**Source:** **Nokia**

**Title:** **ACR characterization of IVAS codec**

**Document for: Agreement**

**Agenda Item: 7.5 – IVAS\_Codec**

# 1. Introduction

IVAS Permanent Document IVAS-8b - Test Plan for Characterization Phase [1], discuss specifics of IVAS Codec characterization testing.

In this document the source proposes test material and conditions for ACR (Absolute Category Rating) listening experiment covering several IVAS spatial operation modes (FOA, HOA, ISM2, MASA and STEREO). ACR test methodology allows a significant number of conditions to be tested at the same time and enables comparison between different signal bandwidths and bitrates. The proposed ACR test also covers several IVAS operation mode and bitrate combinations, otherwise not tested at all.

# 2. Overview

In the current work in progress IVAS Permanent Document IVAS-8b - Test Plan for Characterization Phase [1], at least one ACR test has been reserved. The source proposes specific test material and conditions that can be used to characterize IVAS codec with ACR test methodology.

# 3. Discussion

Some observations and considerations are made relating, e.g., to the size of an experiment. According to ITU-T P.800 – Use cases, P Suppl. 29 DCR experiments can have 36 test conditions [2]. ACR test in principle can have double the amount of conditions, i.e. 72 conditions, for the same amount of listenable audio, since there is no reference sample. However, for the listener there is still double the number of votes to be casted, causing significant mental load. Thus, the number of conditions should be somewhat lower than 72 conditions. In this document the proposal is to have 60 ACR conditions.

* Currently one ACR listening test is reserved for IVAS characterization.
* It is proposed to test all conditions with either IVAS\_dec internal binaural renderer or with IVAS\_rend functionality for Direct references.
* Fixed point version of the IVAS codec (encoder and decoder) and IVAS\_rend should be used, if available.
* It is proposed to test all IVAS operation modes up to the highest available bitrate or up to a bitrate, where quality saturation is expected.
* Some FER and DTX conditions are tested
* bitaFER and DTX operations should be tested with the expected most common operation points.
* No floating point to fixed point interoperation is tested.
* WB and SWB IVAS processing should be done with respective sampling rates 16 and 32, instead of relying on -max\_band functionality. High quality external up- and down-sampling is needed for processing.
* Different IVAS supported spatial audio formats are put to separate signal categories. The proposed IVAS spatial signal categories are:
  + STEREO
  + FOA
  + HOA3
  + MASA 1TC
  + MASA 2TC
  + ISM2
* All proposed signal categories support DTX, and frame error concealment
* All IVAS formats except STEREO are rendered with IVAS\_dec or IVAS\_rend BINAURAL renderer
* STEREO and ISM2 support at maximum bitrate of 256kbit/s
* ISM2 does not support a bitrate of 13.2 kbit/s
* If higher than 256 kbit/s or 13.2 kbit/s are tested the nearest supported bitrate for the particular IVAS coding mode is used instead. This duplicates some conditions for that specific input signal category.
* Test samples
  + There are six sample categories. It is proposed to reuse suitable signal categories from the IVAS P.800 DCR test signals.
  + Sample categories should contain clean speech condition
  + Sample categories should contain speech with background noise.
  + Some sample categories may consist of “mixed content” e.g. have also music and speech or sound effects. e.g. for STEREO and ISM2.
  + In addition to the script generated test material, some real recorded speech material can be used for the listening evaluation, if available.
  + The length of the test samples should be consistent, around 6- 8 seconds. Too short or too varying length samples may cause surprises to the listeners. I.e. listeners are not mentally ready for casting a vote and playback has already ended.

# 4. Proposal for specific test conditions for the ACR test

It is proposed to have following conditions for the IVAS characterization listening test using ACR voting scale.

Table 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 FER5 |
| **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 4 |
| 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 ([ITU-R BS.1770-4]) 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. [3] |
| Languages | [2] | Finnish / [tbd another test conducted with other language is seen beneficial] |
| Listening System | 1 | High-quality headphones (e.g. Sennheiser HD-650), diotic presentation |
| Listening Environment | 1 | No noise |

Table 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 3 Test conditions for extended ACR5 IVAS characterization

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***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 | DTX |  | ISM2 is coded using 16.4kbit/s | 1 |
| C44 | IVAS-WB | 16.4 | DTX |  |  | 2 |
| C45 | IVAS-SWB | 16.4 | DTX |  |  | 3 |
| C46 | IVAS-SWB | 24.4 | DTX |  |  | 4 |
| C47 | IVAS-FB | 24.4 | DTX |  |  | 5 |
| C48 | IVAS-FB | 32 | DTX |  |  | 6 |
| C49 | IVAS-WB | 24.4 |  | FER5 |  | 1 |
| C50 | IVAS-WB | 32 |  | FER5 |  | 2 |
| C51 | IVAS-SWB | 32 |  | FER5 |  | 3 |
| C52 | IVAS-SWB | 48 |  | FER5 |  | 4 |
| C53 | IVAS-FB | 64 |  | FER5 |  | 5 |
| C54 | IVAS-FB | 80 |  | FER5 |  | 6 |
| C55 | IVAS-WB | 16.4 | DTX | FER5 |  | 1 |
| C56 | IVAS-WB | 24.4 | DTX | FER5 |  | 2 |
| C57 | IVAS-SWB | 24.4 | DTX | FER5 |  | 3 |
| C58 | IVAS-SWB | 32 | DTX | FER5 |  | 4 |
| C59 | IVAS-FB | 48 | DTX | FER5 |  | 5 |
| C60 | IVAS-FB | 64 | DTX | FER5 |  | 6 |

# 5. Summary

We have presented some IVAS characterization test plan specifics and options for agreement in clause 4 this input. The specific proposal covers P.800 ACR characterization spatial experiment for IVAS operation with varying signal bandwidth.

# References

[1] S4-250761 “IVAS Permanent Document IVAS-8b - Test Plan for Characterization Phase”, Version v.0.6.1

[2] ITU-T P.800 – Use cases, P Suppl. 29 (01/2023)