**Source: Dolby Laboratories Inc.**

**Title: On IVAS codec performance requirements for FOA audio content**

**Document for: Discussion**

**Agenda Item: 7.5**

# Introduction

Pdoc IVAS-3 [1] will specify the IVAS codec performance requirements. Currently, it contains only suggested and not yet agreed requirements for IVAS stereo operation. The basic principle is that IVAS stereo operation shall be not worse than EVS dual-mono operation at (at least) the next higher bit rate or better than EVS dual-mono operation at the same rate.

The present contribution is a proposal to adopt analogous performance requirements for FOA input audio content.

# Discussion

For FOA audio content, the source provided previously a P.800 test design (included in the Appendix of Pdoc IVAS-6 [2]) and the corresponding test results [3] for multi-mono EVS coding operation. For binaural headphone rendering, it was demonstrated that EVS multi-mono operated on the 4 FOA B-format component signals in Ambix format leads to strong performance close to indistinguishable from the uncoded and directly rendered reference. The experiment was done with speech content, partly with challenging overtalk of concurrent talkers, partly with substantial background noise. Based on the test results, it is clear that multi-mono EVS represents a meaningful and strong reference system for the coding of FOA audio content.

[

The general requirement is that IVAS operated at rate X shall either

* be better than the EVS multi-mono system, where each EVS instance is operated at the closest bit rate to X/4.
* or be no worse than the EVS multi-mono system, where each EVS instance is operated at the next higher available EVS bit rate than the EVS bitrate closest to X/4,

The following table illustrates corresponding detailed performance requirements for FOA audio content (\*\*\*:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Category** | BW | Bitrate (kbit/s) | FER | DTX(\* | Requirements(\*\* |
| **TBD** | WB | 24.4 | All | On/Off | **NWT EVS @ 4x 7.2 kbps (WB) OR****BT EVS @ 4x 5.9 kbps (WB)** |
| 32 | **NWT EVS @ 4x 9.6 kbps (WB) OR****BT EVS @ 4x 8 kbps (WB)** |
| 48 | **NWT EVS @ 4x 16.4 kbps (WB) OR****BT EVS @ 4x 13.2 kbps (WB)** |
| 64 | **NWT EVS @ 4x 24.4 kbps (WB) OR****BT EVS @ 4x 16.4 kbps (WB)** |
| 80 | **NWT EVS @ 4x 24.4 kbps (WB) OR****BT EVS @ 4x 16.4 kbps (WB)** |
| 96 | **NWT EVS @ 4x 32 kbps (WB) OR****BT EVS @ 4x 24.4 kbps (WB)** |
| 128 | **NWT EVS @ 4x 48 kbps (WB) OR****BT EVS @ 4x 32 kbps (WB)** |
| 160 | **NWT EVS @ 4x 48 kbps (WB) OR****BT EVS @ 4x 32 kbps (WB)** |
| 192 | **NWT EVS @ 4x 64 kbps (WB) OR****BT EVS @ 4x 48 kbps (WB)** |
| 256 | **NWT EVS @ 4x 96 kbps (WB) OR****BT EVS @ 4x 64 kbps (WB)** |
| SWBFB | 24.4 | All | On/Off | **NWT EVS @ 4x 7.2 kbps (WB) OR****BT EVS @ 4x 5.9 kbps (WB)** |
| 32 | **NWT EVS @ 4x 9.6 kbps (SWB) OR****BT EVS @ 4x 8 kbps (WB)** |
| 48 | **NWT EVS @ 4x 16.4 kbps (SWB/FB) OR****BT EVS @ 4x 13.2 kbps (SWB)** |
| 64 | **NWT EVS @ 4x 24.4 kbps (SWB/FB) OR****BT EVS @ 4x 16.4 kbps (SWB/FB)** |
| 80 | **NWT EVS @ 4x 24.4 kbps (SWB/FB) OR****BT EVS @ 4x 16.4 kbps (SWB/FB)** |
| 96 | **NWT EVS @ 4x 32 kbps (SWB/FB) OR****BT EVS @ 4x 24.4 kbps (SWB/FB)** |
| 128 | **NWT EVS @ 4x 48 kbps (SWB/FB) OR****BT EVS @ 4x 32 kbps (SWB/FB)** |
| 160 | **NWT EVS @ 4x 48 kbps (SWB/FB) OR****BT EVS @ 4x 32 kbps (SWB/FB)** |
| 192 | **NWT EVS @ 4x 64 kbps (SWB/FB) OR****BT EVS @ 4x 48 kbps (SWB/FB)** |
| 256 | **NWT EVS @ 4x 96 kbps (SWB/FB) OR****BT EVS @ 4x 64 kbps (SWB/FB)** |

(\* DTX will be tested for the rates where DTX must be supported. DTX operation applies also for the multi-mono EVS reference.

(\*\* The multi-mono EVS reference shall be produced by individual EVS coding of the 4 FOA B-format component signals in ACN/SN3D format.

(\*\*\* Editor’s note: The requirements apply for binaural rendering over headphones [and further rendering TBD] using a suitable reference renderer. Such a renderer has for instance been proposed in [4].

Editor’s note: The following should still be for discussion:

* For certain bit rates, the requirement appears very poor. This is for example the case for 24.4 kbps for which the requirement would use multi-mono EVS 4x 5.9 kbps (VBR).
* For certain bit rates, it is difficult to define a different requirement than for the next lowest bit rate. This is for example the case for the requirements at 80 and 160 kbps.
* The content categories for which the requirements should apply is still left TBD. It is for discussion whether to apply the requirements for any speech-based content as well as for music and general audio content.

]

# Conclusion and proposal

The present contribution provides a first proposal for performance requirements for the IVAS codec operated on FOA audio content. It is proposed to include the table in square brackets in Pdoc IVAS-3. The points identified for further discussion should be added as an editor’s note.

# References

[1] Pdoc IVAS-3: IVAS Performance Requirements, v0.0.4

[2] Pdoc IVAS-6: IVAS Selection Deliverables, v0.1.0

[3] Tdoc S4-210836: On reference designs for IVAS codec tests, Dolby Laboratories, Inc.

[4] Tdoc S4-200158: A Reference Audio Renderer for Qualification, Dolby Laboratories, Inc.