**Source: IVAS-8b Editor\***

**Title: IVAS Permanent Document IVAS-8b: Test Plan for Characterization Phase**

**Version: v.0.2.0**

**Agenda Item: 7.5**

Document History:

|  |  |  |
| --- | --- | --- |
| v.0.0.1 | 13 November 2023 | Initial draft |
| v.0.1.0 | 17 November 2023 | Integration of S4-231857 and comments received from the Audio SWG, editorial changes. |
| v.0.1.1 | 15 December 2023 | Editorial changes |
| v.0.1.2 | 15 January 2024 | Integration of agreed text from S4aA230122, S4aA230123 and S4aA230124. Editorial changes. |
| v.0.2.0 | 2 February 2024 | Integration, in brackets, experiment proposals from S4aA230135, S4-240262, and S4-240298. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

1. Introduction

This document contains the Test Plan for the Characterization Phase of the IVAS Codec.

Editor’s note: In general, the IVAS Fixed-point code is understood as the main IVAS Codec CuT, whereas the IVAS floating-point code is considered as Reference. However, for some configurations, the IVAS floating-point code might also play the role of CuT. This is still open.

1. References, Conventions, and Contacts
   1. Permanent Documents

The following documents provide additional information on the IVAS codec development project.

|  |  |
| --- | --- |
| P-doc | Title |
| IVAS-1 | IVAS Codec Development Overview |
| IVAS-2 | IVAS Project Plan |
| IVAS-3 | IVAS Performance Requirements |
| IVAS-4 | IVAS Design Constraints |
| IVAS-5 | Selection Rules for Selection Phase |
| IVAS-6 | Deliverables for Selection Phase |
| IVAS-7a | Processing Plan for Selection Phase |
| IVAS-7b | Processing Plan for Characterization Phase |
| IVAS-8a | Test Plan for Selection Phase |
| IVAS-8b | Test Plan for Characterization Phase |
| 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>

* 1. Reference Documents

1. S4-211523: MESAQIN.com and FORCE Technology - SenseLab expression of interest to participate in IVAS codec selection and characterization work.
2. S4-220152: Interest in participation in IVAS codec selection and characterisation phase.
3. Recommendation ITU-R BS.2051-3 (05/2022): Advanced sound system for programme production.
4. S4-211155: On IVAS example test designs, Source: Nokia Corporation.
5. S4-210848: IVAS MASA spatial speech quality evaluation, Source: Nokia Corporation.
6. S4-191167: Description of the IVAS MASA C Reference Software, Source: Nokia Corporation.
7. S4-210840: Updates to IVAS MASA C Reference Software, Source: Nokia Corporation.
8. Recommendation ITU-T P.800 (08/1996): Methods for subjective determination of transmission quality.
9. Recommendation ITU-T P.811 (01/2019): Subjective test methodology for evaluating Speech oriented stereo communication systems over headphones.
10. S4-211151: Example designs for IVAS codec tests, Source: Dolby Laboratories Inc.
11. S4-210836: On reference designs for IVAS codec tests, Source: Dolby Laboratories Inc.
12. Recommendation ITU-R BS.1770-4 (10/2015): Algorithms to measure audio programme loudness and true-peak audio level.
13. ITU-T Handbook of subjective testing practical procedures, 2011.
14. S4-200158: A Reference Audio Renderer for Qualification, Source: Dolby Laboratories Inc.
15. S4-211160: Experience of P.800 for stereo testing, Source: Ericsson LM.
16. S4-130155: EVS Permanent Document EVS-7a: Processing functions for qualification phase.
17. AFsp Programs and Routines: http://www-mmsp.ece.mcgill.ca/Documents/Software/Packages/AFsp/audio/html/AFsp.html.
18. S4aA220005: On reference designs for IVAS codec tests - Update, Source: Dolby Laboratories Inc.
19. S4aA220007 - DCR test experiments for FOA and HOA3 input in 7.0+4 and binaural listening setup.
20. F. Zotter and M. Frank, “All-Round Ambisonic Panning and Decoding,” in J. Audio Eng. Soc., Vol. 60, No. 10, 2012.
21. T22-SG12-220607-TD-GEN-0138!!MSW-E: Draft new ITU-T P.800-series – Supplement P.SUPPL800: ITU-T Rec. P.800 use case examples.
22. Recommendation ITU-R BS.1534 (10/2015): Method for the subjective assessment of intermediate quality level of audio systems.
23. 3GPP TR 26.952: Codec for Enhanced Voice Services (EVS); Performance characterization.
24. S4-030821: PSS/MMS High-Rate Audio Selection Test and Processing Plan, Version 2.2.
25. Audio File Format Specifications: WAVE, <https://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/WAVE.html>.
26. AFsp Package <https://www-mmsp.ece.mcgill.ca/Documents/Downloads/AFsp/>.
27. Recommendation ITU-T P.191 (03/2023): Software tools for speech and audio coding standardization.
28. S4-230221: Processing updates for IVAS MASA C Reference Software.
29. 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
    1. Key Acronyms

20KPB 20-20k Hz Flat band-pass FIR filter

BIT Beijing Institute of Technology

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 Full Band

FE Frame Erasure

FOA First-Order Ambisonics

GAL Global Analysis Laboratory

HL Host Laboratory

HOA3 Higher-Order Ambisonics, 3rd order

HP50 50 Hz high-pass FIR filter

IVAS Immersive Voice and Audio Services

LKFS Loudness, K-weighted, relative to Full Scale

LL Listening Laboratory

MASA Metadata-Assisted Spatial Audio

MC Material Collection entity

MNRU Modulated Noise Reference Unit

MUSHRA Multi Stimulus test with Hidden Reference and Anchor

PC Proponent Company

SDRU Spatial Distortion Reference Unit

SNR Signal-to-Noise Ratio

SPL Sound Pressure Level

SWB Super Wide Band

WB Wide Band

Editor’s note: The References and Acronyms need to be reviewed when the document gets more stable.

1. Roles and Responsibilities
   1. Overview of the Characterization Test Process

The execution of the IVAS codec Characterization subjective testing is under the responsibility of the LLs participating in the Characterization Phase.

[The execution of the IVAS codec Characterization objective testing is under the responsibility of the PC participating in the Characterization Phase.]

Editor’s note: Whether objective testing shall be part of the Characterization Phase Test Plan is still open.

SA4 shall select both the external LLs and the volunteering LLs (SA4 companies) to perform the subjective listening tests described in this document. ETSI will contract the external laboratories. SA4 shall select the languages used in each experiment conducted by each LL. SA4 shall further select the HL, the CL, the MC, and the GAL to perform respective tasks defined in this document, and ETSI will contract the GAL.

The LLs and volunteering contributors (SA4 companies) shall provide unprocessed 48 kHz sampled clean speech, background material, model parameters, music and mixed content, and critical generic audio content samples to the MC. The format of the material is WAVE [25], 16-bit little endian format. For multi-track audio, the audio tracks are ordered according to Table 5 of IVAS Processing Plan for Characterization Phase (IVAS-7b).

The material collection entity (MC) shall control that the unprocessed raw material (both artificially created and real recorded) and the model parameters meet the requirements defined by SA4, collect a pool of model parameters and sound material and choose the model parameters and sound material to be used in the experiments.

The PC is responsible for delivery of the IVAS executables to HL. This includes retrieving the up-to-date executables from the fixed-point and floating-point code repositories, and x-checking the HL processing.

The LLs shall insert the raw voting data into the workbook provided by the GAL and forward the workbook directly to the GAL. In addition, each LL must provide a report of experiments to SA4 no later than the document submission deadline for the Characterization meeting.

* 1. Allocation of Additional Roles

External LLs: Mesaqin.com, FORCE Technology

Volunteering LLs:

Dolby Laboratories, Inc.

Ericsson LM

Fraunhofer IIS

Huawei Technologies Co Ltd.

Nokia Corporation

NTT

Orange

[Philips International B.V.]

[Qualcomm Incorporated]

VoiceAge Corporation

HL: Contributors of the Public Collaboration

CL: Contributors of the Public Collaboration

MC: Contributors of the Public Collaboration

GAL: [HEAD acoustics GmbH]

* 1. Responsibilities

Many of the procedures to be followed are defined in this test plan, with further information being given in IVAS Processing Plan for Characterization Phase (IVAS-7b).

* + 1. Proponent Companies

The specific responsibilities of PC are:

* Deliver to the HL and to the CL preliminary CuT executables
* Deliver to the HL and the CL final CuT executables
* Develop the processing scripts using the condition lists defined in this document and the processing steps defined in IVAS-7b.
* Provide the randomization playlists for P.SUPPL800 subjective experiments described in this document. In case an experiment is duplicated, the playlists will be different for the two tests of the same experiment conducted in different languages. These playlists will be reused for all experiments. Each LL will receive the randomization playlists only for the experiments to be conducted by that LL. The playlists will be delivered in Excel spreadsheet format.
  + 1. Listening Laboratories
* Provide a listening environment that conforms to the requirements in [8] including:
  + Having a background noise level of less than NR-25.
* For each P.SUPPL800 listening test, use subjects that are native speakers of the tested language.
* Provide a person during the training session of each P.SUPPL800 test that is able to answer questions from the subjects in their native language.
* Provide to SA4 the P.SUPPL800 instructions for subjects in each of the languages to be tested by the LL for the Selection Testing.
* LLs shall record or obtain, if not otherwise available, original clean mono speech material (unprocessed 48 kHz sampled speech) for the P.SUPPL800 tests allocated to them and provide it to the MC.
* LLs shall record or obtain, if not otherwise available, original stereo music and mixed content material (unprocessed 48 kHz sampled signals) for the P.SUPPL800 tests allocated to them and provide it to the MC.
* For any tests, LLs may record or obtain original clean mono speech or stereo/immersive material (unprocessed 48 kHz sampled signals) and provide it to the MC.
* LLs shall have the option to declare their material provided to the MC as not available for use by other LLs.
* Obtain from the HL the processed test materials for all tests to be conducted by the LL.
* Perform the listening tests in accordance with this document.
* Delivery to the GAL of all raw voting data using the data delivery file provided by the GAL for all tests to be conducted by the LL.
* Delivery of a LL report to SA4 which includes:
  + Confirmation that the LL testing environment conforms to the requirements of the Characterization test for all tests conducted by the LL.
  + Listening test instructions for subjects in each of the languages tested by the LL.
  + Age and gender information for the set of subjects used in each listening test, and over all listening tests in each tested language tested by the LL.
  + Discussion of any problems encountered during testing and the solution used to address the problem.

Note: The databases are not assumed pristine.

* + 1. Host Laboratory

The following list defines the tasks expected to be carried out by the Host Laboratory (HL). The tasks have to be conducted and completed following the schedule for the IVAS Characterization phase defined in IVAS-2.

* Get from the [relevant repository] preliminary CuT executables for a cross-check with the CL to identify potential problems.
* Get processing scripts from the [relevant repository].
* Deliver the cross-checked processing scripts to SA4.
* Receive all source materials and model parameters from the LLs and MC.
* Get from the [relevant repository] final CuT executables. Do final cross-check with the CL using the final CuT executables, reference executables, and the clean speech, background material, model parameters, music and mixed content, and critical generic audio content provided by the LLs and MC for each experiment.
* Process and deliver all test files for all experiments to the LLs in phases to meet testing schedules after completion of final cross-check.
* Deliver and present HL report. The report should include a discussion of any problems encountered during the cross-check and processing efforts. The dates for final test material delivery to the LLs should be included.

Editor’s note: The relevant repository for IVAS fixed-point and floating-point codes need to be specified.

* + 1. Cross-check Laboratory

The following list defines the tasks expected to be carried out by the Cross-check Laboratory (CL). The tasks have to be carried out following the schedule for the IVAS Characterization phase defined in IVAS-2.

* Get from the [relevant repository] preliminary CuT executables for a cross-check with the HL to identify potential problems.
* Final cross-check with the HL using the final CuT executables from the HL, reference executables, and the clean speech, background material, model parameters,music and mixed content, and critical generic audio content provided by the LLs and MC for each experiment and available from the HL.
* Get processing scripts from the [relevant repository].
* Receive all source materials and model parameters from the LLs and MC.
* Get from the [relevant repository] final CuT executables. Do final cross-check with the HL using the final CuT executables, reference executables, and the clean speech, background material, model parameters, music and mixed content, and critical generic audio content provided by the LLs and MC for each experiment.
* Process and cross-check all test files for all experiments in phases as needed for the LLs to meet testing schedules.
  + 1. Material Collection Entity (MC)
* Collect the clean mono speech, background material, music and mixed content, real recorded stereo/immersive signals, and a pool of parameters for artificially created stereo/immersive sound material (e.g., impulse responses).
* Verify that the unprocessed material (for both artificially created and real recorded content) and parameters for artificially created stereo/immersive sound material meet the requirements defined by SA4.
* Choose the parameters and sound materials to be used in the experiments.

The proposed procedure for MC tasks is detailed in Annex C:.

