experiment P800- 9 (Speech+Background)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64, 80 |
| DTX | DTX ON and OFF |
| Input level | -26 LKFS [31] |
| Input frequency mask | HP50 |
| Noise | 10, 15dB |
| Error Conditions | No errors |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS + Dual-mono with unquantized metadata |
| Bitrates | 3x7.2, 4x7.2, 4x8, 4x9.6, 4x16.4, 4x24.4 kbps |
|  | 2x7.2, 2x9.6, 2x16.4 kbps |
| Input level  DTX | -26 LKFS [31]  DTX On and OFF |
| Input frequency mask | HP50 |
| Noise | 10, 15dB |
| Error Conditions | No errors |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS [31] |
| P.50 MNRU  ESDRU | Q= 17, 21, 25, 29 dB  *α* = 0.4, 0.6, 0.8 |
| Input frequency mask | HP50 |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Model-based generation according to convolution of raw mono clean speech sentences with FOA Spatial Room Impulse Responses corresponding to the talker positions relative to a capture point and spatial FOA background. MASA format generation from FOA according to MASA analysis |
| Binaural renderer | IVAS MASA C Reference Software binaural rendering |
| Audio sampling frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pair uttered by different talkers and genders (3 male and 3 female) |
| Number of categories | 6 Different environments and talker interactions |
| Number of samples | 6 + 1 (preliminaries) samples per category |
| Listening Level | 73 dB SPL |
| Listeners | Naïve listeners |
| Randomizations | 6 panels of 5 listeners |
| Rating Scale | Following clause 7.1.2.1 |
| Listening System | Headphones, in accordance with clause 7.1.4 |
| Listening Environment | No room noise |
| Languages | Japanese, French |

Table C.9-2: Preliminaries for Experiment P800-9

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **DTX** |
| 1 | c21 |  | EVS | 4x9.6 | on |
| 2 | c02 |  | MNRU Q=29 dB | - | - |
| 3 | c06 |  | ESDRU = 0.8 | - | - |
| 4 | c13 |  | EVS | 4x16.4 | off |
| 5 | c08 |  | ESDRU = 0.4 | - | - |
| 6 | c19 |  | EVS | 4x7.2 | on |
| 7 | c04 |  | MNRU Q=21 dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c14 |  | EVS | 4x24.4 | off |
| 10 | c07 |  | ESDRU = 0.6 | - | - |
| 11 | c05 |  | MNRU Q=17 dB | - | - |
| 12 | c20 |  | EVS | 4x8 | on |

Table C.9-3: Test conditions for Experiment P800-9,  
speech and background under clean channel conditions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **ToR** |
| c01 | Reference | - | - |  |
| c02 | MNRU Q=29 dB | - | - |  |
| c03 | MNRU Q=25 dB | - | - |  |
| c04 | MNRU Q=21 dB | - | - |  |
| c05 | MNRU Q=17 dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU | - | - |  |
| c09 | EVS | 3x7.2 | off |  |
| c10 | EVS | 4x7.2 | off |  |
| c11 | EVS | 4x8 | off |  |
| c12 | EVS | 4x9.6 | off |  |
| c13 | EVS | 4x16.4 | off |  |
| c14 | EVS | 4x24.4 | off |  |
| c15 | EVS(+unquantized metadata) | 2x7.2 | on |  |
| c16 | EVS(+unquantized metadata) | 2x9.6 | on |  |
| c17 | EVS(+unquantized metadata) | 2x16.4 | on |  |
| c18 | EVS | 3x7.2 | on |  |
| c19 | EVS | 4x7.2 | on |  |
| c20 | EVS | 4x8 | on |  |
| c21 | EVS | 4x9.6 | on |  |
| c22 | EVS | 4x16.4 | on |  |
| c23 | EVS | 4x24.4 | on |  |
| c24 | CuT | 13.2 | off | NWT c09 | |
| c25 | CuT | 16.4 | off | NWT c10 | |
| c26 | CuT | 24.4 | off | NWT c11 | |
| c27 | CuT | 32 | off | NWT c12 | |
| c28 | CuT | 48 | off | NWT c13 | |
| c29 | CuT | 64 | off | NWT c14 | |
| c30 | CuT | 80 | off | NWT c14 | |
| c31 | CuT | 13.2 | on | NWT c18 or NWT c15 | |
| c32 | CuT | 16.4 | on | NWT c19 or NWT c15 | |
| c33 | CuT | 24.4 | on | NWT c20 | |
| c34 | CuT | 32 | on | NWT c21 or NWT c16 | |
| c35 | CuT | 48 | on | NWT c22 or NWT c17 | |
| c36 | CuT | 64 | on | NWT c23 | |

