**Source: Editor[[1]](#footnote-1)**

**Title: IVAS Performance Requirements (IVAS-3)**

**Version: 0.2.0**

**Agenda Item: 15.1**

1. **Scope**

This document presents the Performance Requirements of the EVS Codec Extension for Immersive Voice and Audio Services (IVAS). Additional information on the codec development project can be found in the other IVAS permanent documents, for which the latest versions can be found at: <https://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/IVAS_Permanent_Documents>.

1. **Introduction**

The overall objective of the IVAS\_Codec work item is to develop a single general-purpose audio codec for immersive 4G and 5G services and applications. Further details on the objectives can be found in the work item description (SP-220608) .

**3. Reference Codecs**

It is proposed to consider the following legacy codecs for IVAS performance definition and evaluation.

Table 1. List of potential reference codecs

|  |  |  |
| --- | --- | --- |
| **Category** | **Delay** | **Legacy codecs** |
| Mono codecs | Low-delay codecs | AMR-WB, EVS, [G.711, G.719, G.722, G.722.1, G.722.1C, G.722 App. V, G.711.1 App. IV] |
| Higher-delay codecs | [AMR-WB+, e-AAC+] |
| Joint stereo/immersive codecs | Low-delay codecs | [G.722D, G.711.1F] |
| Higher-delay codecs | [AMR-WB+, e-AAC+, MPEG-H 3DA\*, VRStream spAACe\*]  \*Note: pending availability of codec |

# [Editor’s Note: Details on the exact usage of the potential reference codecs (e.g. unmodified or variants on mode synchronization, matrixing, bit allocation) and related processing blocks are to be defined in the processing plan(s)]

1. **Features**

Some of the key codec capabilities considered in developing the IVAS codec performance requirements are listed below.

Table 2. Capabilities of the IVAS codec

|  |  |
| --- | --- |
| **List of parameters** | **Details** |
| Content | Speech (clean, noisy), Mixed/content/Music & Generic Sound |
| Content Bandwidth | WB, SWB and FB |
| Noise type | First proposal could be e.g., Car, street, office; @ 15-20 dB [TBD] |
| FER | [VoLTE profiles (1-10)], [TBD] |
| DTX | Included [TBD] |
| Rate switching | Supported [TBD] |
| [More parameters TBD] |  |

1. **Nomenclature**

In this document, the following conventions apply:

* Notation “Req1 AND Req2” (e.g. NWT EVS @24.4 AND NWT G.719 @48) means that both requirements “Req1” and “Req2” shall be met
* Notation “Req1 OR Req2” (e.g. NWT EVS @24.4 OR BT G.719 @48) means that either requirement “Req1” or requirement “Req2” shall be met

1. **IVAS Performance Requirements**

[Editor’s Note FFS: Tdoc S4-171221 proposes to add high-level performance requirements for IVAS codec modes suitable for a spatial conferencing use-case and if agreed, corresponding updates to the performance requirements would be derived.]

## [

Editor’s note: The following section on performance requirements for stereo operation is a direct adaptation from EVS performance requirements (document EVS-3). These performance requirements do not reflect any expectation of the IVAS stereo performance. Inputs are invited for relevant IVAS stereo performance requirements.

## 6.1 Stereo operation

IVAS stereo operation shall provide a benefit over the case when it is not included, i.e. mono operation.

**High-level definition of stereo requirements**

* For certain correlated stereo content *(e.g.* speech in conferencing use case, mixed/music, binaural signals*)*, IVAS stereo operation shall be not worse than EVS dual-mono operation at (at least) next higher bit rate or better than EVS dual-mono operation at the same rate. Table 1~~3~~ describes the details of these requirements.
* No other content shall be degraded with respect to dual mono operation
* The IVAS stereo operation with mono compatible decoding shall be not worse than the stereo to mono downmix of dual-mono at the same bit rate.

For stereo operation, the IVAS codec will be compared for information to the 3GPP audio codecs in characterization, taking into account complexity, delay, etc.

**Table 1: Stereo requirements.**

Dual mono is abbreviated with a “2x”. It is assumed that the IVAS codec and the EVS reference codec are running in the bandwidth mode as outlined in the BW column, if not otherwise specified.

