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

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

**Version: 0.4.0**

**Agenda Item: 14.2**

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).

It is to be noted that it may not be possible to test all requirements of this Pdoc during IVAS selection. The choice of requirements to be tested is made in the IVAS selection test plan IVAS-8a.

1. **Nomenclature**

In this document, the following conventions apply:

* Notation “Req1 OR Req2” (e.g. NWT 4xEVS @24.4 OR BT 4xEVS @16.4) means that either requirement “Req1” or requirement “Req2” shall be met.

1. **IVAS Performance Requirements**

## 4.0 General remarks

## Empty cells in tables related to requirements or objectives indicate that these are not defined for the corresponding IVAS codec operation point.

## The definition of a requirement for an IVAS codec operation point does not necessarily imply that it will be tested in IVAS codec selection. It is rather specified in the IVAS test plan in what phase of the IVAS standardization an IVAS operation point is tested and how this is done.

## The specification of how to apply frame loss and delay loss profiles is subject to the IVAS test and processing plans.

## 4.1 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 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.

**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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| BW | Bitrate (kbit/s) | FER/ Delay Loss Profile | DTX(\* | Stereo Requirements | Mono compatible requirements |
| WB | 13.2 | All | On/Off | **NWT EVS @ 2x 8 kbps OR**  **BT EVS @ 2x 7.2 kbps** | **NWT stereo downmix of 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 13.2 kbps OR**  **BT EVS @ 2x 9.6 kbps** | **NWT stereo downmix of EVS @ 2x 13.2 kbps** |
| 32 | **NWT EVS @ 2x 16.4 kbps OR**  **BT EVS @ 2x 13.2 kbps** | **NWT stereo downmix of EVS @ 2x 16.4 kbps** |
| 48 | Off | **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 | **NWT EVS-WB @ 2x 8 kbps OR**  **BT EVS-WB @ 2x 7.2 kbps** | **NWT stereo downmix of EVS @ 2x 7.2 kbps** |
| 16.4 | **NWT EVS @ 2x 9.6 kbps OR**  **BT EVS-WB @ 2x 8 kbps** | **NWT stereo downmix of EVS @ 2x 8 kbps** |
| 24.4 | **NWT EVS @ 2x 13.2 kbps OR**  **BT EVS @ 2x 9.6 kbps** | **NWT stereo downmix of EVS @ 2x 13.2 kbps** |
| 32 | **NWT EVS @ 2x 16.4 kbps OR**  **BT EVS @ 2x 13.2 kbps** | **NWT stereo downmix of EVS @ 2x 16.4 kbps** |
| 48 | Off | **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 at rates up to 32 kbps where mandatory DTX operation is available for the multi-mono EVS reference. Other rates with DTX support may be evaluated in IVAS codec characterization. DTX operation applies also for the (multi-mono) EVS references.

* To check Mono compatible requirements, two signals should be compared;
  + CuT: Decoded mono signal from EVS compatible stream generated by stereo downmix for EVS with stereo input.
  + Reference: Downmixed signal of decoded signals from two EVS encoded streams for each channel.

## 4.2 Performance requirements for operation with Scene-Based Audio Content (first-order and higher-order Ambisonics)

The following scene-based audio requirements apply predominantly to speech, ambient sound, and music content. Both natural and synthetic contents is relevant.

**High-level definition of scene-based audio requirements**

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

* be better than the EVS multi-mono system encoding the FOA component channels (while truncating the higher-order channels), where each EVS instance is operated at the closest bit rate to X/4.
* or be no worse than the EVS multi-mono system encoding the FOA component channels (while truncating the higher-order channels), where each EVS instance is operated at the next higher available EVS bit rate than the EVS bitrate closest to X/4.