## C.9.2 Content type categories and scene definitions (Exp P800-9: Speech+Background)

Table C.9-4: Content type categories and scene definitions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Category** | **Environment(1** | **Background(2** | **SNR [dB]** | **Overtalk [s](3** | **Bandwidth** | **Talker positions(4** | **Talker selection by panel(5** |
| cat 1 | car\_1\_MASA | car\_1\_bg\_MASA | 10 | -1 | Max |  | P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3 |
| cat 2 | car\_[1/2]\_MASA | car\_[1/2]\_bg\_MASA | 10 | -1 | Max |  | P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2 |
| cat 3 | out\_1\_MASA | street\_1\_bg\_MASA | 15 | -1 | Max |  | P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1 |
| cat 4 | out\_[1/2]\_MASA | [park\_1\_bg\_MASA / nature\_1\_bg\_MASA / event\_1\_bg\_MASA / street\_[1/2]\_bg\_MASA] | 15 | -1 | Max |  | P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3 |
| cat 5 | room\_[X]\_MASA | office\_1\_bg\_MASA | 15 | -1 | Max |  | P1: f3m3 P2: m1f1 P3: f2m2 P4: m3f3 P5: f1m1 P6: m2f2 |
| cat 6 | room\_[X]\_MASA | [cafeteria\_1\_bg\_MASA / mall\_1\_bg\_MASA/ office[1/2]\_bg\_MASA] | 15 | -1 | Max |  | P1: m2f2 P2: f3m3 P3: m1f1 P4: f2m2 P5: m3f3 P6: f1m1 |
| NOTE 1 Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.  NOTE 2 All sentences by the 6 talkers shall be unique. | | | | | | | |

# C.10 Experiment BS1534-1a: Stereo

Table C.10-1: Conditions (BS1534-1a Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated with stereo audio input at  48 and 64 kbps DTX off at 0% FER |
|  |  |
| **Codec references** |  |
| Codec references | Dual-mono EVS  2\*24.4 kbps, 2\*32 kbps, 2\*48 kbps DTX off at 0% FER |
|  |  |
| **Other references** |  |
| Reference | Direct signal, Nominal input level |
| Hidden Reference | Direct signal, Nominal input level |
| LP7 anchor | 7 kHz lowpass filtered signal, nominal level |
|  |  |
| **Common Conditions** |  |
| Test item generation | According to material collection procedure for IVAS selection BS.1534 tests. |
| Audio sampling frequency/bandwidth | 48 kHz/FB |
| Input frequency mask | 20KBP |
| Nominal output loudness | -26 LKFS [31] |
| Listening Level | Adjusted by listener |
| Listeners | Experienced Listeners |
| Randomizations | Individual per listeners |
| Rating Scale | Continuous BS.1534 scale from 0-100 |
| Listening System | High-quality headphone for diotic presentation, in accordance with clause 7.1.4 |
| Listening Environment | No room noise |

Table C.10-2: Test conditions for Experiment BS1534-1a

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 2x24.4 | - |
| c04 | EVS | 2x32 | - |
| c05 | EVS | 2x48 | - |
| c06 | CuT | 48 | NWT c04 OR BT c03 |
| c07 | CuT | 64 | NWT c05 OR BT c04 |

# C.11 Experiment BS1534-1b: Stereo

Table C.11-1: Conditions (BS1534-1b Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated with stereo audio input at  96 and 128 kbps DTX off at 0% FER |
|  |  |
| **Codec references** |  |
| Codec references | Dual-mono EVS  2\*48 kbps, 2\*64 kbps, 2\*96 kbps DTX off at 0% FER |
|  |  |
| **Other references** |  |
| Reference | Direct signal, Nominal input level |
| Hidden Reference | Direct signal, Nominal input level |
| LP7 anchor | 7 kHz lowpass filtered signal, nominal level |
|  |  |
| **Common Conditions** |  |
| Test item generation | According to material collection procedure for IVAS selection BS.1534 tests. |
| Audio sampling frequency/bandwidth | 48 kHz/FB |
| Input frequency mask | 20KBP |
| Nominal output loudness | -26 LKFS [31] |
| Listening Level | Adjusted by listener |
| Listeners | Experienced Listeners |
| Randomizations | Individual per listeners |
| Rating Scale | Continuous BS.1534 scale from 0-100 |
| Listening System | High-quality headphone for diotic presentation, in accordance with clause 7.1.4 |
| Listening Environment | No room noise |

