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

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

**Version: v.1.0.0**

**Agenda Item: 7.5**

Document History:

|  |  |  |
| --- | --- | --- |
| v.0.0.1 | 13 November 2023 | Initial draft |
| v.0.1.0 | 17 November 2023 | Integration of S4-231857 and comments received from the Audio SWG, editorial changes. |
| v.0.1.1 | 15 December 2023 | Editorial changes. |
| v.0.1.2 | 15 January 2024 | Integration of agreed text from S4aA230122, S4aA230123 and S4aA230124. Editorial changes. |
| v.0.2.0 | 2 February 2024 | Integration, in brackets, experiment proposals from S4aA230135, S4-240262, and S4-240298. |
| 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 |
| v.0.4.1 | 17 February 2025 | Update of experiment allocation to external LLs, update of schedule following S4-241690, adding missing tables for preliminaries for experiments in F.17 and F.20, document harmonization, further editorial changes |
| v.0.5.0 | 20 February 2025 | Update of the list of Permanent documents, integration of conclusion of S4-250212, updated testing time plan, editorial changes |
| v.0.5.1 | 11 April 2025 | Completion of Tables of Preliminaries in Annex F, Correction of bitrate rage in F.9 (experiment P800-9 for 2 ISMs), Insertion of missing Table of Preliminaries in F.20, editorial changes |
| v.0.6.0 | 17 April 2025 | Integration of agreed text from S4-250502, S4-250651, harmonization of preliminaries of experiments P800-17 and P800-20 with the rest of the P.800 experiments, editorial changes |
| v.0.6.1 | 19 May 2025 | Editorial changes |
| v.0.6.2 | 19 May 2025 | Editorial changes, corrections |
| v.0.7.0 | 23 May 2025 | Integration of proposals from S4-250874, integration of proposals from S4-251043 in brackets, assignment of LLs and testing languages, filling bitrate ranges in Tables 7 and 8, initial high-level schedule for the Characterization testing, completion of MNRU and ESDRU values, adjusting bitrates for experiments P800-3, P800-13, P800-18, and P800-19, corrections, editorial changes |
| V.0.7.1 | 17 June 2025 | For P.800 DCR FOA/HOA experiments described in Annex F, or for experiments derived from FOA, the environment and talker position definitions are moved to the processing scripts, edits in Annex E (Testing Timeline) |
| V.0.8.0 | 17 June 2025 | Integration of corrections and proposals from contributions S4aA250053 and S4aA250063, additional editorial changes agreed during the call. |
| V.0.8.1 | 21 July 2025 | Adding clarification to P.800 testing where relevant for DCR experiments only, editorial changes and corrections |
| V.1.0.0 | 24 July 2025 | Integration of agreements from S4-251411, S4-251256, S4-251413, and S4-251258. Editorial corrections. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

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-2b | IVAS\_Codec\_Ph2 |
| 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 |
| IVAS-10 | IVAS BASOP Verification |

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. Recommendation ITU-T P.800 (08/1996): Methods for subjective determination of transmission quality.
3. Recommendation ITU-T P.811 (01/2019): Subjective test methodology for evaluating Speech oriented stereo communication systems over headphones.
4. S4-211151: Example designs for IVAS codec tests, Source: Dolby Laboratories Inc.
5. S4-210836: On reference designs for IVAS codec tests, Source: Dolby Laboratories Inc.
6. Recommendation ITU-R BS.1770-4 (10/2015): Algorithms to measure audio programme loudness and true-peak audio level.
7. ITU-T Handbook of subjective testing practical procedures, 2011.
8. Supplement ITU-T P.Suppl29: "ITU-T P.800 – Use Cases".
9. Recommendation ITU-R BS.1534 (10/2015): Method for the subjective assessment of intermediate quality level of audio systems.
10. S4-030821: PSS/MMS High-Rate Audio Selection Test and Processing Plan, Version 2.2.
11. Audio File Format Specifications: WAVE, <https://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/WAVE.html>.
12. AFsp Package <https://www-mmsp.ece.mcgill.ca/Documents/Downloads/AFsp/>.
13. Recommendation ITU-T P.191 (03/2023): Software tools for speech and audio coding standardization.
14. S4-230221: Processing updates for IVAS MASA C Reference Software.
15. 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

ACR Absolute Category Rating

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

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

HOA2 Higher-Order Ambisonics, 2nd order

HOA3 Higher-Order Ambisonics, 3rd order

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

SNR Signal-to-Noise Ratio

SPL Sound Pressure Level

SWB Super Wide Band

TC Transport Channel

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 [11], 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 and x-checking the HL processing. This includes retrieving the up-to-date executables and processing scripts from the fixed-point and floating-point code repositories.

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

NTT

Orange

Panasonic Holdings Corporation

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). The tasks have to be conducted and completed following the schedule for the IVAS Characterization phase defined in IVAS-2b.

* + 1. Proponent Companies

The specific responsibilities of PC are:

* Get from the Public Collaboration public repository preliminary CuT executables and deliver them to the HL and to the CL.
* Get from the Public Collaboration public repository final CuT executables and deliver them to the HL and the CL
* Get from the Public Collaboration public repository the processing scripts and deliver them to the HL and the CL.
* 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. 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 [2] 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.
* Each LL obtains 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.
* Each LL delivers to the GAL all raw voting data using the data delivery file provided by the GAL for all tests to be conducted by the LL.
* Each LL delivers 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).

* Get from the PC preliminary CuT executables for a cross-check with the CL to identify potential problems.
* Get processing scripts from the PC.
* Deliver the cross-checked processing scripts to SA4.
* Receive all source materials and model parameters from the LLs and MC.
* Get from the PC 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.
* 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.
  + 1. Cross-check Laboratory

The following list defines the tasks expected to be carried out by the Cross-check Laboratory (CL).

* Get from the PC 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 PC.
* Receive all source materials and model parameters from the LLs and MC.
* Get from the PC 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.
* 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 sound 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).

* 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-2b.
  + - 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 DCR 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.

Results of the statistical tests for each P.800 DCR or BS.1534 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" Reference is 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 |

For the room acoustics tests (Annex I) the statistical tests will be based on ANOVA parametric variance analysis followed by Tukey’s HSD post-hoc test for pairwise comparisons between different rendering modes. The reference (anchor) will be the IVAS rendering with anechoic HRTF (IVAS\_Binaural).

Results of the tests will be presented for individual signal types as shown in Figure 1, and averaged for rendering modes, as shown in Figure 2. The markers above the plots in Figure 2 indicate pairwise comparisons considered as statistically significant (p < 0.05).

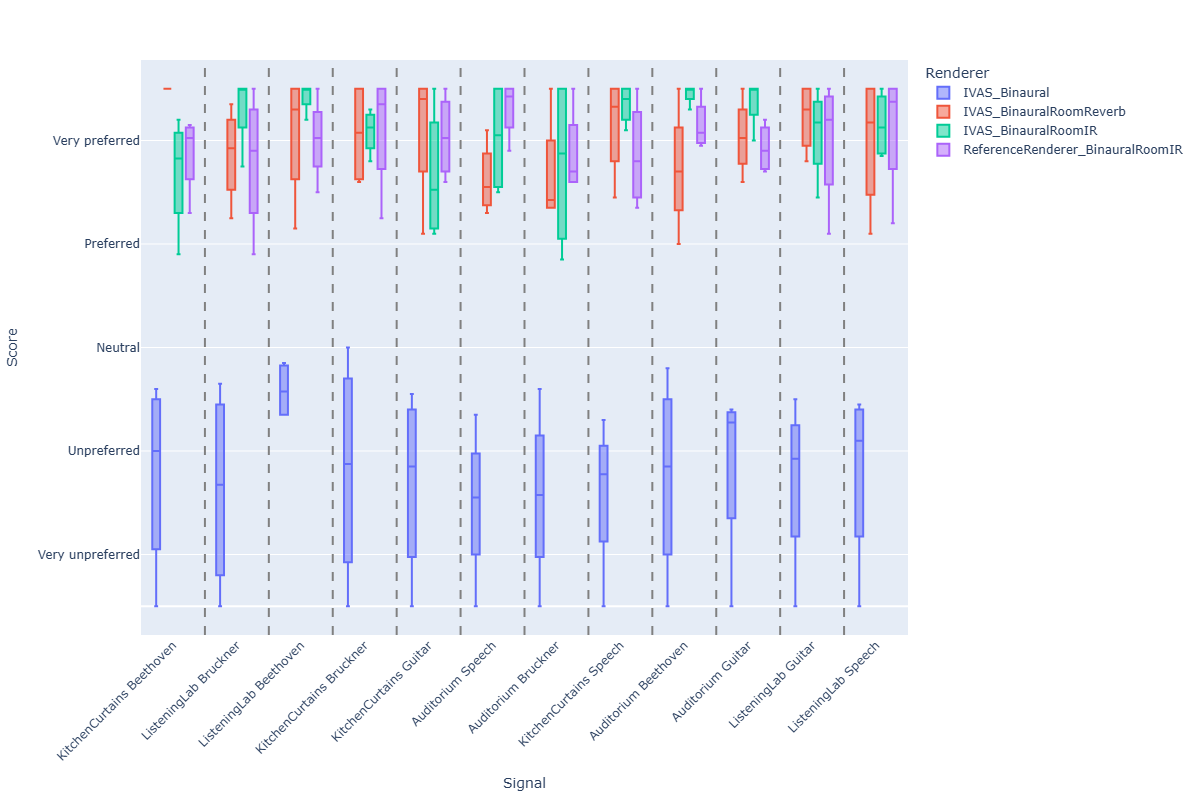


Figure 1. Detailed results per signal type

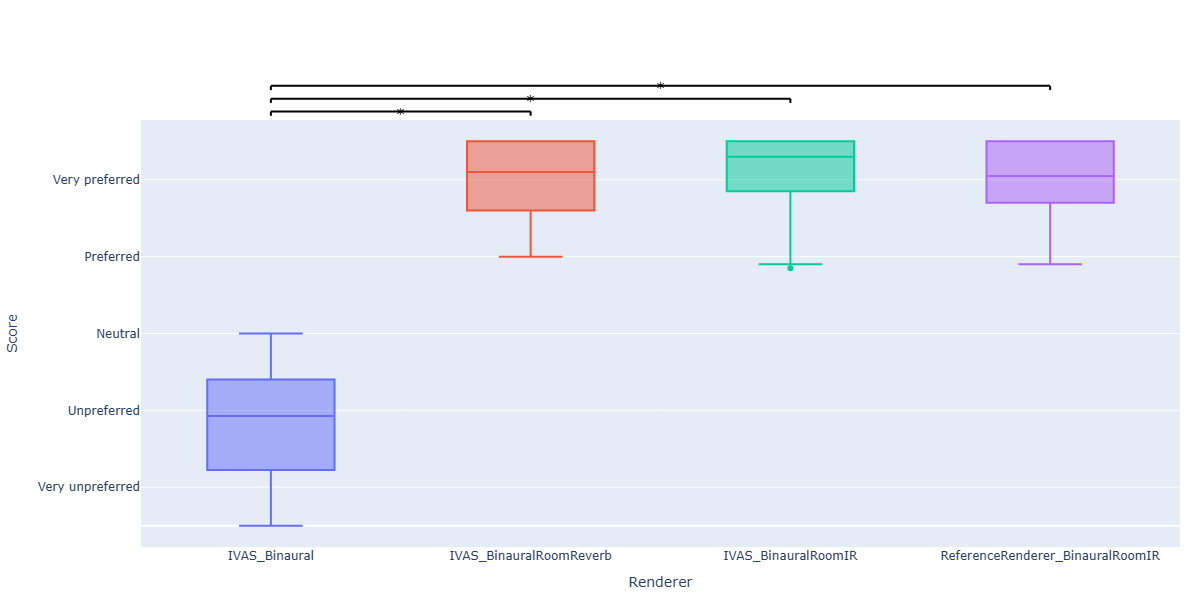


Figure 2. Average results per rendering mode

Next to that, the results of Tukey Honest Squares Difference post-hoc test will be provided as shown in Table 2. The ‘Significant’ column indicates statistically significant differences between the rendering modes. This shows whether there are statistically significant differences in perception of different rendering modes with room acoustics.

Table 2: Tukey HSD post-hoc test results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group 1** | **Group 2** | **Mean diff** | **p-adj** | **Lower** | **Upper** | **Significant** |
| IVAS\_Binaural | IVAS\_BinauralRoomIR | 66.8958 | 0.0 | 60.7568 | 73.0349 | True |
| IVAS\_Binaural | IVAS\_BinauralRoomReverb | 64.625 | 0.0 | 58.4859 | 70.7641 | True |
| IVAS\_Binaural | ReferenceRenderer\_BinauralRoomIR | 64.7292 | 0.0 | 58.5901 | 70.8682 | True |
| IVAS\_BinauralRoomIR | IVAS\_BinauralRoomReverb | -2.2708 | 0.7729 | -8.4099 | 3.8682 | False |
| IVAS\_BinauralRoomIR | ReferenceRenderer\_BinauralRoomIR | -2.1667 | 0.797 | -8.3057 | 3.9724 | False |
| IVAS\_BinauralRoomReverb | ReferenceRenderer\_BinauralRoomIR | 0.1042 | 1.0 | -6.0349 | 6.2432 | False |

P.800 ACR test does not have any requirements. Mean opinion scores with 95% confidence intervals for each condition will be provided.

* + 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 primary test methodologies shall be used in the IVAS Characterization test: P.800 [2] and BS.1534 [9]. Additionally, preference testing methodology shall be used for testing room acoustics synthesis. High-level configuration of experiments for envisaged methodologies is outlined below.

* + 1. P.800 DCR
* Considerations of P.Suppl29 [8] 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 [7].
* 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.
* P.800 experiments are performed with native listeners of the tested language.

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. Opinion Scales

Table 3 defines opinion scale used for ITU-T P.800 DCR test. Instructions in English for the P.800 tests are provided in Annex A:.

Table 3: 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. BS.1534
* Number of items per experiment: 12
* 14 experienced listeners
* Maximum total number of conditions: 8
* Number of anchor conditions: 2
  + Direct
  + 3.5 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 [9].

* + 1. P.800 ACR
* Considerations of P.Suppl29 [8] shall be taken into account as far as they can be interpreted for ACR test methodology.
* Test duration should not exceed 2 hours per listening panel.
* Randomizations constructed under randomized blocks experimental design described in [7].
* 6 sample categories for ACR test.
* 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 the ACR experiment: 60.
* Number of trials: number of test conditions x 6 talkers/categories = 360 trials.
* P.800 ACR experiment is performed with native listeners of the tested language.

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. After practice session all samples are played back in full quality (Spatial Fullband) to give listeners the feeling of the best possible quality. After listening through the material, actual listening test starts. There are comfort pauses after listening every 120 samples.

* + - 1. Opinion Scales

Table 4 defines opinion scale used for ITU-T P.800 ACR test. Instructions in English for the P.800 ACR test are provided in Annex A:.

Table 4: Opinion scale for ITU-T P.800 ACR test

|  |  |
| --- | --- |
| **Quality** | **Scale** |
| Excellent | 5 |
| Good | 4 |
| Fair | 3 |
| Poor | 2 |
| Bad | 1 |

* + 1. Preference testing for room acoustics synthesis

For room acoustics testing preference test methodology shall be used. The listeners are requested to provide preference scoring based on quality rating.

* 4 different rooms
* 4 audio items per room
* 4 conditions per test: IVAS rendering with HRTF (reference), IVAS rendering with BRIR, IVAS rendering with synthetic reverb, rendering with BRIR (Python implementation)
* Total: 64 trials.

Each testing experiment shall be preceded with a training phase in which the listeners should familiarize themselves with the environment and methodology. Unipolar preference scale should be used for rating.

* 1. Material

All audio material shall be sampled at 48 kHz with Full Band (FB) content. The audio material is to be delivered to the HL as 16-bit little endian WAVE format files [11] 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 [12].

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

* 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 4.3.1.3.
* Generic audio - the details are specified in Annex F:.

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 content.

Since it is expected that a significant amount of new material has to be checked and processed for the IVAS Codec Characterization test, test material proponents are encouraged to submit material identical to the Selection testing if possible (i.e. if format, test methodology and test lab match the setup of the Selection phase testing).

* + 1. Material for P.800 DCR testing
       1. Speech Material for P.800 DCR testing

Except for Music and mixed content categories, P.800 DCR 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 [15].

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

* + - 1. Music and Mixed Content Material for P.800 DCR 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 DCR experiments will be collected and selected by MC, similarly as done for the Generic Audio Items Selection for BS.1534 experiments (Clause 4.3.2.2).

* + - 1. Audio Material for 3- and 4-object categories in P.800 DCR 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 4.3.2.2). 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 [10]:

* 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.

In order to streamline the work of the Material Collection entity, the BS.1534 Selection test samples should be used also for the BS.1534 Characterization tests as far as possible, especially for the following formats:

* Stereo
* Ambisonics FOA, HOA2, HOA3
* Multi-channel 5.1, 7.1, 5.1+2, 5.1+4, 7.1+4 (7.1, 5.1+2, 5.1+4 to be derived from 7.1+4)
* Objects: 1-4 objects
* MASA 1-2 TCs 🡨 MASA material expected to be derived from FOA/HOA2 material

In order to facilitate this, the material proponents are asked to re-submit the identical material as for the Selection phase testing.

* + - 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 BS.1534 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.
* VR setup for 3DoF binaural room acoustics synthesis listening comprising of:
  + A gaming-category PC running Unity game engine, capable to also run 3 IVAS decoder instances and a lightweight Python renderer in parallel,
  + A VR headset, e.g., Meta Quest or HTC Vive Pro,
  + Headphones requirements for binaural tests apply.

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, HOA2, HOA3
* MASA 1 TC, 2 TC
* Combined input formats:
  + Objects + MASA (OMASA), 1-4 ISMs
  + Objects + SBA (OSBA), 1-4 ISMs
* JBM
* Packet loss conditions derived from delay and error profiles
* Binaural rendering configurations, e.g.
  + room effects,
  + head rotation,
  + 6 degrees-of-freedom (DoF) and directivity
* EVS-coded mono downmix of stereo input (13.2 and 24.4 kbps)

The subjective material in the P.800 DCR experiments comprising fixed-point validation accommodates all three input signal levels. Low and high signal levels are integrated in different audio samples within categories following Table 5.

