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

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

**Version: v.0.4.0**

**Agenda Item: 14.2**

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. |
| v.0.2.1 | 11 March 2024 | Integration of agreements from S4aA24002 and the following discussion, integration of proposal from S4aA24003, editorial changes. |
| v.0.2.2 | 25 March 2024 | Integration of agreements from S4aA240017 and S4aA240018. Editorial changes, adding tables with Test Conditions in Annex G |
| v.0.2.3 | 8 April 2024 | Integration of agreements from S4aA240025, S4aA240026, and from editing session. Editorial changes. |
| v.0.3.0 | 12 April 2024 | Integration of S4-240697, editorial changes. |
| v.0.4.0 | 23 May 2024 | Integration of agreements from S4-240881, updating term P.SUPPL800, editorial changes |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

1. Introduction

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

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. Recommendation ITU-R BS.2051-3 (05/2022): Advanced sound system for programme production.
2. S4-191167: Description of the IVAS MASA C Reference Software, Source: Nokia Corporation.
3. S4-210840: Updates to IVAS MASA C Reference Software, Source: Nokia Corporation.
4. Recommendation ITU-T P.800 (08/1996): Methods for subjective determination of transmission quality.
5. Recommendation ITU-T P.811 (01/2019): Subjective test methodology for evaluating Speech oriented stereo communication systems over headphones.
6. S4-211151: Example designs for IVAS codec tests, Source: Dolby Laboratories Inc.
7. S4-210836: On reference designs for IVAS codec tests, Source: Dolby Laboratories Inc.
8. Recommendation ITU-R BS.1770-4 (10/2015): Algorithms to measure audio programme loudness and true-peak audio level.
9. ITU-T Handbook of subjective testing practical procedures, 2011.
10. AFsp Programs and Routines: http://www-mmsp.ece.mcgill.ca/Documents/Software/Packages/AFsp/audio/html/AFsp.html.
11. F. Zotter and M. Frank, “All-Round Ambisonic Panning and Decoding,” in J. Audio Eng. Soc., Vol. 60, No. 10, 2012.
12. Supplement ITU-T P.Suppl29: "ITU-T P.800 – Use Cases".
13. Recommendation ITU-R BS.1534 (10/2015): Method for the subjective assessment of intermediate quality level of audio systems.
14. 3GPP TR 26.952: Codec for Enhanced Voice Services (EVS); Performance characterization.
15. S4-030821: PSS/MMS High-Rate Audio Selection Test and Processing Plan, Version 2.2.
16. Audio File Format Specifications: WAVE, <https://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/WAVE.html>.
17. AFsp Package <https://www-mmsp.ece.mcgill.ca/Documents/Downloads/AFsp/>.
18. Recommendation ITU-T P.191 (03/2023): Software tools for speech and audio coding standardization.
19. S4-230221: Processing updates for IVAS MASA C Reference Software.
20. 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

FER Frame Error Rate

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

ISM Independent Streams with Metadata (IVAS Objects)

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

TC Transport Channel

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.

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 [16], 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.800 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 [4] including:
  + Having a background noise level of less than NR-25.
* For each P.800 listening test, use subjects that are native speakers of the tested language.
* Provide a person during the training session of each P.800 test that is able to answer questions from the subjects in their native language.
* Provide to SA4 the P.800 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.800 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.800 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 tests as specified in clause 3.3.6.2. The tests compare the subjective scores of the CuT against the scores for specified reference conditions. Each subjective experiment contains a number of 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 statistical tests using Student's t-test (single-sided at 95% confidence level).

* Student's Independent Groups t-test will be used for P.800 experiments (Annex F:). The test shall be computed for all test conditions comprising the IVAS fixed-point code. The reference shall be the IVAS floating-point code run in the same configuration.
* Student's Dependent Groups t-test will be used for BS.1534 experiments (Annex G:). The test shall be computed for all IVAS conditions. The reference will be the EVS conditions. Thus, typically several statistical tests will be computed for each IVAS condition.

Editor’s note: This section needs to be completed for the remaining methodologies, when available.

Results of the statistical tests for each experiment will be presented in a table as illustrated in Table 1.

In the example below:

* CuT "not worse than" Reference is indicated by **CuT NWT Ref**.
* CuT "better than" Referenceis indicated by **CuT BT Ref**.
* CuT "worse than" Reference is indicated by **CuT WT Ref**.

Table 1: Example of test results



|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LL** | **Language** | **CuT** | **Mean** | **StDev** | **Ref** | **Mean** | **StDev** | **Diff** | **SEMD** | **T-stat** | **Test** |
| b | NAE1 | c14 | 3.458 | 1.015 | c08 | 3.313 | 0.910 | -0.145 | 0.055 | -2.636 | CuT BT Ref |
| b | NAE1 | c15 | 3.499 | 0.870 | c09 | 3.581 | 0.981 | 0.082 | 0.056 | 1.464 | CuT NWT Ref |
| b | NAE1 | c16 | 3.365 | 0.919 | c10 | 3.570 | 0.953 | 0.205 | 0.078 | 2.628 | CuT WT Ref |
| b | NAE1 | c17 | 4.422 | 0.902 | c11 | 4.501 | 0.858 | 0.079 | 0.071 | 1.113 | CuT NWT Ref |
| c12 | 4.521 | 0.712 | 0.099 | 0.072 | 1.375 | CuT NWT Ref |
| c01 | 4.655 | 0.700 | 0.233 | 0.079 | 2.949 | CuT WT Dir |

* + 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. Methodology

The following test methodologies shall be used in the IVAS Characterization test: P.800 [4] and BS.1534 [13]. High-level configuration of experiments for envisaged methodologies is outlined below.

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

* + 1. P.800 DCR
* Considerations of P.Suppl29 [12] shall be taken into account.
* 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 [9].
* 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.
* In case a P.800 experiment is performed twice in two different LLs, it is run in two different languages with native listeners.

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. 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.

Each BS.1534 experiment comprises a training phase in which the subjects familiarize themselves with the testing methodology and environment.

* 1. Opinion Scales

Table 2 defines opinion scale used for ITU-T P.800 DCR test. Instructions in English for the P.800 test are provided in Annex A:.

Table 2: Opinion scale for ITU-T P.800 DCR test

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

* 1. Material

Editor’s note: The Material section should be reviewed.

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 [16] 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 [17].

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

* 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 4.4.1.3.

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. Material for P.800 testing
       1. Speech Material for P.800 testing

Except for Music and mixed content categories, P.800 test experiments will use artificially created immersive audio. LLs shall provide clean speech mono audio 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 [20].

* + - 1. Background Material for P.800 testing
* P.800 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.800 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

LLs shall provide music and mixed content stereo samples for the stereo experiments. This means that LLs shall provide 7 samples per category: 6 for evaluation and 1 for preliminaries. Music and mixed content audio samples for the other P.800 experiments will be collected and selected by MC, similarly as done for the Generic Audio Items Selection for BS.1534 experiments (Clause 4.4.1.4).

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.

* + - 1. Audio Material for 3- and 4-object categories in P.800 testing

Audio material for 3- and 4-object categories will be collected and selected by MC, similarly as done for the Generic Audio Items Selection for BS.1534 experiments (Clause 4.4.1.4). The collected audio material shall consist of complete audio scenes falling in the following categories:

* Speech + effects (scene with 3 objects)
* Speech + music (scene with 3 objects)
* Music or effects (scene with 3 objects)
* Speech + effects (scene with 4 objects)
* Speech + music (scene with 4 objects)
* Music or effects (scene with 4 objects)
  + 1. Critical Generic Audio Items for BS.1534 testing
       1. Steps of Critical Test Item Selection

The following steps are based on [15]:

* 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+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.

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 themselves 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 Characterization Test will use the following listening systems:

* Stereo headphones for binaural listening, e.g.:
  + Beyerdynamic DT 770 Pro for P.800 experiments
  + Sennheiser HD 650 for BS1534 experiments
* Loudspeaker listening system – 5.1, 5.1+2, 5.1+4, 7.1, 7.1+4 loudspeaker setup [1].
* All P.800 tests are carried out via headphones.

1. Subjective Experiments
   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 and validate the fixed-point implementation. 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 is used, rendering to the playout configuration.
* Stereo; it may include binauralized samples (without head tracking).
* Multi-channel configurations 5.1, 5.1+2, 5.1+4, 7.1, and 7.1+4
* Objects (1-4 ISMs including metadata/rendering)
* FOA, HOA3
* MASA 1 TC, 2 TC
* Combined input formats:
  + Objects + MASA (OMASA), 1-4 ISMs
  + Objects + SBA (OSBA), 1-4 ISMs
* 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
* 16 kHz and 32 kHz sampled input
* Different configurations of input and output not tested in the Selection phase
  + Complex inputs to mono 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 clauses and annexes.

[

Editor’s note: The following Generic layouts might not apply to all experiments. In that case they should be reviewed.

* 1. P.800 listening test layout

Table 3: P.800 testing under clean-channel conditions incl. a comparison of fixed-point and floating-point code

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q = xx dB | - | - |
| c03 | MNRU Q = xx dB | - | - |
| c04 | MNRU Q = xx dB | - | - |
| c05 | MNRU Q = xx dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 13.2 |  |
| c11 | IVAS FX enc / FL dec | 16.4 |  |
| c12 | IVAS FL enc / FX dec | 24.4 |  |
| c13 | IVAS FX enc / FL dec | 32.0 |  |
| c14 | IVAS FL enc / FX dec | 48.0 |  |
| c15 | IVAS FX enc / FL dec | 64.0 |  |
| c16 | IVAS FL enc / FX dec | 80.0 |  |
| c17 | IVAS FX enc / FL dec | 96.0 |  |
| c18 | IVAS FL enc / FX dec | 128.0 |  |
| 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. BS.1534 listening test layouts

Table 4: High Bitrate MUSHRA Tests

|  |  |  |  |
| --- | --- | --- | --- |
| **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 |

Table 5: Low Bitrate MUSHRA Tests

|  |  |  |  |
| --- | --- | --- | --- |
| **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 |

Table 6: MUSHRA test characterizing the 6-DoF and directivity feature

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **Extended MD** | **Listener position** |
| c01 | Reference | - | - | Yes | Yes |
| c02 | LP 3.5 kHz | - | - | Yes | Yes |
| c03 | IVAS | 64.0 | Off | No | No (orientation only) |
| c04 | IVAS | 512.0 | Off | No | No (orientation only) |
| c05 | IVAS | 64.0 | Off | Yes | Yes |
| c06 | IVAS | 512.0 | Off | Yes | Yes |

]

* 1. Allocation of experiments

The following tables show high-level overview of the experiments with allocation of experiments to LLs. Table 7 shows overview of P.800 experiments with languages used for each experiment. Table 8 shows overview of BS.1534 experiments. Finally, Table 9 shows overview of experiments run with different methodologies.

Detail conditions for each subjective experiment are defined in Annex F: for P.800 experiments, in Annex G: for BS.1534 experiments, and in Annex H for experiments run with different methodologies.

**Assumptions**

* The same pricing for external LLs and GAL as in Selection phase (IVAS-8a), i.e. 18000 Euros per P.800 test, 10000 Euros per BS.1534 test rendered via headphones, and 12000 Euros for GAL tasks.
* The external LL will run the following experiments:
  + 10 x P.800 (10 x 18000 = 180000 Euros)
  + 1 x BS.1534 headphones test (1 x 10000 Euros)
* Minimum requirements for speech P.800 experiments: 6 talkers (3 male + 3 female) per experiment, 14 single sentences per talker.
* 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. DTX conditions are not tested in BS.1534 experiments.
* Frame erasure conditions are not tested in BS.1534 experiments.
* All experiments except for speech categories of stereo P.800 experiments are assumed Full Band experiments, i.e., the direct reference condition is always FB. P.800 stereo experiments are SWB experiments.
* It is assumed that the subjective material in the P.800 experiments comprising fixed-point validation accommodates all three input signal levels. Levels can be integrated into signal categories or as different audio samples within categories.
* Editor’s note: 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, 2 P.800 tests.
  + Ericsson – 2 P.800 tests, and at minimum 2 BS.1534 tests, both loudspeaker and binaural rendering is possible.
  + FhG – tentatively 2 P.800 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, 2 P.800 tests, and at minimum 3 BS.1534 tests, both loudspeaker and binaural rendering is possible.
  + NTT – 1 P.800 test, 1/2 BS.1534 (shared with Panasonic), binaural rendering.
  + Orange – 2 P.800, 1 BS.1534.
  + Panasonic – 1/2 BS.1534 (shared with NTT), binaural rendering.
  + Philips – tentatively 1 BS.1534, 1 room acoustics testing.
  + QC – 2 BS.1534.
  + VoiceAge – 2 P.800 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:

* 13 x P.800
* 18 x BS.1534, up to 9 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:

* 23 x P.800
* 19 x BS.1534, up to 9 with LS rendering
* 1 x ACR
* 1 x room acoustics testing

For reference, at Selection we ran 18 P.800 and 28 BS.1534 experiments (IVAS-8a).