**Detailed scene-based audio requirements**

The following table illustrates corresponding detailed performance requirements for scene-based audio content (\*\*\*:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| BW | Bitrate (kbit/s) | FER/ Delay Loss Profile | DTX(\* | Requirements(\*\* | Objectives |
| SWB  FB | 13.2 | All | On/Off |  | **NWT EVS @ 3x 7.2 kbps (WB) (planar FOA)** |
| 16.4 | **NWT EVS @ 3x 7.2 kbps (WB) (planar FOA)** | **NWT EVS @ 4x 7.2 kbps (WB)** |
| 24.4 | **NWT EVS @ 4x 7.2 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 | Off | **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)** |  |
| 384 | **NWT EVS @ 4x 128 kbps (SWB/FB) OR**  **BT EVS @ 4x 96 kbps (SWB/FB)** |  |
| 512 | **NWT EVS @ 4x 128 kbps (SWB/FB)** |  |

(\* DTX will be tested for rates up to 80 kbit/s where mandatory DTX operation is available for the multi-mono EVS reference. Other rates with DTX support may be evaluated in IVAS codec characterization. DTX operation applies also for the (multi-mono) EVS references.

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

Further notes:

Note: No performance requirements are defined for WB content.

The FOA requirement will be tested with predominant voice content in naïve listener tests. The HOA3 requirement will be tested with general audio content in experienced listener tests. The HOA2 requirement will not be tested during selection.

The requirements will be tested with binaural rendering over headphones using a suitable reference renderer that will be specified in the selection processing plan IVAS-7a. The requirements may also be tested with rendering over a room loudspeaker system in experienced listener tests with general audio content.

Editor’s note: The SWB requirement will be tested with predominant voice content in naïve listener tests. The FB requirement will be tested with general audio content in experienced listener tests.

## 4.3 Performance requirements for operation with MASA Content

The following MASA audio requirements apply predominantly to speech, ambient sound, and music content.

4.3.1 For Stereo-MASA:

| BW | Bitrate (kbit/s) | FER/ Delay Loss Profile | DTX(\* | Requirements(\*\* |
| --- | --- | --- | --- | --- |
| SWB FB | 13.2 | All | On/Off | **NWT EVS @ 2x 7.2 kbps (WB) + UQ metadataOR**  **NWT EVS @ 3x 7.2 kbps (WB) (planar FOA)** |
| 16.4 | **NWT EVS @ 2x 7.2 kbps (WB) + UQ metadata OR**  **NWT EVS @ 4x 7.2 kbps (WB)** |
| 24.4 | **NWT EVS @ 2x 8 kbps (WB) + UQ metadata OR**  **NWT EVS @ 4x 8.0 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 | Off | **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 rates up to 80 kbit/s where mandatory DTX operation is available for the multi-mono EVS reference. Other rates with DTX support may be evaluated in IVAS codec characterization. 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:

| BW | Bitrate (kbit/s) | FER/ Delay Loss Profile | DTX(\* | Requirements(\*\* |
| --- | --- | --- | --- | --- |
| SWB FB | 13.2 | All | On/Off | **NWT EVS @ 8 kbps (WB) + UQ metadataOR**  **NWT EVS @ 3x 7.2 kbps (WB) (planar FOA)** |
| 16.4 | **NWT EVS @ 9.6 kbps (SWB) + UQ metadata OR**  **NWT EVS @ 4x 7.2 kbps (WB)** |
| 24.4 | **NWT EVS @ 13.2 kbps (SWB) + UQ metadata OR**  **NWT EVS @ 4x 8 kbps (WB)** |
| 32 | **NWT EVS @ 16.4 kbps (SWB/FB) + UQ metadata OR**  **NWT EVS @ 4x 9.6 kbps (SWB)** |
| 48 | Off | **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 rates up to 32 kbit/s where mandatory DTX operation is available for the multi-mono EVS reference. Other rates with DTX support may be evaluated in IVAS codec characterization. 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.

## 6.4 Performance requirements for operation with Multi-channel Audio Content

The following multi-channel audio requirements apply predominantly to cinematic, mixed, and music content.