Table 5: Assignment of input levels to audio samples within categories

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Samples** | **Input Level (in LKFS)** | | | | | |
|  | **cat 1** | **cat 2** | **cat 3** | **cat 4** | **cat 5** | **cat 6** |
| s01 | -16 | -36 | -36 | -26 | -26 | -16 |
| s02 | -16 | -16 | -36 | -36 | -26 | -26 |
| s03 | -26 | -16 | -16 | -36 | -36 | -26 |
| s04 | -26 | -26 | -16 | -16 | -36 | -36 |
| s05 | -36 | -26 | -26 | -16 | -16 | -36 |
| s06 | -36 | -36 | -26 | -26 | -16 | -16 |

Evaluation of different configurations of the IVAS codec require different testing methodologies, outlined in the following clauses and annexes.

* 1. P.800 DCR listening test layout

The following layout is a generic layout. The actual layouts of P.800 DCR experiments may be slightly different. They are described in Annex F:.

Table 6: P.800 DCR 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

The following layouts are generic layouts. The actual layouts of BS.1534 experiments may be slightly different. They are described in Annex G:.

Table 7: 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 8: 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 |

* 1. Allocation of experiments

The following tables show high-level overview of the experiments with allocation of experiments to LLs. Table 9 shows overview of P.800 DCR experiments with languages used for each experiment. Table 10 shows overview of BS.1534 experiments. Finally, Table 11 shows overview of experiments run with different methodologies.

Detail conditions for each subjective experiment are defined in Annex F: for P.800 DCR experiments, in Annex G: for BS.1534 experiments, in Annex H: for P.800 ACR experiments and in Annex I: for room acoustics experiments run with preference testing methodology.

**Assumptions**

* The same pricing for external LLs and GAL as in Selection phase (IVAS-8a), i.e. 18000 Euros per P.800 DCR test, 10000 Euros per BS.1534 test rendered via headphones, and 12000 Euros for GAL tasks.
* The external LL will run the following experiments:
  + Mesaqin: 5 x P.800 DCR (5 x 18000 = 90000 Euros)
  + Force: 5 x P.800 DCR (5 x 18000 = 90000 Euros)
  + Force: 1 x BS.1534 headphones test (1 x 10000 Euros)
* Minimum requirements for speech P.800 DCR experiments: 6 talkers (3 male + 3 female) per experiment, 14 single sentences per talker.
* For inputs 5.1+2, 5.1+4, 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 DCR experiments are assumed Full Band experiments, i.e., the direct reference condition is always FB. Speech categories of P.800 DCR stereo experiments are SWB.

.

Table 9: Allocation of P.800 DCR experiments to LLs and corresponding 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% | 13.2-128 | FR | VA |
| P800-2 | DTX, FE | Stereo | Al | Yes | On | 5% | 13.2-128 | JAP | NTT |
| P800-3 | FX, RD | FOA | All | Yes | Off | 0% | 13.2-256 | FR | Orange |
| P800-4 | DTX, FE | HOA2 | All | Yes | On | 5% | 16.4-384 | FR | Orange |
| P800-5 | FX, RD | HOA3 | All | Yes | Off | 0% | 32-512 | ENG | Dolby |
| P800-6 | FX, RD | MC 5.1, 7.1 | Clean speech, mixed/music | Yes | Off | 0% | 24.4-256 | MAN | Mesaqin |
| P800-7 | FX, RD | MC 5.1+4, 7.11+4 | Clean speech, mixed/music | Yes | Off | 0% | 32-384 | DAN | Force |
| P800-8 | FE | MC (mixed CICP) | Clean speech, mixed/music | Yes | Off | 5% | 16.4-384 | CZ | Mesaqin |
| P800-9 | FX, RD | 1-2 Objects | All | Yes | Off | 0% | 13.2-128 | MAN | Mesaqin |
| P800-10 | FX, RD | 3-4 Objects | speech+ effects,  speech + music, music | Yes | Off | 0% | 24.4-192 | FR | VA |
| P800-11 | DTX, FE | 1-4 Objects | Clean speech, speech+ effects,  speech + music, music | Yes | On | 5% | 24.4-192 | DAN | Force |
| P800-12 | FX, RD | MASA 1 TC | All | Yes | Off | 0% | 13.2-128 | MAN | Mesaqin |
| P800-13 | FX, RD | MASA 2 TC | All | Yes | Off | 0% | 13.2-128 | DAN | Force |
| P800-14 | DTX, FE | MASA 1,2 TC | All | Yes | On | 4, 8% | 13.2-128 | FIN | Nokia |
| P800-15 | FX, RD | OSBA, 1-2 Objects | All | Yes | Off | 0% | 32-512 | ENG | Dolby |
| P800-16 | FX, RD | OSBA, 3-4 Objects | All | Yes | Off | 0% | 32-512 | SWE | Ericsson |
| P800-17 | RD, FE | OSBA, 1-4 Objects | All | Yes | Off | 5% | 13.2-512 | GER | FhG |
| P800-18 | FX, RD | OMASA, 1-2 Objects | All | Yes | Off | 0% | 13.2-128 | SWE | Ericsson |
| P800-19 | FX, RD | OMASA, 3-4 Objects | All | Yes | Off | 0% | 13.2-128 | GER | FhG |
| P800-20 | RD, FE | OMASA, 1-4 Objects | All | Yes | Off | 5% | 13.2-512 | FIN | Nokia |
| P800-21 | JBM/FE | Stereo | All | Yes | On | I1.O1, I1.O2 | 24.4-96 | CZ | Mesaqin |
| P800-22 | JBM/FE | 1-2 Objects | All | Yes | Off | I1.O1, I1.O2 | 24.4-96 | DAN | Force |
| P800-23 | JBM/FE | FOA | All | Yes | Off | I1.O1, I1.O2 | 24.4-96 | DAN | Force |

JAP = Japanese

FR = French

GER = German

MAN = Mandarin

DAN = Danish

ENG = English

FIN = Finnish

SWE = Swedish

CZ = Czech

Table 10: Allocation of BS.1534 experiments to LLs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Exp** | **Objective** | **Input format** | **Source material** | **Bitrates [kb/s]** | **Listening environment** | **LL** |
| BS1534-1 | Compare to EVS | Stereo | Generic audio | 16.4-48 | Headphones | Force |
| BS1534-2 | Compare to EVS | Stereo | Generic audio | 64-256 | Headphones | Huawei |
| BS1534-3 | Compare to EVS | FOA | Generic audio | 16.4-48 | Headphones | Huawei |
| BS1534-4 | Compare to EVS | FOA | Generic audio | 64-256 | Headphones | Panasonic/NTT |
| BS1534-5 | Compare to EVS | HOA3 | Generic audio | 64-256 | 7.1 + 4 | Dolby |
| BS1534-6 | Compare to EVS | 5.1 | Generic audio | 16.4-48 | 5.1 | Ericsson |
| BS1534-7 | Compare to EVS | 5.1, 7.1 | Generic audio | 64-256 | Headphones | Dolby |
| BS1534-8 | Compare to EVS | 5.1+2, 5.1+4 | Generic audio | 64-256 | Headphones | Nokia |
| BS1534-9 | Compare to EVS | 7.1+4 | Generic audio | 64-256 | 7.1 + 4 | FhG |
| BS1534-10 | Compare to EVS | 1-2 Objects | Generic audio | 64-256 | Headphones | Qualcomm |
| BS1534-11 | Compare to EVS | 3-4 Objects | Generic audio | 24.4-64 | Headphones | Dolby |
| BS1534-12 | Compare to EVS | 3-4 Objects | Generic audio | 64-256 | Headphones | Philips |
| BS1534-13 | Compare to EVS | MASA 1 TC | Generic audio | 16.4-48 | Headphones | Qualcomm |
| BS1534-14 | Compare to EVS | MASA 1 TC | Generic audio | 64-256 | Headphones | Dolby |
| BS1534-15 | Compare to EVS | MASA 2 TC | Generic audio | 64-256 | Headphones | Nokia |
| BS1534-16 | Compare to EVS | OSBA 1-4 Objects | Generic audio | 16.4-48 | Headphones | FhG |
| BS1534-17 | Compare to EVS | OSBA 1-4 Objects | Generic audio | 64-256 | Headphones | Dolby |
| BS1534-18 | Compare to EVS | OMASA 1-4 Objects | Generic audio | 64-256 | Headphones | Nokia |
| BS1534-19 | Stereo downmix | Stereo | Generic audio | 13.2, 24.2 | Headphones | Orange |
| BS1534-20 | 6-DoF and directivity | 4 Objects | Generic audio | 64, 512 | Headphones | Ericsson |

Table 11: Allocation of experiments using different methodologies

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Exp** | **Objective** | **Input format** | **Source material** | **Listening environment** | **Bitrates [kb/s]** | **LL** |
| ACR-1 | 16, 32, 48 kHz | Stereo, FOA, HOA3, MASA, ISM2 | All | Headphones | 13.2-256 | Nokia |
| ROOM-1 | Room acoustics | FOA | Generic audio | Headphones | 80 | Philips |
| ROOM-2 | Room acoustics | FOA | Generic audio | Headphones | 96 | Philips |
| ROOM-3 | Room acoustics | MC – 5.1 | Generic audio | Headphones | 96 | Philips |
| ROOM-4 | Room acoustics | MC – 5.1.2 | Generic audio | Headphones | 128 | Philips |

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
* JBM (in the “Objective” column) – including Jitter Buffer Management
* Tan (in the “Objective” column) – Tandeming
* EVS (in the “Objective” column) – comparison to EVS
* P800 – P.800 DCR test
* ACR – P.800 ACR test
* ROOM – room acoustics test
* 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.

* 1. P.800 test instructions for DCR testing

|  |
| --- |
| **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. |

***P.800 DCR 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. BS.1534 test instructions [9]

|  |
| --- |
| **Familiarization or training phase**  The first step in the listening tests is to become familiar with the testing process. This phase is called a training phase, and it precedes the formal evaluation phase.  The purpose of the training phase is to allow you, as an evaluator, to achieve the following two objectives:   * **Part A**: to become familiar with all the sound excerpts under test and their quality level ranges; and * **Part B**:to learn how to use the test equipment and the grading scale.   In Part A of the training phase, you will be able to listen to all sound excerpts that have been selected for the tests in order to illustrate the whole range of possible qualities. The sound items, which you will listen to, will be more or less critical depending on the bit rate and other “conditions” used. Figure 3 shows the user interface. You may click on different buttons to listen to different sound excerpts including the reference excerpts. In this way you can learn to appreciate a range of different levels of quality for different programme items. The excerpts are grouped on the basis of common conditions. Three such groups are identified in this case. Each group includes four processed signals.  In Part B of the training phase, you will learn to use the available playback and scoring equipment that will be used to evaluate the quality of the sound excerpts.  During the training phase you should be able to learn how you, as an individual, interpret the audible impairments in terms of the grading scale. You should not discuss your personal interpretation of the scale with the other assessors at any time during the training phase. However, you are encouraged to explain artefacts to other assessors.  No grades given during the training phase will be taken into account in the true tests.  **Blind grading phase**  The purpose of the blind grading phase is to invite you to assign your grades using the quality scale. Your grades should reflect your subjective judgement of the quality level for each of the sound excerpts presented to you. Each trial will contain 9 signals to be graded. Each of the items is approximately 10 s long. You should listen to the reference, anchor, and all the test conditions by clicking on the respective buttons. You may listen to the signals in any order, any number of times.  Use the slider for each signal to indicate your opinion of its quality. When you are satisfied with your grading of all signals you should click on the “register scores” button at the bottom of the screen.  You will use the following quality scale when assigning your grades. The grading scale is continuous from “excellent” to “bad”. A grade of 0 corresponds to the bottom of the “bad” category, while a grade of 100 corresponds to the top of the “excellent” category.  In evaluating the sound excerpts, please note that you should not necessarily give a grade in the “bad” category to the sound excerpt with the lowest quality in the test. However, one or more excerpts must be given a grade of 100 because the unprocessed reference signal is included as one of the excerpts to be graded. |

* 1. Room acoustics test instructions

|  |
| --- |
| ***INSTRUCTIONS TO THE LISTENERS FOR THE IVAS ROOM ACOUSTICS SYNTHESIS EVALUATION TEST***  *In this experiment you will be evaluating systems that might be used for future immersive telecommunication services using spatial audio. Spatial audio is an audio experience where you can perceive sound as if coming from the space around you, outside and around your head. Spatial audio can be enriched by synthetic reverberation and acoustic reflections that simulate the acoustics of a room, creating a more natural and immersive experience.*  *For this test you will be wearing a VR headset and headphones. The VR scene used in the test allows for three degrees of freedom (3DoF). That means that you can change your position only by rotation, and not by moving around. Please familiarize yourself with the equipment during the training session and adjust it appropriately. It is highly recommended to perform the tests while sitting on a swivel chair. It is advised not to take this test while standing.*  *Please take time to observe the scene around you and listen to all the stimuli representing different CuT (conditions under test). Note that it will take a couple of seconds for the playback to start. The CuT provided have been processed with different reverb pathways from the IVAS decoder. A stimulus rendered without room acoustics is provided as a reference. Please rank your CuT preference regarding room acoustic synthesis quality and matching to the visual environment shown, as compared to the reference without room acoustics. Please consider factors such as timbral quality, spatial depth, and spatial envelopment and how do these relate to the image presented.* |

* 1. P.800 test instructions for ACR testing

|  |
| --- |
| INSTRUCTIONS TO NAÏVE LISTENERS FOR P.800 ACR 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 *test* sample. The *test* sample is possibly impaired after it has passed through a telecommunication system.  Your task is to evaluate the overall quality of the *test* sample, comprising both degradations in the sound quality (e.g., harshness / roughness, loss of bandwidth, unnatural speech, clicks or other distortions), and/or lack of quality in the spatial representation (e.g., unnatural sound source movement, spatial instability, lack of spatial width or other spatial problems).  You should listen carefully the sample within a trial. When sample playback has finished, select the category that best describes your overall impression about the overall quality of the sample you just heard:  5 - Excellent  4 - Good  3 - Fair  2 - Poor  1 - Bad  Note that the level of quality 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. |

***P.800 ACR Voting screen*** (after playback of sample pair)

|  |
| --- |
| Please rate the OVERALL QUALITY of the sample:  5 - Excellent  4 - Good  3 - Fair  2 - Poor  1 - Bad |

1. P.800 DCR Presentation Orders

The PC will provide the Presentation Order for each P.800 DCR 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 References. 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:

* General:
  + All audio material shall be sampled at 48 kHz with Full Band (FB) content. (IVAS-8b, Clause 4.3)
  + 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 4.3)
  + For multi-track audio, the audio tracks are ordered according to Table 5 of IVAS Processing Plan (IVAS-7b). (IVAS-8b, Clause 4.3)
  + Additionally, it should be verified that the audio material can be processed with the AFsp package tools. (IVAS-8b, Clause 4.3)
  + 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 Clause 3, IVAS-7b, Clause 3.8)
  + For Objects, metadata according to IVAS-4 shall be used. (IVAS-4, Annex C.1)
* P.800 DCR testing
  + Speech Material (input material for artificial immersive item creation):
    - Clean speech mono audio samples (IVAS-8b, Clause 4.3.1)
    - The recording SNR should be in accordance with P.800 at least 40 dB but preferably 50 dB or higher. (IVAS-8b, Clause 4.3.1)
    - 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 4.3.1)
    - The length of the sentences should typically correspond to the length of traditional Harvard sentences. (IVAS-8b, Clause 4.3.1)
    - The total length of the generated P.800 DCR 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 4.3.1.2)
    - 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 4.3.1.2)
    - 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 4.3.1.2)
    - 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 4.3.1.2)
    - 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 4.3.1.3)
    - The duration of each sample shall be approximately 8 and at least 7seconds. (IVAS-8b, Clause 4.3.1.3)
    - The following categories shall be used (IVAS-8b, Clause 4.3.1.3):
      * 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,4.3.2.2)
  + 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 4.3)
* 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 4.3)
* Conversion if the audio material can’t be processed with the AFsp package tools. (IVAS-8b, Clause 4.3)
* 20ms Block alignment if the input source material to the IVAS-7b processing stages defined in IVAS-7b 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 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#132].