[

Table 7: Allocation of experiments to LLs and P.800 languages

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Exp** | **Objective** | **Input format** | **Source material** | **FX validation** | **DTX** | **FER** | **Bitrates [kb/s]** | **Language** | **LL** |
| P800-1 | FX, RD | Stereo | All | Yes | Off | 0% |  |  |  |
| P800-2 | DTX, FE | Stereo | Al | Yes | On | 5% |  |  |  |
| P800-3 | FX, RD | FOA | All | Yes | Off | 0% |  |  |  |
| P800-4 | DTX, FE | HOA2 | All | Yes | On | 5% |  |  |  |
| P800-5 | FX, RD | HOA3 | All | Yes | Off | 0% |  |  |  |
| P800-6 | FX, RD | MC 5-1, 7-1 | Clean speech, mixed/music | Yes | Off | 0% |  |  |  |
| P800-7 | FX, RD | MC 5-1-4, 7-1-4 | Clean speech, mixed/music | Yes | Off | 0% |  |  |  |
| P800-8 | FE | MC (mixed CICP) | Clean speech, mixed/music | Yes | Off | 5% |  |  |  |
| P800-9 | FX, RD | 1-2 Objects | Clean speech | Yes | Off | 0% |  |  |  |
| P800-10 | FX, RD | 3-4 Objects | speech+ effects,  speech + music, music | Yes | Off | 0% |  |  |  |
| P800-11 | DTX, FE | 1-4 Objects | Clean speech, speech+ effects,  speech + music, music | Yes | On | 5% |  |  |  |
| P800-12 | FX, RD | MASA 1 TC | All | Yes | Off | 0% |  |  |  |
| P800-13 | FX, RD | MASA 2 TC | All | Yes | Off | 0% |  |  |  |
| P800-14 | DTX, FE | MASA 1,2 TC | All | Yes | On | 5% |  |  |  |
| P800-15 | FX, RD | OSBA, 1-2 Objects | All | Yes | Off | 0% |  |  |  |
| P800-16 | FX, RD | OSBA, 3-4 Objects | All | Yes | Off | 0% |  |  |  |
| P800-17 | RD, FE | OSBA, 1-4 Objects | All | Yes | Off | 5% |  |  |  |
| P800-18 | FX, RD | OMASA, 1-2 Objects | All | Yes | Off | 0% |  |  |  |
| P800-19 | FX, RD | OMASA, 3-4 Objects | All | Yes | Off | 0% |  |  |  |
| P800-20 | RD, FE | OMASA, 1-4 Objects | All | Yes | Off | 5% |  |  |  |
| P800-21 | JBM/FE/Tan | Stereo | Clean speech |  |  |  |  |  |  |
| P800-22 | JBM/FE/Tan | 2 Objects | Clean speech |  |  |  |  |  |  |
| P800-23 | JBM/FE/Tan | FOA [+ MASA] | Clean speech |  |  |  |  |  |  |

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 7stabilizes.

]

Table 8: Allocation of BS.1534 experiments to LLs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Exp** | **Objective** | **Input format** | **Source material** | **Bitrates [kb/s]** | **Listening environment** | **LL** |
| BS1534-1 | EVS | Stereo | Generic audio | 16.4-48 | Headphones |  |
| BS1534-1b | EVS | Stereo | Generic audio | 64-256 | Headphones |  |
| BS1534-2a | EVS | FOA | Generic audio | 16.4-48 | Headphones |  |
| BS1534-2b | EVS | FOA | Generic audio | 64-256 | Headphones |  |
| BS1534-3 | EVS | HOA3 | Generic audio | 64-256 | 7.1 + 4 |  |
| BS1534-4 | EVS | 5.1 | Generic audio | 16.4-48 | 5.1 |  |
| BS1534-5 | EVS | 5.1, 7.1 | Generic audio | 64-256 | Headphones |  |
| BS1534-6 | EVS | 5.1+2, 5.1+4 | Generic audio | 64-256 | Headphones |  |
| BS1534-7 | EVS | 7.1+4 | Generic audio | 64-256 | 7.1 + 4 |  |
| BS1534-8 | EVS | 1-2 Objects | Generic audio | 64-256 | Headphones |  |
| BS1534-9a | EVS | 3-4 Objects | Generic audio | 24.4-64 | Headphones |  |
| BS1534-9b | EVS | 3-4 Objects | Generic audio | 64-256 | Headphones |  |
| BS1534-10a | EVS | MASA 1 TC | Generic audio | 16.4-48 | Headphones |  |
| BS1534-10b | EVS | MASA 1 TC | Generic audio | 64-256 | Headphones |  |
| BS1534-11 | EVS | MASA 2 TC | Generic audio | 64-256 | Headphones |  |
| BS1534-12a | EVS | OSBA 1-4 Objects | Generic audio | 16.4-48 | Headphones |  |
| BS1534-12b | EVS | OSBA 1-4 Objects | Generic audio | 64-256 | Headphones |  |
| BS1534-13 | EVS | OMASA 1-4 Objects | Generic audio | 64-256 | Headphones |  |
| BS1534-14 | Stereo downmix | Stereo | Generic audio | 13.2, 24.2 | Headphones |  |
| BS1534-15 | 6-DoF and directivity | 4 Objects | Generic audio | 64, 512 | Headphones |  |

Editors’s note: The white-background cells for Exp in Table 8 mean that we might not have available capacity at the current state of volunteering LLs’ capacity declarations.

[

Table 9: Allocation of experiments using different methodologies

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Exp** | **Objective** | **Input format** | **Source material** | **Listening environment** | **Bitrates [kb/s]** | **LL** |
| ACR-1 | 16, 32, 48 kHz |  |  | Headphones |  | Nokia |
| ?-1 | Room acoustics |  |  | Headphones |  | Philips |

Editor’s note: The methodology for Room acoustics experiment needs to be decided.

Editor’s note: The tables above need to be completed.

Legend:

* Objective – main objectives of the experiment
* FE (in the “Objective” column) – performance in noisy channel
* DTX (in the “Objective” column) – performance in DTX on
* FX (in the “Objective” column) – fixed-point validation
* RD (in the “Objective” column) – rate-distortion curve
* EVS (in the “Objective” column) – comparison to EVS
* P800 – P.800 DCR test
* bckg – means any background
* All – means clean speech, speech with background, mixed content, and music

]

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

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

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

1. P.800 Presentation Orders

The PC will provide the Presentation Order for each P.800 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 verified wrt up-to-date version of IVAS-7b.

* General:
  + All audio material shall be sampled at 48 kHz with Full Band (FB) content. (IVAS-8b, Clause 4.4)
  + 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, Clause 4.4)
  + For multi-track audio, the audio tracks are ordered according to Table 5 of IVAS Processing Plan (IVAS-7b). (IVAS-8b, Clause 4.4)
  + Additionally, it should be verified that the audio material can be processed with the AFsp package tools. (IVAS-8b, Clause 4.4)
  + All input source material to the IVAS-7b processing stages defined in IVAS-7b Clause 4 files shall be 20 ms block aligned. (IVAS-7b, Clause 4.1.1)
  + For Ambisonics signals, ACN component ordering and SN3D normalization shall be used (IVAS-4 [3] Clause 3, IVAS-7b, Clause 3.8)
  + For Objects, metadata according to IVAS-4 shall be used. (IVAS-4, Annex C.1)
* P.800 testing
  + Speech Material (input material for artificial immersive item creation):
    - Clean speech mono audio samples (IVAS-8b, Clause 4.4.1)
    - The recording SNR should be in accordance with P.800 at least 40 dB but preferably 50 dB or higher. (IVAS-8b, Clause 4.4.1)
    - 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, Clause 4.4.1)
    - The length of the sentences should typically correspond to the length of traditional Harvard sentences. (IVAS-8b, Clause 4.4.1)
    - The total length of the generated P.800 speech samples shall not exceed 10 s. (IVAS-7b, Clause 4.2.1)
  + Background Material
    - The minimum lengths of noise files shall be 80 s. (IVAS-8b, Clause 4.4.1.2)
    - 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, Clause 4.4.1.2)
    - 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, Clause 4.4.1.2)
    - 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, Clause 4.4.1.2)
    - 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, Clause 4.4.1.3)
    - The duration of each sample shall be approximately 8 and at least 7seconds. (IVAS-8b, Clause 4.4.1.3)
    - The following categories shall be used (IVAS-8b, Clause 4.4.1.3):
      * 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,4.4.2.2)
  + 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, Clause 4.4)
* 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, Clause 4.4)
* Conversion if the audio material can’t be processed with the AFsp package tools. (IVAS-8b, Clause 4.4)
* 20ms Block alignment if the input source material to the IVAS-7b processing stages defined in IVAS-7b Clause 4 is not 20ms block aligned. (IVAS-7b, Clause 4.1.1)
* Format conversion for Ambisonics signals to match ACN component ordering and SN3D normalization (IVAS-4 [3] Clause 3, IVAS-7b, Clause 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
* 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.800 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.800 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 E.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  Progress IVAS characterization permanent documents, including:   * IVAS-7b Processing Plan for Characterization Phase * IVAS-8b Test Plan for Characterization Phase |  |
| March-2024 |  | FL-to-FX conversion work finalized according to the contract between ETSI and Ittiam (20 March) |  |
| April-2024 | April 8-12 | 3GPP SA4 e-meeting #127bis  Progress IVAS characterization permanent documents, including:   * IVAS-7b Processing Plan for Characterization Phase * IVAS-8b Test Plan for Characterization Phase |  |
| May-2024 | May 20-24 | 3GPP SA4 meeting #128  Progress IVAS characterization permanent documents, including:  • IVAS-7b Processing Plan for Characterization Phase  • IVAS-8b Test Plan for Characterization Phase |  |
| June-2024 | June 18-21 | 3GPP TSG-SA #104 |  |
| May/June-2024 |  | Characterization processing start date: 27 May.  (Characterization tests start date: 2 weeks afterwards.) |  |
| June-2024 | June 10 | Start of the Characterization listening tests |  |
| Aug-2024 | August 2 | Raw material received by GAL |  |
| Aug-2024 | August 13 | Contributions deadline for SA4#129 |  |
| Aug-2024 | August 19-23 | 3GPP SA4 meeting #129  Characterization test results are available and SA4 reviews them.  Agreement on IVAS TR 26.997. |  |
| Sep-2024 | Sep 10-13 | 3GPP TSG-SA #105  Approval of characterization test results.  Approval of IVAS Technical Report TR 26.997. |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. P.800 Experiments

[

* 1. Experiment P800-1: Stereo

Tables F.1.1 to F.1.5 show conditions to be used for this experiment, list of preliminaries ,full list of conditions, and definition of Speech categories, and Mixed content and Music categories, respectively.

Table F.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, 64, 80, 96, 128 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= 12, 17, 22, 27 dB  *α* = 0.1, 0.3, 0.5, 0.7 |
| Input frequency mask | 20KBP |
| **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 [18] 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 for categories 1-4, up to FB for categories 5-6 |
| 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |
| Languages | TBD |

Table F.1.2: Preliminaries for Experiment P800-1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c26 |  | IVAS FL | 96 | No errors |
| 3 | c06 |  | ESDRU = 0.7 | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c09 |  | ESDRU = 0.1 | - | - |
| 6 | c22 |  | IVAS FL | 32 | No errors |
| 7 | c03 |  | MNRU Q = 27 dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 128 | No errors |
| 10 | c07 |  | ESDRU = 0.5 | - | - |
| 11 | c05 |  | MNRU Q = 12 dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table F.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 | MNRU Q = 27dB | - | - |
| c03 | MNRU Q = 22 dB | - | - |
| c04 | MNRU Q = 17 dB | - | - |
| c05 | MNRU Q = 12 dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 13.2 |  |
| c11 | IVAS FX enc / FL dec | 16.4 |  |
| c12 | IVAS FL enc / FX dec | 24.4 |  |
| c13 | IVAS FX enc / FL dec | 32.0 |  |
| c14 | IVAS FL enc / FX dec | 48.0 |  |
| c15 | IVAS FX enc / FL dec | 64.0 |  |
| c16 | IVAS FL enc / FX dec | 80.0 |  |
| c17 | IVAS FX enc / FL dec | 96.0 |  |
| c18 | IVAS FL enc / FX dec | 128.0 |  |
| 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 |  |

Table F.1.4: Clean and noisy speech categories and scene definitions

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Category | Room | Reverb | Microphone Setup | ***Background*** | SNR  [dB] | Overtalk [s](2 | Bandwidth( | Talker positions(3 | Talker selection by panel |
| cat 1 | small | anechoic | M-S | Low level idle noise | 45 | 1 | Max available up to SWB | 1-7  5-3  2-6  4-1  3-4  7-2 | P1: 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 |

Table F.1.5: Mixed content and music categories

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

* 1. Experiment P800-2: Stereo

Tables F.2.1 to F.2.5 show conditions to be used for this experiment, list of preliminaries ,full list of conditions, and definition of Speech categories, and Mixed content and Music categories, respectively.