Table C.11-2: Test conditions for Experiment BS1534-1b

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 2x48 | - |
| c04 | EVS | 2x64 | - |
| c05 | EVS | 2x96 | - |
| c06 | CuT | 96 | NWT c04 OR BT c03 |
| c07 | CuT | 128 | NWT c05 OR BT c04 |

# C.12 Experiment BS1534-2a: Multi-channel 5.1

Table C.12-1: Conditions (BS1534-2a Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated at 64, 96 kbps at 0% FER |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS operated at 5\*13.2, 5\*16.4, 5\*24.4 kbps with the LFE channel processed using EVS operated at 9.6 kbps NB (for all Codec references) at 0% FER |
|  |  |
| **Other references** |  |
| Reference / Direct | Direct 5.1 signal, nominal level |
| Hidden Reference | Direct 5.1 signal, nominal level |
| LP7 anchor | 7 kHz lowpass filtered direct 5.1 signal, nominal level |
|  |  |
| **Common Conditions** |  |
| Test item generation | According to material collection procedure for IVAS selection *BS.1534* tests. |
| Loudspeaker Rendering | 5.1 channels direct playback |
| Audio sampling frequency / bandwidth | 48 kHz / maximum available audio bandwidth (SWB, FB) |
| Input frequency mask | *20KBP* |
| Nominal output loudness | -26 LKFS [31] |
| Listening Level | Adjusted by listener |
| Listeners | Experienced Listeners |
| Randomizations | Individual per listener |
| Rating Scale | Continuous BS.1534 scale: 0-100 |
| Listening System | 5.1 high-quality loudspeaker setup following IVAS-7a |
| Listening Environment | No room noise |

Table C.12-2: Test conditions for Experiment BS1534-2a

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 5x13.2 | - |
| c04 | EVS | 5x16.4 | - |
| c05 | EVS | 5x24.4 | - |
| c06 | CuT | 64 | NWT c04 OR BT c03 |
| c07 | CuT | 96 | NWT c05 OR BT c04 |

# C.13 Experiment BS1534-2b: Multi-channel 5.1

Table C.13-1: Conditions (BS1534-2b Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated at 128, 160 kbps at 0% FER |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS operated at 5\*24.4, 5\*32, 5\*48 kbps with the LFE channel processed using EVS operated at 9.6 kbps NB (for all Codec references) at 0% FER |
|  |  |
| **Other references** |  |
| Reference | Direct 5.1 signal, nominal input level |
| Hidden Reference | Direct 5.1 signal, nominal input level |
| LP7 anchor | 7 kHz lowpass filtered direct 5.1 signal, nominal level |
|  |  |
| **Common Conditions** | *(see BS1534-2a for full list)* |
| Audio sampling frequency / bandwidth | 48 kHz / FB |

Table C.13-2: Test conditions for Experiment BS1534-2b

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 5x24.4 | - |
| c04 | EVS | 5x32 | - |
| c05 | EVS | 5x48 | - |
| c06 | CuT | 128 | NWT c04 OR BT c03 |
| c07 | CuT | 160 | NWT c05 OR BT c04 |