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 DCR 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

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Date** | **Task** | **Active Parties** |
|  | Feb 17-21, 2025 | **3GPP SA4 #131** |  |
|  | April 7–11, 2025 | **3GPP SA4 #131-bis-e** |  |
|  | May 19–23, 2025 | **3GPP SA4 #132**  Finalization of IVAS characterization permanent documents, including:   * IVAS-7b Processing Plan for Characterization Phase * IVAS-8b Test Plan for Characterization Phase |  |
|  | June 10–13, 2025 | **3GPP SA #108** |  |
|  | July 14, 2025 | • Proper legal framework exists (signed) among proponent companies (includes host lab, cross-check lab), listening labs, GAL to cover use of executables, source codes, audio test material (unprocessed and processed), and test results (raw voting data).  • Make hosting servers for material collection available. |  |
|  |  |  |  |
|  | July 21–25, 2025 | **3GPP SA4 #133-e**  Verification of IVAS fixed-point C-code for TS 26.251 having   * Same functionalities and equivalent performance as the floating-point C-code in TS 26.258. * Full interoperability with floating-point C-code in TS 26.258. * Comparable complexity as the floating-point C-code in TS 26.258.   Agreement on TS 26.251 (IVAS fixed-point C-code) based on the verification reports.  Decision on launching characterization tests |  |
|  | Aug 1, 2025 | Deadline for uploading all required material  Note: Update triggered by verification of material is still allowed. | LLs, MC |
|  | Aug 1, 2025, 16:00 CEST | Pre-release: Submission of IVAS codec executables (floating-point + fixed-point code) for Characterization testing. | IVAS Public Collaboration |
|  | Aug 29, 2025 | MC verified that all the unprocessed material and parameters for artificially created stereo/immersive sound material meet the requirements defined by SA4, and identified missing material and parameters.  Note: Feedback on the material should be provided as early as possible, to allow for updates of the material. | MC |
|  | Sept 5, 2025 | Deadline for uploading updated material | LLs, MC |
|  | Sept 8, 2025 | Provide dry-run material to listening labs, using the pre-release executables. | HL |
|  | Sept 12, 2025,  16:00 CEST | Final release: Submission of IVAS codec executables (floating-point + fixed-point code) for Characterization testing. | IVAS Public Collaboration |
|  | Sept 12, 2025 | * MC verified that all the unprocessed updated material and parameters for artificially created stereo/immersive sound material meet the requirements defined by SA4. * MC shall choose the parameters and sound materials for all experiments. | MC |
|  | Sept 15, 2025 | * Service contract between ETSI MCC and LLs and GAL is signed. * Start of processing audio samples to be delivered to LLs. |  |
|  | Sept 16–19, 2025 | **3GPP SA #109** |  |
|  | Sept 19, 2025 | Delivery of processed audio samples to LLs | HL |
|  | Sept 22 – Oct 31, 2025 | Characterization listening tests |  |
|  | Nov 3-11, 2025 | Global analysis |  |
|  | Nov 11, 2025 | Contribution deadline for SA4 #134 |  |
|  | November 17–21, 2025 | **3GPP SA4 #134**  Characterization test results available for analysis  Agreement on characterization test results to be incorporated into TR 26.997 |  |
|  | December 9-12, 2025 | **3GPP SA #110**  CR to TR 26.997 on IVAS Codec characterization, for approval |  |
|  |  |  |  |

1. P.800 DCR 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 Music and Mixed content 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 | -16, -26, -36 LKFS, as defined in Table 5 |
| Input frequency mask | 20KBP |
| Noise | Idle noise for cat 1,2, 15 dB 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 [13] 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 pairs uttered by different talkers and genders (3 male and 3 female), music and mixed content |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |
|  |  |

Table F.1.2: Preliminaries for Experiment P800-1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 13.2 | No errors |
| 2 | c26 | Cat 2 | IVAS FL | 96 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.7 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 16.4 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.1 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 32 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 27 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 128 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.5 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 12 dB | - | - |
| 12 | c24 | Cat 6 | IVAS FL | 64 | No errors |

Table F.1.3: Test conditions for Experiment P800-1,  
stereo speech and music and mixed content 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 | off |
| c11 | IVAS FX enc / FL dec | 16.4 | off |
| c12 | IVAS FL enc / FX dec | 24.4 | off |
| c13 | IVAS FX enc / FL dec | 32.0 | off |
| c14 | IVAS FL enc / FX dec | 48.0 | off |
| c15 | IVAS FX enc / FL dec | 64.0 | off |
| c16 | IVAS FL enc / FX dec | 80.0 | off |
| c17 | IVAS FX enc / FL dec | 96.0 | off |
| c18 | IVAS FL enc / FX dec | 128.0 | off |
| c19 | IVAS FL | 13.2 | off |
| c20 | IVAS FL | 16.4 | off |
| c21 | IVAS FL | 24.4 | off |
| c22 | IVAS FL | 32.0 | off |
| c23 | IVAS FL | 48.0 | off |
| c24 | IVAS FL | 64.0 | off |
| c25 | IVAS FL | 80.0 | off |
| c26 | IVAS FL | 96.0 | off |
| c27 | IVAS FL | 128.0 | off |
| c28 | IVAS FX | 13.2 | off |
| c29 | IVAS FX | 16.4 | off |
| c30 | IVAS FX | 24.4 | off |
| c31 | IVAS FX | 32.0 | off |
| c32 | IVAS FX | 48.0 | off |
| c33 | IVAS FX | 64.0 | off |
| c34 | IVAS FX | 80.0 | off |
| c35 | IVAS FX | 96.0 | off |
| c36 | IVAS FX | 128.0 | off |

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

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Room*** | ***Reverb*** | ***Microphone Setup*** | ***Background*** | ***SNR[dB]*** | ***Overtalk [s](1*** | ***Bandwidth*** | ***Talker positions by panel(2*** | ***Talker selection by panel*** |
| cat 1 | small | anechoic | M-S | Low level idle noise | 45 | 1 | Max available up to SWB | P1: 1-7  P2: 5-3  P3: 2-6  P4: 4-1  P5: 3-4  P6: 7-2 | P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3 |
| cat 2 | large | echoic | A-B (150 cm) | Low level idle noise | 45 | -1 | max available up to SWB | P1: 5-11  P2: 1-6  P3: 3-7  P4: 5-8  P5: 9-7  P6: 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 | P1: 1-7  P2: 5-3  P3: 2-6  P4: 4-1  P5: 3-4  P6: 7-2 | P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1 |
| cat 4 | car | car | A-B Cardioid pair 20 cm | car | 15 | -1 | Max available up to SWB | P1: Driver-Passenger  P2: BackRight-Driver  P3: Driver-BackCenter  P4: BackLeft-Driver  P5:BackRight-BackLeft  P6: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 |

**Notes:**

**(1** Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.

**(2** The talker positions are part of the scene definition of the different categories. They correspond to the talker positions as depicted in Figures 14.3 and 14.5 of [13] for the large and the small room, respectively.

* 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 Music and Mixed content categories, respectively.

Table F.2.1: Conditions for Experiment P800-2

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64, 80, 96, 128 kbps |
| DTX | DTX on |
| Input level | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Noise | Idle noise for cat 1,2, 15 dB 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 [13] 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 pairs uttered by different talkers and genders (3 male and 3 female), music and mixed content |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |
|  |  |

Table F.2.2 : Preliminaries for Experiment P800-2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** | **DTX** |
| 1 | c19 | Cat 1 | IVAS FL | 13.2 | 5% | on |
| 2 | c26 | Cat 2 | IVAS FL | 96 | 5% | on |
| 3 | c06 | Cat 3 | ESDRU = 0.7 | - | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 16.4 | 5% | on |
| 5 | c09 | Cat 5 | ESDRU = 0.1 | - | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 32 | 5% | on |
| 7 | c02 | Cat 1 | MNRU Q = 27 dB | - | - | - |
| 8 | c01 | Cat 2 | Reference | - | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 128 | 5% | on |
| 10 | c07 | Cat 4 | ESDRU = 0.5 | - | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 12 dB | - | - | - |
| 12 | c24 | Cat 6 | IVAS FL | 64 | 5% | on |

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **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% |

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

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Room*** | ***Reverb*** | ***Microphone Setup*** | ***Background*** | ***SNR***  ***[dB]*** | ***Overtalk [s](1*** | ***Bandwidth*** | ***Talker positions by panel(2*** | ***Talker selection by panel*** |
| cat 1 | small | anechoic | M-S | Low level idle noise | 45 | 1 | Max available up to SWB | P1: 1-7  P2: 5-3  P3: 2-6  P4: 4-1  P5: 3-4  P6: 7-2 | P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3 |
| cat 2 | large | echoic | A-B (150 cm) | Low level idle noise | 45 | -1 | max available up to SWB | P1: 5-11  P2: 1-6  P3: 3-7  P4: 5-8  P5: 9-7  P6: 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 | P1: 1-7  P2: 5-3  P3: 2-6  P4: 4-1  P5: 3-4  P6: 7-2 | P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1 |
| cat 4 | car | car | A-B Cardioid pair 20 cm | car | 15 | -1 | Max available up to SWB | P1: Driver-Passenger  P2: BackRight-Driver  P3:Driver-BackCenter  P4: BackLeft-Driver  P5:BackRight-BackLeft  P6: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 |

**Notes:**

**(1** Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.

**(2** The talker positions are part of the scene definition of the different categories. They correspond to the talker positions as depicted in Figures 14.3 and 14.5 of [13] for the large and the small room, respectively.

* 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 | 13.2, 16.4, 24.4, 32, 48, 64, 96, 128, 256 kbps |
| DTX | DTX off |
| Input level | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | Idle noise for cat 1,2, 15 dB for cat 3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| 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 pairs uttered by different talkers and genders (3 male and 3 female), mixed content, generic audio |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.3.2: Preliminaries for Experiment P800-3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c20 | Cat 1 | IVAS FL | 16.4 | No errors |
| 2 | c25 | Cat 2 | IVAS FL | 96 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c21 | Cat 4 | IVAS FL | 24.4 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 32 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c26 | Cat 3 | IVAS FL | 128 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - |
| 12 | c24 | Cat 6 | 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 = 32 dB | - | - |
| c03 | MNRU Q = 27 dB | - | - |
| c04 | MNRU Q = 22 dB | - | - |
| c05 | MNRU Q = 17 dB | - | - |
| c06 | ESDRU *α* = 0.8 | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - |
| c08 | ESDRU *α* = 0.4 | - | - |
| c09 | ESDRU *α* = 0.2 | - | - |
| c10 | IVAS FL enc / FX dec | 13.2 | off |
| c11 | IVAS FL enc / FX dec | 16.4 | off |
| c12 | IVAS FX enc / FL dec | 24.4 | off |
| c13 | IVAS FL enc / FX dec | 32.0 | off |
| c14 | IVAS FX enc / FL dec | 48.0 | off |
| c15 | IVAS FL enc / FX dec | 64.0 | off |
| c16 | IVAS FL enc / FX dec | 96.0 | off |
| c17 | IVAS FX enc / FL dec | 128.0 | off |
| c18 | IVAS FL enc / FX dec | 256.0 | off |
| c19 | IVAS FL | 13.2 | off |
| c20 | IVAS FL | 16.4 | off |
| c21 | IVAS FL | 24.4 | off |
| c22 | IVAS FL | 32.0 | off |
| c23 | IVAS FL | 48.0 | off |
| c24 | IVAS FL | 64.0 | off |
| c25 | IVAS FL | 96.0 | off |
| c26 | IVAS FL | 128.0 | off |
| c27 | IVAS FL | 256.0 | off |
| c28 | IVAS FX | 13.2 | off |
| c29 | IVAS FX | 16.4 | off |
| c30 | IVAS FX | 24.4 | off |
| c31 | IVAS FX | 32.0 | off |
| c32 | IVAS FX | 48.0 | off |
| c33 | IVAS FX | 64.0 | off |
| c34 | IVAS FX | 96.0 | off |
| c35 | IVAS FX | 128.0 | off |
| c36 | IVAS FX | 256.0 | off |

Table F.3.4: Clean and noisy speech categories and scene definitions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Environment(1*** | ***Background(2*** | ***SNR [dB]*** | ***Overtalk [s](3*** | ***Bandwidth*** | ***Talker position*** | ***Talker selection by panel*** |
| *cat 1* | *env\_1\_FOA* | *env\_1\_cleanbg\_FOA* | *45* | *1* | *Max* | ***(4*** | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *env\_2\_FOA* | *env\_2\_cleanbg\_FOA* | *45* | *-1* | *Max* | ***(4*** | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *env\_3\_FOA* | *env\_3\_noisebg\_1\_FOA* | *15* | *1* | *Max* | ***(4*** | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *env\_4\_FOA* | *env\_4\_noisebg\_1\_FOA* | *15* | *-1* | *Max* | ***(4*** | *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** The specific room/environment characteristic and resulting reverb characteristic are defined by the choice of the specific spatial room impulse responses according to the configuration settings of the characterization processing scripts. “env\_1“, “env\_2“ represent conference rooms while “env\_4“ is another indoor environment such as cafeteria or mall. “env\_3” is an outdoor environment like park, nature, event or street.

**(2** Background is defined by the chosen background noise file according to the configuration settings of the characterization processing scripts. “env\_1\_cleanbg” and “env\_2\_cleanbg” indicate low-noise background noises that can be expected in conference rooms such as low air-conditioning/fan noise. “env\_3\_noisebg” and “env\_4\_noisebg” indicate background noises typical for their respective actual environment.

**(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** The talker positions are part of the scene definition of the different categories. They are defined in the Characterization Processing Plan and the corresponding processing scripts. They are 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 are 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 | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | Idle noise for cat 1,2, 15 dB for cat 3,4 |
| Error Conditions | 5% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| 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 pairs uttered by different talkers and genders (3 male and 3 female), mixed content, generic audio |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.4.2: Preliminaries for Experiment P800-4

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** | **DTX** |
| 1 | c19 | Cat 1 | IVAS FL | 16.4 | 5% | on |
| 2 | c26 | Cat 2 | IVAS FL | 256 | 5% | on |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 24.4 | 5% | on |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |  |
| 6 | c22 | Cat 6 | IVAS FL | 48 | 5% | on |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - | - |
| 8 | c01 | Cat 2 | Reference | - | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 384 | 5% | on |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - | - |
| 12 | c24 | Cat 6 | IVAS FL | 96 | 5% | on |

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **FER** |
| c01 | Reference | - | - | - |
| c02 | MNRU Q = 32 dB | - | - | - |
| c03 | MNRU Q = 27 dB | - | - | - |
| c04 | MNRU Q = 22 dB | - | - | - |
| c05 | MNRU Q = 17 dB | - | - | - |
| c06 | ESDRU *α* = 0.8 | - | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - | - |
| c08 | ESDRU *α* = 0.4 | - | - | - |
| c09 | ESDRU *α* = 0.2 | - | - | - |
| 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% |

Table F.4.4: Clean and noisy speech categories and scene definitions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Environment(1*** | ***Background(2*** | ***SNR [dB]*** | ***Overtalk [s](3*** | ***Bandwidth*** | ***Talker positions*** | ***Talker selection by panel*** |
| *cat 1* | *env\_1\_HOA2* | *env\_1\_cleanbg\_HOA2* | *45* | *1* | *Max* | ***(4*** | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *env\_2\_HOA2* | *env\_2\_cleanbg\_HOA2* | *45* | *-1* | *Max* | ***(4*** | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *env\_3\_HOA2* | *env\_3\_noisebg\_1\_HOA2* | *15* | *1* | *Max* | ***(4*** | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *env\_4\_HOA2* | *env\_4\_noisebg\_1\_HOA2* | *15* | *-1* | *Max* | ***(4*** | *P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3* |

Table F.4.5: Mixed content and Generic audio categories

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

**Notes:**

**(1** The specific room/environment characteristic and resulting reverb characteristic are defined by the choice of the specific spatial room impulse responses according to the configuration settings of the characterization processing scripts. “env\_1“, “env\_2“ represent conference rooms while “env\_4“ is another indoor environment such as cafeteria or mall. “env\_3” is an outdoor environment like park, nature, event or street.

**(2** Background is defined by the chosen background noise file according to the configuration settings of the characterization processing scripts. “env\_1\_cleanbg” and “env\_2\_cleanbg” indicate low-noise background noises that can be expected in conference rooms such as low air-conditioning/fan noise. “env\_3\_noisebg” and “env\_4\_noisebg” indicate background noises typical for their respective actual environment.

**(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** The talker positions are part of the scene definition of the different categories. They are defined in the Characterization Processing Plan and the corresponding processing scripts. They are 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 are 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 | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | Idle noise for cat 1,2, 15 dB for cat 3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| 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 pairs uttered by different talkers and genders (3 male and 3 female), mixed content, generic audio |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.5.2: Preliminaries for Experiment P800-5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 32 | No errors |
| 2 | c26 | Cat 2 | IVAS FL | 384 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 48 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 96 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 512 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - |
| 12 | c24 | Cat 6 | 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 = 32 dB | - | - |
| c03 | MNRU Q = 27 dB | - | - |
| c04 | MNRU Q = 22 dB | - | - |
| c05 | MNRU Q = 17 dB | - | - |
| c06 | ESDRU *α* = 0.8 | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - |
| c08 | ESDRU *α* = 0.4 | - | - |
| c09 | ESDRU *α* = 0.2 | - | - |
| c10 | IVAS FL enc / FX dec | 32.0 | off |
| c11 | IVAS FX enc / FL dec | 48.0 | off |
| c12 | IVAS FL enc / FX dec | 64.0 | off |
| c13 | IVAS FX enc / FL dec | 96.0 | off |
| c14 | IVAS FL enc / FX dec | 128.0 | off |
| c15 | IVAS FX enc / FL dec | 160.0 | off |
| c16 | IVAS FL enc / FX dec | 256.0 | off |
| c17 | IVAS FX enc / FL dec | 384.0 | off |
| c18 | IVAS FL enc / FX dec | 512.0 | off |
| c19 | IVAS FL | 32.0 | off |
| c20 | IVAS FL | 48.0 | off |
| c21 | IVAS FL | 64.0 | off |
| c22 | IVAS FL | 96.0 | off |
| c23 | IVAS FL | 128.0 | off |
| c24 | IVAS FL | 160.0 | off |
| c25 | IVAS FL | 256.0 | off |
| c26 | IVAS FL | 384.0 | off |
| c27 | IVAS FL | 512.0 | off |
| c28 | IVAS FX | 32.0 | off |
| c29 | IVAS FX | 48.0 | off |
| c30 | IVAS FX | 64.0 | off |
| c31 | IVAS FX | 96.0 | off |
| c32 | IVAS FX | 128.0 | off |
| c33 | IVAS FX | 160.0 | off |
| c34 | IVAS FX | 256.0 | off |
| c35 | IVAS FX | 384.0 | off |
| c36 | IVAS FX | 512.0 | off |

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*** | ***Talker selection by panel*** |
| *cat 1* | *env\_1\_HOA3* | *env\_1\_cleanbg\_HOA3* | *45* | *1* | *Max* | ***(4*** | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *env\_2\_HOA3* | *env\_2\_cleanbg\_HOA3* | *45* | *-1* | *Max* | ***(4*** | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *env\_3\_HOA3* | *env\_3\_noisebg\_1\_HOA3* | *15* | *1* | *Max* | ***(4*** | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *env\_4\_HOA3* | *env\_4\_noisebg\_1\_HOA3* | *15* | *-1* | *Max* | ***(4*** | *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** The specific room/environment characteristic and resulting reverb characteristic are defined by the choice of the specific spatial room impulse responses according to the configuration settings of the characterization processing scripts. “env\_1“, “env\_2“ represent conference rooms while “env\_4“ is another indoor environment such as cafeteria or mall. “env\_3” is an outdoor environment like park, nature, event or street.

**(2** Background is defined by the chosen background noise file according to the configuration settings of the characterization processing scripts. “env\_1\_cleanbg” and “env\_2\_cleanbg” indicate low-noise background noises that can be expected in conference rooms such as low air-conditioning/fan noise. “env\_3\_noisebg” and “env\_4\_noisebg” indicate background noises typical for their respective actual environment.

**(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** The talker positions are part of the scene definition of the different categories. They are defined in the Characterization Processing Plan and the corresponding processing scripts. They are 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 are made for the different listener panels.