**6.4.1 Multi-channel 5.1:**

**High-level definition of Multi-Channel 5.1 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/5.
* 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/5.
* For the EVS multi-mono system, the LFE is considered to be encoded with EVS 9.6 kbit/s NB.

**Detailed Multi-Channel 5.1 requirements**

The following table illustrates corresponding detailed performance requirements for Multi-Channel 5.1 audio content:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BW | Bitrate (kbit/s) | FER/ Delay Loss Profile | Requirements (\* | Objectives (\* |
| SWB FB | 13.2 | All |  | **NWT EVS @ 5x 7.2 kbps (WB)** |
| 16.4 |  | **NWT EVS @ 5x 7.2 kbps (WB)** |
| 24.4 | **NWT EVS @ 5x 7.2 kbps (WB)** |  |
| 32 | **NWT EVS @ 5x 8.0 kbps (WB) OR**  **BT EVS @ 5x 7.2 kbps (WB)** |  |
| 48 | **NWT EVS @ 5x 13.2 kbps (SWB) OR**  **BT EVS @ 5x 9.6 kbps (SWB)** |  |
| 64 | **NWT EVS @ 5x 16.4 kbps (SWB) OR**  **BT EVS @ 5x 13.2 kbps (SWB)** |  |
| 80 | **NWT EVS @ 5x 24.4 kbps (SWB/FB) OR**  **BT EVS @ 5x 16.4 kbps (SWB)** |  |
| 96 | **NWT EVS @ 5x 24.4 kbps (SWB/FB) OR**  **BT EVS @ 5x 16.4 kbps (SWB)** |  |
| 128 | **NWT EVS @ 5x 32 kbps (SWB/FB) OR**  **BT EVS @ 5x 24.4 kbps (SWB/FB)** |  |
| 160 | **NWT EVS @ 5x 48 kbps (SWB/FB) OR**  **BT EVS @ 5x 32 kbps (SWB/FB)** |  |
| 192 | **NWT EVS @ 5x 48 kbps (SWB/FB) OR**  **BT EVS @ 5x 32 kbps (SWB/FB)** |  |
| 256 | **NWT EVS @ 5x 64 kbps (SWB/FB) OR**  **BT EVS @ 5x 48 kbps (SWB/FB)** |  |
| 384 | **NWT EVS @ 5x 96 kbps (SWB/FB) OR**  **BT EVS @ 5x 64 kbps (SWB/FB)** |  |
| 512 | **NWT EVS @ 5x 128 kbps (SWB/FB) OR**  **BT EVS @ 5x 96 kbps (SWB/FB)** |  |

(\* The multi-mono EVS reference shall be produced by individual EVS coding of the 5 channels, and the LFE channel shall be encoded with EVS 9.6 kbit/s NB.

Note: No performance requirements are defined for WB content.

**6.4.2 Multi-channel 7.1:**

**High-level definition of Multi-Channel 7.1 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/7.
* 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/7.
* For the EVS multi-mono system, the LFE is considered to be encoded with EVS 9.6 kbit/s NB.