Table F2.1 : Conditions for Experiment P800-2

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64, 80, 96, 128 kbps |
| DTX | DTX on |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 5% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= 12, 17, 22, 27 dB  *α* = 0.1, 0.3, 0.5, 0.7 |
| Input frequency mask | 20KBP |
| **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 [18] 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 for categories 1-4, up to FB for categories 5-6 |
| 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |
| Languages | TBD |

Table F.2.2 : Preliminaries for Experiment P800-2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** | **DTX** |
| 1 | c19 |  | IVAS FL | 13.2 | 5% | on |
| 2 | c26 |  | IVAS FL | 96 | 5% | on |
| 3 | c06 |  | ESDRU = 0.7 | - | - |  |
| 4 | c20 |  | IVAS FL | 16.4 | 5% | on |
| 5 | c09 |  | ESDRU = 0.1 | - | - |  |
| 6 | c22 |  | IVAS FL | 32 | 5% | on |
| 7 | c03 |  | MNRU Q = 27 dB | - | - |  |
| 8 | c01 |  | Reference | - | - |  |
| 9 | c27 |  | IVAS FL | 128 | 5% | on |
| 10 | c07 |  | ESDRU = 0.5 | - | - |  |
| 11 | c05 |  | MNRU Q = 12 dB | - | - |  |
| 12 | c24 |  | IVAS FL | 64 | 5% | on |

Table F.2.3: Test conditions for Experiment P800-2,  
stereo speech and mixed music under impaired channel conditions

Alternative 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **FER** |
| c01 | Reference | - | - |  |
| c02 | MNRU Q = 27 dB | - | - |  |
| c03 | MNRU Q = 22 dB | - | - |  |
| c04 | MNRU Q = 17 dB | - | - |  |
| c05 | MNRU Q = 12 dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU |  |  |  |
| C09 | ESDRU | - | - |  |
| c10 | IVAS FL enc / FX dec | 13.2 | on | 5% |
| c11 | IVAS FX enc / FL dec | 16.4 | on | 5% |
| c12 | IVAS FL enc / FX dec | 24.4 | on | 5% |
| c13 | IVAS FX enc / FL dec | 32.0 | on | 5% |
| c14 | IVAS FL enc / FX dec | 48.0 | on | 5% |
| c15 | IVAS FX enc / FL dec | 64.0 | on | 5% |
| c16 | IVAS FL enc / FX dec | 80.0 | on | 5% |
| c17 | IVAS FX enc / FL dec | 96.0 | on | 5% |
| c18 | IVAS FL enc / FX dec | 128.0 | on | 5% |
| c19 | IVAS FL | 13.2 | on | 5% |
| c20 | IVAS FL | 16.4 | on | 5% |
| c21 | IVAS FL | 24.4 | on | 5% |
| c22 | IVAS FL | 32.0 | on | 5% |
| c23 | IVAS FL | 48.0 | on | 5% |
| c24 | IVAS FL | 64.0 | on | 5% |
| c25 | IVAS FL | 80.0 | on | 5% |
| c26 | IVAS FL | 96.0 | on | 5% |
| c27 | IVAS FL | 128.0 | on | 5% |
| c28 | IVAS FX | 13.2 | on | 5% |
| c29 | IVAS FX | 16.4 | on | 5% |
| c30 | IVAS FX | 24.4 | on | 5% |
| c31 | IVAS FX | 32.0 | on | 5% |
| c32 | IVAS FX | 48.0 | on | 5% |
| c33 | IVAS FX | 64.0 | on | 5% |
| c34 | IVAS FX | 80.0 | on | 5% |
| c35 | IVAS FX | 96.0 | on | 5% |
| c36 | IVAS FX | 128.0 | on | 5% |

Alternative 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **FER** |
| c01 | Reference | - | - |  |
| c02 | MNRU Q=xx dB | - | - |  |
| c03 | MNRU Q=xx dB | - | - |  |
| c04 | MNRU Q=xx dB | - | - |  |
| c05 | MNRU Q=xx dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU |  |  |  |
| C09 | IVAS FL enc / FX dec | 13.2 | on | 0% |
| c10 | IVAS FX enc / FL dec | 16.4 | on | 0% |
| c11 | IVAS FL enc / FL dec | 24.4 | on | 0% |
| c12 | IVAS FX enc / FX dec | 32.0 | on | 0% |
| c13 | IVAS FL enc / FX dec | 48.0 | on | 0% |
| c14 | IVAS FX enc / FL dec | 64.0 | on | 0% |
| c15 | IVAS FL enc / FL dec | 80.0 | on | 0% |
| c16 | IVAS FX enc / FX dec | 13.2 | off | 4% |
| c17 | IVAS FL enc / FX dec | 16.4 | off | 4% |
| c18 | IVAS FX enc / FL dec | 24.4 | off | 4% |
| c19 | IVAS FL enc / FL dec | 32.0 | off | 4% |
| c20 | IVAS FX enc / FX dec | 48.0 | off | 4% |
| c21 | IVAS FL enc / FX dec | 64.0 | off | 4% |
| c22 | IVAS FX enc / FL dec | 80.0 | off | 4% |
| c23 | IVAS FL enc / FL dec | 13.2 | on | 8% |
| c24 | IVAS FX enc / FX dec | 16.4 | on | 8% |
| c25 | IVAS FL enc / FX dec | 24.4 | on | 8% |
| c26 | IVAS FX enc / FL dec | 32.0 | on | 8% |
| c27 | IVAS FL enc / FL dec | 48.0 | on | 8% |
| c28 | IVAS FX enc / FX dec | 64.0 | on | 8% |
| c29 | IVAS FL enc / FX dec | 80.0 | on | 8% |
| c30 | IVAS FX enc / FL dec | 13.2 | off | 8% |
| c31 | IVAS FL enc / FL dec | 16.4 | off | 8% |
| c32 | IVAS FX enc / FX dec | 24.4 | off | 8% |
| c33 | IVAS FL enc / FX dec | 32.0 | off | 8% |
| c34 | IVAS FX enc / FL dec | 48.0 | off | 8% |
| c35 | IVAS FL enc / FL dec | 64.0 | off | 8% |
| c36 | IVAS FX enc / FX dec | 80.0 | off | 8% |

Table F.2.4: Clean and noisy speech categories and scene definitions

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Category | Room | Reverb | Microphone Setup | ***Background*** | SNR  [dB] | Overtalk [s](2 | Bandwidth( | Talker positions(3 | Talker selection by panel |
| cat 1 | small | anechoic | M-S | Low level idle noise | 45 | 1 | Max available up to SWB | 1-7  5-3  2-6  4-1  3-4  7-2 | P1: 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 |

Table F.2.5: Mixed content and music categories

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

* 1. Experiment P800-3: FOA

Tables F.3.1 to F.3.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Mixed content and Generic audio categories, respectively.

Table F.3.1: Conditions for Experiment P800-3

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 16.4, 24.4, 32, 48, 64, 80, 96, 128, 256 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP |
| **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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.3.2: Preliminaries for Experiment P800-3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 16.4 | No errors |
| 2 | c26 |  | IVAS FL | 96 | No errors |
| 3 | c06 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 24.4 | No errors |
| 5 | c09 |  | ESDRU = xx | - | - |
| 6 | c22 |  | IVAS FL | 32 | No errors |
| 7 | c03 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q=xx dB | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 16.4 |  |
| c11 | IVAS FX enc / FL dec | 24.4 |  |
| c12 | IVAS FL enc / FX dec | 32.0 |  |
| c13 | IVAS FX enc / FL dec | 48.0 |  |
| c14 | IVAS FL enc / FX dec | 64.0 |  |
| c15 | IVAS FX enc / FL dec | 80.0 |  |
| c16 | IVAS FL enc / FX dec | 96.0 |  |
| c17 | IVAS FX enc / FL dec | 128.0 |  |
| c18 | IVAS FL enc / FX dec | 256.0 |  |
| c19 | IVAS FL | 16.4 |  |
| c20 | IVAS FL | 24.4 |  |
| c21 | IVAS FL | 32.0 |  |
| c22 | IVAS FL | 48.0 |  |
| c23 | IVAS FL | 64.0 |  |
| c24 | IVAS FL | 80.0 |  |
| c25 | IVAS FL | 96.0 |  |
| c26 | IVAS FL | 128.0 |  |
| c27 | IVAS FL | 256.0 |  |
| c28 | IVAS FX | 16.4 |  |
| c29 | IVAS FX | 24.4 |  |
| c30 | IVAS FX | 32.0 |  |
| c31 | IVAS FX | 48.0 |  |
| c32 | IVAS FX | 64.0 |  |
| c33 | IVAS FX | 80.0 |  |
| c34 | IVAS FX | 96.0 |  |
| c35 | IVAS FX | 128.0 |  |
| c36 | IVAS FX | 256.0 |  |

Table F.3.4: Clean and noisy speech 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* |

Table F.3.5: Mixed content and Generic audio 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-8b.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8b. 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 P800-4: HOA2

Tables F.4.1 to F.4.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Mixed content and Generic audio categories, respectively.

Table F.4.1: Conditions for Experiment P800-4

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 16.4, 24.4, 32, 48, 64, 96, 128, 256, 384 kbps |
| DTX | DTX on |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 5% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP |
| **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 (HOA2) Spatial Room Impulse Responses respective to various talker positions relative to a capture point and spatial (HOA2) ambient noise mixing. Cat. 5-6: Pre-produced content |
| Binaural renderer | HOA2 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.4.2: Preliminaries for Experiment P800-4

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** | **DTX** |
| 1 | c19 |  | IVAS FL | 116.4 | 5% | on |
| 2 | c26 |  | IVAS FL | 256 | 5% | on |
| 3 | c06 |  | ESDRU = xx | - | - |  |
| 4 | c20 |  | IVAS FL | 24.4 | 5% | on |
| 5 | c09 |  | ESDRU = xx | - | - |  |
| 6 | c22 |  | IVAS FL | 48 | 5% | on |
| 7 | c03 |  | MNRU Q=xx dB | - | - |  |
| 8 | c01 |  | Reference | - | - |  |
| 9 | c27 |  | IVAS FL | 384 | 5% | on |
| 10 | c07 |  | ESDRU = xx | - | - |  |
| 11 | c05 |  | MNRU Q=xx dB | - | - |  |
| 12 | c24 |  | IVAS FL | 96 | 5% | on |

Table F.4.3: Test conditions for Experiment P800-4,  
under impaired channel conditions

Alternative 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **FER** |
| c01 | Reference | - | - |  |
| c02 | MNRU Q=xx dB | - | - |  |
| c03 | MNRU Q=xx dB | - | - |  |
| c04 | MNRU Q=xx dB | - | - |  |
| c05 | MNRU Q=xx dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU |  |  |  |
| C09 | ESDRU | - | - |  |
| c10 | IVAS FL enc / FX dec | 16.4 | on | 5% |
| c11 | IVAS FX enc / FL dec | 24.4 | on | 5% |
| c12 | IVAS FL enc / FX dec | 32.0 | on | 5% |
| c13 | IVAS FX enc / FL dec | 48.0 | on | 5% |
| c14 | IVAS FL enc / FX dec | 64.0 | on | 5% |
| c15 | IVAS FX enc / FL dec | 96.0 | on | 5% |
| c16 | IVAS FL enc / FX dec | 128.0 | on | 5% |
| c17 | IVAS FX enc / FL dec | 256.0 | on | 5% |
| c18 | IVAS FL enc / FX dec | 384.0 | on | 5% |
| c19 | IVAS FL | 16.4 | on | 5% |
| c20 | IVAS FL | 24.4 | on | 5% |
| c21 | IVAS FL | 32.0 | on | 5% |
| c22 | IVAS FL | 48.0 | on | 5% |
| c23 | IVAS FL | 64.0 | on | 5% |
| c24 | IVAS FL | 96.0 | on | 5% |
| c25 | IVAS FL | 128.0 | on | 5% |
| c26 | IVAS FL | 256.0 | on | 5% |
| c27 | IVAS FL | 384.0 | on | 5% |
| c28 | IVAS FX | 16.4 | on | 5% |
| c29 | IVAS FX | 24.4 | on | 5% |
| c30 | IVAS FX | 32.0 | on | 5% |
| c31 | IVAS FX | 48.0 | on | 5% |
| c32 | IVAS FX | 64.0 | on | 5% |
| c33 | IVAS FX | 96.0 | on | 5% |
| c34 | IVAS FX | 128.0 | on | 5% |
| c35 | IVAS FX | 256.0 | on | 5% |
| c36 | IVAS FX | 384.0 | on | 5% |