* 1. Experiment P800-6: MC 5.1, 7.1

Tables F.6.1 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 Music and Mixed content 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 | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP excluding the LFE channel when present |
| Background | Idle noise for cat 1,2, 15 dB for cat 3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| 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 | Sentence pairs uttered by different talkers and genders (3 male and 3 female), music and mixed content |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.6.2: Preliminaries for Experiment P800-6

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 24.4 | No errors |
| 2 | c26 | Cat 2 | IVAS FL | 96 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 32 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 48 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 128 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - |
| 12 | c24 | Cat 6 | 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 = 32 dB | - | - |
| c03 | MNRU Q = 27 dB | - | - |
| c04 | MNRU Q = 22 dB | - | - |
| c05 | MNRU Q = 17 dB | - | - |
| c06 | ESDRU *α* = 0.8 | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - |
| c08 | ESDRU *α* = 0.4 | - | - |
| c09 | ESDRU *α* = 0.2 | - | - |
| c10 | IVAS FL enc / FX dec | 24.4 | off |
| c11 | IVAS FX enc / FL dec | 32.0 | off |
| c12 | IVAS FL enc / FX dec | 48.0 | off |
| c13 | IVAS FX enc / FL dec | 64.0 | off |
| c14 | IVAS FL enc / FX dec | 80.0.0 | off |
| c15 | IVAS FX enc / FL dec | 96.0 | Off |
| c16 | IVAS FL enc / FX dec | 128.0 | off |
| c17 | IVAS FX enc / FL dec | 160.0 | Off |
| c18 | IVAS FL enc / FX dec | 256.0 | Off |
| c19 | IVAS FL | 24.4 | Off |
| c20 | IVAS FL | 32.0 | Off |
| c21 | IVAS FL | 48.0 | Off |
| c22 | IVAS FL | 64.0 | Off |
| c23 | IVAS FL | 80.0.0 | Off |
| c24 | IVAS FL | 96.0 | Off |
| c25 | IVAS FL | 128.0 | Off |
| c26 | IVAS FL | 160.0 | Off |
| c27 | IVAS FL | 256.0 | Off |
| c28 | IVAS FX | 24.4 | Off |
| c29 | IVAS FX | 32.0 | Off |
| c30 | IVAS FX | 48.0 | Off |
| c31 | IVAS FX | 64.0 | Off |
| c32 | IVAS FX | 80.0.0 | Off |
| c33 | IVAS FX | 96.0 | Off |
| c34 | IVAS FX | 128.0 | Off |
| c35 | IVAS FX | 160.0 | Off |
| c36 | IVAS FX | 256.0 | Off |

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*** | ***Talker selection by panel*** |
| *cat 1*  *5.1* | *env\_1\_FOA* | *env\_1\_cleanbg\_FOA* | *45* | *1* | *Max* | ***(4*** | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2*  *7.1* | *env\_2\_FOA* | *env\_2\_cleanbg\_FOA* | *45* | *-1* | *Max* | ***(4*** | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3*  *5.1* | *env\_3\_FOA* | *env\_3\_noisebg\_1\_FOA* | *15* | *1* | *Max* | ***(4*** | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4*  *7.1* | *env\_4\_FOA* | *env\_4\_noisebg\_1\_FOA* | *15* | *-1* | *Max* | ***(4*** | *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** The specific room/environment characteristic and resulting reverb characteristic are defined by the choice of the specific spatial room impulse responses according to the configuration settings of the characterization processing scripts. “env\_1“, “env\_2“ represent conference rooms while “env\_4“ is another indoor environment such as cafeteria or mall. “env\_3” is an outdoor environment like park, nature, event or street.

**(2** Background is defined by the chosen background noise file according to the configuration settings of the characterization processing scripts. “env\_1\_cleanbg” and “env\_2\_cleanbg” indicate low-noise background noises that can be expected in conference rooms such as low air-conditioning/fan noise. “env\_3\_noisebg” and “env\_4\_noisebg” indicate background noises typical for their respective actual environment.

**(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** The talker positions are part of the scene definition of the different categories. They are defined in the Characterization Processing Plan and the corresponding processing scripts. They are 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 are 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 Music and Mixed content 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 | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP excluding the LFE channel when present |
| Background | Idle noise for cat 1,2, 15 dB for cat 3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| 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 | Sentence pairs uttered by different talkers and genders (3 male and 3 female), music and mixed content |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.7.2: Preliminaries for Experiment P800-7

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 32 | No errors |
| 2 | c26 | Cat 2 | IVAS FL | 96 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 48 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 32 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 128 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - |
| 12 | c24 | Cat 6 | 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 = 32 dB | - | - |
| c03 | MNRU Q = 27 dB | - | - |
| c04 | MNRU Q = 22 dB | - | - |
| c05 | MNRU Q = 17 dB | - | - |
| c06 | ESDRU *α* = 0.8 | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - |
| c08 | ESDRU *α* = 0.4 | - | - |
| c09 | ESDRU *α* = 0.2 | - | - |
| c10 | IVAS FL enc / FX dec | 32 | off |
| c11 | IVAS FX enc / FL dec | 48 | off |
| c12 | IVAS FL enc / FX dec | 64 | off |
| c13 | IVAS FX enc / FL dec | 96 | off |
| c14 | IVAS FL enc / FX dec | 128.0 | off |
| c15 | IVAS FX enc / FL dec | 160.0 | off |
| c16 | IVAS FL enc / FX dec | 192.0 | off |
| c17 | IVAS FX enc / FL dec | 256.0.0 | off |
| c18 | IVAS FL enc / FX dec | 384.0 | off |
| c19 | IVAS FL | 32 | off |
| c20 | IVAS FL | 48 | off |
| c21 | IVAS FL | 64 | off |
| c22 | IVAS FL | 96 | off |
| c23 | IVAS FL | 128.0 | off |
| c24 | IVAS FL | 160.0 | off |
| c25 | IVAS FL | 192.0 | off |
| c26 | IVAS FL | 256.0.0 | off |
| c27 | IVAS FL | 384.0 | off |
| c28 | IVAS FX | 32 | off |
| c29 | IVAS FX | 48 | off |
| c30 | IVAS FX | 64 | off |
| c31 | IVAS FX | 96 | off |
| c32 | IVAS FX | 128.0 | off |
| c33 | IVAS FX | 160.0 | off |
| c34 | IVAS FX | 192.0 | off |
| c35 | IVAS FX | 256.0.0 | off |
| c36 | IVAS FX | 384.0 | off |

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*** | ***Talker selection by panel*** |
| *cat 1*  *5.1+4* | *env\_1\_FOA* | *env\_1\_cleanbg\_FOA* | *45* | *1* | *Max* | ***(4*** | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2*  *7.1+4* | *env\_2\_FOA* | *env\_2\_cleanbg\_FOA* | *45* | *-1* | *Max* | ***(4*** | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3*  *5.1+4* | *env\_3\_FOA* | *env\_3\_noisebg\_1\_FOA* | *15* | *1* | *Max* | ***(4*** | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4*  *7.1+4* | *env\_4\_FOA* | *env\_4\_noisebg\_1\_FOA* | *15* | *-1* | *Max* | ***(4*** | *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** The specific room/environment characteristic and resulting reverb characteristic are defined by the choice of the specific spatial room impulse responses according to the configuration settings of the characterization processing scripts. “env\_1“, “env\_2“ represent conference rooms while “env\_4“ is another indoor environment such as cafeteria or mall. “env\_3” is an outdoor environment like park, nature, event or street.

**(2** Background is defined by the chosen background noise file according to the configuration settings of the characterization processing scripts. “env\_1\_cleanbg” and “env\_2\_cleanbg” indicate low-noise background noises that can be expected in conference rooms such as low air-conditioning/fan noise. “env\_3\_noisebg” and “env\_4\_noisebg” indicate background noises typical for their respective actual environment.

**(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** The talker positions are part of the scene definition of the different categories. They are defined in the Characterization Processing Plan and the corresponding processing scripts. They are 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 are 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 Music and Mixed content 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 | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP excluding the LFE channel when present |
| Background | Idle noise for cat 1,2, 15 dB for cat 3,4 |
| Error Conditions | 5% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| 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 | Sentence pairs uttered by different talkers and genders (3 male and 3 female), music and mixed content |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.8.2: Preliminaries for Experiment P800-8

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 16.4 | 5% |
| 2 | c26 | Cat 2 | IVAS FL | 256 | 5% |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 24.4 | 5% |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 64 | 5% |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 384 | 5% |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - |
| 12 | c24 | Cat 6 | IVAS FL | 128 | 5% |

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **FER** |
| c01 | Reference | - | - |
| c02 | MNRU Q = 32 dB | - | - |
| c03 | MNRU Q = 27 dB | - | - |
| c04 | MNRU Q = 22 dB | - | - |
| c05 | MNRU Q = 17 dB | - | - |
| c06 | ESDRU *α* = 0.8 | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - |
| c08 | ESDRU *α* = 0.4 | - | - |
| c09 | ESDRU *α* = 0.2 | - | - |
| 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% |

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*** | ***Talker selection by panel*** |
| *cat 1*  *5.1* | *env\_1\_FOA* | *env\_1\_cleanbg\_FOA* | *45* | *1* | *Max* | ***(4*** | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2*  *7.1* | *env\_2\_FOA* | *env\_2\_cleanbg\_FOA* | *45* | *-1* | *Max* | ***(4*** | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3*  *5.1+4* | *env\_3\_FOA* | *env\_3\_noisebg\_1\_FOA* | *15* | *1* | *Max* | ***(4*** | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4*  *7.1+4* | *env\_4\_FOA* | *env\_4\_noisebg\_1\_FOA* | *15* | *-1* | *Max* | ***(4*** | *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** The specific room/environment characteristic and resulting reverb characteristic are defined by the choice of the specific spatial room impulse responses according to the configuration settings of the characterization processing scripts. “env\_1“, “env\_2“ represent conference rooms while “env\_4“ is another indoor environment such as cafeteria or mall. “env\_3” is an outdoor environment like park, nature, event or street.

**(2** Background is defined by the chosen background noise file according to the configuration settings of the characterization processing scripts. “env\_1\_cleanbg” and “env\_2\_cleanbg” indicate low-noise background noises that can be expected in conference rooms such as low air-conditioning/fan noise. “env\_3\_noisebg” and “env\_4\_noisebg” indicate background noises typical for their respective actual environment.

**(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** The talker positions are part of the scene definition of the different categories. They are defined in the Characterization Processing Plan and the corresponding processing scripts. They are 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 are made for the different listener panels.

* 1. Experiment P800-9: 1-2 Objects

Tables F.9.1 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 bacground and Music and mixed content 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 | -16, -26, -36 LKFS, as defined in Table 3 |
| 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. 1-2: Defined scenes, 1 ISM  Cat. 3-4: Defined scenes, 2 ISMs Cat. 5, 6: Pre-produced content |
| Binaural renderer | ISM to binaural internal rendering |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pairs uttered by different talkers and genders (3 male and 3 female), music and mixed content, speech and background |
| Number of categories | 6 |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.9.2: Preliminaries for Experiment P800-9

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 13.2 | No errors |
| 2 | c26 | Cat 2 | IVAS FL | 96 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.7 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 16.4 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.1 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 32 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 36 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 128 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.5 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 18 dB | - | - |
| 12 | c24 | Cat 6 | 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** |
|  |  | **Cat 1, 2, 5** | **Cat 3, 4, 6** |  |
| 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 | 160.0 | off |
| c11 | IVAS FX enc / FL dec | 16.4 | 16.4 | off |
| c12 | IVAS FL enc / FX dec | 24.4 | 24.4 | off |
| c13 | IVAS FX enc / FL dec | 32.0 | 32.0 | off |
| c14 | IVAS FL enc / FX dec | 48.0 | 48.0 | off |
| c15 | IVAS FX enc / FL dec | 64.0 | 64.0 | off |
| c16 | IVAS FL enc / FX dec | 80.0 | 80.0 | off |
| c17 | IVAS FX enc / FL dec | 96.0 | 96.0 | off |
| c18 | IVAS FL enc / FX dec | 128.0 | 128.0 | off |
| c19 | IVAS FL | 13.2 | 160.0 | off |
| c20 | IVAS FL | 16.4 | 16.4 | off |
| c21 | IVAS FL | 24.4 | 24.4 | off |
| c22 | IVAS FL | 32.0 | 32.0 | off |
| c23 | IVAS FL | 48.0 | 48.0 | off |
| c24 | IVAS FL | 64.0 | 64.0 | off |
| c25 | IVAS FL | 80.0 | 80.0 | off |
| c26 | IVAS FL | 96.0 | 96.0 | off |
| c27 | IVAS FL | 128.0 | 128.0 | off |
| c28 | IVAS FX | 13.2 | 160.0 | off |
| c29 | IVAS FX | 16.4 | 16.4 | off |
| c30 | IVAS FX | 24.4 | 24.4 | off |
| c31 | IVAS FX | 32.0 | 32.0 | off |
| c32 | IVAS FX | 48.0 | 48.0 | off |
| c33 | IVAS FX | 64.0 | 64.0 | off |
| c34 | IVAS FX | 80.0 | 80.0 | off |
| c35 | IVAS FX | 96.0 | 96.0 | off |
| c36 | IVAS FX | 128.0 | 128.0 | off |

**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:

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

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

**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. Non-overlapping utterances.
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. Overlapping utterances.
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. Overlapping utterances.
5. Two talkers walking side-by-side around the table (elevation 45°). The azimuth is the same for both talkers and varies continually. Non-overlapping utterances.
6. Two talkers walking around the table in opposite directions (elevation 30°), starting at the same position. Azimuths of both talkers vary continually. Overlapping utterances.

The following table lists the test Categories corresponding to different talkers or talker pairs. Each of the sentence pairs uttered by a certain talker or talker pair is associated to a different scene.

Table F.9.4: Allocation of scenes for each talker or talker pair (category cat 1, cat 2, cat 3, cat 4) and listening panel (P1-P6)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Scene*** | ***Talker initial elevation*** | ***Elevation change*** | ***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 3:***  *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 4:***  *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: Music and mixed content and Speech and background categories

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 | Music and mixed content (1 object) |
| cat 6 | 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, Speech with effects, Speech with music, and Music categories, respectively.

Table F.10.1: Conditions for Experiment P800-10

|  |  |
| --- | --- |
| **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 | -16, -26, -36 LKFS, as defined in Table 3 |
| 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 | Pre-produced speech and audio content |
| Number of categories | 6 |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.10.2: Preliminaries for Experiment P800-10

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 13.2 | No errors |
| 2 | c26 | Cat 2 | IVAS FL | 96 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.7 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 16.4 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.1 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 32 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 36 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 128 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.5 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 18 dB | - | - |
| 12 | c24 | Cat 6 | 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 = 18 dB | - | - |
| c06 | ESDRU | - | - |
| c07 | ESDRU | - | - |
| c08 | ESDRU | - | - |
| c09 | ESDRU | - | - |
| c10 | IVAS FL enc / FX dec | 13.2 | off |
| c11 | IVAS FX enc / FL dec | 16.4 | off |
| c12 | IVAS FL enc / FX dec | 24.4 | off |
| c13 | IVAS FX enc / FL dec | 32.0 | off |
| c14 | IVAS FL enc / FX dec | 48.0 | off |
| c15 | IVAS FX enc / FL dec | 64.0 | off |
| c16 | IVAS FL enc / FX dec | 80.0 | off |
| c17 | IVAS FX enc / FL dec | 96.0 | off |
| c18 | IVAS FL enc / FX dec | 128.0 | off |
| c19 | IVAS FL | 13.2 | off |
| c20 | IVAS FL | 16.4 | off |
| c21 | IVAS FL | 24.4 | off |
| c22 | IVAS FL | 32.0 | off |
| c23 | IVAS FL | 48.0 | off |
| c24 | IVAS FL | 64.0 | off |
| c25 | IVAS FL | 80.0 | off |
| c26 | IVAS FL | 96.0 | off |
| c27 | IVAS FL | 128.0 | off |
| c28 | IVAS FX | 13.2 | off |
| c29 | IVAS FX | 16.4 | off |
| c30 | IVAS FX | 24.4 | off |
| c31 | IVAS FX | 32.0 | off |
| c32 | IVAS FX | 48.0 | off |
| c33 | IVAS FX | 64.0 | off |
| c34 | IVAS FX | 80.0 | off |
| c35 | IVAS FX | 96.0 | off |
| c36 | IVAS FX | 128.0 | off |

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 Speech with music or Music categories, respectively.

Table F.11.1: Conditions for Experiment P800-11

|  |  |
| --- | --- |
| **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 | -16, -26, -36 LKFS, as defined in Table 3 |
| 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 pairs uttered by different talkers and genders (3 male and 3 female), speech and effects, speech and music or music only |
| Number of categories | 6 |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.11.2: Preliminaries for Experiment P800-11

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** | **DTX** |
| 1 | c19 | Cat 1 | IVAS FL | 13.2 | 5% | on |
| 2 | c26 | Cat 2 | IVAS FL | 96 | 5% | on |
| 3 | c06 | Cat 3 | ESDRU = 0.7 | - | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 16.4 | 5% | on |
| 5 | c09 | Cat 5 | ESDRU = 0.1 | - | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 32 | 5% | on |
| 7 | c02 | Cat 1 | MNRU Q = 36 dB | - | - | - |
| 8 | c01 | Cat 2 | Reference | - | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 128 | 5% | on |
| 10 | c07 | Cat 4 | ESDRU = 0.5 | - | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 18 dB | - | - | - |
| 12 | c24 | Cat 6 | IVAS FL | 64 | 5% | on |

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **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% |

**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. Non-overlapping utterances.
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. Overlapping utterances.
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. Overlapping utterances.
5. Two talkers walking side-by-side around the table (elevation 45°). The azimuth is the same for both talkers and varies continually. Non-overlapping utterances.
6. Two talkers walking around the table in opposite directions (elevation 30°), starting at the same position. Azimuths of both talkers vary continually. Overlapping utterances.

The following table lists the test Categories corresponding to different talkers or talker pairs. Each of the sentence pairs uttered by a certain talker or talker pair is associated to a different scene.