* + 1. Global Analysis Laboratory
       1. Tasks

The following list defines the tasks expected to be carried out by the Global Analysis Laboratory (GAL). The tasks have to be carried out following the schedule for the IVAS Characterization phase defined in IVAS-2.

* Provide the raw voting data delivery worksheets to the appropriate LLs. Each LL will receive the data delivery worksheets only for the experiments to be conducted by that LL. The worksheets will be delivered in Excel spreadsheet format.
* Receive the raw voting data from the LLs in the appropriate data delivery worksheets.
* Conduct statistical Terms of Reference (ToR) tests as specified in clause 3.3.6.2. The ToR tests compare the subjective scores of the CuT against the scores for specified reference conditions. Each subjective experiment contains a number of ToR tests to be computed by the GAL.
* Prepare a GAL report to be presented to SA4 as scheduled in the IVAS Project Plan IVAS-2.
  + - 1. Statistical analysis of results

The GAL report will present the results of the Terms of Reference (ToR) tests using Student's Independent Groups t-test (single-sided at 95% confidence level). Results of the Requirement ToR tests for each experiment will be presented in a table as illustrated in Table 1.

In the example below for **Requirement ToR tests**:

* Requirement ToR tests that are passed, (i.e., CuT "not worse than" Requirement) are indicated by **CuT NWT Ref**.
* Requirement ToR tests that are exceeded, (i.e., CuT "better than" Requirement) are indicated by **CuT BT Ref**.
* Requirement ToR tests that are failed (i.e., CuT "worse than" Requirement) are indicated by **CuT WT Ref**.

Table : Example of Requirement ToR test results



* + 1. SA4
* SA4 defines the methods and models for artificial creation of sound material based on original (mono) sound material.
* SA4 defines the stereo/immersive scenes including, e.g., environments/rooms, relative placement of talkers to capture point, and overtalk by talkers.
* SA4 (volunteering members) shall provide the parameter sets for models/methods for artificial creation of sound material based on original (mono) sound material.
* SA4 defines the set of requirements for original sound material (e.g., sampling frequency, formats).
* SA4 (volunteering members) shall record or obtain original stereo/immersive material of Generic audio (unprocessed 48 kHz sampled signals) for BS.1534 experiments.
* SA4 (volunteering members) may record or obtain original clean mono speech material (unprocessed 48 kHz sampled speech).

1. Information relevant to all Experiments
   1. General Technical Notes

Any and all deviations from the specifications contained in this document and the IVAS Processing Plan for Characterization Phase (IVAS-7b) must be documented and submitted to SA4 along with the experimental report.

* 1. General Consideration of Experiments

The main goal of the IVAS Characterization Test is to evaluate the aspects of the IVAS codec that were not tested in the Selection phase. Some of the aspects should be tested in formal subjective evaluation while other aspects can be evaluated informally. The following aspects shall be tested in formal subjective experiments:

* The IVAS fixed-point implementation and the interoperability between the floating-point implementation and the fixed-point implementation
* The integrated IVAS renderer
* Combined input formats:
  + Objects + MASA (OMASA)
  + Objects + SBA (OSBA)
* JBM with 5G profiles
* [Higher packet loss conditions than 5% tested in the Selection phase]
* [Tandem]
* Binaural rendering configurations, e.g.
  + room effects,
  + head rotation,
  + 6 degrees-of-freedom (DoF) and directivity
* 7.1, 5.1+2 and 5.1+4 multi-channel configurations
* 16 kHz and 32 kHz sampled input
* Different configurations of input and output not tested in the Selection phase
  + Complex inputs to mono output
  + [Complex inputs to stereo output]
  + Rendering to an arbitrary LS setup
* EVS-coded mono downmix of stereo input (13.2 and 24.4 kbps)

Evaluation of different configurations of the IVAS codec require different testing methodologies, outlined in the following sections.

* [IVAS Characterization Test is separated into two main use case scenarios, namely speech centric Immersive conversation, and Generic immersive audio. The Immersive conversation use case targets lower bitrates and the evaluation is done by naïve listeners. The Generic immersive audio assumes higher bitrates and the evaluation is done by experienced listeners.]
* In case a P.SUPPL800 experiment is performed twice in two different LLs, it is run in two different languages with native listeners.

[

* + 1. Immersive conversation
* Source material:
  + Clean speech
  + Speech with background
  + Music and mixed content
* Input formats:
  + Stereo, including binaural
  + FOA
  + Object-based audio
  + MASA
* Lower bitrates, up to approximately the bitrate having as reference multi-mono EVS at 24.4 kbps per channel, as specified in IVAS Performance Requirements (IVAS-3).
* Including DTX conditions
* Including FE conditions
* Listening environment: headphones
* Test methodology: P.SUPPL800 [21]
  + 1. Generic immersive audio
* Source material: Generic audio
* Higher bitrates
* No DTX conditions
* Input formats:
  + Stereo, including binaural
  + FOA
  + HOA2
  + HOA3
  + Object-based audio
  + MASA
  + Channel-based audio
* Listening environment:
  + Headphones
  + 7.1 + 4 loudspeaker setup
* Test methodology: BS.1534 (MUSHRA) [22]

]

[

* 1. Alternative Testing against current service quality and untested operation points, comparison against floating-point in same test

This alternative focuses on the following two points:

* Comparison of IVAS against EVS mono at the same bitrate, to evaluate the difference in QoE between the deployed monaural service and the immersive formats that are provided by IVAS.
* Comparison of IVAS floating-point and BASOP code bases at the same bitrate (where applicable).

For this proposal, a combination of testing with P.SUPPL800 and MUSHRA is proposed. It is proposed to use the IVAS fixed-point code for the MUSHRA tests.

The general assumptions for these tests are:

* The internal IVAS renderer is used, rendering to the playout configuration; no head rotation is to be used for these tests.
* P.SUPPL800 tests are carried out via headphones; for these, the decoder output format is set to binaural.
* Given that the BASOP codebase is available and verified, a comparison of BASOP and floating-point codebases is performed in the tests.
* The mono conditions are generated using the IVAS renderer as a preprocessing stage.
* It is proposed to use mixed material for both, P.SUPPL800 and MUSHRA, i.e. a combination of clean-speech, noisy-speech and mixed/music.
  + 1. Testing of format/bitrate combinations under clean-channel conditions

The proposed P.SUPPL800 listening test layout for IVAS formats under clean-channel conditions is as follows:

Table : P.SUPPL800 testing of format/bitrate combinations under clean-channel conditions incl. a comparison of BASOP and Floating-Point code



|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | Mono | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | MNRU Q=xx dB | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU | - | - |
| c09 | ESDRU |  |  |
| c10 | ESDRU | - | - |
| c11 | EVS | 1x13.2 |  |
| c12 | EVS | 1x16.4 |  |
| c13 | EVS | 1x24.4 |  |
| c14 | EVS | 1x32.0 |  |
| c15 | EVS | 1x48.0 |  |
| c16 | EVS | 1x64.0 |  |
| c17 | EVS | 1x96.0 |  |
| c18 | EVS | 1x128.0 |  |
| c19 | IVAS FL | 13.2 |  |
| c20 | IVAS FL | 16.4 |  |
| c21 | IVAS FL | 24.4 |  |
| c22 | IVAS FL | 32.0 |  |
| c23 | IVAS FL | 48.0 |  |
| c24 | IVAS FL | 64.0 |  |
| c25 | IVAS FL | 80.0 |  |
| c26 | IVAS FL | 96.0 |  |
| c27 | IVAS FL | 128.0 |  |
| c28 | IVAS FX | 13.2 |  |
| c29 | IVAS FX | 16.4 |  |
| c30 | IVAS FX | 24.4 |  |
| c31 | IVAS FX | 32.0 |  |
| c32 | IVAS FX | 48.0 |  |
| c33 | IVAS FX | 64.0 |  |
| c34 | IVAS FX | 80.0 |  |
| c35 | IVAS FX | 96.0 |  |
| c36 | IVAS FX | 128.0 |  |

It is proposed to conduct in total 6 P.SUPPL800 clean-channel test of this type for the following formats

* Stereo
* FOA
* ISM (3-4 objects in same test; ISM incl. metadata/rendering, EVS w/o metadata/rendering)
* MASA 1TC
* OSBA (1-4 objects in same test)







* + 1. High Bitrate MUSHRA Tests

It is proposed to set up the high bitrate MUSHRA tests using the following conditions:

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | LP 3.5 kHz | - | - |
| c03 | EVS | 1x 64.0 | Off |
| c04 | EVS | 1x128.0 | Off |
| c05 | IVAS | 64.0 | Off |
| c06 | IVAS | 96.0 | Off |
| c07 | IVAS | 128.0 | Off |
| c08 | IVAS | 256.0 | Off |

The tests are foreseen for the following formats:

* Stereo (see Anex ‎G)
* FOA (see Anex ‎G)
* HOA3 (see Anex ‎G)
* 5.1, 7.1 (both formats in same test, binaural rendering) (see Anex ‎G)
* 5.1+2, 5.1+4 (both formats in same test, binaural rendering) (see Anex ‎G)
* 7.1+4 (see Anex ‎G)
* ISM (1-2 objects in same test; ISM incl. metadata/rendering, EVS w/o metadata/rendering) (see Anex ‎G)
* ISM (3-4 objects in same test; ISM incl. metadata/rendering, EVS w/o metadata/rendering) (see Anex ‎G)
* MASA 1TC (see Anex ‎G)
* MASA 2TC (see Anex ‎G)
* OSBA (1-4 objects in same test) (see Anex ‎G)
* OMASA (1-4 objects in same test) (see Anex ‎G)
  + 1. Low Bitrate MUSHRA Tests:

It is proposed to set up the low bitrate MUSHRA tests using the following conditions:

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | LP 3.5 kHz | - | - |
| c03 | EVS | 1x 16.4 | Off |
| c04 | EVS | 1x 32.0 | Off |
| c05 | IVAS | 16.4 | Off |
| c06 | IVAS | 24.4 | Off |
| c07 | IVAS | 32.0 | Off |
| c08 | IVAS | 48.0 | Off |

The tests are foreseen for the following formats:

* Stereo (see Anex ‎G)
* FOA (see Anex ‎G)
* 5.1 (see Anex ‎G)
* ISM (3-4 objects in same test; ISM incl. metadata/rendering, EVS w/o metadata/rendering); test starting at 24.4 kbit/s, possibly adding 64 kbit/s (see Anex ‎G)
* MASA 1TC (see Anex ‎G)
* OSBA (1-4 objects in same test) (see Anex ‎G)

The proposed formats can be seen as a super-set of all supported IVAS formats.

* + 1. Summary

The testing scheme as proposed above would consume 6/19 P.SUPPL800 tests and 12/(estimated) 30 MUSHRA tests. The remaining tests could be used for testing other untested features such as EVS compatible mono downmix, rendering aspects, etc.

Editor’s note: As the proposal comprises different special rendering (mono, stereo, binaurally rendered immersive formats) in a single experiment, some evidence is needed to show that the proposed methodology can provide consistent results.

Editor’s note: As different input formats are included in an experiment, it is proposed that different input formats would correspond to different categories.

Editor’s note: As different input formats are included in an experiment, also the EVS compatible mono downmix could be integrated in the Tables above if there are available slots.

Editor’s note: For Multi-channel (MC) input formats, it is assumed that input material for speech samples will be generated artificially while music and mixed content will use multi-channel recorded content. If MC-recorded content is not available for some LLs, samples from the Generic Audio Items can be provided to the LLs.

Editor’s note: The number of EVS testing points might be reduced.

Editor’s note: If this alternative proposal is agreed, better integration within the rest of the Test Plan is needed. In particular, the proposal does not assume anymore the split we used in Selection between the P.SUPPL800-tested Immersive conversation use case and audio material, and the MUSHRA-tested Generic Audio use case

]

* 1. Methodology

The following test methodologies shall be used in the IVAS Characterization test. [P.SUPPL800 [21] shall be used in experiments designed to evaluate the Immersive conversation use case scenarios. BS.1534 [22] shall be used in experiments designed to evaluate Generic immersive audio use case scenarios.] High-level configuration of experiments for envisaged methodologies is outlined below.

Editor’s note: additional methodologies might be needed to evaluate some configurations.

* + 1. P.SUPPL800
* Test duration should not exceed 2 hours per listening panel. The 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 [13].
* 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.
  + 1. 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 conditions may vary depending on actual experiment.

* 1. Opinion Scales

Table 5 defines opinion scale used for ITU-T P.SUPPL800 DCR test. Instructions in English for the P.SUPPL800 test are provided in Annex A:.

Table : Opinion scale for ITU-T P.SUPPL800 DCR test

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

* 1. Material

All audio material shall be sampled at 48 kHz with Full Band (FB) content. The audio material is to be delivered to the HL as 16-bit little endian WAVE format files [25] following the naming convention provided in the IVAS Processing Plan for Characterization Phase (IVAS-7b). For multi-track audio, the audio tracks are ordered according to Table 5 of IVAS Processing Plan for Characterization Phase (IVAS-7b). Additionally, it should be verified that the audio material can be processed with the AFsp package tools [26].