Alternative 2:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **FER** |
| c01 | Reference | - | - |  |
| c02 | MNRU Q=xx dB | - | - |  |
| c03 | MNRU Q=xx dB | - | - |  |
| c04 | MNRU Q=xx dB | - | - |  |
| c05 | MNRU Q=xx dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU |  |  |  |
| C09 | IVAS FL enc / FX dec | 13.2 | on | 0% |
| c10 | IVAS FX enc / FL dec | 16.4 | on | 0% |
| c11 | IVAS FL enc / FL dec | 24.4 | on | 0% |
| c12 | IVAS FX enc / FX dec | 32.0 | on | 0% |
| c13 | IVAS FL enc / FX dec | 48.0 | on | 0% |
| c14 | IVAS FX enc / FL dec | 64.0 | on | 0% |
| c15 | IVAS FL enc / FL dec | 80.0 | on | 0% |
| c16 | IVAS FX enc / FX dec | 13.2 | off | 4% |
| c17 | IVAS FL enc / FX dec | 16.4 | off | 4% |
| c18 | IVAS FX enc / FL dec | 24.4 | off | 4% |
| c19 | IVAS FL enc / FL dec | 32.0 | off | 4% |
| c20 | IVAS FX enc / FX dec | 48.0 | off | 4% |
| c21 | IVAS FL enc / FX dec | 64.0 | off | 4% |
| c22 | IVAS FX enc / FL dec | 80.0 | off | 4% |
| c23 | IVAS FL enc / FL dec | 13.2 | on | 8% |
| c24 | IVAS FX enc / FX dec | 16.4 | on | 8% |
| c25 | IVAS FL enc / FX dec | 24.4 | on | 8% |
| c26 | IVAS FX enc / FL dec | 32.0 | on | 8% |
| c27 | IVAS FL enc / FL dec | 48.0 | on | 8% |
| c28 | IVAS FX enc / FX dec | 64.0 | on | 8% |
| c29 | IVAS FL enc / FX dec | 80.0 | on | 8% |
| c30 | IVAS FX enc / FL dec | 13.2 | off | 8% |
| c31 | IVAS FL enc / FL dec | 16.4 | off | 8% |
| c32 | IVAS FX enc / FX dec | 24.4 | off | 8% |
| c33 | IVAS FL enc / FX dec | 32.0 | off | 8% |
| c34 | IVAS FX enc / FL dec | 48.0 | off | 8% |
| c35 | IVAS FL enc / FL dec | 64.0 | off | 8% |
| c36 | IVAS FX enc / FX dec | 80.0 | off | 8% |

Table 4.4: Clean and noisy speech 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\_HOA2* | *room\_1\_cleanbg\_HOA2* | *45* | *1* | *Max* |  | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *room\_4\_HOA2* | *room\_[1/4]\_cleanbg\_HOA2* | *45* | *-1* | *Max* |  | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *out\_[X]\_HOA2* | *[park\_1\_bg\_HOA2 / nature\_1\_bg\_HOA2 / event\_1\_bg\_HOA2 / street\_[1/2]\_bg\_HOA2]* | *15* | *1* | *Max* |  | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *room\_[X]\_HOA2* | *[cafeteria\_1\_bg\_HOA2 / mall\_1\_bg\_HOA2/ office[1/2]\_bg\_HOA2]* | *15* | *-1* | *Max* |  | *P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3* |

Table 4.5: Mixed content and Generic audio 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-8b.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8b. Backround name ‘clean\_bg\_[X]\_HOA2’ 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 P800-5: HOA3

Tables F.5.1 to F.5.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Mixed content and Generic audio categories, respectively.

Table F.5.1: Conditions for Experiment P800-5

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 32, 48, 64, 96, 128, 160, 256, 384, 512 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP |
| **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 (HOA3) Spatial Room Impulse Responses respective to various talker positions relative to a capture point and spatial (HOA3) ambient noise mixing. Cat. 5-6: Pre-produced content |
| Binaural renderer | HOA3 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.5.2: Preliminaries for Experiment P800-5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 32 | No errors |
| 2 | c26 |  | IVAS FL | 384 | No errors |
| 3 | c06 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 48 | No errors |
| 5 | c09 |  | ESDRU = xx | - | - |
| 6 | c22 |  | IVAS FL | 96 | No errors |
| 7 | c03 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 512 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 160 | No errors |

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q=xx dB | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 32.0 |  |
| c11 | IVAS FX enc / FL dec | 48.0 |  |
| c12 | IVAS FL enc / FX dec | 64.0 |  |
| c13 | IVAS FX enc / FL dec | 96.0 |  |
| c14 | IVAS FL enc / FX dec | 128.0 |  |
| c15 | IVAS FX enc / FL dec | 160.0 |  |
| c16 | IVAS FL enc / FX dec | 256.0 |  |
| c17 | IVAS FX enc / FL dec | 384.0 |  |
| c18 | IVAS FL enc / FX dec | 512.0 |  |
| c19 | IVAS FL | 32.0 |  |
| c20 | IVAS FL | 48.0 |  |
| c21 | IVAS FL | 64.0 |  |
| c22 | IVAS FL | 96.0 |  |
| c23 | IVAS FL | 128.0 |  |
| c24 | IVAS FL | 160.0 |  |
| c25 | IVAS FL | 256.0 |  |
| c26 | IVAS FL | 384.0 |  |
| c27 | IVAS FL | 512.0 |  |
| c28 | IVAS FX | 32.0 |  |
| c29 | IVAS FX | 48.0 |  |
| c30 | IVAS FX | 64.0 |  |
| c31 | IVAS FX | 96.0 |  |
| c32 | IVAS FX | 128.0 |  |
| c33 | IVAS FX | 160.0 |  |
| c34 | IVAS FX | 256.0 |  |
| c35 | IVAS FX | 384.0 |  |
| c36 | IVAS FX | 512.0 |  |

Table F.5.4: Clean and noisy speech 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\_HOA3* | *room\_1\_cleanbg\_HOA3* | *45* | *1* | *Max* |  | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *room\_4\_HOA3* | *room\_[1/4]\_cleanbg\_HOA3* | *45* | *-1* | *Max* |  | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *out\_[X]\_HOA3* | *[park\_1\_bg\_HOA3 / nature\_1\_bg\_HOA3 / event\_1\_bg\_HOA3 / street\_[1/2]\_bg\_HOA3]* | *15* | *1* | *Max* |  | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *room\_[X]\_HOA3* | *[cafeteria\_1\_bg\_HOA3 / mall\_1\_bg\_HOA3/ office[1/2]\_bg\_HOA3]* | *15* | *-1* | *Max* |  | *P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3* |

Table F.5.5: Mixed content and Generic audio 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-8b.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8b. Backround name ‘clean\_bg\_[X]\_HOA3’ 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: MC 5.1, 7.1

Tables F.61 to F.6.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Mixed content and Music categories, respectively.

Table F.6.1: Conditions for Experiment P800-6

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 24.4, 32, 48, 64, 80, 96, 128, 160, 256 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP excluding the LFE channel when present |
| Noise | No noise |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP excluding the LFE channel when present |
| **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 and 7.1 using IVAS Pre-renderer. Cat. 5-6: Pre-produced native 5.1 and 7.1 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 4.4.1.4 |
| 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.6.2: Preliminaries for Experiment P800-6

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 24.4 | No errors |
| 2 | c26 |  | IVAS FL | 96 | No errors |
| 3 | c06 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 32 | No errors |
| 5 | c09 |  | ESDRU = xx | - | - |
| 6 | c22 |  | IVAS FL | 48 | No errors |
| 7 | c03 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q=xx dB | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 24.4 |  |
| c11 | IVAS FX enc / FL dec | 32.0 |  |
| c12 | IVAS FL enc / FX dec | 48.0 |  |
| c13 | IVAS FX enc / FL dec | 64.0 |  |
| c14 | IVAS FL enc / FX dec | 80.0.0 |  |
| c15 | IVAS FX enc / FL dec | 96.0 |  |
| c16 | IVAS FL enc / FX dec | 128.0 |  |
| c17 | IVAS FX enc / FL dec | 160.0 |  |
| c18 | IVAS FL enc / FX dec | 256.0 |  |
| c19 | IVAS FL | 24.4 |  |
| c20 | IVAS FL | 32.0 |  |
| c21 | IVAS FL | 48.0 |  |
| c22 | IVAS FL | 64.0 |  |
| c23 | IVAS FL | 80.0.0 |  |
| c24 | IVAS FL | 96.0 |  |
| c25 | IVAS FL | 128.0 |  |
| c26 | IVAS FL | 160.0 |  |
| c27 | IVAS FL | 256.0 |  |
| c28 | IVAS FX | 24.4 |  |
| c29 | IVAS FX | 32.0 |  |
| c30 | IVAS FX | 48.0 |  |
| c31 | IVAS FX | 64.0 |  |
| c32 | IVAS FX | 80.0.0 |  |
| c33 | IVAS FX | 96.0 |  |
| c34 | IVAS FX | 128.0 |  |
| c35 | IVAS FX | 160.0 |  |
| c36 | IVAS FX | 256.0 |  |

Table F.6.4: Clean and noisy speech 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*  *5.1* | *room\_1\_FOA* | *room\_1\_cleanbg\_FOA* | *45* | *1* | *Max* |  | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2*  *7.1* | *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*  *5.1* | *out\_[X]\_FOA* | *[park\_1\_bg\_FOA / nature\_1\_bg\_FOA / event\_1\_bg\_FOA / street\_[1/2]\_bg\_FOA]* | *15* | *1* | *Max* |  | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4*  *7.1* | *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* |

Table F.6.5: Mixed content and Music categories

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 – 5.1 | music and mixed content |
| cat 6 – 7.1 | music and mixed content |

**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-8b.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8b. 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 P800-7: MC 5.1+4, 7.1+4

Tables F.7.1 to F.7.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Mixed content and Music categories, respectively.

Table F.7.1: Conditions for Experiment P800-7

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 32, 48, 64, 96, 128,160, 192, 256, 384 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP excluding the LFE channel when present |
| Noise | No noise |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP excluding the LFE channel when present |
| **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+4 and 7.1+4 using IVAS Pre-renderer. Cat. 5-6: Pre-produced native 5.1+4 and 7.1+4 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 4.4.1.4 |
| 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.7.2: Preliminaries for Experiment P800-7

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 32 | No errors |
| 2 | c26 |  | IVAS FL | 96 | No errors |
| 3 | c06 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 48 | No errors |
| 5 | c09 |  | ESDRU = xx | - | - |
| 6 | c22 |  | IVAS FL | 32 | No errors |
| 7 | c03 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table F.7.3: Test conditions for Experiment P800-7,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q=xx dB | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 32 |  |
| c11 | IVAS FX enc / FL dec | 48 |  |
| c12 | IVAS FL enc / FX dec | 64 |  |
| c13 | IVAS FX enc / FL dec | 96 |  |
| c14 | IVAS FL enc / FX dec | 128.0 |  |
| c15 | IVAS FX enc / FL dec | 160.0 |  |
| c16 | IVAS FL enc / FX dec | 192.0 |  |
| c17 | IVAS FX enc / FL dec | 256.0.0 |  |
| c18 | IVAS FL enc / FX dec | 384.0 |  |
| c19 | IVAS FL | 32 |  |
| c20 | IVAS FL | 48 |  |
| c21 | IVAS FL | 64 |  |
| c22 | IVAS FL | 96 |  |
| c23 | IVAS FL | 128.0 |  |
| c24 | IVAS FL | 160.0 |  |
| c25 | IVAS FL | 192.0 |  |
| c26 | IVAS FL | 256.0.0 |  |
| c27 | IVAS FL | 384.0 |  |
| c28 | IVAS FX | 32 |  |
| c29 | IVAS FX | 48 |  |
| c30 | IVAS FX | 64 |  |
| c31 | IVAS FX | 96 |  |
| c32 | IVAS FX | 128.0 |  |
| c33 | IVAS FX | 160.0 |  |
| c34 | IVAS FX | 192.0 |  |
| c35 | IVAS FX | 256.0.0 |  |
| c36 | IVAS FX | 384.0 |  |

Table F.7.4: Clean and noisy speech 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*  *5.1+4* | *room\_1\_FOA* | *room\_1\_cleanbg\_FOA* | *45* | *1* | *Max* |  | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2*  *7.1+4* | *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*  *5.1+4* | *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*  *7.1+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* |

Table F.7.5: Mixed content and Music categories

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 – 5.1+4 | music and mixed content |
| cat 6 – 7.1+4 | music and mixed content |

**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-8b.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8b. 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 P800-8: MC - Mixed CICP

Tables F.8.1 to F.8.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Mixed content and Music categories, respectively.