Note: for all bit rates / audio bandwidth with no requirement, the following applies: If provided it should be characterized; characterization results shall be used for the decision of inclusion

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Category** | BW | Bitrate (kbit/s) | FER | DTX\* | Stereo Requirements | Mono compatible requirements |
| **Certain correlated content** | NB | 7.2 | All |  |  | |
| 8 |
| 9.6 |
| 13.2 |
| WB | 7.2 | All | On/Off |  |  |
| 8 |  |  |
| 9.6 |  |  |
| 13.2 | **NWT EVS @ 2x 8 kbps OR**  **BT EVS @ 2x 7.2 kbps** |  |
| 16.4 | **NWT EVS @ 2x 9.6 kbps OR**  **BT EVS @ 2x 8 kbps** | **NWT stereo downmix of EVS @ 2x 8 kbps** |
| 24.4 | **NWT EVS @ 2x 16.4 kbps OR**  **BT EVS @ 2x 13.2 kbps** | **NWT stereo downmix of EVS @ 2x 13.2 kbps** |
| 32 | **NWT EVS @ 2x 24.4 kbps OR**  **BT EVS @ 2x 16.4 kbps** | **NWT stereo downmix of EVS @ 2x 16.4 kbps** |
| 48 | **NWT EVS @ 2x 32 kbps OR**  **BT EVS @ 2x 24.4 kbps** | **NWT stereo downmix of EVS @ 2x 24.4 kbps** |
| 64 | **NWT EVS @ 2x 48 kbps OR**  **BT EVS @ 2x 32 kbps** | **NWT stereo downmix of EVS @ 2x 32 kbps** |
| 96 | **NWT EVS @ 2x 64 kbps OR**  **BT EVS @ 2x 48 kbps** | **NWT stereo downmix of EVS @ 2x 48 kbps** |
| SWB  FB | 13.2 | All | On/Off |  |  |
| 16.4 | **NWT EVS-WB @ 2x 9.6 kbps OR**  **BT EVS-WB @ 2x 8 kbps** | **NWT stereo downmix of EVS @ 2x 8 kbps** |
| 24.4 | **NWT EVS @ 2x 16.4 kbps OR**  **BT EVS @ 2x 13.2 kbps** | **NWT stereo downmix of EVS @ 2x 13.2 kbps** |
| 32 | **NWT EVS @ 2x 24.4 kbps OR**  **BT EVS @ 2x 16.4 kbps** | **NWT stereo downmix of EVS @ 2x 16.4 kbps** |
| 48 | **NWT EVS @ 2x 32 kbps OR**  **BT EVS @ 2x 24.4 kbps** | **NWT stereo downmix of EVS @ 2x 24.4 kbps** |
| 64 | **NWT EVS @ 2x 48 kbps OR**  **BT EVS @ 2x 32 kbps** | **NWT stereo downmix of EVS @ 2x 32 kbps** |
| 96 | **NWT EVS @ 2x 64 kbps OR**  **BT EVS @ 2x 48 kbps** | **NWT stereo downmix of EVS @ 2x 48 kbps** |
| 128 | **NWT EVS @ 2x 96 kbps OR**  **BT EVS @ 2x 64 kbps** | **NWT stereo downmix of EVS @ 2x 64 kbps** |

\* DTX will be tested for stereo if provided.

]

[

## 6.2 Performance requirements for operation with Scene-Based Audio Content

6.2.1 First Order Ambisonics (FOA)

**High-level definition of FOA requirements**

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.

**Detailed FOA requirements**

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)** |
| SWB  FB | 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.

]