# C.14 Experiment BS1534-3a: Multi-channel 7.1+4

Table C.14-1: Conditions (BS1534-3a Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated with multi-channel 7.1+4 input at  128 and 160 kbps DTX off at 0% FER |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS coding multi-channel 7.1+4 input at  11\*9.6 kbps, 11\*13.2 kbps, 11\*16.4 kbps DTX off at 0% FER  LFE coded with 9.6kbps NB (IVAS-3) |
|  |  |
| **Other references** |  |
| Reference | Direct signal, Nominal input level |
| Hidden Reference | Direct signal, Nominal input level |
| LP7 anchor | 7 kHz lowpass filtered signal, nominal level |
|  |  |
| **Common Conditions** |  |
| Test item generation | According to material collection procedure for IVAS selection BS.1534 tests. |
| Audio sampling frequency/bandwidth | 48 kHz/FB |
| Input frequency mask | 20KBP |
| Nominal output loudness | -26 LKFS [31] |
| Listening Level | Adjusted by listener |
| Listeners | Experienced Listeners |
| Randomizations | Individual per listeners |
| Rating Scale | Continuous BS.1534 scale from 0-100 |
| Listening System | High-quality loudspeaker: 7.1+4 overhead speaker setup with the configuration following IVAS-7a |
| Listening Environment | No room noise |

Table C.14-2: Test conditions for Experiment BS1534-3a

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 11x9.6 | - |
| c04 | EVS | 11x13.2 | - |
| c05 | EVS | 11x16.4 | - |
| c06 | CuT | 128 | NWT c04 OR BT c03 |
| c07 | CuT | 160 | NWT c05 OR BT c04 |

# C.15 Experiment BS1534-3b: Multi-channel 7.1+4

Table C.15-1: Conditions (BS1534-3b Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated with multi-channel 7.1+4 audio input at  384 and 512 kbps DTX off at 0% FER |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS coding multi-channel 7.1+4 input at  11\*32 kbps, 11\*48 kbps, 11\*64 kbps DTX off at 0% FER  LFE coded with 9.6kbps NB (IVAS-3) |
|  |  |
| **Other references** |  |
| Reference | Direct signal, Nominal input level |
| Hidden Reference | Direct signal, Nominal input level |
| LP7 anchor | 7 kHz lowpass filtered signal, nominal level |
|  |  |
| **Common Conditions** |  |
| Test item generation | According to material collection procedure for IVAS selection BS.1534 tests. |
| Audio sampling frequency/bandwidth | 48 kHz/FB |
| Input frequency mask | 20KBP |
| Nominal output loudness | -26 LKFS [31] |
| Listening Level | Adjusted by listener |
| Listeners | Experienced Listeners |
| Randomizations | Individual per listeners |
| Rating Scale | Continuous BS.1534 scale from 0-100 |
| Listening System | High-quality loudspeaker: 7.1+4 overhead speaker setup with the configuration following IVAS-7a |
| Listening Environment | No room noise |

Table C.15-2: Test conditions for Experiment BS1534-3b

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 11x32 | - |
| c04 | EVS | 11x48 | - |
| c05 | EVS | 11x64 | - |
| c06 | CuT | 384 | NWT c04 OR BT c03 |
| c07 | CuT | 512 | NWT c05 OR BT c04 |

# C.16 Experiment BS1534-4a: FOA

Table C.16-1: Conditions (BS1534-4a, Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated with audio input truncated to FOA, HOA3 output, at 96 kbps, 128 kbps, 160 kbps |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS operated with audio input truncated to FOA at  4\*24.4 kbps, 4\*32 kbps, 4\*48 kbps |
|  |  |
| **Other references** |  |
| Reference | Direct rendering of HOA3 signal, Nominal input level |
| Hidden Reference | Direct rendering of HOA3 signal, Nominal input level |
| LP7 anchor | 7 kHz lowpass filtered direct rendered HOA3 signal: nominal level |
|  |  |
| **Common Conditions** |  |
| Test item generation | According to material collection procedure for IVAS selection BS.1534 tests. |
| **Binaural renderer** | Ambisonics to binaural (external) rendering |
| Audio sampling frequency/bandwidth | 48 kHz/FB |
| Input frequency mask | 20KBP |
| Nominal output loudness | -26 LKFS [31] |
| Listening Level | Adjusted by listener |
| Listeners | Experienced Listeners |
| Randomizations | Individual per listeners |
| Rating Scale | Continuous BS.1534 scale from 0-100 |
| Listening System | High-quality headphone for diotic presentation, in accordance with 7.1.4 |
| Listening Environment | No room noise |

Table C.16-2: Test conditions for Experiment BS1534-4a

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 4x24.4 | - |
| c04 | EVS | 4x32 | - |
| c05 | EVS | 4x48 | - |
| c06 | CuT | 96 | NWT c04 OR BT c03 |
| c07 | CuT | 128 | NWT c05 OR BT c04 |
| c08 | CuT | 160 | NWT c05 OR BT c04 |