Table F.8.1: Conditions for Experiment P800-8

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 16.4, 24.4, 48, 64, 80, 128, 160, 192, 384 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP excluding the LFE channel when present |
| Noise | No noise |
| Error Conditions | 5% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP excluding the LFE channel when present |
| **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 MC using IVAS Pre-renderer. Cat. 5-6: Pre-produced native MC 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 4.4.1.4 |
| 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.8.2: Preliminaries for Experiment P800-8

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 16.4 | 5% |
| 2 | c26 |  | IVAS FL | 256 | 5% |
| 3 | c06 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 24.4 | 5% |
| 5 | c09 |  | ESDRU = xx | - | - |
| 6 | c22 |  | IVAS FL | 64 | 5% |
| 7 | c03 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 384 | 5% |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 128 | 5% |

Table F.8.3: Test conditions for Experiment P800-8,  
under impaired channel conditions

Alternative 1:

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **FER** |
| c01 | Reference | - | - |
| c02 | MNRU Q=xx dB | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 16.4 | 5% |
| c11 | IVAS FX enc / FL dec | 24.4 | 5% |
| c12 | IVAS FL enc / FX dec | 48.0 | 5% |
| c13 | IVAS FX enc / FL dec | 64.0 | 5% |
| c14 | IVAS FL enc / FX dec | 80.0 | 5% |
| c15 | IVAS FX enc / FL dec | 128.0 | 5% |
| c16 | IVAS FL enc / FX dec | 160.0 | 5% |
| c17 | IVAS FX enc / FL dec | 256.0 | 5% |
| c18 | IVAS FL enc / FX dec | 384.0 | 5% |
| c19 | IVAS FL | 16.4 | 5% |
| c20 | IVAS FL | 24.4 | 5% |
| c21 | IVAS FL | 48.0 | 5% |
| c22 | IVAS FL | 64.0 | 5% |
| c23 | IVAS FL | 80.0 | 5% |
| c24 | IVAS FL | 128.0 | 5% |
| c25 | IVAS FL | 160.0 | 5% |
| c26 | IVAS FL | 256.0 | 5% |
| c27 | IVAS FL | 384.0 | 5% |
| c28 | IVAS FX | 16.4 | 5% |
| c29 | IVAS FX | 24.4 | 5% |
| c30 | IVAS FX | 48.0 | 5% |
| c31 | IVAS FX | 64.0 | 5% |
| c32 | IVAS FX | 80.0 | 5% |
| c33 | IVAS FX | 128.0 | 5% |
| c34 | IVAS FX | 160.0 | 5% |
| c35 | IVAS FX | 256.0 | 5% |
| c36 | IVAS FX | 384.0 | 5% |

Alternative 2:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bandwidth** | **Bitrate [kbps]** | **DTX** | **FER** |
| c01 | Reference | FB | - | - |  |
| c02 | MNRU Q=xx dB | FB | - | - |  |
| c03 | MNRU Q=xx dB | FB | - | - |  |
| c04 | MNRU Q=xx dB | FB | - | - |  |
| c05 | MNRU Q=xx dB | FB | - | - |  |
| c06 | ESDRU | FB | - | - |  |
| c07 | ESDRU | FB | - | - |  |
| c08 | ESDRU | FB |  |  |  |
| C09 | IVAS FL enc / FX dec | FB | 16.4 | off | 3% |
| c10 | IVAS FX enc / FL dec | FB | 24.4 | off | 3% |
| c11 | IVAS FL enc / FL dec | FB | 32.0 | off | 3% |
| c12 | IVAS FX enc / FX dec | FB | 48.0 | off | 3% |
| c13 | IVAS FL enc / FX dec | FB | 80.0 | off | 3% |
| c14 | IVAS FX enc / FL dec | FB | 96.0 | off | 3% |
| c15 | IVAS FL enc / FL dec | FB | 160.0 | off | 3% |
| c16 | IVAS FX enc / FX dec | SWB (32k IO) | 16.4 | off | 3% |
| c17 | IVAS FL enc / FX dec | SWB (32k IO) | 24.4 | off | 3% |
| c18 | IVAS FX enc / FL dec | SWB (32k IO) | 32.0 | off | 3% |
| c19 | IVAS FL enc / FL dec | SWB (32k IO) | 48.0 | off | 3% |
| c20 | IVAS FX enc / FX dec | SWB (32k IO) | 80.0 | off | 3% |
| c21 | IVAS FL enc / FX dec | SWB (32k IO) | 96.0 | off | 3% |
| c22 | IVAS FX enc / FL dec | SWB (32k IO) | 160.0 | off | 3% |
| c23 | IVAS FL enc / FL dec | FB | 16.4 | off | 6% |
| c24 | IVAS FX enc / FX dec | FB | 24.4 | off | 6% |
| c25 | IVAS FL enc / FX dec | FB | 32.0 | off | 6% |
| c26 | IVAS FX enc / FL dec | FB | 48.0 | off | 6% |
| c27 | IVAS FL enc / FL dec | FB | 80.0 | off | 6% |
| c28 | IVAS FX enc / FX dec | FB | 96.0 | off | 6% |
| c29 | IVAS FL enc / FX dec | FB | 160.0 | off | 6% |
| c30 | IVAS FX enc / FL dec | WB (16k IO) | 16.4 | off | 6% |
| c31 | IVAS FL enc / FL dec | WB (16k IO) | 24.4 | off | 6% |
| c32 | IVAS FX enc / FX dec | WB (16k IO) | 32.0 | off | 6% |
| c33 | IVAS FL enc / FX dec | WB (16k IO) | 64.0 | off | 6% |
| c34 | IVAS FX enc / FL dec | WB (16k IO) | 80.0 | off | 6% |
| c35 | IVAS FL enc / FL dec | WB (16k IO) | 96.0 | off | 6% |
| c36 | IVAS FX enc / FX dec | WB (16k IO) | 160.0 | off | 6% |

Table F.8.4: Clean and noisy speech 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*  *5.1* | *room\_1\_FOA* | *room\_1\_cleanbg\_FOA* | *45* | *1* | *Max* |  | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2*  *7.1* | *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*  *5.1+4* | *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*  *7.1+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* |

Table F.8.5: Mixed content and Music categories

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 – 5.1/7.1 | music and mixed content |
| cat 6 – 5.1+4/7.1+4 | music and mixed content |

**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-8b.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8b. 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 P800-9: 1-2 Objects

Tables F.91 to F.9.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Speech with effects and Music categories, respectively.

Table F.9.1: Conditions for Experiment P800-9

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64, 80, 96, 128 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 18, 24, 30, 36 dB  *α* = 0.1, 0.3, 0.5, 0.7 |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: | Cat. 1-3: Defined scenes, 1 ISM  Cat. 4-6: Defined scenes, 2 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Editor’s note: 13.2 kbps is not supported for 2-ISMs

Editor’s note: Using categories with different number of ISMs per audio sample within one experiment might be revisited, for simplicity.

Table F.9.2: Preliminaries for Experiment P800-9

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c26 |  | IVAS FL | 96 | No errors |
| 3 | c06 |  | ESDRU = 0.7 | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c09 |  | ESDRU = 0.1 | - | - |
| 6 | c22 |  | IVAS FL | 32 | No errors |
| 7 | c03 |  | MNRU Q = 36 dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 128 | No errors |
| 10 | c07 |  | ESDRU = 0.5 | - | - |
| 11 | c05 |  | MNRU Q = 18 dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table F.9.3: Test conditions for Experiment P800-9,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q = 36 dB | - | - |
| c03 | MNRU Q = 30 dB | - | - |
| c04 | MNRU Q = 24 dB | - | - |
| c05 | MNRU Q = 18 dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 13.2 |  |
| c11 | IVAS FX enc / FL dec | 16.4 |  |
| c12 | IVAS FL enc / FX dec | 24.4 |  |
| c13 | IVAS FX enc / FL dec | 32.0 |  |
| c14 | IVAS FL enc / FX dec | 48.0 |  |
| c15 | IVAS FX enc / FL dec | 64.0 |  |
| c16 | IVAS FL enc / FX dec | 80.0 |  |
| c17 | IVAS FX enc / FL dec | 96.0 |  |
| c18 | IVAS FL enc / FX dec | 128.0 |  |
| 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 |  |

**Scene definitions categories 1-2**

A leading and trailing silence is present for each sample, in accordance with IVAS-7b. 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:

* Talker sitting at a table (elevation 0°), at different azimuths.
* Standing talker (elevation 35°), at different azimuths.
* Smaller talker (child) walking around a table in the positive sense (counterclockwise), elevation 0°. Azimuth varies continuously for the sentence pair.
* Adult talker walking around a table in the negative sense (clockwise), elevation 35°. Azimuth varies continuously for the sentence pair.
* Elevation displacement: Elevation varies continuously for the sentence pair . Azimuth is constant for a sentence pair, but different for each sentence pair.
* Azimuth and elevation displacement: Azimuth and elevation vary continuously.

Each of the sentences uttered by a certain talker is encoded using different scene. Allocation of scenes to each panel is given in the Table F.9.4

**Scene definitions categories 4-5**

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-7b. 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 **counterclockwise**.
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.

Table F.9.4: Allocation of scenes for each talker pair (category cat 1, cat 2, cat 4, cat 5) and listening panel (P1-P6)







|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Scene*** | ***Talker initial elevation*** | ***Elevation change(1*** | ***Talker initial azimuth*** | ***Azimuth change(2*** | ***Panel*** |
| ***cat 1:***  *M1* | a  b  e  f  c  d | 0°  35°  -90°  35°  0°  35° | static  static  0.3°/ frame  -0.2°/ frame  static  static | 270°  180°  120°  0°  240°  180° | static  static  static  0.5°/ frame  1°/ frame  -1°/ frame | P1  P2  P3  P4  P5  P6 |
| ***cat 2:***  *F1* | f  c  d  a  b  e | 35°  0°  35°  0°  35°  -90° | -0.2°/ frame  static  static  static  static  0.3°/ frame | 300°  60°  120°  60°  300°  60° | 0.5°/ frame  1°/ frame  -1°/ frame  static  static  static | P1  P2  P3  P4  P5  P6 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Scene*** | ***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 4:***  *M2 + F2* | *a*  *b*  *c*  *d*  *e*  *f* | *-1*  *1*  *-1*  *1*  *-1*  *1* | *0°*  *35°*  *0°*  *0°*  *45°*  *30°* | *0°*  *35°*  *45°*  *45°*  *45°*  *30°* | *0°*  *10°*  *20°*  *200°*  *340°*  *120°* | *static*  *static*  *static*  *static*  *-1°/ frame*  *1°/ frame* | *50°*  *110°*  *170°*  *30°*  *340°*  *120°* | *static*  *static*  *static*  *-1°/ frame*  *-1°/ frame*  *-1°/ frame* | *P1*  *P2*  *P3*  *P4*  *P5*  *P6* |
| ***cat 5:***  *M3 + F3* | *d*  *e*  *f*  *a*  *b*  *c* | *1*  *-1*  *1*  *-1*  *1*  *-1* | *0°*  *45°*  *30°*  *0°*  *35°*  *0°* | *45°*  *45°*  *30°*  *0°*  *35°*  *45°* | *50°*  *130°*  *300°*  *30°*  *40°*  *50°* | *static*  *1°/ frame*  *1°/ frame*  *static*  *static*  *static* | *180°*  *130°*  *300°*  *230°*  *290°*  *350°* | *1°/ frame*  *1°/ frame*  *-1°/ frame*  *static*  *static*  *static* | *P1*  *P2*  *P3*  *P4*  *P5*  *P6* |

**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

Table F.9.5: Mixed content and Generic audio categories

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 3 | Mixed content & Music (1 object) |
| cat 6 | 2 ISMs: Speech + background (2 objects) |

* 1. Experiment P800-10: 3-4 Objects

Tables F.10.1 to F.10.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Speech with effects and Music categories, respectively.

Table F.10.1: Conditions for Experiment P800-10

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 24.4, 32, 48, 64, 80, 96, 128, 160, 192 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 18, 24, 30, 36 dB  *α* = 0.1, 0.3, 0.5, 0.7 |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: | Cat. 1-3: Pre-produced content, 3 ISMs  Cat. 4-6: Pre-produced content, 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.10.2: Preliminaries for Experiment P800-10

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c26 |  | IVAS FL | 96 | No errors |
| 3 | c06 |  | ESDRU = 0.7 | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c09 |  | ESDRU = 0.1 | - | - |
| 6 | c22 |  | IVAS FL | 32 | No errors |
| 7 | c03 |  | MNRU Q = 36 dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 128 | No errors |
| 10 | c07 |  | ESDRU = 0.5 | - | - |
| 11 | c05 |  | MNRU Q = 18 dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table F.10.3: Test conditions for Experiment P800-10,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q = 36 dB | - | - |
| c03 | MNRU Q = 30 dB | - | - |
| c04 | MNRU Q = 24 dB | - | - |
| c05 | MNRU Q = 28 dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 13.2 |  |
| c11 | IVAS FX enc / FL dec | 16.4 |  |
| c12 | IVAS FL enc / FX dec | 24.4 |  |
| c13 | IVAS FX enc / FL dec | 32.0 |  |
| c14 | IVAS FL enc / FX dec | 48.0 |  |
| c15 | IVAS FX enc / FL dec | 64.0 |  |
| c16 | IVAS FL enc / FX dec | 80.0 |  |
| c17 | IVAS FX enc / FL dec | 96.0 |  |
| c18 | IVAS FL enc / FX dec | 128.0 |  |
| 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 |  |

Table F.10.4: Categories for Speech with effects, Speech with music, Music or Effects

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 1 | speech + effects (scene with 3 objects) |
| cat 2 | speech + music (scene with 3 objects) |
| cat 3 | Music or effects (scene with 3 objects) |
| cat 4 | speech + effects (scene with 4 objects) |
| cat 5 | speech + music (scene with 4 objects) |
| cat 6 | Music or effects (scene with 4 objects) |

* 1. Experiment P800-11: 1-4 Objects

Tables F.11.1 to F.11.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Speech with effects and Music categories, respectively.