**Detailed Multi-Channel 7.1 requirements**

The following table illustrates corresponding detailed performance requirements for Multi-channel 7.1 audio content:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BW | Bitrate (kbit/s) | FER/ Delay Loss Profile | Requirements (\* | Objectives (\* |
| SWB FB | 13.2 | All |  | **NWT EVS @ 7x 7.2 kbps (WB)** |
| 16.4 |  | **NWT EVS @ 7x 7.2 kbps (WB)** |
| 24.4 |  | **NWT EVS @ 7x 7.2 kbps (WB)** |
| 32 | **NWT EVS @ 7x 7.2 kbps (WB)** |  |
| 48 | **NWT EVS @ 7x 8.0 kbps (WB) OR**  **BT EVS @ 7x 7.2 kbps (WB)** |  |
| 64 | **NWT EVS @ 7x 13.2 kbps (SWB) OR**  **BT EVS @ 7x 9.6 kbps (SWB)** |  |
| 80 | **NWT EVS @ 7x 16.4 kbps (SWB) OR**  **BT EVS @ 7x 13.2 kbps (SWB)** |  |
| 96 | **NWT EVS @ 7x 16.4 kbps (SWB) OR**  **BT EVS @ 7x 13.2 kbps (SWB)** |  |
| 128 | **NWT EVS @ 7x 24.4 kbps (SWB/FB) OR**  **BT EVS @ 7x 16.4 kbps (SWB)** |  |
| 160 | **NWT EVS @ 7x 32 kbps (SWB/FB) OR**  **BT EVS @ 7x 24.4 kbps (SWB/FB)** |  |
| 192 | **NWT EVS @ 7x 32 kbps (SWB/FB) OR**  **BT EVS @ 7x 24.4 kbps (SWB/FB)** |  |
| 256 | **NWT EVS @ 7x 48 kbps (SWB/FB) OR**  **BT EVS @ 7x 32 kbps (SWB/FB)** |  |
| 384 | **NWT EVS @ 7x 64 kbps (SWB/FB) OR**  **BT EVS @ 7x 48 kbps (SWB/FB)** |  |
| 512 | **NWT EVS @ 7x 96 kbps (SWB/FB) OR**  **BT EVS @ 7x 64 kbps (SWB/FB)** |  |

(\* The multi-mono EVS reference shall be produced by individual EVS coding of the 7 channels, and the LFE channel shall be encoded with EVS 9.6 kbit/s NB.

Note: No performance requirements are defined for WB content.

**6.4.3 Multi-channel 5.1+4:**

**High-level definition of Multi-Channel 5.1+4 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/9.
* 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/9.
* For the EVS multi-mono system, the LFE is considered to be encoded with EVS 9.6 kbit/s NB.

**Detailed Multi-Channel 5.1+4 requirements**

The following table illustrates corresponding detailed performance requirements for Multi-Channel 5.1+4 audio content:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BW | Bitrate (kbit/s) | FER/ Delay Loss Profile | Requirements (\* | Objectives (\* |
| SWB FB | 13.2 | All |  | **NWT EVS @ 9x 7.2 kbps (WB)** |
| 16.4 |  | **NWT EVS @ 9x 7.2 kbps (WB)** |
| 24.4 |  | **NWT EVS @ 9x 7.2 kbps (WB)** |
| 32 |  | **NWT EVS @ 9x 7.2 kbps (WB)** |
| 48 | **NWT EVS @ 9x 7.2 kbps (WB)** |  |
| 64 | **NWT EVS @ 9x 8.0 kbps (WB) OR**  **BT EVS @ 9x 7.2 kbps (WB)** |  |
| 80 | **NWT EVS @ 9x 9.6 kbps (SWB) OR**  **BT EVS @ 9x 8.0 kbps (WB)** |  |
| 96 | **NWT EVS @ 9x 13.2 kbps (SWB) OR**  **BT EVS @ 9x 9.6 kbps (SWB)** |  |
| 128 | **NWT EVS @ 9x 16.4 kbps (SWB) OR**  **BT EVS @ 9x 13.2 kbps (SWB)** |  |
| 160 | **NWT EVS @ 9x 24.4 kbps (SWB/FB) OR**  **BT EVS @ 9x 16.4 kbps (SWB)** |  |
| 192 | **NWT EVS @ 9x 32 kbps (SWB/FB) OR**  **BT EVS @ 9x 24.4 kbps (SWB/FB)** |  |
| 256 | **NWT EVS @ 9x 48 kbps (SWB/FB) OR**  **BT EVS @ 9x 32 kbps (SWB/FB)** |  |
| 384 | **NWT EVS @ 9x 64 kbps (SWB/FB) OR**  **BT EVS @ 9x 48 kbps (SWB/FB)** |  |
| 512 | **NWT EVS @ 9x 96 kbps (SWB/FB) OR**  **BT EVS @ 9x 64 kbps (SWB/FB)** |  |

(\* The multi-mono EVS reference shall be produced by individual EVS coding of the 5+4 channels, and the LFE channel shall be encoded with EVS 9.6 kbit/s NB.