# C.17 Experiment BS1534-4b: HOA2

Table C.17-1: Conditions (BS1534-4b, Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated with audio input truncated to HOA2, HOA3 output, at 160, 192 kbps kbps. |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS operated with audio input truncated to FOA at  4\*32 kbps, 4\*48 kbps, 4\*64 kbps. |
|  |  |
| **Other references** |  |
| Reference | Direct rendering of HOA3 signal, Nominal input level |
| Hidden Reference | Direct rendering of HOA3 signal, Nominal input level |
| LP7 anchor | 7 kHz lowpass filtered direct rendered HOA3 signal: nominal level |
|  |  |
| **Common Conditions** |  |
| Test item generation | According to material collection procedure for IVAS selection BS.1534 tests. |
| Binaural renderer | Ambisonics to binaural (external) rendering |
| Audio sampling frequency/bandwidth | 48 kHz/FB |
| Input frequency mask | 20KBP |
| Nominal output loudness | -26 LKFS [31] |
| Listening Level | Adjusted by listener |
| Listeners | Experienced Listeners |
| Randomizations | Individual per listeners |
| Rating Scale | Continuous BS.1534 scale from 0-100 |
| Listening System | High-quality headphone for diotic presentation, in accordance with clause 7.1.4 |
| Listening Environment | No room noise |

Table C.17-2: Test conditions for Experiment BS1534-4b

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 4x32 | - |
| c04 | EVS | 4x48 | - |
| c05 | EVS | 4x64 | - |
| c06 | CuT | 160 | NWT c04 OR BT c03 |
| c07 | CuT | 192 | NWT c05 OR BT c04 |

# C.18 Experiment BS1534-5a: HOA3

Table C.18-1: Conditions (BS1534-5a, Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated with HOA3 input at  192 kbps, 256 kbps. |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS operated with audio input truncated to FOA at  4\*48 kbps, 4\*64, 4\*96 kbps. |
|  |  |
| **Other references** |  |
| Reference | Direct rendering of HOA3 signal, Nominal input level |
| Hidden Reference | Direct rendering of HOA3 signal, Nominal input level |
| LP7 anchor | 7 kHz lowpass filtered direct rendered HOA3 signal: nominal level |
|  |  |
| **Common Conditions** |  |
| Test item generation | According to material collection procedure for IVAS selection BS.1534 tests. |
| Binaural renderer | Ambisonics to binaural (external) rendering |
| Audio sampling frequency/bandwidth | 48 kHz/FB |
| Input frequency mask | 20KBP |
| Nominal output loudness | -26 LKFS [31] |
| Listening Level | Adjusted by listener |
| Listeners | Experienced Listeners |
| Randomizations | Individual per listeners |
| Rating Scale | Continuous BS.1534 scale from 0-100 |
| Listening System | High-quality headphone for diotic presentation, in accordance with clause 7.1.4. |
| Listening Environment | No room noise |

Table C.18-2: Test conditions for Experiment BS1534-5a

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 4x48 | - |
| c04 | EVS | 4x64 | - |
| c05 | EVS | 4x96 | - |
| c06 | CuT | 192 | NWT c04 OR BT c03 |
| c07 | CuT | 256 | NWT c05 OR BT c04 |

# C.19 Experiment BS1534-5b: HOA3

Table C.19-1: Conditions (BS1534-5b, Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated with HOA3 input at  384 kbps, 512 kbps. |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS operated with audio input truncated to FOA at  4\*96 kbps, 4\*128 kbps. |
|  |  |
| **Other references** |  |
| Reference | Direct rendering of HOA3 signal, Nominal input level |
| Hidden Reference | Direct rendering of HOA3 signal, Nominal input level |
| LP7 anchor | 7 kHz lowpass filtered direct rendered HOA3 signal: nominal level |
|  |  |
| **Common Conditions** |  |
| Test item generation | According to material collection procedure for IVAS selection BS.1534 tests. |
| Renderer | Ambisonics to loudspeaker (external) rendering |
| Audio sampling frequency/bandwidth | 48 kHz/FB |
| Input frequency mask | 20KBP |
| Nominal output loudness | -26 LKFS [31] |
| Listening Level | Adjusted by listener |
| Listeners | Experienced Listeners |
| Randomizations | Individual per listeners |
| Rating Scale | Continuous BS.1534 scale from 0-100 |
| Listening System | Calibrated and conformant 7.1.4 listening room with single listener in center. |
| Listening Environment | No room noise |