The following categories of audio content will be used in IVAS Selection Test using P.SUPPL800:

* Clean speech: [Except for experiment evaluating 1 object input,] 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 Annex F:.
* Music and Mixed content – categories specified in section 4.6.3.

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

* Generic audio – critical generic audio items including speech with and/or without background, music, mixed.
  + 1. Speech Material for P.SUPPL800 testing

Except for the experiment evaluating Music and mixed content stereo input, P.SUPPL800 test experiments will use artificially created immersive audio. LLs shall provide 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 with 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 [29].

* + 1. Background Material
* [Immersive conversation use case scenario,] P.SUPPL800 testing-: A mix-based approach using separate background recordings will be 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.
  + 1. Music and Mixed Content Material for P.SUPPL800 testing

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

* + 1. Critical Generic Audio Items for BS.1534 testing
       1. Steps of Critical Test Item Selection

The following steps are based on [24]:

* Call for test material according to the generic audio signal categories described below.
* MC collects candidate material submitted in response to the call and selects a number of critical items to be used in the Characterization test.
* MC selects a limited set of training items to be used in a training phase.
  + - 1. Test Material

First, a call will be sent out for test material according to a number of generic audio signal categories as specified below. All 3GPP members are invited to submit test material to MC. The submitting organization shall assign the items to the below-mentioned audio signal categories. Then, MC will identify 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, 5.1+4, 7.1 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.

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

MC will further maintain and report to SA4 a list indicating the number of proposed items per submitting organization.

In case the submitted material is insufficient/inadequate to conduct the tests, MC will add the missing test items.

* + - 1. Training material

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

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

* 1. Listening Systems and Listening Environments

The IVAS Selection Test will use the following listening systems:

* Stereo headphones for binaural listening, e.g.:
  + Beyerdynamic DT 770 Pro for P.SUPPL800 experiments
  + Sennheiser HD 650 for BS1534 experiments
* Loudspeaker listening system – 5.1, 5.1+2, 5.1+4, 7.1, 7.1+4 loudspeaker setup [3].
  1. Experimental Procedure
     1. Experimental Procedure for P.SUPPL800 experiments

Initially the experimenter should provide a written copy of the experiment instructions to the listeners. When the listeners have acknowledged that they understand the instructions, they will be presented with a practice session to rate the preliminary conditions. After the practice session has been completed, the experimenter should ask 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.

* 1. Results and Analysis

On completion of the experiments, the LLs must provide the raw voting data to the GAL for the purpose of performing a global analysis. The raw voting data for each experiment shall be delivered in the spreadsheet provided by the GAL for that purpose.

1. Subjective Experiments

The purpose of the Characterization phase experiments is to evaluate the performances of the IVAS codec algorithm [ for the configurations not covered in the Selection phase and to verify its fixed-point conversion].

The details provided in this section and in corresponding Annexes are those that are specific to each particular experiment. Generic information can be found in Section 4. Therefore, the LLs should use the information in Section 4 in conjunction with the information given in this section and Annexes.

Table 6 shows high-level overview of P.SUPPL800 experiments. Table 7 shows high-level overview of BS.1534 experiments. Finally, Table 8 shows allocation of experiments to LLs, and the languages for each P.SUPPL800 experiment.

Detail conditions for each subjective experiment are defined in Annex F: for P.SUPPL800 experiments and in Annex G: for BS.1534 experiments.

Table : High-level overview of P.SUPPL800 experiments

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Exp | Input format | Source material | Listening environment | Bitrates kbps | FER/jitter | DTX |
|  |  |  | Headphones |  |  |  |
|  |  |  | Headphones |  |  |  |
|  |  |  | Headphones |  |  |  |
|  |  |  | Headphones |  |  |  |
|  |  |  | Headphones |  |  |  |
|  |  |  | Headphones |  |  |  |
|  |  |  | Headphones |  |  |  |
|  |  |  | Headphones |  |  |  |
|  |  |  | Headphones |  |  |  |

Table : High-level overview of BS.1534 experiments

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Exp** | **Input format** | **Source material** | **Listening environment** | **Bitrates kbps** |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |
|  |  | Generic Audio |  |  |

Notes:

* Stereo may include binauralized samples (without head tracking).
* For inputs 5.1, 5.1+2, 5.1+4, 7.1, 7.1+4, FOA, HOA2, HOA3, Objects, MASA, OSBA, and OMASA 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.SUPPL800 experiments are assumed Full Band experiments, i.e., the direct reference condition is always FB. P.SUPPL800 stereo experiments are SWB experiments.

Editor’s note: Tables 3 and 4 may be removed if not relevant (i.e. if all information fits in Table 5)

Note: the assumption is to have [about 5] weeks for subjective testing, from receiving the processed samples to delivering the listening results, assuming a dry run could be available a week before.

Minimum requirements for P.SUPPL800 experiments: 6 talkers (3 male + 3 female) per experiment, 14 single sentences per talker.

[

Under the following preliminary assumptions, Table 8 shows an example allocation of experiments.

**Assumptions**

* The same pricing for external LLs and GAL as in Selection phase (IVAS-8a), i.e. 18000 Euros per P.SUPPL800 test, 13000 Euros per BS.1534 test rendered via loudspeakers, 10000 Euros per BS.1534 test rendered via headphones, and 12000 Euros for GAL tasks.
* Assuming similar ratio of PS.SUPPL800 and BS.1534 experiments as in Selection phase, the following test methodology distribution would fit exactly in 192000 Euros:
  + 5 x P.SUPPL800 (5 x 18000 = 90000 Euros)
  + 4 x BS.1534, LS rendering (4 x 13000 = 52000 Euros)
  + 5 x BS.1534, headphones rendering (5 x 10000 = 50000 Euros)
* Fixed-point code evaluation does not require any EVS reference. This should allow for testing all the combinations of floating-point and fixed-point encoder and decoder.
* It is assumed that the subjective material in the P.SUPPL800 fixed-point validation experiments accommodates all three input signal levels

Editor’s note: Practicality of this still needs to be confirmed.

* No replication of experiments
* Number of experiments as indicated by the volunteering LLs. At the last Audio SWG call, the volunteering LLs labs gave the following preliminary test availabilities:
  + Dolby – 4-6 BS.1534 tests, both loudspeaker and binaural rendering is possible, tentatively 2 P.SUPPL800 tests
  + Ericsson – 2 P.SUPPL800 tests, and at minimum 2 BS.1534 tests, both loudspeaker and binaural rendering is possible.
  + FhG – tentatively 2 P.SUPPL800 tests, and at minimum 2 BS.1534 tests, both loudspeaker and binaural rendering is possible.
  + Huawei – 2 BS.1534 tests, binaural rendering.
  + Nokia – 1 ACR, 3 P.SUPPL800 tests, and 6 BS.1534 tests, both loudspeaker and binaural rendering is possible.
  + NTT – 1 P.SUPPL800 test, 1/2 BS.1534 (shared with Panasonic), binaural rendering.
  + Orange –2 P.SUPPL800, 1 BS.1534.
  + Panasonic – 1/2 BS.1534 (shared with NTT), binaural rendering.
  + Philips – tentatively 1 BS.1534, 1 room acoustics testing.
  + QC – tentatively 2 BS.1534.
  + VoiceAge – 2 P.SUPPL800 tests.

Taking the conservative minimum number wherever a range was indicated, but assuming that LLs will be able to test the experiments marked “tentatively”, we get:

* 14 x P.SUPPL800
* 21 x BS.1534, up to 12 with LS rendering
* 1 x ACR
* 1 x room acoustics testing

Together with the experiments allocated to external listening laboratories, we can thus test overall at minimum:

* 19 x P.SUPPL800
* 30 x BS.1534, up to 12 with LS rendering
* 1 x ACR
* 1 x room acoustics testing

For reference, at Selection we ran 18 P.SUPPL800 and 28 BS.1534 experiments (IVAS-8a).

]

Table : Allocation of experiments to LLs and P.SUPPL800 languages

[

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Exp** | **Objectif** | **Input format** | **Source material** | **Listening environment** | **Language** | **LL 1** | **LL 2** | **Pricing [Euros]** |
| P800-1 | Fixed-point | Stereo | All | Headphones |  |  |  |  |
| P800-2 | Fixed-point | FOA | Speech w/wo bckg | Headphones |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| P800-4 | Fixed-point | 2 Objects | Clean speech | Headphones |  |  |  |  |
| P800-5 | Fixed-point | Stereo MASA | Speech w/wo bckg | Headphones |  |  |  |  |
| P800-6 |  | monoMASA | Clean speech | Headphones |  |  |  |  |
| P800-7 |  | monoMASA | Speech+bckg | Headphones |  |  |  |  |
| P800-8 |  | OSBA | Speech+bckg | Headphones |  |  |  |  |
| P800-9 |  | OMASA | Speech+bckg | Headphones |  |  |  |  |
| P800-10 | JBM/FER/Tan | Stereo | Clean speech | Headphones |  |  |  |  |
| P800-11 | JBM/FER/Tan | 2 Objects | Clean speech | Headphones |  |  |  |  |
| P800-12 | Stereo dmix | Stereo | All | Headphones |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| BS1534-1 | Fixed-point | Stereo | Generic audio | Headphones |  |  |  |  |
| BS1534-2 | Fixed-point | Stereo | Generic audio | Headphones |  |  |  |  |
| BS1534-3 | Fixed-point | 5.1 | Generic audio | 5.1 |  |  |  |  |
| BS1534-4 | Fixed-point | 5.1 | Generic audio | 5.1 |  |  |  |  |
| BS1534-5 | Fixed-point | 7.1.4 | Generic audio | 7.1 + 4 |  |  |  |  |
| BS1534-6 | Fixed-point | 7.1.4 | Generic audio | 7.1 + 4 |  |  |  |  |
| BS1534-7 | Fixed-point | FOA | Generic audio | Headphones |  |  |  |  |
| BS1534-8 | Fixed-point | HOA2 | Generic audio | Headphones |  |  |  |  |
| BS1534-9 | Fixed-point | HOA3 | Generic audio | Headphones |  |  |  |  |
| BS1534-10 | Fixed-point | HOA3 | Generic audio | 7.1 + 4 |  |  |  |  |
| BS1534-11 | Fixed-point | Objects | Generic audio | Headphones |  |  |  |  |
| BS1534-12 | Fixed-point | Objects | Generic audio | Headphones |  |  |  |  |
| BS1534-13 | Fixed-point | Stereo MASA | Generic audio | Headphones |  |  |  |  |
| BS1534-14 | Fixed-point | Stereo MASA | Generic audio | Headphones |  |  |  |  |
| BS1534-15 |  | OSBA | Generic audio | Headphones |  |  |  |  |
| BS1534-16 |  | OSBA | Generic audio | 7.1+4 |  |  |  |  |
| BS1534-17 |  | OMASA | Generic audio | Headphones |  |  |  |  |
| BS1534-18 |  | OMASA | Generic audio | Headphones |  |  |  |  |
| BS1534-19 |  | 5.1+2 | Generic audio | 5.1+2 |  |  |  |  |
| BS1534-20 |  | 5.1+2 | Generic audio | 5.1+2 |  |  |  |  |
| BS1534-21 |  | 5.1+4 | Generic audio | 5.1+4 |  |  |  |  |
| BS1534-22 |  | 5.1+4 | Generic audio | 5.1+4 |  |  |  |  |
| BS1534-23 |  | 7.1 | Generic audio | 7.1 |  |  |  |  |
| BS1534-24 |  | 7.1 | Generic audio | 7.1 |  |  |  |  |
| BS1534-25 | Renderer |  | Generic audio |  |  |  |  |  |
| BS1534-26 | Renderer |  | Generic audio |  |  |  |  |  |
| BS1534-27 | Renderer |  | Generic audio |  |  |  |  |  |
| BS1534-28 | Renderer |  | Generic audio |  |  |  |  |  |
| BS1534-29 | Renderer |  | Generic audio |  |  |  |  |  |
| BS1534-30 | Renderer |  | Generic audio |  |  |  |  |  |
| ACR | 16, 32, 48 kHz |  |  |  |  |  |  |  |
|  |  | OSBA, OMASA | Speech+bckg |  |  |  |  |  |
|  | JBM/FER/Tan | FOA/MASA | Clean speech | Headphones |  |  |  |  |
|  | Room acoustics |  |  |  |  |  |  |  |
|  | 6 DoF,  directivity |  |  |  |  |  |  |  |
| Total price | | | | | | | | **192000** |

Editor’s note: The Table above is currently understood as an example list of experiments that can be accommodated in the Characterization Test given the collective minimum testing capabilities and allocated budget. In particular, the question whether to replicate all or some of the experiments in two LLs needs to be answered. This will affect the list substantially.

Legend:

* P.800 – P.SUPPL800 DCR test
* bckg – means any background
* All – means clean speech, speech with background, mixed content, and music

Editor’s note: Proposal made that at least some floating-point conditions should be inserted in experiments evaluating new configurations.

Editor's note: Proposal made to test only fxd-fxd vs. flt-flt, i.e. excluding or limiting testing of fxd-flt and flt-fxd.