Table F.11.4: Allocation of scenes for each talker or talker pair (category cat 1, cat 2, cat 3, cat 4) and listening panel (P1-P6)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Scene*** | ***Talker initial elevation*** | ***Elevation change*** | ***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 3:***  *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 4:***  *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 P800-12

|  |  |
| --- | --- |
| **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 | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | Idle noise for cat 1,2, 15 dB for cat 3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| 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 [14]. |
| Binaural renderer | IVAS MASA internal binaural rendering |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pairs uttered by different talkers and genders (3 male and 3 female), mixed content and generic audio |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.12.2: Preliminaries for Experiment P800-12

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 13.2 | No errors |
| 2 | c26 | Cat 2 | IVAS FL | 96 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 16.4 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 32 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 128 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - |
| 12 | c24 | Cat 6 | IVAS FL | 64 | No errors |

Table F.12.3: Test conditions for Experiment P800-12,  
clean speech under clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** |
| c01 | Reference | - | - |
| c02 | MNRU Q = 32 dB | - | - |
| c03 | MNRU Q = 27 dB | - | - |
| c04 | MNRU Q = 22 dB | - | - |
| c05 | MNRU Q = 17 dB | - | - |
| c06 | ESDRU *α* = 0.8 | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - |
| c08 | ESDRU *α* = 0.4 | - | - |
| c09 | ESDRU *α* = 0.2 | - | - |
| c10 | IVAS FL enc / FX dec | 13.2 | off |
| c11 | IVAS FX enc / FL dec | 16.4 | off |
| c12 | IVAS FL enc / FX dec | 24.4 | off |
| c13 | IVAS FX enc / FL dec | 32.0 | off |
| c14 | IVAS FL enc / FX dec | 48.0 | off |
| c15 | IVAS FX enc / FL dec | 64.0 | off |
| c16 | IVAS FL enc / FX dec | 80.0 | off |
| c17 | IVAS FX enc / FL dec | 96.0 | off |
| c18 | IVAS FL enc / FX dec | 128.0 | off |
| c19 | IVAS FL | 13.2 | off |
| c20 | IVAS FL | 16.4 | off |
| c21 | IVAS FL | 24.4 | off |
| c22 | IVAS FL | 32.0 | off |
| c23 | IVAS FL | 48.0 | off |
| c24 | IVAS FL | 64.0 | off |
| c25 | IVAS FL | 80.0 | off |
| c26 | IVAS FL | 96.0 | off |
| c27 | IVAS FL | 128.0 | off |
| c28 | IVAS FX | 13.2 | off |
| c29 | IVAS FX | 16.4 | off |
| c30 | IVAS FX | 24.4 | off |
| c31 | IVAS FX | 32.0 | off |
| c32 | IVAS FX | 48.0 | off |
| c33 | IVAS FX | 64.0 | off |
| c34 | IVAS FX | 80.0 | off |
| c35 | IVAS FX | 96.0 | off |
| c36 | IVAS FX | 128.0 | off |

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*** | ***Talker selection by panel*** |
| *cat 1* | *env\_1\_MASA* | *env\_1\_cleanbg\_MASA* | *45* | *1* | *Max* | ***(4*** | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *env\_2\_MASA* | *env\_2\_cleanbg\_MASA* | *45* | *-1* | *Max* | ***(4*** | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *env\_3\_MASA* | *env\_3\_noisebg\_1\_MASA* | *15* | *-1* | *Max* | ***(4*** | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *env\_4\_MASA* | *env\_4\_noisebg\_1\_MASA* | *15* | *-1* | *Max* | ***(4*** | *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** The specific room/environment characteristic and resulting reverb characteristic are defined by the choice of the specific spatial room impulse responses according to the configuration settings of the characterization processing scripts. “env\_1“, “env\_2“ represent conference rooms while “env\_4“ is another indoor environment such as cafeteria or mall. “env\_3” is an outdoor environment like park, nature, event or street.

**(2** Background is defined by the chosen background noise file according to the configuration settings of the characterization processing scripts. “env\_1\_cleanbg” and “env\_2\_cleanbg” indicate low-noise background noises that can be expected in conference rooms such as low air-conditioning/fan noise. “env\_3\_noisebg” and “env\_4\_noisebg” indicate background noises typical for their respective actual environment.

**(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** The talker positions are part of the scene definition of the different categories. They are defined in the Characterization Processing Plan and the corresponding processing scripts. They are 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 are 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 | 16.4, 24.4, 32, 48, 64, 96, 128, 192, 384 kbps |
| DTX | DTX off |
| Input level | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | Idle noise for cat 1,2, 15 dB for cat 3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| 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 [14]. |
| Binaural renderer | IVAS MASA internal binaural rendering |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pairs uttered by different talkers and genders (3 male and 3 female), mixed content, generic audio |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.13.2: Preliminaries for Experiment P800-13

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 16.4 | No errors |
| 2 | c26 | Cat 2 | IVAS FL | 192 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 24.4 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 48 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 384 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - |
| 12 | c24 | Cat 6 | IVAS FL | 96 | 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 = 32 dB | - | - |
| c03 | MNRU Q = 27 dB | - | - |
| c04 | MNRU Q = 22 dB | - | - |
| c05 | MNRU Q = 17 dB | - | - |
| c06 | ESDRU *α* = 0.8 | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - |
| c08 | ESDRU *α* = 0.4 | - | - |
| c09 | ESDRU *α* = 0.2 | - | - |
| c10 | IVAS FL enc / FX dec | 16.4 | off |
| c11 | IVAS FX enc / FL dec | 24.4 | off |
| c12 | IVAS FL enc / FX dec | 32.0 | off |
| c13 | IVAS FX enc / FL dec | 48.0 | off |
| c14 | IVAS FL enc / FX dec | 64.0 | off |
| c15 | IVAS FX enc / FL dec | 96.0 | off |
| c16 | IVAS FL enc / FX dec | 128.0 | off |
| c17 | IVAS FX enc / FL dec | 192 | off |
| c18 | IVAS FL enc / FX dec | 384 | off |
| c19 | IVAS FL | 16.4 | off |
| c20 | IVAS FL | 24.4 | off |
| c21 | IVAS FL | 32.0 | off |
| c22 | IVAS FL | 48.0 | off |
| c23 | IVAS FL | 64.0 | off |
| c24 | IVAS FL | 96.0 | off |
| c25 | IVAS FL | 128.0 | off |
| c26 | IVAS FL | 192 | off |
| c27 | IVAS FL | 384 | off |
| c28 | IVAS FX | 16.4 | off |
| c29 | IVAS FX | 24.4 | off |
| c30 | IVAS FX | 32.0 | off |
| c31 | IVAS FX | 48.0 | off |
| c32 | IVAS FX | 64.0 | off |
| c33 | IVAS FX | 96.0 | off |
| c34 | IVAS FX | 128.0 | off |
| c35 | IVAS FX | 192 | off |
| c36 | IVAS FX | 384 | off |

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*** | ***Talker selection by panel*** |
| *cat 1* | *env\_1\_MASA* | *env\_1\_cleanbg\_MASA* | *45* | *1* | *Max* | ***(4*** | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *env\_2\_MASA* | *env\_2\_cleanbg\_MASA* | *45* | *-1* | *Max* | ***(4*** | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *env\_3\_MASA* | *env\_3\_noisebg\_1\_MASA* | *15* | *-1* | *Max* | ***(4*** | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *env\_4\_MASA* | *env\_4\_noisebg\_1\_MASA* | *15* | *-1* | *Max* | ***(4*** | *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** The specific room/environment characteristic and resulting reverb characteristic are defined by the choice of the specific spatial room impulse responses according to the configuration settings of the characterization processing scripts. “env\_1“, “env\_2“ represent conference rooms while “env\_4“ is another indoor environment such as cafeteria or mall. “env\_3” is an outdoor environment like park, nature, event or street.

**(2** Background is defined by the chosen background noise file according to the configuration settings of the characterization processing scripts. “env\_1\_cleanbg” and “env\_2\_cleanbg” indicate low-noise background noises that can be expected in conference rooms such as low air-conditioning/fan noise. “env\_3\_noisebg” and “env\_4\_noisebg” indicate background noises typical for their respective actual environment.

**(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** The talker positions are part of the scene definition of the different categories. They are defined in the Characterization Processing Plan and the corresponding processing scripts. They are 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 are 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 | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | Idle noise for cat 1,2, 15 dB for cat 3,4 |
| Error Conditions | 4, 8% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 18, 25, 32 dB  *α* = 0.4, 0.6, 0.8 |
| 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 [14]. |
| Binaural renderer | IVAS MASA internal binaural rendering |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pairs uttered by different talkers and genders (3 male and 3 female), mixed content, generic audio |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.14.2: Preliminaries for Experiment P800-14

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** | **DTX** |
| 1 | c23 | Cat 1 | IVAS FL 1TC | 13.2 | 8% | on |
| 2 | c15 | Cat 2 | IVAS FL 1TC | 80 | 4% | on |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - | - |
| 4 | c31 | Cat 4 | IVAS FL 2TC | 16.4 | 8% | on |
| 5 | c08 | Cat 5 | ESDRU = 0.4 | - | - | - |
| 6 | c19 | Cat 6 | IVAS FL 2TC | 32 | 4% | on |
| 7 | c03 | Cat 1 | MNRU Q = 32 dB | - | - | - |
| 8 | c01 | Cat 2 | Reference | - | - | - |
| 9 | c35 | Cat 3 | IVAS FL 2TC | 64 | 8% | on |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 18 dB | - | - | - |
| 12 | c11 | Cat 6 | IVAS FL 1TC | 24.4 | 4% | 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 = 32 dB | 1 TC | - | - | - |
| c04 | MNRU Q = 25 dB | 1 TC | - | - | - |
| c05 | MNRU Q = 18 dB | 1 TC | - | - | - |
| c06 | ESDRU *α* = 0.8 | 2 TC | - | - | - |
| c07 | ESDRU  *α* = 0.6 | 2 TC | - | - | - |
| c08 | ESDRU *α* = 0.4 | 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 F.14.4: Clean and noisy speech categories and scene definitions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Environment(1*** | ***Background(2*** | ***SNR [dB]*** | ***Overtalk [s](3*** | ***Bandwidth*** | ***Talker positions*** | ***Talker selection by panel*** |
| *cat 1* | *env\_1\_MASA* | *env\_1\_cleanbg\_MASA* | *45* | *1* | *Max* | ***(4*** | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *env\_2\_MASA* | *env\_2\_cleanbg\_MASA* | *45* | *-1* | *Max* | ***(4*** | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *env\_3\_MASA* | *env\_3\_noisebg\_1\_MASA* | *15* | *-1* | *Max* | ***(4*** | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *env\_4\_MASA* | *env\_4\_noisebg\_1\_MASA* | *15* | *-1* | *Max* | ***(4*** | *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** The specific room/environment characteristic and resulting reverb characteristic are defined by the choice of the specific spatial room impulse responses according to the configuration settings of the characterization processing scripts. “env\_1“, “env\_2“ represent conference rooms while “env\_4“ is another indoor environment such as cafeteria or mall. “env\_3” is an outdoor environment like park, nature, event or street.

**(2** Background is defined by the chosen background noise file according to the configuration settings of the characterization processing scripts. “env\_1\_cleanbg” and “env\_2\_cleanbg” indicate low-noise background noises that can be expected in conference rooms such as low air-conditioning/fan noise. “env\_3\_noisebg” and “env\_4\_noisebg” indicate background noises typical for their respective actual environment.

**(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** The talker positions are part of the scene definition of the different categories. They are defined in the Characterization Processing Plan and the corresponding processing scripts. They are 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 are 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 | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | 15 dB for cat 1,2,3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Cat. 1-2: Defined scenes, 1 ISM + FOA or HOA3 background  Cat. 3-4: Defined scenes, 2 ISMs + FOA or HOA3 background Cat. 5-6: Pre-producded content |
| Binaural renderer | OSBA to binaural internal rendering |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pairs uttered by different talkers and genders (3 male and 3 female) with scene background, objects with generic audio background |
| Number of categories | 6 |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.15.2: Preliminaries for Experiment P800-15

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 32 | No errors |
| 2 | c26 | Cat 2 | IVAS FL | 384 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 48 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 96 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 512 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - |
| 12 | c24 | Cat 6 | 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 = 32 dB | - | - |
| c03 | MNRU Q = 27 dB | - | - |
| c04 | MNRU Q = 22 dB | - | - |
| c05 | MNRU Q = 17 dB | - | - |
| c06 | ESDRU *α* = 0.8 | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - |
| c08 | ESDRU *α* = 0.4 | - | - |
| c09 | ESDRU *α* = 0.2 | - | - |
| c10 | IVAS FL enc / FX dec | 32.0 | Off |
| c11 | IVAS FX enc / FL dec | 48.0 | Off |
| c12 | IVAS FL enc / FX dec | 64.0 | Off |
| c13 | IVAS FX enc / FL dec | 96.0 | Off |
| c14 | IVAS FL enc / FX dec | 128.0 | Off |
| c15 | IVAS FX enc / FL dec | 192.0 | Off |
| c16 | IVAS FL enc / FX dec | 256.0 | Off |
| c17 | IVAS FX enc / FL dec | 384.0 | Off |
| c18 | IVAS FL enc / FX dec | 512.0 | Off |
| c19 | IVAS FL | 32.0 | Off |
| c20 | IVAS FL | 48.0 | Off |
| c21 | IVAS FL | 64.0 | Off |
| c22 | IVAS FL | 96.0 | Off |
| c23 | IVAS FL | 128.0 | Off |
| c24 | IVAS FL | 192.0 | Off |
| c25 | IVAS FL | 256.0 | Off |
| c26 | IVAS FL | 384.0 | Off |
| c27 | IVAS FL | 512.0 | Off |
| c28 | IVAS FX | 32.0 | Off |
| c29 | IVAS FX | 48.0 | Off |
| c30 | IVAS FX | 64.0 | Off |
| c31 | IVAS FX | 96.0 | Off |
| c32 | IVAS FX | 128.0 | Off |
| c33 | IVAS FX | 192.0 | Off |
| c34 | IVAS FX | 256.0 | Off |
| c35 | IVAS FX | 384.0 | Off |
| c36 | IVAS FX | 512.0 | Off |

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Background(1*** | ***SNR [dB]*** | ***Overtalk [s](2*** | ***Bandwidth*** | ***Talker positions(3*** | ***Talker selection by panel*** |
| *cat 1* | *env\_3\_noisebg\_1\_HOA3* | *15* | *-* | *Max* | *as defined in Table F.9.4* | *P1: M1 P2: M1*  *P3: M1 P4: M1*  *P5: M1*  *P6: M1* |
| *cat 2* | *env\_4\_noisebg\_1\_HOA3* | *15* | *-* | *Max* | *as defined in Table F.9.4* | *P1: F1 P2: F1 P3: F1*  *P4: F1*  *P5: F1*  *P6: F1* |
| *cat 3* | *env\_3\_noisebg\_2\_HOA3* | *15* | *1* | *Max* | *as defined in Table F.9.4* | *P1: M2F2 P2: M2F2 P3: M2F2 P4: M2F2*  *P5: M2F2 P6: M2F2* |
| *cat 4* | *env\_4\_noisebg\_2\_HOA3* | *15* | *-1* | *Max* | *as defined in Table F.9.4* | *P1: M3F3 P2: M3F3 P3: M3F3 P4: M3F3 P5: M3F3 P6: M3F3* |

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** The selected background noise is representative for a given environment. It is defined by the chosen background noise file according to the configuration settings of the characterization processing scripts. “env\_3” is an outdoor environment like park, nature, event or street, while “env\_4” corresponds to indoor environments such as cafeteria or mall.

**(2** 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.

**(3** Panel’s ISM positions for categories cat 1-4 are as defined respectively in Table F.9.4.

* 1. Experiment P800-16: OSBA (3-4 objects)

Tables F.16.1 to F.16.3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively. Table F.16.4 shows 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 | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | About 15 dB |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Cat. 1-3: Pre-produced content, 3 ISMs + HOA3 background  Cat. 4-6: Pre-produced content, 4 ISMs + HOA3 background |
| Binaural renderer | OSBA to binaural internal rendering |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | ISMs with ambisonics background |
| Number of categories | 6 |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.16.2: Preliminaries for Experiment P800-16

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 32 | No errors |
| 2 | c26 | Cat 2 | IVAS FL | 384 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 48 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 96 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 512 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - |
| 12 | c24 | Cat 6 | 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 = 32 dB | - | - |
| c03 | MNRU Q = 27 dB | - | - |
| c04 | MNRU Q = 22 dB | - | - |
| c05 | MNRU Q = 17 dB | - | - |
| c06 | ESDRU *α* = 0.8 | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - |
| c08 | ESDRU *α* = 0.4 | - | - |
| c09 | ESDRU *α* = 0.2 | - | - |
| c10 | IVAS FL enc / FX dec | 32.0 | Off |
| c11 | IVAS FX enc / FL dec | 48.0 | Off |
| c12 | IVAS FL enc / FX dec | 64.0 | Off |
| c13 | IVAS FX enc / FL dec | 96.0 | Off |
| c14 | IVAS FL enc / FX dec | 128.0 | Off |
| c15 | IVAS FX enc / FL dec | 192.0 | Off |
| c16 | IVAS FL enc / FX dec | 256.0 | Off |
| c17 | IVAS FX enc / FL dec | 384.0 | Off |
| c18 | IVAS FL enc / FX dec | 512.0 | Off |
| c19 | IVAS FL | 32.0 | Off |
| c20 | IVAS FL | 48.0 | Off |
| c21 | IVAS FL | 64.0 | Off |
| c22 | IVAS FL | 96.0 | Off |
| c23 | IVAS FL | 128.0 | Off |
| c24 | IVAS FL | 192.0 | Off |
| c25 | IVAS FL | 256.0 | Off |
| c26 | IVAS FL | 384.0 | Off |
| c27 | IVAS FL | 512.0 | Off |
| c28 | IVAS FX | 32.0 | Off |
| c29 | IVAS FX | 48.0 | Off |
| c30 | IVAS FX | 64.0 | Off |
| c31 | IVAS FX | 96.0 | Off |
| c32 | IVAS FX | 128.0 | Off |
| c33 | IVAS FX | 192.0 | Off |
| c34 | IVAS FX | 256.0 | Off |
| c35 | IVAS FX | 384.0 | Off |
| c36 | IVAS FX | 512.0 | Off |