Table F.11.1: Conditions for Experiment P800-11

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 24.4, 32, 48, 64, 80, 96, 128, 160, 192 kbps |
| DTX | DTX on |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise |
| Error Conditions | 5% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 18, 24, 30, 36 dB  *α* = 0.1, 0.3, 0.5, 0.7 |
| Input frequency mask | 20KBP |
| **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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.11.2: Preliminaries for Experiment P800-11

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** | **DTX** |
| 1 | c19 |  | IVAS FL | 13.2 | 5% | on |
| 2 | c26 |  | IVAS FL | 96 | 5% | on |
| 3 | c06 |  | ESDRU = 0.7 | - | - |  |
| 4 | c20 |  | IVAS FL | 16.4 | 5% | on |
| 5 | c09 |  | ESDRU = 0.1 | - | - |  |
| 6 | c22 |  | IVAS FL | 32 | 5% | on |
| 7 | c03 |  | MNRU Q = 36 dB | - | - |  |
| 8 | c01 |  | Reference | - | - |  |
| 9 | c27 |  | IVAS FL | 128 | 5% | on |
| 10 | c07 |  | ESDRU = 0.5 | - | - |  |
| 11 | c05 |  | MNRU Q = 18 dB | - | - |  |
| 12 | c24 |  | IVAS FL | 64 | 5% | on |

Table F.11.3: Test conditions for Experiment P800-11,  
 under impaired channel conditions

Alternative 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **FER** |
| c01 | Reference | - | - |  |
| c02 | MNRU Q = 36 dB | - | - |  |
| c03 | MNRU Q = 30 dB | - | - |  |
| c04 | MNRU Q = 24 dB | - | - |  |
| c05 | MNRU Q = 18 dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU |  |  |  |
| C09 | ESDRU | - | - |  |
| c10 | IVAS FL enc / FX dec | 13.2 | on | 5% |
| c11 | IVAS FX enc / FL dec | 16.4 | on | 5% |
| c12 | IVAS FL enc / FX dec | 24.4 | on | 5% |
| c13 | IVAS FX enc / FL dec | 32.0 | on | 5% |
| c14 | IVAS FL enc / FX dec | 48.0 | on | 5% |
| c15 | IVAS FX enc / FL dec | 64.0 | on | 5% |
| c16 | IVAS FL enc / FX dec | 80.0 | on | 5% |
| c17 | IVAS FX enc / FL dec | 96.0 | on | 5% |
| c18 | IVAS FL enc / FX dec | 128.0 | on | 5% |
| c19 | IVAS FL | 13.2 | on | 5% |
| c20 | IVAS FL | 16.4 | on | 5% |
| c21 | IVAS FL | 24.4 | on | 5% |
| c22 | IVAS FL | 32.0 | on | 5% |
| c23 | IVAS FL | 48.0 | on | 5% |
| c24 | IVAS FL | 64.0 | on | 5% |
| c25 | IVAS FL | 80.0 | on | 5% |
| c26 | IVAS FL | 96.0 | on | 5% |
| c27 | IVAS FL | 128.0 | on | 5% |
| c28 | IVAS FX | 13.2 | on | 5% |
| c29 | IVAS FX | 16.4 | on | 5% |
| c30 | IVAS FX | 24.4 | on | 5% |
| c31 | IVAS FX | 32.0 | on | 5% |
| c32 | IVAS FX | 48.0 | on | 5% |
| c33 | IVAS FX | 64.0 | on | 5% |
| c34 | IVAS FX | 80.0 | on | 5% |
| c35 | IVAS FX | 96.0 | on | 5% |
| c36 | IVAS FX | 128.0 | on | 5% |

Alternative 2:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **FER** |
| c01 | Reference | - | - |  |
| c02 | MNRU Q=xx dB | - | - |  |
| c03 | MNRU Q=xx dB | - | - |  |
| c04 | MNRU Q=xx dB | - | - |  |
| c05 | MNRU Q=xx dB | - | - |  |
| c06 | ESDRU | - | - |  |
| c07 | ESDRU | - | - |  |
| c08 | ESDRU |  |  |  |
| C09 | IVAS FL enc / FX dec | 13.2 | on | 0% |
| c10 | IVAS FX enc / FL dec | 16.4 | on | 0% |
| c11 | IVAS FL enc / FL dec | 24.4 | on | 0% |
| c12 | IVAS FX enc / FX dec | 32.0 | on | 0% |
| c13 | IVAS FL enc / FX dec | 48.0 | on | 0% |
| c14 | IVAS FX enc / FL dec | 64.0 | on | 0% |
| c15 | IVAS FL enc / FL dec | 80.0 | on | 0% |
| c16 | IVAS FX enc / FX dec | 13.2 | off | 4% |
| c17 | IVAS FL enc / FX dec | 16.4 | off | 4% |
| c18 | IVAS FX enc / FL dec | 24.4 | off | 4% |
| c19 | IVAS FL enc / FL dec | 32.0 | off | 4% |
| c20 | IVAS FX enc / FX dec | 48.0 | off | 4% |
| c21 | IVAS FL enc / FX dec | 64.0 | off | 4% |
| c22 | IVAS FX enc / FL dec | 80.0 | off | 4% |
| c23 | IVAS FL enc / FL dec | 13.2 | on | 8% |
| c24 | IVAS FX enc / FX dec | 16.4 | on | 8% |
| c25 | IVAS FL enc / FX dec | 24.4 | on | 8% |
| c26 | IVAS FX enc / FL dec | 32.0 | on | 8% |
| c27 | IVAS FL enc / FL dec | 48.0 | on | 8% |
| c28 | IVAS FX enc / FX dec | 64.0 | on | 8% |
| c29 | IVAS FL enc / FX dec | 80.0 | on | 8% |
| c30 | IVAS FX enc / FL dec | 13.2 | off | 8% |
| c31 | IVAS FL enc / FL dec | 16.4 | off | 8% |
| c32 | IVAS FX enc / FX dec | 24.4 | off | 8% |
| c33 | IVAS FL enc / FX dec | 32.0 | off | 8% |
| c34 | IVAS FX enc / FL dec | 48.0 | off | 8% |
| c35 | IVAS FL enc / FL dec | 64.0 | off | 8% |
| c36 | IVAS FX enc / FX dec | 80.0 | off | 8% |

**Scene definitions categories 1-2**

A leading and trailing silence is present for each sample, in accordance with IVAS-7b. 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:

* Talker sitting at a table (elevation 0°), at different azimuths.
* Standing talker (elevation 35°), at different azimuths.
* Smaller talker (child) walking around a table in the positive sense (counterclockwise), elevation 0°. Azimuth varies continuously for the sentence pair.
* Adult talker walking around a table in the negative sense (clockwise), elevation 35°. Azimuth varies continuously for the sentence pair.
* Elevation displacement: Elevation varies continuously for the sentence pair. Azimuth is constant for a sentence pair, but different for each sentence pair.
* Azimuth and elevation displacement: Azimuth and elevation vary continuously.

Each of the sentences uttered by a certain talker is encoded using different scene. Allocation of scenes to each panel is given in the Table F.11.4

**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-7b. 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.

Table F.11.4: Allocation of scenes for each talker pair (category cat 1, cat 2, cat 4, cat 5) and listening panel (P1-P6)







|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Scene*** | ***Talker initial elevation*** | ***Elevation change(1*** | ***Talker initial azimuth*** | ***Azimuth change(2*** | ***Panel*** |
| ***cat 1:***  *M1* | a  b  e  f  c  d | 0°  35°  -90°  35°  0°  35° | static  static  0.3°/ frame  -0.2°/ frame  static  static | 270°  180°  120°  0°  240°  180° | static  static  static  0.5°/ frame  1°/ frame  -1°/ frame | P1  P2  P3  P4  P5  P6 |
| ***cat 2:***  *F1* | f  c  d  a  b  e | 35°  0°  35°  0°  35°  -90° | -0.2°/ frame  static  static  static  static  0.3°/ frame | 300°  60°  120°  60°  300°  60° | 0.5°/ frame  1°/ frame  -1°/ frame  static  static  static | P1  P2  P3  P4  P5  P6 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Scene*** | ***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 4:***  *M2 + F2* | *a*  *b*  *c*  *d*  *e*  *f* | *-1*  *1*  *-1*  *1*  *-1*  *1* | *0°*  *35°*  *0°*  *0°*  *45°*  *30°* | *0°*  *35°*  *45°*  *45°*  *45°*  *30°* | *0°*  *10°*  *20°*  *200°*  *340°*  *120°* | *static*  *static*  *static*  *static*  *-1°/ frame*  *1°/ frame* | *50°*  *110°*  *170°*  *30°*  *340°*  *120°* | *static*  *static*  *static*  *-1°/ frame*  *-1°/ frame*  *-1°/ frame* | *P1*  *P2*  *P3*  *P4*  *P5*  *P6* |
| ***cat 5:***  *M3 + F3* | *d*  *e*  *f*  *a*  *b*  *c* | *1*  *-1*  *1*  *-1*  *1*  *-1* | *0°*  *45°*  *30°*  *0°*  *35°*  *0°* | *45°*  *45°*  *30°*  *0°*  *35°*  *45°* | *50°*  *130°*  *300°*  *30°*  *40°*  *50°* | *static*  *1°/ frame*  *1°/ frame*  *static*  *static*  *static* | *180°*  *130°*  *300°*  *230°*  *290°*  *350°* | *1°/ frame*  *1°/ frame*  *-1°/ frame*  *static*  *static*  *static* | *P1*  *P2*  *P3*  *P4*  *P5*  *P6* |

Table F.11.5: Categories for Speech with effects, and Speech with music or Music

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 | Speech + effects (3 or 4 objects) |
| cat 6 | Speech + music or music only (3 or 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 P800-12: MASA 1 TC

Tables F.12.1 to F.12.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Mixed content and Generic audio categories, respectively.

Table F.12.1: Conditions for Experiment P800-5

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64, 80, 96, 128 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP |
| **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 [19]. |
| 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.12.2: Preliminaries for Experiment P800-5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c26 |  | IVAS FL | 96 | No errors |
| 3 | c06 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c09 |  | ESDRU = xx | - | - |
| 6 | c22 |  | IVAS FL | 32 | No errors |
| 7 | c03 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table F.12.3: Test conditions for Experiment P800-5,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q=xx dB | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 13.2 |  |
| c11 | IVAS FX enc / FL dec | 16.4 |  |
| c12 | IVAS FL enc / FX dec | 24.4 |  |
| c13 | IVAS FX enc / FL dec | 32.0 |  |
| c14 | IVAS FL enc / FX dec | 48.0 |  |
| c15 | IVAS FX enc / FL dec | 64.0 |  |
| c16 | IVAS FL enc / FX dec | 80.0 |  |
| c17 | IVAS FX enc / FL dec | 96.0 |  |
| c18 | IVAS FL enc / FX dec | 128.0 |  |
| 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 |  |

Table F.12.4: Clean and noisy speech 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* |

Table F.12.5: Mixed content and Generic audio 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-8b.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8b. 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 P800-13: MASA 2 TCs

Tables F.13.1 to F.13.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Mixed content and Generic audio categories, respectively.

Table F.13.1: Conditions for Experiment P800-13

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64, 80, 96, 128 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP |
| **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 [19]. |
| 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.13.2: Preliminaries for Experiment P800-13

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c26 |  | IVAS FL | 96 | No errors |
| 3 | c06 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c09 |  | ESDRU = xx | - | - |
| 6 | c22 |  | IVAS FL | 32 | No errors |
| 7 | c03 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table F.13.3: Test conditions for Experiment P800-13,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q=xx dB | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 13.2 |  |
| c11 | IVAS FX enc / FL dec | 16.4 |  |
| c12 | IVAS FL enc / FX dec | 24.4 |  |
| c13 | IVAS FX enc / FL dec | 32.0 |  |
| c14 | IVAS FL enc / FX dec | 48.0 |  |
| c15 | IVAS FX enc / FL dec | 64.0 |  |
| c16 | IVAS FL enc / FX dec | 80.0 |  |
| c17 | IVAS FX enc / FL dec | 96.0 |  |
| c18 | IVAS FL enc / FX dec | 128.0 |  |
| 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 |  |

Table F.13.4: Clean and noisy speech 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* |

Table F.13.5: Mixed content and Generic audio 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-8b.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8b. 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 P800-14: MASA 1-2 TC

Tables F.14.1 to F.14.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Mixed content and Generic audio categories, respectively.