Note: No performance requirements are defined for WB content.

**6.4.4 Multi-channel 7.1+4:**

**High-level definition of Multi-Channel 7.1+4 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/11.
* 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/11.
* For the EVS multi-mono system, the LFE is considered to be encoded with EVS 9.6 kbit/s NB.

**Detailed Multi-Channel 7.1+4 requirements**

The following table illustrates corresponding detailed performance requirements for Multi-Channel 7.1+4 audio content:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BW | Bitrate (kbit/s) | FER/ Delay Loss Profile | Requirements (\* | Objectives (\* |
| SWB FB | 13.2 | All |  | **NWT EVS @ 11x 7.2 kbps (WB)** |
| 16.4 |  | **NWT EVS @ 11x 7.2 kbps (WB)** |
| 24.4 |  | **NWT EVS @ 11x 7.2 kbps (WB)** |
| 32 |  | **NWT EVS @ 11x 7.2 kbps (WB)** |
| 48 |  | **NWT EVS @ 11x 7.2 kbps (WB)** |
| 64 | **NWT EVS @ 11x 7.2 kbps (WB)** |  |
| 80 | **NWT EVS @ 11x 8.0 kbps (WB) OR**  **BT EVS @ 11x 7.2 kbps (WB)** |  |
| 96 | **NWT EVS @ 11x 9.6 kbps (SWB) OR**  **BT EVS @ 11x 8.0 kbps (WB)** |  |
| 128 | **NWT EVS @ 11x 13.2 kbps (SWB) OR**  **BT EVS @ 11x 9.6 kbps (SWB)** |  |
| 160 | **NWT EVS @ 11x 16.4 kbps (SWB) OR**  **BT EVS @ 11x 13.2 kbps (SWB)** |  |
| 192 | **NWT EVS @ 11x 24.4 kbps (SWB/FB) OR**  **BT EVS @ 11x 16.4 kbps (SWB)** |  |
| 256 | **NWT EVS @ 11x 32 kbps (SWB/FB) OR**  **BT EVS @ 11x 24.4 kbps (SWB/FB)** |  |
| 384 | **NWT EVS @ 11x 48 kbps (SWB/FB) OR**  **BT EVS @ 11x 32 kbps (SWB/FB)** |  |
| 512 | **NWT EVS @ 11x 64 kbps (SWB/FB) OR**  **BT EVS @ 11x 48 kbps (SWB/FB)** |  |

(\* The multi-mono EVS reference shall be produced by individual EVS coding of the 7+4 channels, and the LFE channel shall be encoded with EVS 9.6 kbit/s NB.

Note: No performance requirements are defined for WB content.

## 6.5 Performance requirements for operation with Object-based Audio Content

The following object-based audio requirements apply predominantly to conferencing scenarios. In experienced listener tests other critical material may be used, too.

**6.5.1 Object-based Audio, 1 Object:**

**High-level definition of Object-based Audio, 1 Object requirements**

The general requirement is that IVAS operated at rate X shall

* be no worse than the EVS multi-mono system, where each EVS instance is operated at the closest bit rate to X.
* For the EVS multi-mono system, the object metadata is fed without any quantization directly to the renderer.
* The identical set of minimal IVAS object metadata (as defined in IVAS-4, Annex C) is used for both, IVAS and the EVS multi-mono system.