Editor's note: Different signal levels should be tested for fixed-point validation (all 3 levels, possibly not for all bitrates). Possibly, levels could be integrated into signal categories.

Editor’s note: Table 5 might be better split into several tables, one per methodology

]

JAP = Japanese

FR = French

GER = German

MAN = Mandarin

DAN = Danish

ENG = English

Editor’s note: the list of languages needs to be reviewed when Table 5 stabilizes.

1. Sample Instructions to Subjects and Data Collection

These instructions shall be translated properly to the LL’s language and be given to the listeners. The instructions given to the listeners shall be provided for information in the LL report.

|  |
| --- |
| **INSTRUCTIONS TO NAÏVE LISTENERS FOR P.SUPPL800 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 |

1. P.SUPPL800 Presentation Orders

The PC will provide the Presentation Order for each P.SUPPL800 experiment to the Listening Lab assigned to conduct the test. The presentation order for each experiment has been developed by the PC using a randomized-blocks experimental design and sample allocation for conducting Independent Groups Student T-tests for the specified Terms of Reference tests. Each Presentation Order includes six blocks, corresponding to six categories and includes a separate presentation sequence for each of 6 panels of subjects. The Presentation Orders will be delivered to the Listening Labs in the form of Data Delivery Excel spreadsheets which are attached to this document. Presentation Orders will be cross-checked before the actual listening tests start.

1. Proposed Procedure for MC Tasks
   1. Control that the unprocessed material matches the requirements defined by SA4

The following requirements have been identified:

Editor’s note: the references to sections need to be updated if the structure of IVAS-7b or OVAS-8b changes.

* General:
  + All audio material shall be sampled at 48 kHz with Full Band (FB) content. (IVAS-8b, Section 4.5)
  + The audio material is to be delivered to the HL as 16-bit little endian WAVE format files following the naming convention provided in the IVAS Processing Plan (IVAS-7b). (IVAS-8b, Section 4.5)
  + For multi-track audio, the audio tracks are ordered according to Table 5 of IVAS Processing Plan (IVAS-7b). (IVAS-8b, Section 4.5)
  + Additionally, it should be verified that the audio material can be processed with the AFsp package tools. (IVAS-8b, Section 4.5)
  + All input source material to the IVAS-7b processing stages defined in IVAS-7b Section 4 files shall be 20 ms block aligned. (IVAS-7b, Section 4.1.1)
  + For Ambisonics signals, ACN component ordering and SN3D normalization shall be used (IVAS-4 [3] Section 3, IVAS-7b, Section 3.8)
  + For Objects, metadata according to IVAS-4 shall be used. (IVAS-4, Annex C.1)
* P.SUPPL800 testing
  + Speech Material (input material for artificial immersive item creation):
    - Clean speech mono audio samples (IVAS-8b, Section 4.5.1)
    - The recording SNR should be in accordance with P.800 at least 40 dB but preferably 50 dB or higher. (IVAS-8b, Section 4.5.1)
    - 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. (IVAS-8b, Section 4.5.1)
    - The length of the sentences should typically correspond to the length of traditional Harvard sentences. (IVAS-8b, Section 4.5.1)
    - The total length of the generated P.SUPPL800 speech samples shall not exceed 10 s. (IVAS-7b, Section 4.2.1)
  + Background Material
    - The minimum lengths of noise files shall be 80 s. (IVAS-8b, Section 4.5.2)
    - Car noise: 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. (IVAS-8b, Section 4.5.2)
    - Office noise: This noise type should also contain typical office sounds, such as keyboard noise, computer fans, telephones ringing, printers, air conditioner, etc. (IVAS-8b, Section 4.5.2)
    - Street noise: 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. (IVAS-8b, Section 4.5.2)
    - Additional background noise types as defined in IVAS-8b, Annex F:.
  + Model Parameters:
    - Submitted model parameters have to be reviewed and checked for their suitability.
    - The number of talker source and microphone capture positions have to be documented.
  + Music and Mixed:
    - Shall contain meaningful contents. (IVAS-8b, Section 4.5.3)
    - The duration of each sample shall be approximately 8 and at least 7seconds. (IVAS-8b, Section 4.5.3)
    - The following categories shall be used (IVAS-8b, Section 4.5.3):
      * Classical music
      * Modern instrumental music
      * Modern vocal music
      * Radio Jingle
      * Movie Trailer
      * Advertisement
* BS.1534 Testing
  + General:
    - Length in time of the items will be 10 s at a maximum. (IVAS-8b, 4.5.4.2)
  + 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, 5.1+2, 5.1+4, 7.1 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 control tasks should be performed on a best effort basis, since for certain requirements no clear objective measures or other means for controlling are available on MC side.

In coordination with the material contributors, the MC may perform technical corrections to the submitted material, as long as these don’t change the subjective characteristics of the submitted material. Such Corrections could for example include:

* Conversion, if the audio material is not delivered as 16-bit little endian WAVE format files following the naming convention provided in the IVAS Processing Plan (IVAS-7b). (IVAS-8b, Section 4.5)
* Renaming in order to match the naming conventions according to IVAS-7b.
* Track re-ordering for multi-track audio, if the audio tracks are ordered according to Table 5 of IVAS Processing Plan (IVAS-7b). (IVAS-8b, Section 4.5)
* Conversion if the audio material can’t be processed with the AFsp package tools. (IVAS-8b, Section 4.5)
* 20ms Block alignment if the input source material to the IVAS-7b processing stages defined in IVAS-7b Section 4 is not 20ms block aligned. (IVAS-7b, Section 4.1.1)
* Format conversion for Ambisonics signals to match ACN component ordering and SN3D normalization (IVAS-4 [3] Section 3, IVAS-7b, Section 3.8)
* Conversion of objects metadata, if metadata is not according to IVAS-4. (IVAS-4, Annex C.1)
* File cropping or padding if felt necessary.
* Level adjustment if felt necessary in order to avoid clipping in one of the processing steps.
* Corrections of similar technical impact.
  1. Selection of parameters and sound materials to be used in the experiments

The following guidelines for the selection of parameters and sound materials shall be applied:

* Material matches the content listed in IVAS-8b
* Material is representative for assumed typical IVAS application scenarios (IVAS-8b[, 4.5])
* Material exploits the audio format sufficiently (i.e. generally the channels/objects/Ambisonics orders provided by the format shall be sufficiently in use)
* Material is distinguishable from anchor conditions (for BS.1534 tests: audio bandwidth >> 7 kHz)
* Material provides sufficient variability and balance in test coverage:
  + Variability and balance in signal characteristics
  + Variability and balance in spatial characteristics
  + Variability and balance in criticality of material
* In case not a sufficient amount of suitable parameters and sound materials is submitted per experiment, the MC may select material submitted to other experiments, possibly after format conversion.
* In case more than the required amount of parameters and sound materials is submitted per experiment, the MC may preferably select materials based on a ranking of suitability in terms of variability and balance (see above bullets and sub-bullets) and ultimately make a random choice among parameters/materials found equally suitable.
* For P.SUPPL800 experiments with model-based test sample generation, after selection of rooms/environments and assignment to categories, the MC shall define the specific scenes to be applied for the different samples of a given room/environment (category). The definition of scenes shall be based on the principle of variability and balance and may ultimately also be based on a random selection in case more equivalent scenes than necessary are available.
  1. Documentation

Control of the unprocessed material and the selection of parameters and sound materials shall be documented by the MC and a report shall be submitted to [SA4#128].

For a transparent documentation for any material provided under NDA, the MC will create a mapping table from the original filename to a more generic alias filename as follows:

Experiment\_[Type]\_Company\_FileXX.ext, where

* Experiment: P.SUPPL800 Experiment (e.g. P800-1) or BS.1534 experiment category (e.g. BS-1534-Stereo)
* Type (if applicable): Model Parameter or Background Noise
* Company: Submitting Company
* XX: Numerator
* ext: File Extension

MD5 hashes allow for a mapping between original and alias filename.

1. Data to be Provided by LL

The GAL will provide a Data Delivery spreadsheet for each experiment to the Listening Lab assigned to conduct the test. For each trial, the table shows both the Reference file (condition c01) and the Test file (conditions c01-c36) followed by 5 data cells, one per subject, to be filled by the Listening Lab with the raw voting data provided by the subjects. The file naming convention is as follows:

***leee*a*y*s*zz*.cnn.wav** ***l*** = Listening Lab, ***eee*** = Experiment, **a*y*** = Category, **s*zz*** = Sample, **c*nn*** = Condition (see IVAS-7b)

1. Characterization Testing Timeline

Table H.1: Testing Timeline

|  |  |  |  |
| --- | --- | --- | --- |
| **Month** | **Meeting/date** | **Task** | **Active Parties** |
| Nov-2023 | Nov 13-17 | 3GPP SA4 meeting #126  Progress on IVAS characterization permanent documents, including:   * IVAS-7b Processing Plan for Characterization Phase * IVAS-8b Test Plan for Characterization Phase |  |
| Feb-2024 | Jan 29 – Feb 2 | 3GPP SA4 meeting #127  Finalization of IVAS characterization permanent documents, including:   * IVAS-7b Processing Plan for Characterization Phase * IVAS-8b Test Plan for Characterization Phase |  |
| April-2024 | April 8-12 | 3GPP SA4 e-meeting #127bis |  |
| May-2024 | May 20-24 | 3GPP SA4 meeting #128  Characterization test results are available and SA4 reviews them. |  |
| June-2024 | June 18-21 | 3GPP TSG-SA #104  Approval of characterization test results. |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. P.SUPPL800 Experiments

[

* 1. Experiment P800-1: Mono MASA Clean Speech Test

It is proposed for mono-MASA inputs to have following conditions for clean speech test.

Table F.1. Overview of the conditions

|  |  |  |
| --- | --- | --- |
| **Main Codec Conditions** |  |  |
| Codec under Test (CuT) | 7 | IVAS operated at 13.2, 16.4, 24.4, 32, 48, 64, 80 kbps with DTX off at 0% FER |
| 3 | IVAS operated at 16.4, 32, 64 kbps with DTX off at 5% FER |
| 3 | IVAS in frame rate switching at low (13.2 – 32 kbit/s), medium (24.4- - 64 kbit/s) and high (48 – 128 kbit/s) bitrate range. Bitrate change happens every ~20th frame to random bitrate within range. |
| 6 | IVAS fixed point / floating point interoperability conditions |
|  |  |  |
| **Codec references** |  |  |
| Codec references | 6 | Multi-mono EVS operated (via FoA) at 3\*7.2, 4\*7.2, 4\*8, 4\*9.6, 4\*16.4, 4\*24.4 kbps with DTX off at 0% FER |
| 3 | mono EVS with unquantized metadata operated at 9.6, 16.4, and 48 kbps with DTX off at 0% FER |
|  |  |  |
| **Other references** |  |  |
| Direct | 2 | Floating and fixed point IVAS\_rend. Nominal input level |
| P.50 MNRU (applied to MASA transport streams) | 3 | Q = 35, 30, 25 dB (all: nominal input level) |
| ESDRU [ITU-T P.811] | 3 | α = 0.8, 0.6, 0.4 (output loudness set to nominal level) |
|  |  |  |
| **Common Conditions** |  |  |
| Test item generation | 1 | Model-based generation according to processing scripts |
| Binaural rendering | 1 | IVAS codec internal binaural renderer and for references IVAS external renderer (IVAS\_rend) |
| Audio sampling frequency / bandwidth | 1 | 48 kHz / maximum available audio bandwidth (WB, SWB, FB) |
| Content types / categories | 6 | Spatial scenes as in selection test P800-8 |
| Number of talkers | 6 | Sentence pairs uttered by different talkers (one each of 3 male and 3 female talkers) |
| Number of speech samples | 7 | 6 for tests + 1 for preliminaries per category |
| Input frequency mask | 1 | Flat |
| Nominal output loudness | 1 | -26 LKFS ([ITU-R BS.1770-4]) |
| Listening Level | 1 | 73 dB SPL |
| Listeners | 30 | Naïve Listeners |
| Randomizations | 6 | 6 panels of 5 listeners |
| Rating Scale | 1 | DCR with instructions according to [P Suppl. 29] |
| Languages | 1 | [tbd] |
| Listening System | 1 | High-quality headphones, diotic presentation |
| Listening Environment | 1 | No noise |

Table F.1. Sample categories

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Environment (impulse responses, idle noise background)*** | ***Level [dB]*** | ***Overtalk [s]*** | ***Talker positions*** | ***Talker selection by panel*** |
| *cat 1* | *room\_1\_MASA (FOA)* | *-26* | *Same as in selection P800-8* | *Same as in selection P800-8* | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *room\_2\_MASA (HOA2)* | *-26* | *Same as in selection P800-8* | *Same as in selection P800-8* | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *room\_3\_MASA (FOA)* | *-26* | *Same as in selection P800-8* | *Same as in selection P800-8* | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *room\_4\_MASA (HOA2)* | *-26* | *Same as in selection P800-8* | *Same as in selection P800-8* | *P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3* |
| *cat 5* | *room\_5\_MASA (FOA)* | *-26* | *Same as in selection P800-8* | *Same as in selection P800-8* | *P1: f3m3 P2: m1f1 P3: f2m2 P4: m3f3 P5: f1m1 P6: m2f2* |
| *cat 6* | *room\_6\_MASA (HOA2)* | *-26* | *Same as in selection P800-8* | *Same as in selection P800-8* | *P1: m2f2 P2: f3m3 P3: m1f1 P4: f2m2 P5: m3f3 P6: f1m1* |