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

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 1, 4 | 3-object, 4-objects respectively + ambient background (office, nature, outdoor/indoor environment), where objects are speech (including overlap with max. 2 talkers). Nevertheless, for 4-objects one object may be music or effects. |
| cat 2, 5 | 3-object, 4-objects respectively + mixed and music where objects are speech (including overlap with max. 2 talkers). Nevertheless, for 4-objects one object may be music or effects. |
| 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 F.17.1: Overview of test conditions

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64, 80, 96, 128, 160, 192, 256, 384, 512 kbps |
| DTX | DTX off |
| Input level | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | About 10 dB |
| Error Conditions | 0%, 5% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 22, 26, 30, 34 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Pre-produced content of defined scenes, ISMs + HOA3 |
| Binaural renderer | OSBA to binaural internal rendering |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pairs uttered by different talkers and genders (3 male and 3 female) with scene background |
| Number of categories | 6 |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.17.2: Preliminaries for Experiment P800-17

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c18 | Cat 1 | IVAS FL | 16.4 | No errors |
| 2 | c16 | Cat 2 | IVAS FL | 256 | 5% |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c17 | Cat 4 | IVAS FL | 13.2 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c12 | Cat 6 | IVAS FL | 128 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 34 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c14 | Cat 3 | IVAS FL | 32 | 5% |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 22 dB | - | - |
| 12 | c30 | Cat 6 | IVAS FL | 512 | No errors |

Table F.17.3: Test conditions for P800-17 OSBA, speech and background environments.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **FER** | **Reference** |
| c01 | Reference | - | - | - |  |
| 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 | IVAS FL OSBA | 32 | off | 0% |  |
| c11 | IVAS FL OSBA | 64 | off | 0% |  |
| c12 | IVAS FL OSBA | 128 | off | 0% |  |
| c13 | IVAS FL OSBA | 256 | off | 0% |  |
| c14 | IVAS FX OSBA | 32 | off | 5% |  |
| c15 | IVAS FX OSBA | 64 | off | 5% |  |
| c16 | IVAS FX OSBA | 256 | off | 5% |  |
| c17 | IVAS FX OSBA | 13.2 | off | 0% |  |
| c18 | IVAS FX OSBA | 16.4 | off | 0% |  |
| c19 | IVAS FX OSBA | 24.4 | off | 0% |  |
| c20 | IVAS FX OSBA | 32 | off | 0% |  |
| c21 | IVAS FX OSBA | 48 | off | 0% | c31 |
| c22 | IVAS FX OSBA | 64 | off | 0% | c32 |
| c23 | IVAS FX OSBA | 80 | off | 0% |  |
| c24 | IVAS FX OSBA | 96 | off | 0% | c33 |
| c25 | IVAS FX OSBA | 128 | off | 0% | c34 |
| c26 | IVAS FX OSBA | 160 | off | 0% |  |
| c27 | IVAS FX OSBA | 192 | off | 0% | c35 |
| c28 | IVAS FX OSBA | 256 | off | 0% | c36 |
| c29 | IVAS FX OSBA | 384 | off | 0% |  |
| c30 | IVAS FX OSBA | 512 | off | 0% |  |
| c31 | IVAS FX ISM + HOA3 | 24.4 + 24.4 | off | 0% |  |
| c32 | IVAS FX ISM + HOA3 | 32 + 32 | off | 0% |  |
| c33 | IVAS FX ISM + HOA3 | 48 + 48 | off | 0% |  |
| c34 | IVAS FX ISM + HOA3 | 64 + 64 | off | 0% |  |
| c35 | IVAS FX ISM + HOA3 | 96 + 96 | off | 0% |  |
| c36 | IVAS FX ISM + HOA3 | 128 + 128 | off | 0% |  |

Table F.17.4: Sample Categories

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

* 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. Table F.18.4 shows 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, 128, 256 kbps |
| DTX | DTX off |
| Input level | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | 15 dB for cat 1,2,3,4 |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Cat. 1-2: Defined scenes, 1 ISM + MASA 2TC background  Cat. 3-4: Defined scenes, 2 ISMs + MASA 2TC background Cat. 5-6: Pre-produced content |
| Binaural renderer | OMASA to binaural internal rendering |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pairs uttered by different talkers and genders (3 male and 3 female) |
| Number of categories | 6 Different environments with 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.18.2: Preliminaries for Experiment P800-18

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 13.2 | No errors |
| 2 | c26 | Cat 2 | IVAS FL | 128 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 16.4 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 32 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 256 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - |
| 12 | c24 | Cat 6 | 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 = 32 dB | - | - |
| c03 | MNRU Q = 27 dB | - | - |
| c04 | MNRU Q = 22 dB | - | - |
| c05 | MNRU Q = 17 dB | - | - |
| c06 | ESDRU *α* = 0.8 | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - |
| c08 | ESDRU *α* = 0.4 | - | - |
| c09 | ESDRU *α* = 0.2 | - | - |
| c10 | IVAS FL enc / FX dec | 13.2 | Off |
| c11 | IVAS FX enc / FL dec | 16.4 | Off |
| c12 | IVAS FL enc / FX dec | 24.4 | Off |
| c13 | IVAS FX enc / FL dec | 32.0 | Off |
| c14 | IVAS FL enc / FX dec | 48.0 | Off |
| c15 | IVAS FX enc / FL dec | 64.0 | Off |
| c16 | IVAS FL enc / FX dec | 80.0 | Off |
| c17 | IVAS FX enc / FL dec | 128.0 | Off |
| c18 | IVAS FL enc / FX dec | 256.0 | Off |
| c19 | IVAS FL | 13.2 | off |
| c20 | IVAS FL | 16.4 | off |
| c21 | IVAS FL | 24.4 | off |
| c22 | IVAS FL | 32.0 | off |
| c23 | IVAS FL | 48.0 | off |
| c24 | IVAS FL | 64.0 | off |
| c25 | IVAS FL | 80.0 | off |
| c26 | IVAS FL | 128.0 | off |
| c27 | IVAS FL | 256.0 | off |
| c28 | IVAS FX | 13.2 | off |
| c29 | IVAS FX | 16.4 | off |
| c30 | IVAS FX | 24.4 | off |
| c31 | IVAS FX | 32.0 | off |
| c32 | IVAS FX | 48.0 | off |
| c33 | IVAS FX | 64.0 | off |
| c34 | IVAS FX | 80.0 | off |
| c35 | IVAS FX | 128.0 | off |
| c36 | IVAS FX | 256.0 | off |

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Background(1*** | ***SNR [dB]*** | ***Overtalk [s](2*** | ***Bandwidth*** | ***Talker positions(3*** | ***Talker selection by panel*** |
| *cat 1* | *env\_3\_noisebg\_1\_MA*SA | *15* | *-* | *Max* | *as defined in Table F.9.4* | *P1: M1 P2: M1 P3: M1 P4: M1 P5: M1 P6: M1* |
| *cat 2* | *env\_4\_noisebg\_1\_MA*SA | *15* | *-* | *Max* | *as defined in Table F.9.4* | *P1: F1 P2: F1 P3: F1 P4: F1 P5: F1 P6: F1* |
| *cat 3* | *env\_3\_noisebg\_2\_MA*SA | *15* | *1* | *Max* | *as defined in Table F.9.4* | *P1: M2F2 P2: M2F2 P3: M2F2 P4: M2F2 P5: M2F2 P6: M2F2* |
| *cat 4* | *env\_4\_noisebg\_2\_MA*SA | *15* | *-1* | *Max* | *as defined in Table F.9.4* | *P1: M3F3 P2: M3F3 P3: M3F3 P4: M3F3 P5: M3F3 P6: M3F3* |

Table F.18.5: 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** The selected background noise is representative for a given environment. It is defined by the chosen background noise file according to the configuration settings of the characterization processing scripts. “env\_3” is an outdoor environment like park, nature, event or street, while “env\_4” corresponds to indoor environments such as cafeteria or mall.

**(2** 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.

**(3** Panel’s ISM positions for categories cat 1-4 are as defined respectively in Table F.9.4.

* 1. Experiment P800-19: OMASA (3-4 objects)

Tables F.19.1 to F.19.3 show conditions to be used for this experiment, list of preliminaries and full list of conditions, respectively. Table F.19.4 shows 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 | 16.4, 24.4, 48, 64, 80, 96, 128, 256, 512 kbps |
| DTX | DTX off |
| Input level | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | About 15 dB |
| Error Conditions | 0% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Cat. 1-3: Pre-produced content, 3 ISMs + MASA 2TC background  Cat. 4-6: Pre-produced content, 4 ISMs + MASA 2TC background |
| Binaural renderer | OMASA to binaural internal rendering |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | ISMs with MASA background |
| Number of categories | 6 |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.19.2: Preliminaries for Experiment P800-19

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 | Cat 1 | IVAS FL | 16.4 | No errors |
| 2 | c26 | Cat 2 | IVAS FL | 256 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c20 | Cat 4 | IVAS FL | 24.4 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c22 | Cat 6 | IVAS FL | 64 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c27 | Cat 3 | IVAS FL | 512 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - |
| 12 | c24 | Cat 6 | IVAS FL | 96 | 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 = 32 dB | - | - |
| c03 | MNRU Q = 27 dB | - | - |
| c04 | MNRU Q = 22 dB | - | - |
| c05 | MNRU Q = 17 dB | - | - |
| c06 | ESDRU *α* = 0.8 | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - |
| c08 | ESDRU *α* = 0.4 | - | - |
| c09 | ESDRU *α* = 0.2 | - | - |
| c10 | IVAS FL enc / FX dec | 16.4 | off |
| c11 | IVAS FX enc / FL dec | 24.4 | off |
| c12 | IVAS FL enc / FX dec | 48.0 | off |
| c13 | IVAS FX enc / FL dec | 64.0 | off |
| c14 | IVAS FL enc / FX dec | 80.0 | off |
| c15 | IVAS FX enc / FL dec | 96.0 | off |
| c16 | IVAS FL enc / FX dec | 128.0 | off |
| c17 | IVAS FX enc / FL dec | 256.0 | off |
| c18 | IVAS FL enc / FX dec | 512.0 | off |
| c19 | IVAS FL | 16.4 | off |
| c20 | IVAS FL | 24.4 | off |
| c21 | IVAS FL | 48.0 | off |
| c22 | IVAS FL | 64.0 | off |
| c23 | IVAS FL | 80.0 | off |
| c24 | IVAS FL | 96.0 | off |
| c25 | IVAS FL | 128.0 | off |
| c26 | IVAS FL | 256.0 | off |
| c27 | IVAS FL | 512.0 | off |
| c28 | IVAS FX | 16.4 | off |
| c29 | IVAS FX | 24.4 | off |
| c30 | IVAS FX | 48.0 | off |
| c31 | IVAS FX | 64.0 | off |
| c32 | IVAS FX | 80.0 | off |
| c33 | IVAS FX | 96.0 | off |
| c34 | IVAS FX | 128.0 | off |
| c35 | IVAS FX | 256.0 | off |
| c36 | IVAS FX | 512.0 | off |

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

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 1-3 | 3-objects + 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** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 13.2, 16.4, 24.4, 32, 48, 64, 80, 96, 128, 160, 192, 256, 384, 512 kbps |
| DTX | DTX off |
| Input level | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | About 10 dB |
| Error Conditions | 0%, 5% |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 22, 26, 30, 34 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Pre-produced content of defined scenes, ISMs + MASA 2TC |
| Binaural renderer | OMASA to binaural internal rendering |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pairs uttered by different talkers and genders (3 male and 3 female) with scene background |
| Number of categories | 6 |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.20.2: Preliminaries for Experiment P800-20

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c18 | Cat 1 | IVAS FL | 16.4 | No errors |
| 2 | c16 | Cat 2 | IVAS FL | 256 | No errors |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c17 | Cat 4 | IVAS FL | 13.2 | No errors |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c12 | Cat 6 | IVAS FL | 128 | No errors |
| 7 | c02 | Cat 1 | MNRU Q = 34 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c14 | Cat 3 | IVAS FL | 32 | No errors |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 22 dB | - | - |
| 12 | c30 | Cat 6 | IVAS FL | 512 | No errors |

Table F.20.3: Test conditions for P800-20 OMASA, speech and background environments.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **FER** | **Reference** |
| c01 | Reference | - | - | - |  |
| 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 | IVAS FL OMASA 2TC | 32 | off | 0% |  |
| c11 | IVAS FL OMASA 2TC | 64 | off | 0% |  |
| c12 | IVAS FL OMASA 2TC | 128 | off | 0% |  |
| c13 | IVAS FL OMASA 2TC | 256 | off | 0% |  |
| c14 | IVAS FX OMASA 2TC | 32 | off | 5% |  |
| c15 | IVAS FX OMASA 2TC | 64 | off | 5% |  |
| c16 | IVAS FX OMASA 2TC | 256 | off | 5% |  |
| c17 | IVAS FX OMASA 2TC | 13.2 | off | 0% |  |
| c18 | IVAS FX OMASA 2TC | 16.4 | off | 0% |  |
| c19 | IVAS FX OMASA 2TC | 24.4 | off | 0% |  |
| c20 | IVAS FX OMASA 2TC | 32 | off | 0% |  |
| c21 | IVAS FX OMASA 2TC | 48 | off | 0% | c31 |
| c22 | IVAS FX OMASA 2TC | 64 | off | 0% | c32 |
| c23 | IVAS FX OMASA 2TC | 80 | off | 0% |  |
| c24 | IVAS FX OMASA 2TC | 96 | off | 0% | c33 |
| c25 | IVAS FX OMASA 2TC | 128 | off | 0% | c34 |
| c26 | IVAS FX OMASA 2TC | 160 | off | 0% |  |
| c27 | IVAS FX OMASA 2TC | 192 | off | 0% | c35 |
| c28 | IVAS FX OMASA 2TC | 256 | off | 0% | c36 |
| c29 | IVAS FX OMASA 2TC | 384 | off | 0% |  |
| c30 | IVAS FX OMASA 2TC | 512 | off | 0% |  |
| c31 | IVAS FX ISM + MASA 2TC | 24.4 + 24.4 | off | 0% |  |
| c32 | IVAS FX ISM + MASA 2TC | 32 + 32 | off | 0% |  |
| c33 | IVAS FX ISM + MASA 2TC | 48 + 48 | off | 0% |  |
| c34 | IVAS FX ISM + MASA 2TC | 64 + 64 | off | 0% |  |
| c35 | IVAS FX ISM + MASA 2TC | 96 + 96 | off | 0% |  |
| c36 | IVAS FX ISM + MASA 2TC | 128 + 128 | off | 0% |  |

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* | *Approx. -36* | *No overtalk* | *2 samples with fixed positions, 4 samples with movement* |
| *cat 2* | *2* | *-26* | *Indoors 2* | *Approx. -36* | *Overtalk* | *2 samples with fixed positions, 4 samples with movement\** |
| *cat 3* | *3* | *-26* | *Outdoors 1* | *Approx. -36* | *Overtalk* | *2 samples with fixed positions, 4 samples with movement\** |
| *cat 4* | *4* | *-26* | *Outdoors 2* | *Approx. -36* | *Overtalk* | *2 samples with fixed positions, 4 samples with movement\** |
| *cat 5* | *2* | *-26* | *Background with music 1* | *Approx. -36* | *No overtalk* | *2 samples with fixed positions, 4 samples with movement\** |
| *cat 6* | *3* | *-26* | *Background with music 2* | *Approx. -36* | *Overtalk* | *2 samples with fixed positions, 4 samples 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.

* 1. Experiment P800-21: JBM with Stereo

Tables F.21.1 to F.21.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Music and Mixed content categories, respectively.

Table F.21.1: Conditions for Experiment P800-21

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 24.4, 48, 96 kbps |
| DTX | DTX on |
| Input level | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Noise | Idle noise for cat 1,2, 15 dB for cat 3,4 |
| Error Conditions | JBM Profiles I1. O1, I1.O2 (offset O1 is a random number and O2=(O1+4000)%8000)  Error patterns Error I1.O1, Error I1.O2 (error pattern files derived from the respective delay and error profiles) |
| **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 [13] 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 pairs uttered by different talkers and genders (3 male and 3 female), music and mixed content |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |
|  |  |

Table F.21.2 : Preliminaries for Experiment P800-21

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **Profile*(3*** | **DTX** |
| 1 | c31 | Cat 1 | IVAS FL | 24.4 | I1.O1 | on |
| 2 | c24 | Cat 2 | IVAS FL | 48 | no error | on |
| 3 | c06 | Cat 3 | ESDRU = 0.7 | - | - | - |
| 4 | c30 | Cat 4 | IVAS FL | 96 | Error I1.O2 | on |
| 5 | c09 | Cat 5 | ESDRU = 0.1 | - | - | - |
| 6 | c35 | Cat 6 | IVAS FL | 48 | I1.O2 | on |
| 7 | c02 | Cat 1 | MNRU Q = 27 dB | - | - | - |
| 8 | c01 | Cat 2 | Reference | - | - | - |
| 9 | c26 | Cat 3 | IVAS FL | 24.4 | Error I1.O1 | on |
| 10 | c07 | Cat 4 | ESDRU = 0.5 | - | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 12 dB | - | - | - |
| 12 | c25 | Cat 6 | IVAS FL | 96 | no error | on |

Table F.21.3: Test conditions for Experiment P800-21,  
stereo speech and music and mixed content under impaired and clean channel conditions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **Profile*(3*** |
| 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 | 24.4 | on | No error |
| c11 | IVAS FL enc / FX dec | 48.0 | on | No error |
| c12 | IVAS FL enc / FX dec | 96.0 | on | No error |
| c13 | IVAS FL enc / FX dec | 24.4 | on | Error I1. O1 |
| c14 | IVAS FL enc / FX dec | 48.0 | on | Error I1. O1 |
| c15 | IVAS FL enc / FX dec | 24.4 | on | Error I1. O2 |
| c16 | IVAS FL enc / FX dec | 48.0 | on | Error I1. O2 |
| c17 | IVAS FL enc / FX dec | 24.4 | on | I1. O1 |
| c18 | IVAS FL enc / FX dec | 48.0 | on | I1. O1 |
| c19 | IVAS FL enc / FX dec | 96.0 | on | I1. O1 |
| c20 | IVAS FL enc / FX dec | 24.4 | on | I1. O2 |
| c21 | IVAS FL enc / FX dec | 48.0 | on | I1. O2 |
| c22 | IVAS FL enc / FX dec | 96.0 | on | I1. O2 |
| c23 | IVAS FL enc / FL dec | 24.4 | on | No error |
| c24 | IVAS FL enc / FL dec | 48.0 | on | No error |
| c25 | IVAS FL enc / FL dec | 96.0 | on | No error |
| c26 | IVAS FL enc / FL dec | 24.4 | on | Error I1. O1 |
| c27 | IVAS FL enc / FL dec | 48.0 | on | Error I1. O1 |
| c28 | IVAS FL enc / FL dec | 24.4 | on | Error I1. O2 |
| c29 | IVAS FL enc / FL dec | 48.0 | on | Error I1. O2 |
| c30 | IVAS FL enc / FL dec | 96.0 | on | Error I1. O2 |
| c31 | IVAS FL enc / FL dec | 24.4 | on | I1. O1 |
| c32 | IVAS FL enc / FL dec | 48.0 | on | I1. O1 |
| c33 | IVAS FL enc / FL dec | 96.0 | on | I1. O1 |
| c34 | IVAS FL enc / FL dec | 24.4 | on | I1. O2 |
| c35 | IVAS FL enc / FL dec | 48.0 | on | I1. O2 |
| c36 | IVAS FL enc / FL dec | 96.0 | on | I1. O2 |

Table F.21.4: Clean and noisy speech categories and scene definitions

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Room*** | ***Reverb*** | ***Microphone Setup*** | ***Background*** | ***SNR***  ***[dB]*** | ***Overtalk [s](1*** | ***Bandwidth*** | ***Talker positions by panel(2*** | ***Talker selection by panel*** |
| cat 1 | small | anechoic | M-S | Low level idle noise | 45 | 1 | Max available up to SWB | P1: 1-7  P2: 5-3  P3: 2-6  P4: 4-1  P5: 3-4  P6: 7-2 | P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3 |
| cat 2 | large | echoic | A-B (150 cm) | Low level idle noise | 45 | -1 | max available up to SWB | P1: 5-11  P2: 1-6  P3: 3-7  P4: 5-8  P5: 9-7  P6: 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 | P1: 1-7  P2: 5-3  P3: 2-6  P4: 4-1  P5: 3-4  P6: 7-2 | P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1 |
| cat 4 | car | car | A-B Cardioid pair 20 cm | car | 15 | -1 | Max available up to SWB | P1: Driver-Passenger  P2: BackRight-Driver  P3: Driver-BackCenter  P4: BackLeft-Driver  P5:BackRight-BackLeft  P6:BackCenter-BackRight | P1: m1f1  P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3 |

Table F.21.5: Mixed content and music categories

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

**Notes:**

**(1** Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.

**(2** The talker positions are part of the scene definition of the different categories. They correspond to the talker positions as depicted in Figures 14.3 and 14.5 of [13] for the large and the small room, respectively.