Table F.14.1: Conditions for Experiment P800-14

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64, 80, 96, 128 kbps |
| DTX | DTX on |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise for cat 1,2,5,6, 15dB for cat 3,4 |
| Error Conditions | 5% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP |
| **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 [19]. |
| 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.14.2: Preliminaries for Experiment P800-14

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** | **DTX** |
| 1 | c19 |  | IVAS FL | 13.2 | 5% | on |
| 2 | c26 |  | IVAS FL | 96 | 5% | on |
| 3 | c06 |  | ESDRU = xx | - | - |  |
| 4 | c20 |  | IVAS FL | 16.4 | 5% | on |
| 5 | c09 |  | ESDRU = xx | - | - |  |
| 6 | c22 |  | IVAS FL | 32 | 5% | on |
| 7 | c03 |  | MNRU Q=xx dB | - | - |  |
| 8 | c01 |  | Reference | - | - |  |
| 9 | c27 |  | IVAS FL | 128 | 5% | on |
| 10 | c07 |  | ESDRU = xx | - | - |  |
| 11 | c05 |  | MNRU Q=xx dB | - | - |  |
| 12 | c24 |  | IVAS FL | 64 | 5% | on |

Table F.14.3: Test conditions for Experiment P800-14,  
under impaired channel conditions

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Label** | **Condition** | **MASA TC mode** | **Bitrate [kbps]** | **DTX** | **FER** |
| c01 | Reference 1 TC | 1 TC | - | - |  |
| c02 | Reference 2 TC | 2 TC | - | - |  |
| c03 | MNRU Q=xx dB | 1 TC | - | - |  |
| c04 | MNRU Q=xx dB | 1 TC | - | - |  |
| c05 | MNRU Q=xx dB | 1 TC | - | - |  |
| c06 | ESDRU | 2 TC | - | - |  |
| c07 | ESDRU | 2 TC | - | - |  |
| c08 | ESDRU | 2 TC |  |  |  |
| C09 | IVAS FL enc / FX dec | 1 TC | 13.2 | on | 4% |
| c10 | IVAS FX enc / FL dec | 1 TC | 16.4 | on | 4% |
| c11 | IVAS FL enc / FL dec | 1 TC | 24.4 | on | 4% |
| c12 | IVAS FX enc / FX dec | 1 TC | 32.0 | on | 4% |
| c13 | IVAS FL enc / FX dec | 1 TC | 48.0 | on | 4% |
| c14 | IVAS FX enc / FL dec | 1 TC | 64.0 | on | 4% |
| c15 | IVAS FL enc / FL dec | 1 TC | 80.0 | on | 4% |
| c16 | IVAS FX enc / FX dec | 2 TC | 13.2 | on | 4% |
| c17 | IVAS FL enc / FX dec | 2 TC | 16.4 | on | 4% |
| c18 | IVAS FX enc / FL dec | 2 TC | 24.4 | on | 4% |
| c19 | IVAS FL enc / FL dec | 2 TC | 32.0 | on | 4% |
| c20 | IVAS FX enc / FX dec | 2 TC | 48.0 | on | 4% |
| c21 | IVAS FL enc / FX dec | 2 TC | 64.0 | on | 4% |
| c22 | IVAS FX enc / FL dec | 2 TC | 80.0 | on | 4% |
| c23 | IVAS FL enc / FL dec | 1 TC | 13.2 | on | 8% |
| c24 | IVAS FX enc / FX dec | 1 TC | 16.4 | on | 8% |
| c25 | IVAS FL enc / FX dec | 1 TC | 24.4 | on | 8% |
| c26 | IVAS FX enc / FL dec | 1 TC | 32.0 | on | 8% |
| c27 | IVAS FL enc / FL dec | 1 TC | 48.0 | on | 8% |
| c28 | IVAS FX enc / FX dec | 1 TC | 64.0 | on | 8% |
| c29 | IVAS FL enc / FX dec | 1 TC | 80.0 | on | 8% |
| c30 | IVAS FX enc / FL dec | 2 TC | 13.2 | on | 8% |
| c31 | IVAS FL enc / FL dec | 2 TC | 16.4 | on | 8% |
| c32 | IVAS FX enc / FX dec | 2 TC | 24.4 | on | 8% |
| c33 | IVAS FL enc / FX dec | 2 TC | 32.0 | on | 8% |
| c34 | IVAS FX enc / FL dec | 2 TC | 48.0 | on | 8% |
| c35 | IVAS FL enc / FL dec | 2 TC | 64.0 | on | 8% |
| c36 | IVAS FX enc / FX dec | 2 TC | 80.0 | on | 8% |

Table 14.4: Clean and noisy speech 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* |

Table F.14.5: Mixed content and Generic audio 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-8b.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8b. 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 P800-15: OSBA (1-2 objects)

Tables F.15.1 to F.15.3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively. Tables F.15.4 to F.15.5 show definition of categories and scenes.

Table F.15.1: Conditions for Experiment P800-15

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 32, 48, 64, 96, 128, 192, 256, 384, 512 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | 15 dB for cat 1,2,3,4, tbd for cat 5,6 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Cat. 1-3: Defined scenes, 1 ISM + FOA [or/and HOA3 if available] background  Cat. 4-6: Defined scenes, 2 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.15.2: Preliminaries for Experiment P800-15

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 32 | No errors |
| 2 | c26 |  | IVAS FL | 384 | No errors |
| 3 | c06 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 48 | No errors |
| 5 | c09 |  | ESDRU = xx | - | - |
| 6 | c22 |  | IVAS FL | 96 | No errors |
| 7 | c03 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 512 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 192 | No errors |

Table F.15.3: Test conditions for Experiment P800-15,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q=xx dB | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 32.0 |  |
| c11 | IVAS FX enc / FL dec | 48.0 |  |
| c12 | IVAS FL enc / FX dec | 64.0 |  |
| c13 | IVAS FX enc / FL dec | 96.0 |  |
| c14 | IVAS FL enc / FX dec | 128.0 |  |
| c15 | IVAS FX enc / FL dec | 192.0 |  |
| c16 | IVAS FL enc / FX dec | 256.0 |  |
| c17 | IVAS FX enc / FL dec | 384.0 |  |
| c18 | IVAS FL enc / FX dec | 512.0 |  |
| c19 | IVAS FL | 32.0 |  |
| c20 | IVAS FL | 48.0 |  |
| c21 | IVAS FL | 64.0 |  |
| c22 | IVAS FL | 96.0 |  |
| c23 | IVAS FL | 128.0 |  |
| c24 | IVAS FL | 192.0 |  |
| c25 | IVAS FL | 256.0 |  |
| c26 | IVAS FL | 384.0 |  |
| c27 | IVAS FL | 512.0 |  |
| c28 | IVAS FX | 32.0 |  |
| c29 | IVAS FX | 48.0 |  |
| c30 | IVAS FX | 64.0 |  |
| c31 | IVAS FX | 96.0 |  |
| c32 | IVAS FX | 128.0 |  |
| c33 | IVAS FX | 192.0 |  |
| c34 | IVAS FX | 256.0 |  |
| c35 | IVAS FX | 384.0 |  |
| c36 | IVAS FX | 512.0 |  |

Table F.15.4: Clean and noisy speech categories and scene definitions for SBA

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***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 4* | *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 5* | *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* |
|  |  |  |  |  |  |  |  |

Note: ISM positions for categories cat 1-4 are as defined respectively in Table F.9.4.

Table F.15.5: Categories for Objects with Generic audio background

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 | 1-object + General audio background |
| cat 6 | 2-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-8b.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8b. 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 P800-16: OSBA (3-4 objects)

Tables F.16.1 to F.16.4 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively. Tables F.16.4 to F.16.5 show definition of categories and scenes.

Table F.16.1: Conditions for Experiment P800-16

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 32, 48, 64, 96, 128, 192, 256, 384, 512 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | 15 dB for cat 1,2,3,4, tbd for cat 5,6 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Cat. 1-3: Pre-produced content, 3 ISMs + FOA [or/and HOA3 if available] background  Cat. 4-6: Pre-produced content, 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.16.2: Preliminaries for Experiment P800-16

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 32 | No errors |
| 2 | c26 |  | IVAS FL | 384 | No errors |
| 3 | c06 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 48 | No errors |
| 5 | c09 |  | ESDRU = xx | - | - |
| 6 | c22 |  | IVAS FL | 96 | No errors |
| 7 | c03 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 512 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 192 | No errors |

Table F.16.3: Test conditions for Experiment P800-16,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q=xx dB | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 32.0 |  |
| c11 | IVAS FX enc / FL dec | 48.0 |  |
| c12 | IVAS FL enc / FX dec | 64.0 |  |
| c13 | IVAS FX enc / FL dec | 96.0 |  |
| c14 | IVAS FL enc / FX dec | 128.0 |  |
| c15 | IVAS FX enc / FL dec | 192.0 |  |
| c16 | IVAS FL enc / FX dec | 256.0 |  |
| c17 | IVAS FX enc / FL dec | 384.0 |  |
| c18 | IVAS FL enc / FX dec | 512.0 |  |
| c19 | IVAS FL | 32.0 |  |
| c20 | IVAS FL | 48.0 |  |
| c21 | IVAS FL | 64.0 |  |
| c22 | IVAS FL | 96.0 |  |
| c23 | IVAS FL | 128.0 |  |
| c24 | IVAS FL | 192.0 |  |
| c25 | IVAS FL | 256.0 |  |
| c26 | IVAS FL | 384.0 |  |
| c27 | IVAS FL | 512.0 |  |
| c28 | IVAS FX | 32.0 |  |
| c29 | IVAS FX | 48.0 |  |
| c30 | IVAS FX | 64.0 |  |
| c31 | IVAS FX | 96.0 |  |
| c32 | IVAS FX | 128.0 |  |
| c33 | IVAS FX | 192.0 |  |
| c34 | IVAS FX | 256.0 |  |
| c35 | IVAS FX | 384.0 |  |
| c36 | IVAS FX | 512.0 |  |

Table F.16.4: Clean and noisy speech categories and scene definitions for SBA

|  |  |  |
| --- | --- | --- |
| **Category** | **Type** | |
| cat 1, 4 | 3-object, 4-objects respectively + ambient background (office, nature, outdoor/indoor enviroment), where objects are speech (including overlap with max. 2 talkers) | |
| cat 2, 5 | 3-object, 4-objects respectively + mixed and music where objects are speech (including overlap with max. 2 talkers) | |
| cat 3, 6 | 3-object, 4-objects respectively + General audio background, where objects may be speech, music or/and effects |

* 1. Experiment P800-17: OSBA (1-4 objects)

Table F17.1: 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 [5] | 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 Table F.17.2 |
| 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 ([8]) |
| Background signal (SBA) input loudness | 1 | -36 LKFS ([8]) |
| ISM speech input loudness | 1 | -26 LKFS ([8]) |
| 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 [12] |
| Languages | 1 | [tbd] |
| Listening System | 1 | High-quality headphones, diotic presentation |
| Listening Environment | 1 | No noise |

Table F.17.2: 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 F.17.3: Test conditions for P.800 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 P800-18: OMASA (1-2 objects)

Tables F.18.1 to F.18.3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively. Tables F.18.4 to F.18.5 show definition of categories and scenes.

Table F.18.1: Conditions for Experiment P800-18

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64, 80, 96, 128 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | 15 dB for cat 1,2,3,4, tbd for cat 5,6 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Cat. 1-3: Defined scenes, 1 ISM + MASA background  Cat. 3-6: Defined scenes, 2 ISMs + MASA background |
| Binaural renderer | OMASA 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.18.2: Preliminaries for Experiment P800-18

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c26 |  | IVAS FL | 96 | No errors |
| 3 | c06 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c09 |  | ESDRU = xx | - | - |
| 6 | c22 |  | IVAS FL | 32 | No errors |
| 7 | c03 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table F.18.3: Test conditions for Experiment P800-18,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q=xx dB | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 13.2 |  |
| c11 | IVAS FX enc / FL dec | 16.4 |  |
| c12 | IVAS FL enc / FX dec | 24.4 |  |
| c13 | IVAS FX enc / FL dec | 32.0 |  |
| c14 | IVAS FL enc / FX dec | 48.0 |  |
| c15 | IVAS FX enc / FL dec | 64.0 |  |
| c16 | IVAS FL enc / FX dec | 80.0 |  |
| c17 | IVAS FX enc / FL dec | 96.0 |  |
| c18 | IVAS FL enc / FX dec | 128.0 |  |
| 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 |  |

Table F.18.4: Clean and noisy speech categories and scene definitions for SBA

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Environment(1*** | ***Background(2*** | ***SNR [dB]*** | ***Overtalk [s](3*** | ***Bandwidth*** | ***Talker positions(4*** | ***Talker selection by panel(5*** |
| *cat 1* | *out\_[X]\_MASA* | *[park\_1\_bg\_MASA / nature\_1\_bg\_MASA / event\_1\_bg\_MASA / street\_[1/2]\_bg\_MASA]* | *15* | *1* | *Max* |  | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *room\_[X]\_MASA* | *[cafeteria\_1\_bg\_MASA / mall\_1\_bg\_MASA/ office[1/2]\_bg\_MASA]* | *15* | *-1* | *Max* |  | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* |  |  |  |  |  |  |  |
| *cat 4* | *out\_[X]\_MASA* | *[park\_1\_bg\_MASA / nature\_1\_bg\_MASA / event\_1\_bg\_MASA / street\_[1/2]\_bg\_MASA]* | *15* | *1* | *Max* |  | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 5* | *room\_[X]\_MASA* | *[cafeteria\_1\_bg\_MASA / mall\_1\_bg\_MASA/ office[1/2]\_bg\_MASA]* | *15* | *-1* | *Max* |  | *P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3* |
| *cat 6* |  |  |  |  |  |  |  |

Note: ISM positions for categories cat 1-4 are as defined respectively in Table F.9.4.