Table F.1. Test conditions for P.SUPPL800 mono MASA in clean speech under clean and impaired channel conditions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **FER/Profile** | **ToR** |
| c01 | Reference Float | - | - |  |
| c02 | Reference Fixed | - | - |  |
| c03 | MNRU Q=35 dB | - | - |  |
| c04 | MNRU Q=30 dB | - | - |  |
| c05 | MNRU Q=25 dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU | - | - |  |
| c09 | IVAS 1TC MASA | 80 | No errors | NWT c18 OR NWT c12 |
| c10 | EVS FoA path | 4x7.2 | No errors |  |
| c11 | EVS FoA path | 4x9.6 | No errors |  |
| c12 | EVS FoA path | 4x24.4 | No errors |  |
| c13 | EVS FoA path | 3x7.2 | No errors |  |
| c14 | EVS FoA path | 4x8 | No errors |  |
| c15 | EVS FoA path | 4x16.4 | No errors |  |
| c16 | EVS (+unquantized metadata) | 9.6 | No errors |  |
| c17 | EVS (+unquantized metadata) | 16.4 | No errors |  |
| c18 | EVS (+unquantized metadata) | 48 | No errors |  |
| c19 | IVAS 1TC MASA | 13.2 | No errors | NWT c13 |
| c20 | IVAS 1TC MASA | 24.4 | No errors | NWT c14 |
| c21 | IVAS 1TC MASA | 48 | No errors | NWT c15 |
| c22 | IVAS 1TC MASA | 16.4 | No errors | NWT c16 OR NWT c10 |
| c23 | IVAS 1TC MASA | 32 | No errors | NWT c17 OR NWT c11 |
| c24 | IVAS 1TC MASA | 64 | No errors | NWT c18 OR NWT c12 |
| [c25 | IVAS 1TC MASA | 13.2 – 32 | Bitrate switching | - |
| c26 | IVAS 1TC MASA | 24.4 – 64 | Bitrate switching | - |
| c27 | IVAS 1TC MASA | 48 – 96 | Bitrate switching | - |
| c28 | IVAS 1TC MASA | 32 | FER 5% | - |
| c29 | IVAS 1TC MASA | 48 | FER 5% | - |
| c30 | IVAS 1TC MASA | 64 | FER 5% | - |
| c31 | IVAS 1TC MASA flt / fxd | 13.2 – 32 | Bitrate switching | NWT c25 |
| c32 | IVAS 1TC MASA fxd / flt | 24.4 – 64 | Bitrate switching | NWT c26 |
| c33 | IVAS 1TC MASA flt / fxd | 48 – 96 | Bitrate switching | NWT c27 |
| c34 | IVAS 1TC MASA fxd / flt | 32 | FER 5% | NWT c28 |
| c35 | IVAS 1TC MASA flt / fxd | 48 | FER 5% | NWT c29 |
| C36 | IVAS 1TC MASA fxd / flt | 64 | FER 5% | NWT c30] |

Editor’s note: Conditions 25-36 need further discussion.

Editor’s note: It has been agreed in the Audio SWG not test bitrate switching in the formal Characterization test. If bitrate switching is included in mono MASA testing, some harmonization should be done with the other experiments.

* 1. Experiment P800-2: Mono MASA experiment Speech + Background

It is proposed for mono-MASA inputs to have following conditions for speech + background test.

Table F.2.1: Overview of test conditions

|  |  |  |
| --- | --- | --- |
| **Main Codec Conditions** |  |  |
| Codec under Test (CuT) | 7 | IVAS operated at 13.2, 16.4, 24.4, 32, 48, 64, 80 kbps with DTX off at 0% FER |
| 3 | IVAS operated at 16.4, 32, 64 kbps with DTX on at 0% FER |
| 3 | IVAS with 16 kHz IO bandwidth at 13.2, 24.4 and 48 kbps |
| 6 | IVAS fixed point / floating point interoperability conditions |
|  |  |  |
| **Codec references** |  |  |
| Codec references | 6 | Multi-mono EVS operated (via FoA) at 3\*7.2, 4\*7.2, 4\*8, 4\*9.6, 4\*16.4, 4\*24.4 kbps with DTX off at 0% FER |
| 3 | mono EVS with unquantized metadata operated at 9.6, 16.4, and 48 kbps with DTX off at 0% FER |
|  |  |  |
| **Other references** |  |  |
| Direct | 2 | Floating and fixed point IVAS\_rend. Nominal input level |
| P.50 MNRU (applied to MASA transport streams) | 3 | Q = 33, 28, 23 dB (all: nominal input level) |
| ESDRU [ITU-T P.811] | 3 | α = 0.8, 0.6, 0.4 (output loudness set to nominal level) |
|  |  |  |
| **Common Conditions** |  |  |
| Test item generation | 1 | Model-based generation according to processing scripts |
| Binaural rendering | 1 | IVAS codec internal binaural renderer and for references IVAS external renderer (IVAS\_rend) |
| Audio sampling frequency / bandwidth | 2 | 48 kHz / maximum available audio bandwidth (WB, SWB, FB) and 16 kHz with WB bandwidth |
| Content types / categories | 6 | Spatial scenes as in selection test P800-9 |
| Number of talkers | 6 | Sentence pairs uttered by different talkers (one each of 3 male and 3 female talkers) |
| Number of speech samples | 7 | 6 for tests + 1 for preliminaries per category |
| Input frequency mask | 1 | Flat |
| Nominal output loudness | 1 | -26 LKFS ([ITU-R BS.1770-4]) |
| Listening Level | 1 | 73 dB SPL |
| Listeners | 30 | Naïve Listeners |
| Randomizations | 6 | 6 panels of 5 listeners |
| Rating Scale | 1 | DCR with instructions according to [P Suppl. 29] |
| Languages | 1 | [tbd] |
| Listening System | 1 | High-quality headphones, diotic presentation |
| Listening Environment | 1 | No noise |

Table F.2.2: Sample Categories

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Environment (Both impulse responses and background noise)*** | ***Level [dB]*** | ***Background*** | ***SNR [dB]*** | ***Overtalk [s]*** | ***Talker positions*** | ***Talker selection by panel*** |
| *cat 1* | *car\_1\_MASA (HOA2)* | *-26* | *car\_1\_bg\_MASA* | *10* | *Same as in selection P800-9* | *Same as in selection P800-9* | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *car\_2\_MASA (FOA)* | *-26* | *car\_2\_bg\_MASA* | *10* | *Same as in selection P800-9* | *Same as in selection P800-9* | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *outdoors\_1\_MASA (FOA)* | *-26* | *outdoors\_1\_bg\_MASA* | *15* | *Same as in selection P800-9* | *Same as in selection P800-9* | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *outdoors\_2\_MASA (HOA2)* | *-26* | *outdoors\_2\_bg\_MASA* | *15* | *Same as in selection P800-9* | *Same as in selection P800-9* | *P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3* |
| *cat 5* | *indoors\_1\_MASA (HOA2)* | *-26* | *indoors\_1\_bg\_MASA* | *15* | *Same as in selection P800-9* | *Same as in selection P800-9* | *P1: f3m3 P2: m1f1 P3: f2m2 P4: m3f3 P5: f1m1 P6: m2f2* |
| *cat 6* | *indoors\_2\_MASA (FOA)* | *-26* | *indoors\_2\_bg\_MASA* | *15* | *Same as in selection P800-9* | *Same as in selection P800-9* | *P1: m2f2 P2: f3m3 P3: m1f1 P4: f2m2 P5: m3f3 P6: f1m1* |

Table F.2.3: Test conditions for P.SUPPL800 mono MASA, speech and background under clean channel conditions with and without DTX

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **FER/Profile/DTX** | **ToR** |
| c01 | Reference Float | - | - |  |
| c02 | Reference Fixed | - | - |  |
| c03 | MNRU Q=33 dB | - | - |  |
| c04 | MNRU Q=28 dB | - | - |  |
| c05 | MNRU Q=23 dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU | - | - |  |
| c09 | IVAS | 80 | No errors | NWT c18 OR NWT c12 |
| c10 | EVS FoA path | 4x7.2 | No errors |  |
| c11 | EVS FoA path | 4x9.6 | No errors |  |
| c12 | EVS FoA path | 4x24.4 | No errors |  |
| c13 | EVS FoA path | 3x7.2 | No errors |  |
| c14 | EVS FoA path | 4x8 | No errors |  |
| c15 | EVS FoA path | 4x16.4 | No errors |  |
| c16 | EVS (+unquantized metadata) | 9.6 | No errors |  |
| c17 | EVS (+unquantized metadata) | 16.4 | No errors |  |
| c18 | EVS (+unquantized metadata) | 48 | No errors |  |
| c19 | IVAS 1TC MASA | 13.2 | No errors | NWT c13 |
| c20 | IVAS 1TC MASA | 24.4 | No errors | NWT c14 |
| c21 | IVAS 1TC MASA | 48 | No errors | NWT c15 |
| c22 | IVAS 1TC MASA | 16.4 | No errors | NWT c16 OR NWT c10 |
| c23 | IVAS 1TC MASA | 32 | No errors | NWT c17 OR NWT c11 |
| c24 | IVAS 1TC MASA | 64 | No errors | NWT c18 OR NWT c12 |
| [c25 | IVAS 1TC MASA | 13.2 | WB IO | - |
| c26 | IVAS 1TC MASA | 24.4 | WB IO | - |
| c27 | IVAS 1TC MASA | 48 | WB IO | - |
| c28 | IVAS 1TC MASA | 32 | DTX | - |
| c29 | IVAS 1TC MASA | 48 | DTX | - |
| c30 | IVAS 1TC MASA | 64 | DTX | - |
| c31 | IVAS 1TC MASA fxd / flt | 13.2 | WB IO | NWT c25 |
| c32 | IVAS 1TC MASA flt / fxd | 24.4 | WB IO | NWT c26 |
| c33 | IVAS 1TC MASA fxd / flt | 48 | WB IO | NWT c27 |
| c34 | IVAS 1TC MASA flt / fxd | 32 | DTX | NWT c28 |
| c35 | IVAS 1TC MASA fxd / flt | 48 | DTX | NWT c29 |
| c36 | IVAS 1TC MASA flt / fxd | 64 | DTX | NWT c30] |

Editor’s note: Conditions 25-36 need further discussion.

Editor’s note: Concerning the WB IO conditions, it might be better to define a WB bandwidth input as a separate sample category to avoid assessing WB conditions wrt FB reference.

Editor’s note: If the proposal in section 4.3 is agreed, the description of these experiments has to be aligned with the proposal.

]

[

* 1. Experiment P800-1: Stereo
     1. Experiment setup

Tables E.1.1 to E.1.3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table E.1.1: Conditions for Experiment P800-1

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 46, 80, 96, 128 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | HP50 |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 0% |
|  |  |
| **Codec references** |  |
| Codec references  Bitrates | EVS - Mono signal generated with IVAS Pre-renderer  13.2, 16.4, 24.4, 32, 48, 64, 96, 128 kbps |
| Input level  DTX | -26 LKFS  DTX off |
| Input frequency mask | HP50 |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 0% |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx, xx |
| Input frequency mask | HP50 |
| **Common Conditions** |  |
| Test item generation: | Cat. 1-4: 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 [27] and impulse responses provided by MC. Cat. 5-6: Pre-produced content |
| 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 XX |
| Listening System | Headphones, in accordance with clause XX |
| Listening Environment | No room noise, in accordance with clause XX |
| Languages | TBD |

Table E.1.2: Preliminaries for Experiment P800-1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c02 |  | Mono | - | - |
| 3 | c07 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c10 |  | ESDRU = xx | - | - |
| 6 | c11 |  | EVS | 13.2 | No errors |
| 7 | c04 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c18 |  | EVS | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table E.1.3: Test conditions for Experiment P800-1,  
stereo speech and mixed music under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | Mono | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | MNRU Q=xx dB | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU | - | - |
| c09 | ESDRU |  |  |
| c10 | ESDRU | - | - |
| c11 | EVS | 1x13.2 |  |
| c12 | EVS | 1x16.4 |  |
| c13 | EVS | 1x24.4 |  |
| c14 | EVS | 1x32.0 |  |
| c15 | EVS | 1x48.0 |  |
| c16 | EVS | 1x64.0 |  |
| c17 | EVS | 1x96.0 |  |
| c18 | EVS | 1x128.0 |  |
| c19 | IVAS FL | 13.2 |  |
| c20 | IVAS FL | 16.4 |  |
| c21 | IVAS FL | 24.4 |  |
| c22 | IVAS FL | 32.0 |  |
| c23 | IVAS FL | 48.0 |  |
| c24 | IVAS FL | 64.0 |  |
| c25 | IVAS FL | 80.0 |  |
| c26 | IVAS FL | 96.0 |  |
| c27 | IVAS FL | 128.0 |  |
| c28 | IVAS FX | 13.2 |  |
| c29 | IVAS FX | 16.4 |  |
| c30 | IVAS FX | 24.4 |  |
| c31 | IVAS FX | 32.0 |  |
| c32 | IVAS FX | 48.0 |  |
| c33 | IVAS FX | 64.0 |  |
| c34 | IVAS FX | 80.0 |  |
| c35 | IVAS FX | 96.0 |  |
| c36 | IVAS FX | 128.0 |  |