## 6.3 Performance requirements for operation with MASA Content

[

6.3.1 For Stereo-MASA:

| **Category** | BW | Bitrate (kbit/s) | FER | DTX(\* | Requirements(\*\* |
| --- | --- | --- | --- | --- | --- |
|  | SWB FB | 13.2 | All | On/Off | **NWT EVS @ 2x 7.2 kbps (WB) + UQ metadata OR**  **NWT EVS @ 3x 5.9 kbps (WB)** |
| 16.4 | **NWT EVS @ 2x 7.2 kbps (WB) + UQ metadata OR**  **NWT EVS @ 3x 7.2 kbps (WB)** |
| 24.4 | **NWT EVS @ 2x 8 kbps (WB) + UQ metadata OR**  **NWT EVS @ 4x 7.2 kbps (WB)** |
| 32 | **NWT EVS @ 2x 9.6 kbps (SWB) + UQ metadata OR**  **NWT EVS @ 4x 9.6 kbps (SWB)** |
| 48 | **NWT EVS @ 2x 16.4 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 16.4 kbps (SWB/FB)** |
| 64 | **NWT EVS @ 2x 24.4 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 24.4 kbps (SWB/FB)** |
| 80 | **NWT EVS @ 2x 24.4 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 24.4 kbps (SWB/FB)** |
| 96 | **NWT EVS @ 2x 32 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 32 kbps (SWB/FB)** |
| 128 | **NWT EVS @ 2x 48 kbps (WB/SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 48 kbps (SWB/FB)** |
| 160 | **NWT EVS @ 2x 48 kbps (WB/SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 48 kbps (SWB/FB)** |
| 192 | **NWT EVS @ 2x 64 kbps (WB/SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 64 kbps (SWB/FB)** |
| 256 | **NWT EVS @ 2x 96 kbps (WB/SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 96 kbps (SWB/FB)** |

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

(\*\* The multi-mono 2x EVS + UQ metadata reference shall be produced by individual EVS coding of the stereo MASA format transport channels and passthrough of the original MASA metadata. The multi-mono 3x EVS reference shall be produced by individual EVS coding of the 3 “planar FOA” component signals in ACN/SN3D format. The multi-mono 4x EVS reference shall be produced by individual EVS coding of the 4 FOA component signals in ACN/SN3D format.

6.3.2 For Mono-MASA:

| **Category** | BW | Bitrate (kbit/s) | FER | DTX(\* | Requirements(\*\* |
| --- | --- | --- | --- | --- | --- |
|  | SWB FB | 13.2 | All | On/Off | **NWT EVS @ 8 kbps (WB) + UQ metadata OR**  **NWT EVS @ 3x 5.9 kbps (WB)** |
| 16.4 | **NWT EVS @ 9.6 kbps (SWB) + UQ metadata OR**  **NWT EVS @ 3x 7.2 kbps (WB)** |
| 24.4 | **NWT EVS @ 13.2 kbps (SWB) + UQ metadata OR**  **NWT EVS @ 4x 7.2 kbps (WB)** |
| 32 | **NWT EVS @ 16.4 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 9.6 kbps (SWB)** |
| 48 | **NWT EVS @ 32 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 16.4 kbps (SWB/FB)** |
| 64 | **NWT EVS @ 48 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 24.4 kbps (SWB/FB)** |
| 80 | **NWT EVS @ 48 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 24.4 kbps (SWB/FB)** |
| 96 | **NWT EVS @ 64 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 32 kbps (SWB/FB)** |
| 128 | **NWT EVS @ 64 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 48 kbps (SWB/FB)** |
| 160 | **NWT EVS @ 96 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 48 kbps (SWB/FB)** |
| 192 | **NWT EVS @ 128 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 64 kbps (SWB/FB)** |
| 256 | **NWT EVS @ 128 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 96 kbps (SWB/FB)** |

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

(\*\* The EVS + UQ metadata reference shall be produced by EVS coding of the mono MASA format transport channel and passthrough of the original MASA metadata. The multi-mono 3x EVS reference shall be produced by individual EVS coding of the 3 “planar FOA” component signals in ACN/SN3D format. The multi-mono 4x EVS reference shall be produced by individual EVS coding of the 4 FOA component signals in ACN/SN3D format.

]

# **X. Revision history**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Meeting** | **Subject/Comment** | **Old** | **New** |
| 2019-08-16 | SA4#105 | Agreement of Initial Skeleton of Performance Requirements (IVAS-3) including draft stereo requirements | N/A | 0.0.1…  0.0.4 |
| 2021-11-16 | SA4#116-e | Inclusion of draft requirements for FOA | 0.0.4 | 0.1.0 |
| 2022-08-22 | SA4#120-e | Updates of Scope and Introduction Inclusion of draft requirements for MASA | 0.1.0 | 0.2.0 |

1. Stefan Bruhn, Dolby Laboratories Inc. [↑](#footnote-ref-1)