**(3 ”**Error I1. Ox**”** means an error pattern derived from the Profile I1.Ox.

* 1. Experiment P800-22: JBM with 1-2 objects

Tables F.22.1 to F.22.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Speech with background and Music and mixed content categories, respectively.

Table F.22.1: Conditions for Experiment P800-22

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 24.4, 48, 96kbps |
| DTX | DTX off |
| Input level | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Noise | No noise |
| Error Conditions | JBM Profiles I1.O1, I1.O2 (offset O1 is a random number and O2=(O1+4000)%8000)  Error patterns Error I1.O1, Error I1.O2 (error pattern files derived from the respective delay and error profiles) |
| **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 |
| Binaural renderer | ISM to binaural internal rendering |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pairs uttered by different talkers and genders (3 male and 3 female), music and mixed content, speech and background |
| Number of categories | 6 |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.22.2: Preliminaries for Experiment P800-22

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **Profile*(3*** |
| 1 | c31 | Cat 1 | IVAS FL | 24.4 | I1.O1 |
| 2 | c24 | Cat 2 | IVAS FL | 48 | no error |
| 3 | c06 | Cat 3 | ESDRU = 0.7 | - | - |
| 4 | c30 | Cat 4 | IVAS FL | 96 | Error I1.O2 |
| 5 | c09 | Cat 5 | ESDRU = 0.1 | - | - |
| 6 | c35 | Cat 6 | IVAS FL | 48 | I1.O2 |
| 7 | c02 | Cat 1 | MNRU Q = 36 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c26 | Cat 3 | IVAS FL | 24.4 | Error I1.O1 |
| 10 | c07 | Cat 4 | ESDRU = 0.5 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 18 dB | - | - |
| 12 | c25 | Cat 6 | IVAS FL | 96 | no error |

Table F.22.3: Test conditions for Experiment P800-22,  
speech under clean and impaired channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **Profile*(3*** |
| 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 | 24.4 | No error |
| c11 | IVAS FL enc / FX dec | 48.0 | No error |
| c12 | IVAS FL enc / FX dec | 96.0 | No error |
| c13 | IVAS FL enc / FX dec | 24.4 | Error I1. O1 |
| c14 | IVAS FL enc / FX dec | 48.0 | Error I1. O1 |
| c15 | IVAS FL enc / FX dec | 24.4 | Error I1. O2 |
| c16 | IVAS FL enc / FX dec | 48.0 | Error I1. O2 |
| c17 | IVAS FL enc / FX dec | 24.4 | I1. O1 |
| c18 | IVAS FL enc / FX dec | 48.0 | I1. O1 |
| c19 | IVAS FL enc / FX dec | 96.0 | I1. O1 |
| c20 | IVAS FL enc / FX dec | 24.4 | I1. O2 |
| c21 | IVAS FL enc / FX dec | 48.0 | I1. O2 |
| c22 | IVAS FL enc / FX dec | 96.0 | I1. O2 |
| c23 | IVAS FL enc / FL dec | 24.4 | No error |
| c24 | IVAS FL enc / FL dec | 48.0 | No error |
| c25 | IVAS FL enc / FL dec | 96.0 | No error |
| c26 | IVAS FL enc / FL dec | 24.4 | Error I1. O1 |
| c27 | IVAS FL enc / FL dec | 48.0 | Error I1. O1 |
| c28 | IVAS FL enc / FL dec | 24.4 | Error I1. O2 |
| c29 | IVAS FL enc / FL dec | 48.0 | Error I1. O2 |
| c30 | IVAS FL enc / FL dec | 96.0 | Error I1. O2 |
| c31 | IVAS FL enc / FL dec | 24.4 | I1. O1 |
| c32 | IVAS FL enc / FL dec | 48.0 | I1. O1 |
| c33 | IVAS FL enc / FL dec | 96.0 | I1. O1 |
| c34 | IVAS FL enc / FL dec | 24.4 | I1. O2 |
| c35 | IVAS FL enc / FL dec | 48.0 | I1. O2 |
| c36 | IVAS FL enc / FL dec | 96.0 | I1. O2 |

**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:

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

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

**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. Non-overlapping utterances.
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. Overlapping utterances.
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. Overlapping utterances.
5. Two talkers walking side-by-side around the table (elevation 45°). The azimuth is the same for both talkers and varies continually. Non-overlapping utterances.
6. Two talkers walking around the table in opposite directions (elevation 30°), starting at the same position. Azimuths of both talkers vary continually. Overlapping utterances.

The following table lists the test Categories corresponding to different talkers or talker pairs. Each of the sentence pairs uttered by a certain talker or talker pair is associated to a different scene.

Table F.22.4: Allocation of scenes for each talker or talker pair (category cat 1, cat 2, cat 3, cat 4) and listening panel (P1-P6)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Scene*** | ***Talker initial elevation*** | ***Elevation change*** | ***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 3:***  *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 4:***  *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

**(3 ”**Error I1. Ox**”** means an error pattern derived from the Profile I1.Ox.

Table F.22.5: Music and mixed content and Speech and background categories

|  |  |
| --- | --- |
| **Category** | **Type** |
| cat 5 | Music and mixed content (1 object) |
| cat 6 | speech + background (2 objects) |

* 1. Experiment P800-23: JBM with FOA

Tables F.23.1 to F.23.5 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.23.1: Conditions for Experiment P800-23

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 24.4, 48, 96 kbps |
| DTX | DTX off |
| Input level | -16, -26, -36 LKFS, as defined in Table 3 |
| Input frequency mask | 20KBP |
| Background | Idle noise for cat 1,2, 15 dB for cat 3,4 |
| Error Conditions | JBM Profiles I1.O1, I1.O2 (offset O1 is a random number and O2=(O1+4000)%8000)  Error patterns Error I1.O1, Error I1.O2 (error pattern files derived from the respective delay and error profiles) |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRU  ESDRU | Q = 17, 22, 27, 32 dB  *α* = 0.2, 0.4, 0.6, 0.8 |
| 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 pairs uttered by different talkers and genders (3 male and 3 female), mixed content, generic audio |
| 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 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.23.2: Preliminaries for Experiment P800-23

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **Profile*(5*** |
| 1 | c31 | Cat 1 | IVAS FL | 24.4 | I1.O1 |
| 2 | c24 | Cat 2 | IVAS FL | 48 | no error |
| 3 | c06 | Cat 3 | ESDRU = 0.8 | - | - |
| 4 | c30 | Cat 4 | IVAS FL | 96 | Error I1.O2 |
| 5 | c09 | Cat 5 | ESDRU = 0.2 | - | - |
| 6 | c35 | Cat 6 | IVAS FL | 48 | I1.O2 |
| 7 | c02 | Cat 1 | MNRU Q = 32 dB | - | - |
| 8 | c01 | Cat 2 | Reference | - | - |
| 9 | c26 | Cat 3 | IVAS FL | 24.4 | Error I1.O1 |
| 10 | c07 | Cat 4 | ESDRU = 0.6 | - | - |
| 11 | c05 | Cat 5 | MNRU Q = 17 dB | - | - |
| 12 | c25 | Cat 6 | IVAS FL | 96 | no error |

Table F.23.3: Test conditions for Experiment P800-23,  
speech and mixed-music under impaired and clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **Profile*(5*** |
| c01 | Reference | - | - |
| c02 | MNRU Q = 32 dB | - | - |
| c03 | MNRU Q = 27 dB | - | - |
| c04 | MNRU Q = 22 dB | - | - |
| c05 | MNRU Q = 17 dB | - | - |
| c06 | ESDRU *α* = 0.8 | - | - |
| c07 | ESDRU  *α* = 0.6 | - | - |
| c08 | ESDRU *α* = 0.4 | - | - |
| c09 | ESDRU *α* = 0.2 | - | - |
| c10 | IVAS FL enc / FX dec | 24.4 | No error |
| c11 | IVAS FL enc / FX dec | 48.0 | No error |
| c12 | IVAS FL enc / FX dec | 96.0 | No error |
| c13 | IVAS FL enc / FX dec | 24.4 | Error I1. O1 |
| c14 | IVAS FL enc / FX dec | 48.0 | Error I1. O1 |
| c15 | IVAS FL enc / FX dec | 24.4 | Error I1. O2 |
| c16 | IVAS FL enc / FX dec | 48.0 | Error I1. O2 |
| c17 | IVAS FL enc / FX dec | 24.4 | I1. O1 |
| c18 | IVAS FL enc / FX dec | 48.0 | I1. O1 |
| c19 | IVAS FL enc / FX dec | 96.0 | I1. O1 |
| c20 | IVAS FL enc / FX dec | 24.4 | I1. O2 |
| c21 | IVAS FL enc / FX dec | 48.0 | I1. O2 |
| c22 | IVAS FL enc / FX dec | 96.0 | I1. O2 |
| c23 | IVAS FL enc / FL dec | 24.4 | No error |
| c24 | IVAS FL enc / FL dec | 48.0 | No error |
| c25 | IVAS FL enc / FL dec | 96.0 | No error |
| c26 | IVAS FL enc / FL dec | 24.4 | Error I1. O1 |
| c27 | IVAS FL enc / FL dec | 48.0 | Error I1. O1 |
| c28 | IVAS FL enc / FL dec | 24.4 | Error I1. O2 |
| c29 | IVAS FL enc / FL dec | 48.0 | Error I1. O2 |
| c30 | IVAS FL enc / FL dec | 96.0 | Error I1. O2 |
| c31 | IVAS FL enc / FL dec | 24.4 | I1. O1 |
| c32 | IVAS FL enc / FL dec | 48.0 | I1. O1 |
| c33 | IVAS FL enc / FL dec | 96.0 | I1. O1 |
| c34 | IVAS FL enc / FL dec | 24.4 | I1. O2 |
| c35 | IVAS FL enc / FL dec | 48.0 | I1. O2 |
| c36 | IVAS FL enc / FL dec | 96.0 | I1. O2 |

Table F.23.4: Clean and noisy speech categories and scene definitions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | ***Environment(1*** | ***Background(2*** | ***SNR [dB]*** | ***Overtalk [s](3*** | ***Bandwidth*** | ***Talker position*** | ***Talker selection by panel*** |
| *cat 1* | *env\_1\_FOA* | *env\_1\_cleanbg\_FOA* | *45* | *1* | *Max* | ***(4*** | *P1: f1m1 P2: m2f2 P3: f3m3 P4: m1f1 P5: f2m2 P6: m3f3* |
| *cat 2* | *env\_2\_FOA* | *env\_2\_cleanbg\_FOA* | *45* | *-1* | *Max* | ***(4*** | *P1: m3f3 P2: f1m1 P3: m2f2 P4: f3m3 P5: m1f1 P6: f2m2* |
| *cat 3* | *env\_3\_FOA* | *env\_3\_noisebg\_1\_FOA* | *15* | *1* | *Max* | ***(4*** | *P1: f2m2 P2: m3f3 P3: f1m1 P4: m2f2 P5: f3m3 P6: m1f1* |
| *cat 4* | *env\_4\_FOA* | *env\_4\_noisebg\_1\_FOA* | *15* | *-1* | *Max* | ***(4*** | *P1: m1f1 P2: f2m2 P3: m3f3 P4: f1m1 P5: m2f2 P6: f3m3* |

Table F.23.5: Mixed content and Generic audio categories

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

**Notes:**

**(1** The specific room/environment characteristic and resulting reverb characteristic are defined by the choice of the specific spatial room impulse responses according to the configuration settings of the characterization processing scripts. “env\_1“, “env\_2“ represent conference rooms while “env\_4“ is another indoor environment such as cafeteria or mall. “env\_3” is an outdoor environment like park, nature, event or street.

**(2** Background is defined by the chosen background noise file according to the configuration settings of the characterization processing scripts. “env\_1\_cleanbg” and “env\_2\_cleanbg” indicate low-noise background noises that can be expected in conference rooms such as low air-conditioning/fan noise. “env\_3\_noisebg” and “env\_4\_noisebg” indicate background noises typical for their respective actual environment.

**(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** The talker positions are part of the scene definition of the different categories. They are defined in the Characterization Processing Plan and the corresponding processing scripts. They are 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 are made for the different listener panels.

**(5 ”**Error I1. Ox**”** means an error pattern derived from the Profile I1.Ox.

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

Table G.1.1: Conditions (BS1534-1 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 4.4* |
| *Listening Environment* | *No room noise* |

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

|  |  |  |
| --- | --- | --- |
| **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 BS1534-2: Stereo

Table G.2.1: Conditions (BS1534-2 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.2.2: Test conditions for Experiment BS1534-2

|  |  |  |
| --- | --- | --- |
| **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: FOA

Table G.3.1: Conditions (BS1534-3 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.3.2: Test conditions for Experiment BS1534-3

|  |  |  |
| --- | --- | --- |
| **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 BS1534-4: FOA

Table G.4.1: Conditions (BS1534-4 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.4.2: Test conditions for Experiment BS1534-4

|  |  |  |
| --- | --- | --- |
| **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 BS1534-5: HOA3

Table G.5.1: Conditions (BS1534-5 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* *4.4* |
| *Listening Environment* |  | *No room noise* |

Table G.5.2: 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: Multichannel 5.1

Table G.6.1: Conditions (BS1534-6 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* *4.4* |
| *Listening Environment* | *No room noise* |

Table G.6.2: Test conditions for Experiment BS1534-6

|  |  |  |
| --- | --- | --- |
| **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 BS1534-7: Multi-channel 5.1, 7.1

Table G.7.1: Conditions (BS1534-7 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.7.2: 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: Multi-channel 5.1+2, 5.1+4

Table G.8.1: Conditions (BS1534-8 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.8.2: 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 BS1534-9: Multi-channel 7.1+4

Table G.9.1: Conditions (BS1534-9 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.9.2: Test conditions for Experiment BS1534-9

|  |  |  |
| --- | --- | --- |
| **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 BS1534-10: ISM 1-2

Table G.10.1: Conditions (BS1534-10 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.10.2: Test conditions for Experiment BS1534-10

|  |  |  |
| --- | --- | --- |
| **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: ISM 3-4

Table G.11.1: Conditions (BS1534-11 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.11.2: Test conditions for Experiment BS1534-11

|  |  |  |
| --- | --- | --- |
| **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 BS1534-12: ISM 3-4

Table G.12.1: Conditions (BS1534-12 Generic Audio)

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under Test (CuT)* | *IVAS FX operated with ISM 3-4 audio input at*  *80 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.12.2: Test conditions for Experiment BS1534-12

|  |  |  |
| --- | --- | --- |
| **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: MASA (1TC)

Table G.13.1: Conditions (BS1534-13 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.13.2: Test conditions for Experiment BS1534-13

|  |  |  |
| --- | --- | --- |
| **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 BS1534-14: MASA (1TC)

Table G.14.1: Conditions (BS1534-14 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.14.2: Test conditions for Experiment BS1534-14

|  |  |  |
| --- | --- | --- |
| **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 BS1534-15: MASA (2TC)

Table G.15.1: Conditions (BS1534-15 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.15.2: Test conditions for Experiment BS1534-15

|  |  |  |
| --- | --- | --- |
| **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 BS1534-16: OSBA (1-4 obj.)

Table G.16.1: Conditions (BS1534-16 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.16.2: Test conditions for Experiment BS1534-16

|  |  |  |
| --- | --- | --- |
| **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 BS1534-17: OSBA (1-4 obj.)

Table G.17.1: Conditions (BS1534-17 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.17.2: Test conditions for Experiment BS1534-17

|  |  |  |
| --- | --- | --- |
| **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 BS1534-18: OMASA (1-4 obj.)