* + 1. Content type categories and scene definitions

**Clean and noisy speech categories**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Category | Room | Reverb | Microphone Setup | ***Background*** | SNR  [dB] | Overtalk [s](2 | Bandwidth( | Talker positions(3 | 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 | echoic | 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 | echoic | Binaural | 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 | 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: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3 |

**Mixed music categories**

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 | mixed content |
| cat 6 | music |

* 1. Experiment P800-2: FOA
     1. Experiment setup

Tables E.2.1 to E.2.3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table E.2.1: Conditions for Experiment P800-2

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 46, 80, 96, 128 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | HP50 |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 0% |
|  |  |
| **Codec references** |  |
| Codec references  Bitrates | EVS - Mono signal generated with IVAS Pre-renderer  13.2, 16.4, 24.4, 32, 48, 64, 96, 128 kbps |
| Input level  DTX | -26 LKFS  DTX off |
| Input frequency mask | HP50 |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 0% |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | HP50 |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Cat. 1-4: 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. Cat. 5-6: Pre-produced content |
| Binaural renderer | FOA to binaural internal 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 or without background) 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 XX |
| Listening System | Headphones, in accordance with clause XX |
| Listening Environment | No room noise, in accordance with clause XX |

Table E.2.2: Preliminaries for Experiment P800-2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c02 |  | Mono | - | - |
| 3 | c07 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c10 |  | ESDRU = xx | - | - |
| 6 | C11 |  | EVS | 13.2 | No errors |
| 7 | c04 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c18 |  | EVS | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table E.2.3: Test conditions for Experiment P800-2,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | Mono | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | MNRU Q=xx dB | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU | - | - |
| c09 | ESDRU |  |  |
| c10 | ESDRU | - | - |
| c11 | EVS | 1x13.2 |  |
| c12 | EVS | 1x16.4 |  |
| c13 | EVS | 1x24.4 |  |
| c14 | EVS | 1x32.0 |  |
| c15 | EVS | 1x48.0 |  |
| c16 | EVS | 1x64.0 |  |
| c17 | EVS | 1x96.0 |  |
| c18 | EVS | 1x128.0 |  |
| c19 | IVAS FL | 13.2 |  |
| c20 | IVAS FL | 16.4 |  |
| c21 | IVAS FL | 24.4 |  |
| c22 | IVAS FL | 32.0 |  |
| c23 | IVAS FL | 48.0 |  |
| c24 | IVAS FL | 64.0 |  |
| c25 | IVAS FL | 80.0 |  |
| c26 | IVAS FL | 96.0 |  |
| c27 | IVAS FL | 128.0 |  |
| c28 | IVAS FX | 13.2 |  |
| c29 | IVAS FX | 16.4 |  |
| c30 | IVAS FX | 24.4 |  |
| c31 | IVAS FX | 32.0 |  |
| c32 | IVAS FX | 48.0 |  |
| c33 | IVAS FX | 64.0 |  |
| c34 | IVAS FX | 80.0 |  |
| c35 | IVAS FX | 96.0 |  |
| c36 | IVAS FX | 128.0 |  |

* + 1. 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* | *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* | *out\_[X]\_FOA* | *[park\_1\_bg\_FOA / nature\_1\_bg\_FOA / event\_1\_bg\_FOA / street\_[1/2]\_bg\_FOA]* | *15* | *1* | *Max* |  | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *room\_[X]\_FOA* | *[cafeteria\_1\_bg\_FOA / mall\_1\_bg\_FOA/ office[1/2]\_bg\_FOA]* | *15* | *-1* | *Max* |  | *P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3* |

**Mixed music categories**

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 | mixed content |
| cat 6 | Generic audio |

**Notes:**

**(1** Editor’s note: The specific room/environment characteristic and resulting reverb characteristic will be defined by the choice of the specific Spatial Room Impulse Responses used in the convolution process with the raw mono sentences, according to the pertaining stipulations of the test plan IVAS-8a.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8a. Backround name ‘clean\_bg\_[X]\_FOA’ indicates a low-noise background corresponding to environment [X], e.g., with low air-conditioning/fan noise.

**(3** 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.

**(4** Editor’s note: The talker positions are part of the scene definition of the different categories. They should be chosen in a way from the available set of SRIRs for the used room making sure that there is a good coverage of different possible positions. Different selections should be made for the different listener panels.

* 1. Experiment P800-3: MC 5-1
     1. Experiment setup

Tables E.3.1 to E.3.3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table E.3.1: Conditions for Experiment P800-3

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 46, 80, 96, 128 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0% |
|  |  |
| **Codec references** |  |
| Codec references  Bitrates | EVS - Mono signal generated with IVAS Pre-renderer  13.2, 16.4, 24.4, 32, 48, 64, 96, 128 kbps |
| Input level  DTX | -26 LKFS  DTX off |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0% |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | HP50 |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Cat. 1-4: 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, converted to 5.1 using IVAS Pre-renderer. Cat. 5-6: Pre-produced native 5.1 content |
| Binaural renderer | MC to binaural internal rendering |
| Audio sampling frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Generic audio content as described in clause XX |
| Number of categories | 6 Different environments (with or without background) 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 XX |
| Listening System | Headphones, in accordance with clause XX |
| Listening Environment | No room noise, in accordance with clause XX |

Table E.3.2: Preliminaries for Experiment P800-3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c02 |  | Mono | - | - |
| 3 | c07 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c10 |  | ESDRU = xx | - | - |
| 6 | c11 |  | EVS | 13.2 | No errors |
| 7 | c04 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c18 |  | EVS | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table E.3.3: Test conditions for Experiment P800-3,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | Mono | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | MNRU Q=xx dB | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU | - | - |
| c09 | ESDRU |  |  |
| c10 | ESDRU | - | - |
| c11 | EVS | 1x13.2 |  |
| c12 | EVS | 1x16.4 |  |
| c13 | EVS | 1x24.4 |  |
| c14 | EVS | 1x32.0 |  |
| c15 | EVS | 1x48.0 |  |
| c16 | EVS | 1x64.0 |  |
| c17 | EVS | 1x96.0 |  |
| c18 | EVS | 1x128.0 |  |
| c19 | IVAS FL | 13.2 |  |
| c20 | IVAS FL | 16.4 |  |
| c21 | IVAS FL | 24.4 |  |
| c22 | IVAS FL | 32.0 |  |
| c23 | IVAS FL | 48.0 |  |
| c24 | IVAS FL | 64.0 |  |
| c25 | IVAS FL | 80.0 |  |
| c26 | IVAS FL | 96.0 |  |
| c27 | IVAS FL | 128.0 |  |
| c28 | IVAS FX | 13.2 |  |
| c29 | IVAS FX | 16.4 |  |
| c30 | IVAS FX | 24.4 |  |
| c31 | IVAS FX | 32.0 |  |
| c32 | IVAS FX | 48.0 |  |
| c33 | IVAS FX | 64.0 |  |
| c34 | IVAS FX | 80.0 |  |
| c35 | IVAS FX | 96.0 |  |
| c36 | IVAS FX | 128.0 |  |

* + 1. 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* | *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* | *out\_[X]\_FOA* | *[park\_1\_bg\_FOA / nature\_1\_bg\_FOA / event\_1\_bg\_FOA / street\_[1/2]\_bg\_FOA]* | *15* | *1* | *Max* |  | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *room\_[X]\_FOA* | *[cafeteria\_1\_bg\_FOA / mall\_1\_bg\_FOA/ office[1/2]\_bg\_FOA]* | *15* | *-1* | *Max* |  | *P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3* |

**Mixed music categories**

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 | mixed content |
| cat 6 | music |

**Notes:**

**(1** Editor’s note: The specific room/environment characteristic and resulting reverb characteristic will be defined by the choice of the specific Spatial Room Impulse Responses used in the convolution process with the raw mono sentences, according to the pertaining stipulations of the test plan IVAS-8a.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8a. Background name ‘clean\_bg\_[X]\_FOA’ indicates a low-noise background corresponding to environment [X], e.g., with low air-conditioning/fan noise.

**(3** 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.

**(4** Editor’s note: The talker positions are part of the scene definition of the different categories. They should be chosen in a way from the available set of SRIRs for the used room making sure that there is a good coverage of different possible positions. Different selections should be made for the different listener panels.

* 1. Experiment P800-4: 1-4 Objects
     1. Experiment setup

Tables E.4.1 to E.4.3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table E.4.1: Conditions for Experiment P800-4

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 46, 80, 96, 128 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0% |
|  |  |
| **Codec references** |  |
| Codec references  Bitrates | EVS - Mono signal generated with IVAS Pre-renderer  13.2, 16.4, 24.4, 32, 48, 64, 96, 128 kbps |
| Input level  DTX | -26 LKFS  DTX off |
| Input frequency mask | HP50 |
| Noise | No noise |
| Error Conditions | 0% |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | HP50 |
| **Common Conditions** |  |
| Test item generation: | Cat. 1-2: Defined scenes, 1 ISM  Cat. 3-4: Defined scenes, 2 ISMs Cat. 5-6: Pre-produced content, 3-4 ISMs |
| Binaural renderer | ISM to binaural internal 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 or without background) 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 XX |
| Listening System | Headphones, in accordance with clause XX |
| Listening Environment | No room noise, in accordance with clause XX |

Table E.4.2: Preliminaries for Experiment P800-4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c02 |  | Mono | - | - |
| 3 | c07 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c10 |  | ESDRU = xx | - | - |
| 6 | c11 |  | EVS | 13.2 | No errors |
| 7 | c04 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c18 |  | EVS | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table E.4.3: Test conditions for Experiment P800-4,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | Mono | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | MNRU Q=xx dB | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU | - | - |
| c09 | ESDRU |  |  |
| c10 | ESDRU | - | - |
| c11 | EVS | 1x13.2 |  |
| c12 | EVS | 1x16.4 |  |
| c13 | EVS | 1x24.4 |  |
| c14 | EVS | 1x32.0 |  |
| c15 | EVS | 1x48.0 |  |
| c16 | EVS | 1x64.0 |  |
| c17 | EVS | 1x96.0 |  |
| c18 | EVS | 1x128.0 |  |
| c19 | IVAS FL | 13.2 |  |
| c20 | IVAS FL | 16.4 |  |
| c21 | IVAS FL | 24.4 |  |
| c22 | IVAS FL | 32.0 |  |
| c23 | IVAS FL | 48.0 |  |
| c24 | IVAS FL | 64.0 |  |
| c25 | IVAS FL | 80.0 |  |
| c26 | IVAS FL | 96.0 |  |
| c27 | IVAS FL | 128.0 |  |
| c28 | IVAS FX | 13.2 |  |
| c29 | IVAS FX | 16.4 |  |
| c30 | IVAS FX | 24.4 |  |
| c31 | IVAS FX | 32.0 |  |
| c32 | IVAS FX | 48.0 |  |
| c33 | IVAS FX | 64.0 |  |
| c34 | IVAS FX | 80.0 |  |
| c35 | IVAS FX | 96.0 |  |
| c36 | IVAS FX | 128.0 |  |

* + 1. Content type categories and scene definitions

**Scene definitions categories 1-2**

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:

1. Talker sitting at a table (elevation 0°), at different azimuths.
2. Standing talker (elevation 35°), at different azimuths.
3. Smaller talker (child) walking around a table in the positive sense (counterclockwise), elevation 0°. Azimuth varies continuously for the sentence pair.
4. Adult talker walking around a table in the negative sense (clockwise), elevation 35°. Azimuth varies continuously for the sentence pair.
5. Elevation displacement: Elevation varies continuously for the sentence pair . Azimuth is constant for a sentence pair, but different for each sentence pair.
6. 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 E 4.4

**Scene definitions categories 3-4**

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:

1. 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.
2. 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.
3. One talker sitting at a table (elevation 0°), second talker standing beside the table (elevation 45°). Non-overlapping utterances.
4. 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.
5. Two talkers walking side-by-side around the table (elevation 45°). The azimuth is the same for both talkers and varies continually.
6. 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 E.4.4: Allocation of scenes for each talker pair (category cat 1 – cat 4) and listening panel (P1-P6)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Talker initial elevation*** | ***Elevation change(1*** | ***Talker initial azimuth*** | ***Azimuth change(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 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Overtalk***  ***[s]* (1** | ***1st talker elevation*** | ***2nd talker elevation*** | ***1st talker initial azimuth*** | ***1st talker azimuth change(2*** | ***2nd talker initial azimuth*** | ***2nd talker azimuth change(2*** | ***Panel*** |
| ***cat 3:***  *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 4:***  *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* |

**Mixed music categories**

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 | Speech + effects (3-4 objects) |
| cat 6 | Speech + music or music only (3-4 objects) |

**Notes:**

(1Overtalk [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.