Table C.19-2: Test conditions for Experiment BS1534-5b

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 4x96 | - |
| c04 | EVS | 4x128 | - |
| c05 | CuT | 384 | NWT c04 OR BT c03 |
| c06 | CuT | 512 | NWT c04 |

# C.20 Experiment BS1534-6a: Objects

Test with 3 Objects in BS1534-6a at mid bitrates.

Table C.20-1: Conditions (BS1534-6a Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated at 48, 64, 96 kbps |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS operated at 3\*16.4, 3\*24.4, 3\*32 kbps |
|  |  |
| **Other references** |  |
| Reference / Direct | Direct Object signal(s), nominal input level |
| Hidden Reference | Direct Object signal(s), nominal input level |
| LP7 anchor | 7 kHz lowpass filtered direct Object signal(s), nominal level |
|  |  |
| **Common Conditions** |  |
| Test item generation | According to material collection procedure for IVAS selection *BS.1534* tests. |
| Binaural rendering | Objects to binaural (external) rendering |
| Audio sampling frequency / bandwidth | 48 kHz / FB |
| Input frequency mask | 20KBP |
| Nominal output loudness | -26 LKFS [31] |
| Listening Level | Adjusted by listener |
| Listeners | Experienced Listeners |
| Randomizations | Individual per listener |
| Rating Scale | Continuous BS.1534 scale: 0-100 |
| Listening System | High-quality headphones, diotic presentation, in accordance with clause 7.1.4 |
| Listening Environment | No noise |

Table C.20-2: Test conditions for Experiment BS1534-6a

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 3x16.4 | - |
| c04 | EVS | 3x24.4 | - |
| c05 | EVS | 3x32 | - |
| c06 | CuT | 48 | NWT c03 |
| c07 | CuT | 64 | NWT c04 |
| c08 | CuT | 96 | NWT c05 |

# C.21 Experiment BS1534-6b: Objects

Test with 4 Objects in BS1534-6b at mid/high bitrates.

Table C.21-1: Conditions (BS1534-6b Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated at 96, 128, 256 kbps with DTX off at 0% FER |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS operated at 4\*24.4, 4\*32, 4\*64 kbps with DTX off at 0% FER |
|  |  |
| **Other references** |  |
| Reference / Direct | Direct Object signal(s), nominal input level |
| Hidden Reference | Direct Object signal(s), nominal input level |
| LP7 anchor | 7 kHz lowpass filtered direct Object signal(s), nominal level |
|  |  |
| **Common Conditions** | *(see BS1534-6a for full list)* |

Table C.21-2: Test conditions for Experiment BS1534-6b

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 4x24.4 | - |
| c04 | EVS | 4x32 | - |
| c05 | EVS | 4x64 | - |
| c06 | CuT | 96 | NWT c03 |
| c07 | CuT | 128 | NWT c04 |
| c08 | CuT | 256 | NWT c05 |

# C.22 Experiment BS1534-7a: MASA

Table C.22-1: Conditions (BS1534-7a Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated at 96, 128 kbps |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS operated at 4\*32, 4\*48 kbps with DTX off at 0% FER |
| Multi-mono EVS with unquantized metadata operated at 2\*32, 2\*48 kbps with DTX off at 0% FER |
| **Other references** |  |
| Reference / Direct | Direct MASA signal(s), nominal input level |
| Hidden Reference | Direct MASA signal(s), nominal input level |
| LP7 anchor | 7 kHz lowpass filtered direct MASA signal(s), nominal level |
|  |  |
| **Common Conditions** |  |
| Test item generation | According to material collection procedure for IVAS selection *BS.1534* tests |
| Binaural rendering | MASA to binaural (external) rendering |
| Audio sampling frequency / bandwidth | 48 kHz / FB |
| Input frequency mask | 20KBP |
| Nominal output loudness | -26 LKFS [31] |
| Listening Level | Adjusted by listener |
| Listeners | Experienced Listeners |
| Randomizations | Individual per listener |
| Rating Scale | Continuous BS.1534 scale: 0-100 |
| Listening System | High-quality headphones, diotic presentation, in accordance with clause 7.1.4 |
| Listening Environment | No noise |