Table G.18.1: Conditions (BS1534-18 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.18.2: Test conditions for Experiment BS1534-18

|  |  |  |
| --- | --- | --- |
| **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 BS1534-19: Stereo downmix for EVS

Table G.19.1: Conditions (BS1534-19 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 BS1534-19

|  |  |  |
| --- | --- | --- |
| **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 BS1534-20: ISM 6 DoF (4 objects)

Table G.20.1: Conditions (BS1534-20 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 4.4* |
| *Listening Environment* | *No room noise* |

Table G.20.2: Test conditions for Experiment BS1534-20

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

1. P.800 ACR experiment
   1. Experiment ACR-1

Table H.1.1 Overview of test conditions

|  |  |  |
| --- | --- | --- |
| **Main Codec Conditions** |  |  |
| Codec under Test (CuT) | 8 | IVAS-WB spatial 13.2- 96 kbps single bitrates |
|  | 8 | IVAS-SWB spatial 13.2- 96 kbps single bitrates |
|  | 10 | IVAS-FB spatial 24.4- 256 kbps single bitrates |
|  | 12 | Conditions with DTX enabled |
|  | 12 | Conditions with 5% of frame errors |
| **Other references** |  |  |
| Directs | 8 | Varying bandwidth with spatial rendering: 4, 6, 8, 10, 12, 14, 16, and 20 kHz |
| P.50 MNRU (applied to MASA transport streams) | 4 | Q = 30, 26, 22 and 18 dB (all: nominal input level) |
| ESDRU [ITU-T P.811] | 4 | α = 0.8, 0.6, 0.4 and 0.2 (output loudness set to nominal level) |
| **Common Conditions** |  |  |
| Test item generation | 2 | 3 categories using model-based generation according to processing scripts. 3 categories using true recordings in Eigenmike, HOA3 or STEREO format. |
| Binaural rendering | 1 | IVAS codec internal binaural renderer and for references IVAS external renderer (IVAS\_rend). Diotic playback for mono conditions. Stereo conditions use stereo in and out. |
| Audio sampling frequency | 3 | 16 / 32 / 48 kHz for IVAS codec in WB / SWB / FB |
| Content types / categories | 6 | Scenes as described in Table H.1.2 |
| Number of talkers | 3 | In each sample there may be 2, 3, or 4 different talkers |
| Number of speech samples | 7 | 6 for tests + 1 for preliminaries per category |
| Input frequency mask | 1 | Flat |
| Nominal output loudness | 1 | -26 LKFS [6] measured from each listenable sample (post-scaling). |
| Listening Level | 1 | 73 dB SPL |
| Listeners | 30 | Naïve |
| Randomizations | 6 | 6 panels of 5 listeners |
| Rating Scale | 1 | 5- scale ACR with instructions according to P Suppl. 29, with some modifications needed for ACR testing methodology [2]. |
| Language | 1 | Finnish |
| Listening System | 1 | High-quality headphones (e.g. Sennheiser HD-650), diotic presentation |
| Listening Environment | 1 | No noise |

Table H.1.2 Sample Categories

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Category*** | Format | ***Overall Level [dB]*** | ***Background signal type*** | ***Background Level*** | ***Overtalk [s]*** | ***Talker positions*** |
| *cat 1* | *STEREO* | *-26* | *Mixed Music* | *Na* | *Na* | *Same as in P800-1 cat5* |
| *cat 2* | *FOA* | *-26* | *Clean speech* | *Na* | *No overtalk* | *Same as in P800-3 cat2* |
| *cat 3* | *HOA3* | *-26* | *Outdoors* | *15 db SNR* | *Overtalk* | *Same as in P800-5 cat3* |
| *cat 4* | *MASA 1TC* | *-26* | *Clean speech (real recording)* | *Over 30 dB* | *Overtalk* | *Frontal area 4 talkers (indoors)* |
| *cat 5* | *MASA 2TC* | *-26* | *Outdoors (real recording)* | *Na* | *No overtalk* | *2 talkers (outdoors)* |
| *cat 6* | *ISM2* | *-26* | *Clean speech* | *Na* | *Half* | *Same as in P800-9 cat3* |

Table H.1.3 Preliminaries for P.800 ACR experiment

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Category** | **Condition** | **Bitrate** | **FER*(3*** | **DTX** |
| 1 | C26 | Cat 1 | IVAS-SWB | 16.4 | no error | - |
| 2 | C21 | Cat 2 | IVAS-WB | 48 | no error | - |
| 3 | C14 | Cat 3 | ESDRU = 0.6 | - | - | - |
| 4 | C60 | Cat 4 | IVAS-FB | 64 | 5% | on |
| 5 | C07 | Cat 5 | Spatial MB (6 kHz) | - | - | - |
| 6 | C32 | Cat 6 | IVAS-SWB | 96 | no error | - |
| 7 | C10 | Cat 1 | MNRU Q = 26 dB | - | - | - |
| 8 | C01 | Cat 2 | Spatial FB | - | - | - |
| 9 | C50 | Cat 3 | IVAS-WB | 32 | 5% | - |
| 10 | C16 | Cat 4 | ESDRU = 0.2 | - | - | - |
| 11 | C05 | Cat 5 | Spatial (10 kHz) | - | - | - |
| 12 | C48 | Cat 6 | IVAS-FB | 32 | no error | on |
| 13 | C12 | Cat 1 | MNRU Q = 18 dB | - | - | - |
| 14 | C31 | Cat 2 | IVAS-SWB | 80 | no error | - |
| 15 | C43 | Cat 3 | IVAS-WB | 13.2 | no error | on |
| 16 | C08 | Cat 4 | Spatial NB (4 kHz) | - | - | - |
| 17 | C57 | Cat 5 | IVAS-SWB | 24.4 | 5% | on |
| 18 | C40 | Cat 6 | IVAS-FB | 160 | no error | - |

Table H.1.4 Test conditions for extended P.800 ACR experiment experiment

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Label*** | ***Condition*** | ***Bitrate MNRU ESDRU*** | ***DTX*** | ***FER*** | ***Notes*** | ***set*** |
| C01 | Spatial FB |  |  |  | Spatial reference signal is created with | 1 |
| C02 | Spatial SWB (16 kHz) |  |  |  | IVAS\_rend BINAURAL | 2 |
| C03 | Spatial (14 kHz) |  |  |  | Except STEREO is played back as is | 3 |
| C04 | Spatial SSWB (12 kHz) |  |  |  | Low pass filtered reference conditions | 4 |
| C05 | Spatial (10 kHz) |  |  |  | are created with high quality filters | 5 |
| C06 | Spatial WB (8 kHz) |  |  |  |  | 6 |
| C07 | Spatial MB (6 kHz) |  |  |  |  | 1 |
| C08 | Spatial NB (4 kHz) |  |  |  |  | 2 |
| C09 | MNRU | Q= 30 dB |  |  |  | 3 |
| C10 | MNRU | Q= 26 dB |  |  |  | 4 |
| C11 | MNRU | Q= 22 dB |  |  |  | 5 |
| C12 | MNRU | Q= 18 dB |  |  |  | 6 |
| C13 | ESDRU | α=0.8 |  |  |  | 1 |
| C14 | ESDRU | α=0.6 |  |  |  | 2 |
| C15 | ESDRU | α=0.4 |  |  |  | 3 |
| C16 | ESDRU | α=0.2 |  |  |  | 4 |
| C17 | IVAS-WB | 13.2 |  |  | ISM2 is coded using 16.4kbit/s | 5 |
| C18 | IVAS-WB | 16.4 |  |  | IVAS\_dec BINAURAL for IVAS rendering | 6 |
| C19 | IVAS-WB | 24.4 |  |  | Except STEREO is used for STEREO input | 1 |
| C20 | IVAS-WB | 32 |  |  |  | 2 |
| C21 | IVAS-WB | 48 |  |  |  | 3 |
| C22 | IVAS-WB | 64 |  |  |  | 4 |
| C23 | IVAS-WB | 80 |  |  |  | 5 |
| C24 | IVAS-WB | 96 |  |  |  | 6 |
| C25 | IVAS-SWB | 13.2 |  |  | ISM2 is coded using 16.4kbit/s | 1 |
| C26 | IVAS-SWB | 16.4 |  |  |  | 2 |
| C27 | IVAS-SWB | 24.4 |  |  |  | 3 |
| C28 | IVAS-SWB | 32 |  |  |  | 4 |
| C29 | IVAS-SWB | 48 |  |  |  | 5 |
| C30 | IVAS-SWB | 64 |  |  |  | 6 |
| C31 | IVAS-SWB | 80 |  |  |  | 1 |
| C32 | IVAS-SWB | 96 |  |  |  | 2 |
| C33 | IVAS-FB | 24.4 |  |  |  | 3 |
| C34 | IVAS-FB | 32 |  |  |  | 4 |
| C35 | IVAS-FB | 48 |  |  |  | 5 |
| C36 | IVAS-FB | 64 |  |  |  | 6 |
| C37 | IVAS-FB | 80 |  |  |  | 1 |
| C38 | IVAS-FB | 96 |  |  |  | 2 |
| C39 | IVAS-FB | 128 |  |  |  | 3 |
| C40 | IVAS-FB | 160 |  |  |  | 4 |
| C41 | IVAS-FB | 192 |  |  |  | 5 |
| C42 | IVAS-FB | 256 |  |  |  | 6 |
| C43 | IVAS-WB | 13.2 | on |  | ISM2 is coded using 16.4kbit/s | 1 |
| C44 | IVAS-WB | 16.4 | on |  |  | 2 |
| C45 | IVAS-SWB | 16.4 | on |  |  | 3 |
| C46 | IVAS-SWB | 24.4 | on |  |  | 4 |
| C47 | IVAS-FB | 24.4 | on |  |  | 5 |
| C48 | IVAS-FB | 32 | on |  |  | 6 |
| C49 | IVAS-WB | 24.4 |  | 5% |  | 1 |
| C50 | IVAS-WB | 32 |  | 5% |  | 2 |
| C51 | IVAS-SWB | 32 |  | 5% |  | 3 |
| C52 | IVAS-SWB | 48 |  | 5% |  | 4 |
| C53 | IVAS-FB | 64 |  | 5% |  | 5 |
| C54 | IVAS-FB | 80 |  | 5% |  | 6 |
| C55 | IVAS-WB | 16.4 | on | 5% |  | 1 |
| C56 | IVAS-WB | 24.4 | on | 5% |  | 2 |
| C57 | IVAS-SWB | 24.4 | on | 5% |  | 3 |
| C58 | IVAS-SWB | 32 | on | 5% |  | 4 |
| C59 | IVAS-FB | 48 | on | 5% |  | 5 |
| C60 | IVAS-FB | 64 | on | 5% |  | 6 |

1. Room acoustics testing experiments
   1. Experiment ROOM-1: Parametric binaural renderer

Table I.1.1 Conditions

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under test (CuT)* | *IVAS FX* |
| *Input format* | *FOA* |
| *Bitrate* | *80 kbps* |
| *DTX* | *DTX off* |
| *Input level* | *-26 LKFS* |
| *Input frequency mask* | *20KBP* |
| *Noise* | *No noise* |
| *Error Conditions* | *0%* |
| ***References*** |  |
| *Reference (anechoic)* | *IVAS decoding with anechoic HRIR rendering* |
| ***Other references*** |  |
| *Reference (Python)* | *Lightweight Python renderer with BRIR support* |
| ***Common Conditions*** |  |
| *Audio sampling frequency/bandwidth* | *48 kHz/FB* |
| *Kind of samples* | *Clean speech recordings:*   * *Single talker* * *Multiple talkers*   *Music recordings:*   * *Single instrument* * *Multiple instruments*   *All recordings should be made in a non-reverberant environment.* |
| *Number of categories* | *4 different room environments* |
| *Number of samples* | *3 + 1 samples per category* |
| *Listening Level* | *Adjusted by listener* |
| *Listeners* | *Experienced listeners* |
| *Randomizations* | *Individual per listener* |
| *Rating Scale* | *Unipolar preference scale following clause 4.2.3* |
| *Listening System* | *Headphones, VR setup in accordance with clause 4.4* |
| *Listening Environment* | *No room noise* |

Table I.1.2 Test conditions for Experiment ROOM-1

|  |  |  |
| --- | --- | --- |
| ***Label*** | ***Condition – output format*** | ***Bitrate [kbps]*** |
| *c01* | *Reference (anechoic)* | *80* |
| *c02* | *Reference (Python)* | *80* |
| *c03* | *IVAS – BINAURAL\_ROOM\_IR* | *80* |
| *c04* | *IVAS – BINAURAL\_ROOM\_REVERB* | *80* |

* 1. Experiment ROOM-2: FastConv Binaural Renderer

Table I.2.1 Conditions

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under test (CuT)* | *IVAS FX* |
| *Input format* | *FOA* |
| *Bitrate* | *96 kbps* |
| *DTX* | *DTX off* |
| *Input level* | *-26 LKFS* |
| *Input frequency mask* | *20KBP* |
| *Noise* | *No noise* |
| *Error Conditions* | *0%* |
| ***References*** |  |
| *Reference (anechoic)* | *IVAS decoding with anechoic HRIR rendering* |
| ***Other references*** |  |
| *Reference (Python)* | *Lightweight Python renderer with BRIR support* |
| ***Common Conditions*** |  |
| *Audio sampling frequency/bandwidth* | *48 kHz/FB* |
| *Kind of samples* | *Clean speech recordings:*   * *Single talker* * *Multiple talkers*   *Music recordings:*   * *Single instrument* * *Multiple instruments*   *All recordings should be made in a non-reverberant environment.* |
| *Number of categories* | *4 different room environments* |
| *Number of samples* | *3 + 1 samples per category* |
| *Listening Level* | *Adjusted by listener* |
| *Listeners* | *Experienced listeners* |
| *Randomizations* | *Individual per listener* |
| *Rating Scale* | *Unipolar preference scale following clause 4.2.3* |
| *Listening System* | *Headphones, VR setup in accordance with clause 4.4* |
| *Listening Environment* | *No room noise* |

Table I.2.2 Test conditions for Experiment ROOM-2

|  |  |  |
| --- | --- | --- |
| ***Label*** | ***Condition – output format*** | ***Bitrate [kbps]*** |
| *c01* | *Reference (anechoic)* | *96* |
| *c02* | *Reference (Python)* | *96* |
| *c03* | *IVAS – BINAURAL\_ROOM\_IR* | *96* |
| *c04* | *IVAS – BINAURAL\_ROOM\_REVERB* | *96* |

* 1. Experiment ROOM-3: Time-domain object renderer / Crend binaural renderer

Table I.3.1 Conditions

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under test (CuT)* | *IVAS FX* |
| *Input format* | *Multichannel 5.1* |
| *Bitrate* | *96 kbps* |
| *DTX* | *DTX off* |
| *Input level* | *-26 LKFS* |
| *Input frequency mask* | *20KBP* |
| *Noise* | *No noise* |
| *Error Conditions* | *0%* |
| ***References*** |  |
| *Reference (anechoic)* | *IVAS decoding with anechoic HRIR rendering* |
| ***Other references*** |  |
| *Reference (Python)* | *Lightweight Python renderer with BRIR support* |
| ***Common Conditions*** |  |
| *Audio sampling frequency/bandwidth* | *48 kHz/FB* |
| *Kind of samples* | *Clean speech recordings:*   * *Single talker* * *Multiple talkers*   *Music recordings:*   * *Single instrument* * *Multiple instruments*   *All recordings should be made in a non-reverberant environment.* |
| *Number of categories* | *4 different room environments* |
| *Number of samples* | *3 + 1 samples per category* |
| *Listening Level* | *Adjusted by listener* |
| *Listeners* | *Experienced listeners* |
| *Randomizations* | *Individual per listener* |
| *Rating Scale* | *Unipolar preference scale following clause 4.2.3* |
| *Listening System* | *Headphones, VR setup in accordance with clause 4.4* |
| *Listening Environment* | *No room noise* |

Table I.3.2 Test conditions for Experiment ROOM-3

|  |  |  |
| --- | --- | --- |
| ***Label*** | ***Condition – output format*** | ***Bitrate [kbps]*** |
| *c01* | *Reference (anechoic)* | *96* |
| *c02* | *Reference (Python)* | *96* |
| *c03* | *IVAS – BINAURAL\_ROOM\_IR* | *96* |
| *c04* | *IVAS – BINAURAL\_ROOM\_REVERB* | *96* |

* 1. Experiment ROOM-4: Crend binaural renderer

Table I.4.1 Conditions

|  |  |
| --- | --- |
| ***Main Codec Conditions*** |  |
| *Codec under test (CuT)* | *IVAS FX* |
| *Input format* | *Multichannel 5.1.2* |
| *Bitrate* | *128 kbps* |
| *DTX* | *DTX off* |
| *Input level* | *-26 LKFS* |
| *Input frequency mask* | *20KBP* |
| *Noise* | *No noise* |
| *Error Conditions* | *0%* |
| ***References*** |  |
| *Reference (anechoic)* | *IVAS decoding with anechoic HRIR rendering* |
| ***Other references*** |  |
| *Reference (Python)* | *Lightweight Python renderer with BRIR support* |
| ***References*** |  |
| *Audio sampling frequency/bandwidth* | *48 kHz/FB* |
| *Kind of samples* | *Clean speech recordings:*   * *Single talker* * *Multiple talkers*   *Music recordings:*   * *Single instrument* * *Multiple instruments*   *All recordings should be made in a non-reverberant environment.* |
| *Number of categories* | *4 different room environments* |
| *Number of samples* | *3 + 1 samples per category* |
| *Listening Level* | *Adjusted by listener* |
| *Listeners* | *Experienced listeners* |
| *Randomizations* | *Individual per listener* |
| *Rating Scale* | *Unipolar preference scale following clause 4.2.3* |
| *Listening System* | *Headphones, VR setup in accordance with clause 4.4* |
| *Listening Environment* | *No room noise* |

Table I.4.2 Test conditions for Experiment ROOM-4

|  |  |  |
| --- | --- | --- |
| ***Label*** | ***Condition – output format*** | ***Bitrate [kbps]*** |
| *c01* | *Reference (anechoic)* | *128* |
| *c02* | *Reference (Python)* | *128* |
| *c03* | *IVAS – BINAURAL\_ROOM\_IR* | *128* |
| *c04* | *IVAS – BINAURAL\_ROOM\_REVERB* | *128* |