(2 The positive sense for azimuth is counterclockwise

* 1. Experiment P800-5: MASA 1 TC
     1. Experiment setup

Tables E.5.1 to E.5.3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table E.5.1: Conditions for Experiment P800-5

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 46, 80, 96, 128 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | HP50 |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 0% |
|  |  |
| **Codec references** |  |
| Codec references  Bitrates | EVS - Mono signal generated with IVAS Pre-renderer  13.2, 16.4, 24.4, 32, 48, 64, 96, 128 kbps |
| Input level  DTX | -26 LKFS  DTX off |
| Input frequency mask | HP50 |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 0% |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| 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 [28]. |
| Binaural renderer | IVAS MASA internal 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 4.4 |
| Listening System | Headphones, in accordance with clause 4.6 |
| Listening Environment | No room noise, in accordance with clause 4.6 |

Table E.5.2: Preliminaries for Experiment P800-5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c02 |  | Mono | - | - |
| 3 | c07 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c10 |  | ESDRU = xx | - | - |
| 6 | c11 |  | EVS | 13.2 | No errors |
| 7 | c04 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c18 |  | EVS | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table E.5.3: Test conditions for Experiment P800-5,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | Mono | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | MNRU Q=xx dB | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU | - | - |
| c09 | ESDRU |  |  |
| c10 | ESDRU | - | - |
| c11 | EVS | 1x13.2 |  |
| c12 | EVS | 1x16.4 |  |
| c13 | EVS | 1x24.4 |  |
| c14 | EVS | 1x32.0 |  |
| c15 | EVS | 1x48.0 |  |
| c16 | EVS | 1x64.0 |  |
| c17 | EVS | 1x96.0 |  |
| c18 | EVS | 1x128.0 |  |
| c19 | IVAS FL | 13.2 |  |
| c20 | IVAS FL | 16.4 |  |
| c21 | IVAS FL | 24.4 |  |
| c22 | IVAS FL | 32.0 |  |
| c23 | IVAS FL | 48.0 |  |
| c24 | IVAS FL | 64.0 |  |
| c25 | IVAS FL | 80.0 |  |
| c26 | IVAS FL | 96.0 |  |
| c27 | IVAS FL | 128.0 |  |
| c28 | IVAS FX | 13.2 |  |
| c29 | IVAS FX | 16.4 |  |
| c30 | IVAS FX | 24.4 |  |
| c31 | IVAS FX | 32.0 |  |
| c32 | IVAS FX | 48.0 |  |
| c33 | IVAS FX | 64.0 |  |
| c34 | IVAS FX | 80.0 |  |
| c35 | IVAS FX | 96.0 |  |
| c36 | IVAS FX | 128.0 |  |

* + 1. 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* | *room\_1\_MASA* | *room\_1\_cleanbg\_MASA* | *45* | *1* | *Max* | *tbd* | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *room\_4\_MASA* | *room\_4\_cleanbg\_MASA* | *45* | *-1* | *Max* | *tbd* | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *out\_[1/2]\_MASA* | *[park\_1\_bg\_MASA / nature\_1\_bg\_MASA / event\_1\_bg\_MASA / street\_[1/2]\_bg\_MASA]* | *15* | *-1* | *Max* | *tbd* | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *room\_[X]\_MASA* | *[cafeteria\_1\_bg\_MASA / mall\_1\_bg\_MASA/ office[1/2]\_bg\_MASA]* | *15* | *-1* | *Max* | *tbd* | *P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3* |

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 | mixed content |
| cat 6 | Generic audio |

**Notes:**

**(1** Editor’s note: The specific room/environment characteristic and resulting reverb characteristic will be defined by the choice of the specific Spatial Room Impulse Responses used in the convolution process with the raw mono sentences, according to the pertaining stipulations of the test plan IVAS-8a.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8a. Backround name ‘clean\_bg\_[X]\_FOA’ indicates a low-noise background corresponding to environment [X], e.g., with low air-conditioning/fan noise.

**(3** 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.

**(4** Editor’s note: The talker positions are part of the scene definition of the different categories. They should be chosen in a way from the available set of SRIRs for the used room making sure that there is a good coverage of different possible positions. Different selections should be made for the different listener panels.

* 1. Experiment P800-6: OSBA (1-4 objects)
     1. Experiment setup

Tables E.6.1 to E.6.3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively.

Table E.6.1: Conditions for Experiment P800-6

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 46, 80, 96, 128 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | HP50 |
| Noise | 15 dB for cat 1,2,3,4, tbd for cat 5,6 |
| Error Conditions | 0% |
|  |  |
| **Codec references** |  |
| Codec references  Bitrates | EVS - Mono signal generated with IVAS Pre-renderer  13.2, 16.4, 24.4, 32, 48, 64, 96, 128 kbps |
| Input level  DTX | -26 LKFS  DTX off |
| Input frequency mask | HP50 |
| Noise | 15 dB for cat 1,2,3,4, tbd for cat 5,6 |
| Error Conditions | 0% |
|  |  |
| **Other references** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | HP50 |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Cat. 1-2: Defined scenes, 1 ISM + FOA [or/and HOA3 if available] background  Cat. 3-4: Defined scenes, 2 ISMs + FOA [or/and HOA3 if available] background Cat. 5-6: Pre-produced content, 3-4 ISMs + FOA [or/and HOA3 if available] background |
| Binaural renderer | OSBA to binaural internal 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 or without background) 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 XX |
| Listening System | Headphones, in accordance with clause XX |
| Listening Environment | No room noise, in accordance with clause XX |

Table E.6.2: Preliminaries for Experiment P800-6

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c02 |  | Mono | - | - |
| 3 | c07 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c10 |  | ESDRU = xx | - | - |
| 6 | c11 |  | EVS | 13.2 | No errors |
| 7 | c04 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c18 |  | EVS | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table E.6.3: Test conditions for Experiment P800-6,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | Mono | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | MNRU Q=xx dB | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU | - | - |
| c09 | ESDRU |  |  |
| c10 | ESDRU | - | - |
| c11 | EVS | 1x13.2 |  |
| c12 | EVS | 1x16.4 |  |
| c13 | EVS | 1x24.4 |  |
| c14 | EVS | 1x32.0 |  |
| c15 | EVS | 1x48.0 |  |
| c16 | EVS | 1x64.0 |  |
| c17 | EVS | 1x96.0 |  |
| c18 | EVS | 1x128.0 |  |
| c19 | IVAS FL | 13.2 |  |
| c20 | IVAS FL | 16.4 |  |
| c21 | IVAS FL | 24.4 |  |
| c22 | IVAS FL | 32.0 |  |
| c23 | IVAS FL | 48.0 |  |
| c24 | IVAS FL | 64.0 |  |
| c25 | IVAS FL | 80.0 |  |
| c26 | IVAS FL | 96.0 |  |
| c27 | IVAS FL | 128.0 |  |
| c28 | IVAS FX | 13.2 |  |
| c29 | IVAS FX | 16.4 |  |
| c30 | IVAS FX | 24.4 |  |
| c31 | IVAS FX | 32.0 |  |
| c32 | IVAS FX | 48.0 |  |
| c33 | IVAS FX | 64.0 |  |
| c34 | IVAS FX | 80.0 |  |
| c35 | IVAS FX | 96.0 |  |
| c36 | IVAS FX | 128.0 |  |

* + 1. Content type categories and scene definitions

ISM positions for categories cat 1-4 as defined respectively in Table E.4.4.

SBA scenes as follows:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Environment(1*** | ***Background(2*** | ***SNR [dB]*** | ***Overtalk [s](3*** | ***Bandwidth*** | ***Talker positions(4*** | ***Talker selection by panel(5*** |
| *cat 1* | *out\_[X]\_FOA* | *[park\_1\_bg\_FOA / nature\_1\_bg\_FOA / event\_1\_bg\_FOA / street\_[1/2]\_bg\_FOA]* | *15* | *1* | *Max* |  | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *room\_[X]\_FOA* | *[cafeteria\_1\_bg\_FOA / mall\_1\_bg\_FOA/ office[1/2]\_bg\_FOA]* | *15* | *-1* | *Max* |  | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *out\_[X]\_FOA* | *[park\_1\_bg\_FOA / nature\_1\_bg\_FOA / event\_1\_bg\_FOA / street\_[1/2]\_bg\_FOA]* | *15* | *1* | *Max* |  | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *room\_[X]\_FOA* | *[cafeteria\_1\_bg\_FOA / mall\_1\_bg\_FOA/ office[1/2]\_bg\_FOA]* | *15* | *-1* | *Max* |  | *P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3* |

**Mixed music categories**

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 | 3-object + General audio background |
| cat 6 | 4-objects + General audio background |

**Notes:**

**(1** Editor’s note: The specific room/environment characteristic and resulting reverb characteristic will be defined by the choice of the specific Spatial Room Impulse Responses used in the convolution process with the raw mono sentences, according to the pertaining stipulations of the test plan IVAS-8a.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8a. Backround name ‘clean\_bg\_[X]\_FOA’ indicates a low-noise background corresponding to environment [X], e.g., with low air-conditioning/fan noise.

**(3** 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.

**(4** Editor’s note: The talker positions are part of the scene definition of the different categories. They should be chosen in a way from the available set of SRIRs for the used room making sure that there is a good coverage of different possible positions. Different selections should be made for the different listener panels.

]

[

* 1. Experiment P800-x: OSBA

Table 1 Overview of test conditions

|  |  |  |
| --- | --- | --- |
| **Main Codec Conditions** |  |  |
| Codec under Test (CuT) | 15 | IVAS OSBA operated at all bitrates 13.2- 512 kbps with DTX off at 0% FER |
| 3 | IVAS OSBA operated at 32, 64, 256 kbps with 5% FER |
| 4 | IVAS OSBA fixed point / floating point interoperability conditions at 32, 64, 128 and 256 kbps |
| **Codec references** |  |  |
| Codec references | 6 | IVAS operation in two separate instances (SBA + ISM) |
| **Other references** |  |  |
| Direct | 1 | Fixed point IVAS\_rend. Nominal input level |
| P.50 MNRU (applied to SBA transport streams) | 4 | Q = 34, 30, 26, 22 dB (all: nominal input level) |
| ESDRU [ITU-T P.811] | 4 | α = 0.8, 0.6, 0.4, 0.2 (output loudness set to nominal level) |
| **Common Conditions** |  |  |
| Test item generation | 1 | Model-based generation according to processing scripts. |
| Binaural rendering | 1 | IVAS codec internal binaural renderer and for references IVAS external renderer (IVAS\_rend) |
| Audio sampling frequency / bandwidth | 1 | 48 kHz / maximum available audio bandwidth (WB, SWB, FB) |
| Content types / categories | 6 | Scenes as described in Table2 of [1] |
| Number of talkers | 6 | ISM signals are generated so that different objects are uttered by different talkers (one each of 3 male and 3 female talkers). All talkers are tried to be used equal amount. |
| Number of speech samples | 7 | 6 for tests + 1 for preliminaries per category |
| Input frequency mask | 1 | Flat |
| Nominal output loudness | 1 | -26 LKFS ([ITU-R BS.1770-4]) |
| Background signal (SBA) input loudness | 1 | -36 LKFS ([ITU-R BS.1770-4]) |
| ISM speech input loudness | 1 | -26 LKFS ([ITU-R BS.1770-4]) |
| Listening Level | 1 | 73 dB SPL |
| Listeners | 30 | Naïve Listeners |
| Randomizations | 6 | 6 panels of 5 listeners |
| Rating Scale | 1 | DCR with instructions according to IVAS 8a [3] |
| Languages | 1 | [tbd] |
| Listening System | 1 | High-quality headphones, diotic presentation |
| Listening Environment | 1 | No noise |

Table 2 Sample Categories

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Number of objects*** | ***Speech Level [dB]*** | ***Background signal type\*\**** | ***Background Level*** | ***Overtalk [s]*** | ***Talker positions*** |
| *cat 1* | *1* | *-26* | *Indoors 1* | *[-36]* | *No overtalk* | *2 fixed, 4 with movement* |
| *cat 2* | *2* | *-26* | *Indoors 2* | *[-36]* | *Overtalk* | *2 fixed, 4 with movement\** |
| *cat 3* | *3* | *-26* | *Outdoors 1* | *[-36]* | *Overtalk* | *2 fixed, 4 with movement\** |
| *cat 4* | *4* | *-26* | *Outdoors 2* | *[-36]* | *Overtalk* | *2 fixed, 4 with movement\** |
| *cat 5* | *2* | *-26* | *Background with music 1* | *[-36]* | *No overtalk* | *2 fixed, 4 with movement\** |
| *cat 6* | *3* | *-26* | *Background with music 2* | *[-36]* | *Overtalk* | *2 fixed, 4 with movement\** |

\*for 2 samples one ISM is moving, for the last 2 samples two or more objects are moving. For practice sample one ISM is moving.