**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-8b.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8b. 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 P800-19: OMASA (3-4 objects)

Tables F.19.1 to F.19.4 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively. Tables F.19.4 to F.19.5 show definition of categories and scenes.

Table F.19.1: Conditions for Experiment P800-19

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64, 80, 96, 128 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | 15 dB for cat 1,2,3,4, tbd for cat 5,6 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q= xx, xx, xx, xx dB  *α* = xx, xx, xx |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Cat. 1-3: Pre-produced content, 3 ISMs + MASA background  Cat. 4-6: Pre-produced content, 4 ISMs + MASA background |
| Binaural renderer | OMASA 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 4.3 |
| Listening System | Headphones, in accordance with clause 4.5 |
| Listening Environment | No room noise |

Table F.19.2: Preliminaries for Experiment P800-19

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c26 |  | IVAS FL | 96 | No errors |
| 3 | c06 |  | ESDRU = xx | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c09 |  | ESDRU = xx | - | - |
| 6 | c22 |  | IVAS FL | 32 | No errors |
| 7 | c03 |  | MNRU Q=xx dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 128 | No errors |
| 10 | c07 |  | ESDRU = xx | - | - |
| 11 | c05 |  | MNRU Q=xx dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

Table F.19.3: Test conditions for Experiment P800-19,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q=xx dB | - | - |
| c03 | MNRU Q=xx dB | - | - |
| c04 | MNRU Q=xx dB | - | - |
| c05 | MNRU Q=xx dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU |  |  |
| C09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 13.2 |  |
| c11 | IVAS FX enc / FL dec | 16.4 |  |
| c12 | IVAS FL enc / FX dec | 24.4 |  |
| c13 | IVAS FX enc / FL dec | 32.0 |  |
| c14 | IVAS FL enc / FX dec | 48.0 |  |
| c15 | IVAS FX enc / FL dec | 64.0 |  |
| c16 | IVAS FL enc / FX dec | 80.0 |  |
| c17 | IVAS FX enc / FL dec | 96.0 |  |
| c18 | IVAS FL enc / FX dec | 128.0 |  |
| 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 |  |

Table F.19.4: Clean and noisy speech categories and scene definitions for SBA

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 1-3 | 3-object + General audio background |
| cat 4-6 | 4-objects + General audio background |

* 1. Experiment P800-20: OMASA (1-4 objects)

Table F.20.1: 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 [5] | 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 ([8]) |
| Background signal (MASA) input loudness | 1 | -36 LKFS ([8]) |
| ISM speech input loudness | 1 | -26 LKFS ([8]) |
| 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 [12] |
| Languages | 1 | [tbd] |
| Listening System | 1 | High-quality headphones, diotic presentation |
| Listening Environment | 1 | No noise |

Table F.20.2: 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 F.20.3: Test conditions for P.800 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. Experiment P800-21:
  2. Experiment P800-21:
  3. Experiment P800-21:

]

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

Table G.1.1: Conditions (BS1534-1a Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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* |
|  |  |
| ***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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.1.2: Test conditions for Experiment BS1534-1a

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 16.4 |
| c04 | Mono EVS | 32 |
| c05 | IVAS | 16.4 |
| c06 | IVAS | 24.4 |
| c07 | IVAS | 32 |
| c08 | IVAS | 48 |

* 1. Experiment BS1534-1b: Stereo

Table G.2.1: Conditions (BS1534-1b Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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* |
|  |  |
| ***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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.2.2: Test conditions for Experiment BS1534-1b

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 64 |
| c04 | Mono EVS | 128 |
| c05 | IVAS | 64 |
| c06 | IVAS | 96 |
| c07 | IVAS | 128 |
| c08 | IVAS | 256 |

* 1. Experiment BS1534-2a: FOA

Table G.3.1: Conditions (BS1534-2a Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.3.2: Test conditions for Experiment BS1534-2a

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 16.4 |
| c04 | Mono EVS | 32 |
| c05 | IVAS | 16.4 |
| c06 | IVAS | 24.4 |
| c07 | IVAS | 32 |
| c08 | IVAS | 48 |

* 1. Experiment BS1534-2b: FOA

Table G.4.1: Conditions (BS1534-2b Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.4.2: Test conditions for Experiment BS1534-2b

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 64 |
| c04 | Mono EVS | 128 |
| c05 | IVAS | 64 |
| c06 | IVAS | 96 |
| c07 | IVAS | 128 |
| c08 | IVAS | 256 |

* 1. Experiment BS1534-3: HOA3

Table G.5.1: Conditions (BS1534-3 Generic Audio)

|  |  |  |
| --- | --- | --- |
| ***Main Codec Conditions*** |  |  |
| *Codec under Test (CuT)* |  | *IVAS FX 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 loudspeaker: 7.1+4 overhead speaker setup with the configuration following clause* *4.5* |
| *Listening Environment* |  | *No room noise* |

Table G.5.2: Test conditions for Experiment BS1534-3

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 64 |
| c04 | Mono EVS | 128 |
| c05 | IVAS | 64 |
| c06 | IVAS | 96 |
| c07 | IVAS | 128 |
| c08 | IVAS | 256 |

* 1. Experiment BS1534-4: Multichannel 5.1

Table G.6.1: Conditions (BS1534-4 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 loudspeaker: 5.1,* *following clause* *4.5* |
| *Listening Environment* | *No room noise* |

Table G.6.2: Test conditions for Experiment BS1534-4

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 16.4 |
| c04 | Mono EVS | 32 |
| c05 | IVAS | 16.4 |
| c06 | IVAS | 24.4 |
| c07 | IVAS | 32 |
| c08 | IVAS | 48 |

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

Table G.7.1: Conditions (BS1534-5 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.7.2: Test conditions for Experiment BS1534-5

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 64 |
| c04 | Mono EVS | 128 |
| c05 | IVAS | 64 |
| c06 | IVAS | 96 |
| c07 | IVAS | 128 |
| c08 | IVAS | 256 |

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

Table G.8.1: Conditions (BS1534-6 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.8.2: Test conditions for Experiment BS1534-6

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 64 |
| c04 | Mono EVS | 128 |
| c05 | IVAS | 64 |
| c06 | IVAS | 96 |
| c07 | IVAS | 128 |
| c08 | IVAS | 256 |

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

Table G.9.1: Conditions (BS1534-7 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 loudspeaker: 7.1+4 overhead speaker setup with the configuration following clause 4.5* |
| *Listening Environment* | *No room noise* |

Table G.9.2: Test conditions for Experiment BS1534-7

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 64 |
| c04 | Mono EVS | 128 |
| c05 | IVAS | 64 |
| c06 | IVAS | 96 |
| c07 | IVAS | 128 |
| c08 | IVAS | 256 |

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

Table G.10.1: Conditions (BS1534-8 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.10.2: Test conditions for Experiment BS1534-8

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 64 |
| c04 | Mono EVS | 128 |
| c05 | IVAS | 64 |
| c06 | IVAS | 96 |
| c07 | IVAS | 128 |
| c08 | IVAS | 256 |

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

Table G.11.1: Conditions (BS1534-9a Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.11.2: Test conditions for Experiment BS1534-9a

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 24.4 |
| c04 | Mono EVS | 48 |
| c05 | IVAS | 24.4 |
| c06 | IVAS | 32 |
| c07 | IVAS | 48 |
| c08 | IVAS | 64 |

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

Table G.12.1: Conditions (BS1534-9b Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.12.2: Test conditions for Experiment BS1534-9b

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 64 |
| c04 | Mono EVS | 128 |
| c05 | IVAS | 64 |
| c06 | IVAS | 96 |
| c07 | IVAS | 128 |
| c08 | IVAS | 256 |

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

Table G.13.1: Conditions (BS1534-10a Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.13.2: Test conditions for Experiment BS1534-10a

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 16.4 |
| c04 | Mono EVS | 32 |
| c05 | IVAS | 16.4 |
| c06 | IVAS | 24.4 |
| c07 | IVAS | 32 |
| c08 | IVAS | 48 |

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

Table G.14.1: Conditions (BS1534-10b Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.14.2: Test conditions for Experiment BS1534-10b

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 64 |
| c04 | Mono EVS | 128 |
| c05 | IVAS | 64 |
| c06 | IVAS | 96 |
| c07 | IVAS | 128 |
| c08 | IVAS | 256 |

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

Table G.15.1: Conditions (BS1534-11 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.15.2: Test conditions for Experiment BS1534-11

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 64 |
| c04 | Mono EVS | 128 |
| c05 | IVAS | 64 |
| c06 | IVAS | 96 |
| c07 | IVAS | 128 |
| c08 | IVAS | 256 |

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

Table G.16.1: Conditions (BS1534-12a Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.16.2: Test conditions for Experiment BS1534-12a

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 16.4 |
| c04 | Mono EVS | 32 |
| c05 | IVAS | 16.4 |
| c06 | IVAS | 24.4 |
| c07 | IVAS | 32 |
| c08 | IVAS | 48 |

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

Table G.17.1: Conditions (BS1534-12b Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.17.2: Test conditions for Experiment BS1534-12b

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 64 |
| c04 | Mono EVS | 128 |
| c05 | IVAS | 64 |
| c06 | IVAS | 96 |
| c07 | IVAS | 128 |
| c08 | IVAS | 256 |

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

Table G.18.1: Conditions (BS1534-13 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX 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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.18.2: Test conditions for Experiment BS1534-13

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Mono EVS | 64 |
| c04 | Mono EVS | 128 |
| c05 | IVAS | 64 |
| c06 | IVAS | 96 |
| c07 | IVAS | 128 |
| c08 | IVAS | 256 |

* 1. Experiment BS1534-14: Sterep downmix for EVS

Table G.19.1: Conditions (BS1534-14 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *EVS encode-decode after stereo (dynamic) downmix*  *13.2 kbps, 24.4 kbps DTX off at 0% FER* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *Dual EVS encode-decode* ***before*** *stereo (static) downmix*  *7,2 kbps\*2, 13.2 kbps\*2, DTX off at 0% FER* |
|  |  |
| ***Other references*** |  |
| *Reference* | *EVS encode-decode* ***after*** *stereo (static) downmix*  *13.2 kbps, 24.4 kbps DTX off at 0% FER* |
| *Hidden Reference* | *Direct stereo signal,* |
| *LP3k5 anchor* | *3.5 kHz lowpass filtered stereo signal,* |
|  |  |
| ***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* |
| *Listening Environment* | *No room noise* |

Table G.19.2: Test conditions for Experiment BS1534-14

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | LP 3.5 kHz | - |
| c03 | Dual EVS + static dmx | 7.2 \*2 |
| c04 | Dual EVS + static dmx | 13.2 \*2 |
| c05 | Static dmx +EVS | 13.2 |
| c06 | Static dmx +EVS | 24.4 |
| c07 | CuT (dynamic dmx +EVS) | 13.2 |
| c08 | CuT (dynamic dmx +EVS) | 24.4 |

* 1. Experiment BS1534-15: ISM 6 DoF (4 objects)

Table G.20.1: Conditions (BS1534-15 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX operated with ISM (4 obj.) audio input with extended metadata enabled at*  *64 kbps, 512 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Codec references*** |  |
| *Codec references* | *IVAS operation with extended metadata disabled in the encoder and listener position set to zero in the decoder.*  *64 kbps, 512 kbps DTX off at 0% FER*  *Rendering via the IVAS-internal rendering* |
|  |  |
| ***Other references*** |  |
| *Reference* | *Direct signal, non-quantized (direct) extended metadata, rendered with IVAS external renderer* |
| *Hidden Reference* | *Direct signal, non-quantized (direct) extended metadata, rendered with IVAS external renderer* |
| *LP3k5 anchor* | *3.5 kHz lowpass filtered signal, non-quantized (direct) extended metadata, rendered with IVAS external renderer* |
|  |  |
| ***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 4.5* |
| *Listening Environment* | *No room noise* |

Table G.20.2: Test conditions for Experiment BS1534-15

|  |  |  |
| --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** |
| c01 | Reference | - |
| c02 | Hidden Reference |  |
| c03 | LP 3.5 kHz | - |
| c04 | IVAS, no ext. metadata | 64 |
| c05 | IVAS, no ext. metadata | 512 |
| c06 | IVAS, extended metadata | 64 |
| c07 | IVAS, extended metadata | 512 |