Table C.22-2: Test conditions for Experiment BS1534-7a

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 4x32 | - |
| c04 | EVS | 4x48 | - |
| c05 | EVS | 2x32, unquantized MD | - |
| c06 | EVS | 2x48, unquantized MD | - |
| c07 | CuT | 96 | NWT c03 OR NWT c05 |
| c08 | CuT | 128 | NWT c04 OR NWT c06 |

# C.23 Experiment BS1534-7b: MASA

Table C.23-1: Conditions (BS1534-7b Generic Audio)

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Codec under Test (CuT) | IVAS candidate operated at 192, 256 kbps with DTX off at 0% FER |
|  |  |
| **Codec references** |  |
| Codec references | Multi-mono EVS operated at 4\*64, 4\*96 kbps with DTX off at 0% FER |
| Multi-mono EVS with unquantized metadata operated at 2\*64, 2\*96 kbps with DTX off at 0% FER |
|  |  |
| **Other references** |  |
| Reference / Direct | Direct MASA signal(s), nominal input level |
| Hidden Reference | Direct MASA signal(s), nominal input level |
| LP7 anchor | 7 kHz lowpass filtered direct MASA signal(s), nominal level |
|  |  |
| **Common Conditions** | *(see BS1534-7a for full list)* |

Table C.23-2: Test conditions for Experiment BS1534-7b

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 | EVS | 4x64 | - |
| c04 | EVS | 4x96 | - |
| c05 | EVS | 2x64, unquantized MD | - |
| c06 | EVS | 2x96, unquantized MD | - |
| c07 | CuT | 192 | NWT c03 OR NWT c05 |
| c08 | CuT | 256 | NWT c04 OR NWT c06 |

Annex D:  
IVAS Characterization Phase Testplan

NOTE: The IVAS Characterization Phase Testplan will be added after completion of the characterization phase.

Annex E:   
IVAS Permanent Documents in 3GPP FTP-site

The standardization of the IVAS codec is described in a series of permanent project documents. They contain the most important guidelines, rules and decisions. The following permanent project documents can be found in a specific location on the 3GPP FTP site:

Table E.1: IVAS Permanent Project Documents

|  |  |  |
| --- | --- | --- |
| SA4 TDoc number | P-doc | Title |
| S4-231032 | IVAS-1 | IVAS Codec Development Overview |
|  | IVAS-2 | IVAS Project Plan |
| S4-231049 | IVAS-3 | IVAS Performance Requirements |
| S4-231031 | IVAS-4 | IVAS design constraints |
| S4-231086 | IVAS-5 | Selection Rules |
| S4-231087 | IVAS-6 | Selection Deliverables |
| S4-231306 | IVAS-7a | Processing plan for selection phase |
|  | IVAS-7b | Processing plan for characterization phase |
| S4-231115 | IVAS-8a | Test Plan for Selection Phase |
|  | IVAS-8b | Test plan for characterization phase |
| S4-231523 | IVAS-9 | IVAS Usage Scenarios |

The latest version of these documents can be found in the following link:  
https://www.3gpp.org/ftp/TSG\_SA/WG4\_CODEC/IVAS\_Permanent\_Documents

Annex F:  
Attachments

Attachments to the present document include:

1) 3GPP S4-231573: "Global Analysis Laboratory report – IVAS Selection Phase", including Excel sheets containing the IVAS Selection Phase Test Results

Annex G (informative):  
Change history

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Change history | | | | | | | |
| Date | Meeting | TDoc | CR | Rev | Cat | Subject/Comment | New version |
| 2024-04 | SA4#127-bis-e | S4-240626 |  |  |  | Initial Version | v0.0.1 |
| 2024-04 | SA4#127-bis-e | S4-240792 |  |  |  | Endorsed by SA4 | v0.1.0 |
| 2024-05 | S4#128 | S4-240982 |  |  |  | New input by rapporteurs, lots of added text | V0.1.1 |
| 2024-05 | S4#128 | S4-241245 |  |  |  | Submitted to SA4 for agreement | V0.2.0 |