\*\* Background type signal is HOA3

Table 3 Test conditions for P.SUPPL800 OSBA, speech and background environments.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **FER/float** | **Notes** |
| c01 | Reference IVAS\_rend | - | - |  |
| c02 | MNRU Q=34 dB | - | - |  |
| c03 | MNRU Q=30 dB | - | - |  |
| c04 | MNRU Q=26 dB | - | - |  |
| c05 | MNRU Q=22 dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU | - | - |  |
| c09 | ESDRU | - | - |  |
| c10 | OSBA | 32 | Float |  |
| c11 | OSBA | 64 | Float |  |
| c12 | OSBA | 128 | Float |  |
| c13 | OSBA | 256 | Float |  |
| c14 | OSBA | 32 | 5% FER |  |
| c15 | OSBA | 64 | 5% FER |  |
| c16 | OSBA | 256 | 5% FER |  |
| c17 | OSBA | 13.2 | No errors |  |
| c18 | OSBA | 16.4 | No errors |  |
| c19 | OSBA | 24.4 | No errors |  |
| c20 | OSBA | 32 | No errors |  |
| c21 | OSBA | 80 | No errors |  |
| c22 | OSBA | 160 | No errors |  |
| c23 | OSBA | 384 | No errors |  |
| c24 | OSBA | 512 | No errors |  |
| c25 | OSBA | 48 | No errors | C31 |
| c26 | OSBA | 64 | No errors | C32 |
| c27 | OSBA | 96 | No errors | C33 |
| c28 | OSBA | 128 | No errors | C34 |
| c29 | OSBA | 192 | No errors | C35 |
| c30 | OSBA | 256 | No errors | C36 |
| c31 | ISM + SBA (HOA3) | 24.4 + 24.4 | No errors |  |
| c32 | ISM + SBA (HOA3) | 32 + 32 | No errors |  |
| c33 | ISM + SBA (HOA3) | 48 + 48 | No errors |  |
| c34 | ISM + SBA (HOA3) | 64 + 64 | No errors |  |
| c35 | ISM + SBA (HOA3) | 96 + 96 | No errors |  |
| c36 | ISM + SBA (HOA3) | 128 + 128 | No errors |  |

* 1. Experiment P800-x: OMASA

Table 1 Overview of test conditions

|  |  |  |
| --- | --- | --- |
| **Main Codec Conditions** |  |  |
| Codec under Test (CuT) | 15 | IVAS oMASA operated at all bitrates 13.2- 512 kbps with DTX off at 0% FER |
| 3 | IVAS oMASA operated at 32, 64, 128 kbps with 5% FER |
| 4 | IVAS oMASA fixed point / floating point interoperability conditions at 32, 64, 128 and 256 kbps |
| **Codec references** |  |  |
| Codec references | 6 | IVAS operation in two separate instances (MASA + ISM) |
| **Other references** |  |  |
| Direct | 1 | Fixed point IVAS\_rend. Nominal input level |
| P.50 MNRU (applied to MASA transport streams) | 4 | Q = 34, 30, 26, 22 dB (all: nominal input level) |
| ESDRU [ITU-T P.811] | 4 | α = 0.8, 0.6, 0.4, 0.2 (output loudness set to nominal level) |
| **Common Conditions** |  |  |
| Test item generation | 1 | Model-based generation according to processing scripts. |
| Binaural rendering | 1 | IVAS codec internal binaural renderer and for references IVAS external renderer (IVAS\_rend) |
| Audio sampling frequency / bandwidth | 1 | 48 kHz / maximum available audio bandwidth (WB, SWB, FB) |
| Content types / categories | 6 | Scenes as described in Table2 |
| Number of talkers | 6 | ISM signals are generated so that different objects are uttered by different talkers (one each of 3 male and 3 female talkers). All talkers are tried to be used equal amount. |
| Number of speech samples | 7 | 6 for tests + 1 for preliminaries per category |
| Input frequency mask | 1 | Flat |
| Nominal output loudness | 1 | -26 LKFS ([ITU-R BS.1770-4]) |
| Background signal (MASA) input loudness | 1 | -36 LKFS ([ITU-R BS.1770-4]) |
| ISM speech input loudness | 1 | -26 LKFS ([ITU-R BS.1770-4]) |
| Listening Level | 1 | 73 dB SPL |
| Listeners | 30 | Naïve Listeners |
| Randomizations | 6 | 6 panels of 5 listeners |
| Rating Scale | 1 | DCR with instructions according to [P Suppl. 29] |
| Languages | 1 | [tbd] |
| Listening System | 1 | High-quality headphones, diotic presentation |
| Listening Environment | 1 | No noise |

Table 2 Sample Categories

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Number of objects*** | ***Speech Level [dB]*** | ***Background signal type*** | ***Background Level*** | ***Overtalk [s]*** | ***Talker positions*** |
| *cat 1* | *1* | *-26* | *Indoors 1* | *[-36]* | *No overtalk* | *2 fixed, 4 with movement* |
| *cat 2* | *2* | *-26* | *Indoors 2* | *[-36]* | *Overtalk* | *2 fixed, 4 with movement\** |
| *cat 3* | *3* | *-26* | *Outdoors 1* | *[-36]* | *Overtalk* | *2 fixed, 4 with movement\** |
| *cat 4* | *4* | *-26* | *Outdoors 2* | *[-36]* | *Overtalk* | *2 fixed, 4 with movement\** |
| *cat 5* | *2* | *-26* | *Background with music 1* | *[-36]* | *No overtalk* | *2 fixed, 4 with movement\** |
| *cat 6* | *3* | *-26* | *Background with music 2* | *[-36]* | *Overtalk* | *2 fixed, 4 with movement\** |

\*for 2 samples one ISM is moving, for the last 2 samples two or more objects are moving. For practice sample one ISM is moving.

Table 3 Test conditions for P.SUPPL800 oMASA, speech and background environments.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **FER/float** | **Notes** |
| c01 | Reference IVAS\_rend | - | - |  |
| c02 | MNRU Q=34 dB | - | - |  |
| c03 | MNRU Q=30 dB | - | - |  |
| c04 | MNRU Q=26 dB | - | - |  |
| c05 | MNRU Q=22 dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU | - | - |  |
| c09 | ESDRU | - | - |  |
| c10 | OMASA 2TC | 32 | Float |  |
| c11 | OMASA 2TC | 64 | Float |  |
| c12 | OMASA 2TC | 128 | Float |  |
| c13 | OMASA 2TC | 256 | Float |  |
| c14 | OMASA 2TC | 32 | 5% FER |  |
| c15 | OMASA 2TC | 64 | 5% FER |  |
| c16 | OMASA 2TC | 128 | 5% FER |  |
| c17 | OMASA 2TC | 13,2 | No errors |  |
| c18 | OMASA 2TC | 16,4 | No errors |  |
| c19 | OMASA 2TC | 24,4 | No errors |  |
| c20 | OMASA 2TC | 32 | No errors |  |
| c21 | OMASA 2TC | 80 | No errors |  |
| c22 | OMASA 2TC | 160 | No errors |  |
| c23 | OMASA 2TC | 384 | No errors |  |
| c24 | OMASA 2TC | 512 | No errors |  |
| c25 | OMASA 2TC | 48 | No errors | C31 |
| c26 | OMASA 2TC | 64 | No errors | C32 |
| c27 | OMASA 2TC | 96 | No errors | C33 |
| c28 | OMASA 2TC | 128 | No errors | C34 |
| c29 | OMASA 2TC | 192 | No errors | C35 |
| c30 | OMASA 2TC | 256 | No errors | C36 |
| c31 | ISM + MASA | 24,4 + 24,4 | No errors |  |
| c32 | ISM + MASA | 32 + 32 | No errors |  |
| c33 | ISM + MASA | 48 + 48 | No errors |  |
| c34 | ISM + MASA | 64 + 64 | No errors |  |
| c35 | ISM + MASA | 96 + 96 | No errors |  |
| c36 | ISM + MASA | 128 + 128 | No errors |  |

]

1. BS.1534 Experiments
   1. Experiment BS1534-1a: Stereo



Conditions (BS1534-1a Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with stereo audio input at*  *16.4 kbps, 24.4 kbps, 32 kbps, 48 kbps DTX off at 0% FER* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *16.4 kbps, 32 kbps DTX off at 0% FER*  *Mono signal generated with [IVAS Pre-renderer or EVS compatible Mono DMX]* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

Test conditions for Experiment BS1534-1a

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **ToR** |
| c01 | Reference | - | - |
| c02 | LP7 anchor | - | - |
| c03 |  |  |  |
| c04 |  |  |  |
| c05 |  |  |  |
| c06 |  |  |  |
| c07 |  |  |  |

* 1. Experiment BS1534-1b: Stereo

Conditions (BS1534-1b Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with stereo audio input at*  *64 kbps, 96 kbps, 128 kbps, 256 kbps DTX off at 0% FER* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *64 kbps, 128 kbps DTX off at 0% FER*  *Mono signal generated with [IVAS Pre-renderer or EVS compatible Mono DMX]* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-2a: FOA

Conditions (BS1534-2a Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with FOA audio input at*  *16.4 kbps, 24.4 kbps, 32 kbps, 48 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *16.4 kbps, 32 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-2b: FOA

Conditions (BS1534-2b Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with FOA audio input at*  *64 kbps, 96 kbps, 128 kbps, 256 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *64 kbps, 128 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-3: HOA3

Conditions (BS1534-3 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with HOA3 audio input at*  *64 kbps, 96 kbps, 128 kbps, 256 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *64 kbps, 128 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX or High-quality loudspeaker: 7.1+4 overhead speaker setup with the configuration following IVAS-8a]* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-4: Multichannel 5.1

Conditions (BS1534-4 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with multichannel 5.1 audio input at*  *16.4 kbps, 24.4 kbps, 32 kbps, 48 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *16.4 kbps, 32 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX or High-quality loudspeaker: 5.1]* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-5: Multi-channel 5.1, 7.1

Conditions (BS1534-5 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with multichannel 5.1, 7.1 audio input at*  *64 kbps, 96 kbps, 128 kbps, 256 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *64 kbps, 128 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-6: Multi-channel 5.1+2, 5.1+4

Conditions (BS1534-6 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with multichannel 5.1+2, 5.1+4 audio input at*  *64 kbps, 96 kbps, 128 kbps, 256 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *64 kbps, 128 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-7: Multi-channel 7.1+4

Conditions (BS1534-7 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with multichannel 7.1+4 audio input at*  *64 kbps, 96 kbps, 128 kbps, 256 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *64 kbps, 128 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX or High-quality loudspeaker: 7.1+4 overhead speaker setup with the configuration following IVAS-8a]* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-8: ISM 1-2

Conditions (BS1534-8 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with ISM 1-2 audio input at*  *64 kbps, 96 kbps, 128 kbps, 256 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *64 kbps, 128 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-9a: ISM 3-4

Conditions (BS1534-9a Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with ISM 3-4 audio input at*  *24.4 kbps, 32 kbps, 48 kbps, 64 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *24.4 kbps, 48 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-9b: ISM 3-4

Conditions (BS1534-9b Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with ISM 3-4 audio input at*  *64 kbps, 96 kbps, 128 kbps, 256 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *64 kbps, 128 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-10a: MASA (1TC)

Conditions (BS1534-10a Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with MASA (1TC) audio input at*  *16.4 kbps, 24.4 kbps, 32 kbps, 48 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *16.4 kbps, 32 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-10b: MASA (1TC)

Conditions (BS1534-10b Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with MASA (1TC) audio input at*  *64 kbps, 96 kbps, 128 kbps, 256 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *64 kbps, 128 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-11: MASA (2TC)

Conditions (BS1534-11 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with MASA (2TC) audio input at*  *64 kbps, 96 kbps, 128 kbps, 256 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *64 kbps, 128 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-12a: OSBA (1-4 obj.)

Conditions (BS1534-12a Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with OSBA (1-4 obj.) audio input at*  *16.4 kbps, 24.4 kbps, 32 kbps, 48 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *16.4 kbps, 32 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-12b: OSBA (1-4 obj.)

Conditions (BS1534-12b Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with OSBA (1-4 obj.) audio input at*  *64 kbps, 96 kbps, 128 kbps, 256 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *64 kbps, 128 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

* 1. Experiment BS1534-13: OMASA (1-4 obj.)

Conditions (BS1534-1a Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS [FX/FL] operated with OMASA (1-4 obj.) audio input at*  *64 kbps, 96 kbps, 128 kbps, 256 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Mono EVS*  *64 kbps, 128 kbps DTX off at 0% FER*  *Mono signal generated with IVAS Pre-renderer* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, Nominal input level* |
| *Hidden Reference* | *Direct signal, Nominal input level* |
| *LP3k5 anchor* | *3.5 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* |
| *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 XXX* |
| *Listening Environment* | *No room noise* |

]