**Detailed Object-based Audio, 1 Object requirements**

The following table illustrates corresponding detailed performance requirements for Object-based audio content with 1 object:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| BW | Bitrate (kbit/s) | FER/ Delay Loss Profile | DTX(\* | Requirements(\*\* | Objectives |
| SWB  FB | 13.2 | All | On/Off | **NWT EVS @ 1x 13.2 kbps (SWB)** |  |
| 16.4 | **NWT EVS @ 1x 16.4 kbps (SWB/FB)** |  |
| 24.4 | **NWT EVS @ 1x 24.4 kbps (SWB/FB)** |  |
| 32 | Off | **NWT EVS @ 1x 32 kbps (SWB/FB)** |  |
| 48 | **NWT EVS @ 1x 48 kbps (SWB/FB)** |  |
| 64 | **NWT EVS @ 1x 64 kbps (SWB/FB)** |  |
| 80 | **NWT EVS @ 1x 64 kbps (SWB/FB)** |  |
| 96 | **NWT EVS @ 1x 96 kbps (SWB/FB)** |  |
| 128 | **NWT EVS @ 1x 128 kbps (SWB/FB)** |  |

(\* DTX will be tested for rates up to 24.4 kbit/s where mandatory DTX operation is available for the multi-mono EVS reference.

(\*\* The EVS reference shall be produced by 1 individual EVS coding of the object signals using unquantized object metadata.

The requirements will be tested with binaural rendering over headphones. The requirements may also be tested with rendering over a room loudspeaker system in experienced listener tests.

Note: No performance requirements are defined for WB content.

**6.5.2 Object-based Audio, 2 Objects:**

**High-level definition of Object-based Audio, 2 Objects requirements**

The general requirement is that IVAS operated at rate X shall

* be no worse than the EVS multi-mono system, where each EVS instance is operated at the closest bit rate to X/2.
* For the EVS multi-mono system, the object metadata is fed without any quantization directly to the renderer.
* The identical set of minimal IVAS object metadata (as defined in IVAS-4, Annex C) is used for both, IVAS and the EVS multi-mono system.

**Detailed Object-based Audio, 2 Object requirements**

The following table illustrates corresponding detailed performance requirements for Object-based audio content with 2 objects:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| BW | Bitrate (kbit/s) | FER/ Delay Loss Profile | DTX(\* | Requirements(\*\* | Objectives |
| SWB  FB | 13.2 | All | On/Off |  |  |
| 16.4 | **NWT EVS @ 2x 8.0 kbps (WB)** |  |
| 24.4 | **NWT EVS @ 2x 13.2 kbps (SWB)** |  |
| 32 | **NWT EVS @ 2x 16.4 kbps (SWB/FB)** |  |
| 48 | **NWT EVS @ 2x 24.4 kbps (SWB/FB)** |  |
| 64 | Off | **NWT EVS @ 2x 32 kbps (SWB/FB)** |  |
| 80 | **NWT EVS @ 2x 32 kbps (SWB/FB)** |  |
| 96 | **NWT EVS @ 2x 48 kbps (SWB/FB)** |  |
| 128 | **NWT EVS @ 2x 64 kbps (SWB/FB)** |  |
| 160 | **NWT EVS @ 2x 64 kbps (SWB/FB)** |  |
| 192 | **NWT EVS @ 2x 96 kbps (SWB/FB)** |  |
| 256 | **NWT EVS @ 2x 128 kbps (SWB/FB)** |  |

(\* DTX will be tested for rates up to 48 kbit/s where mandatory DTX operation is available for the multi-mono EVS reference.

(\*\* The EVS reference shall be produced by 2 individual EVS codings of the object signals using unquantized object metadata.

The requirements will be tested with binaural rendering over headphones. The requirements may also be tested with rendering over a room loudspeaker system in experienced listener tests.

Note: No performance requirements are defined for WB content.

**6.5.3 Object-based Audio, 3 Objects:**

**High-level definition of Object-based Audio, 3 Objects requirements**

The general requirement is that IVAS operated at rate X shall

* be no worse than the EVS multi-mono system, where each EVS instance is operated at the closest bit rate to X/3.
* For the EVS multi-mono system, the object metadata is fed without any quantization directly to the renderer.
* The identical set of minimal IVAS object metadata (as defined in IVAS-4, Annex C) is used for both, IVAS and the EVS multi-mono system.

**Detailed Object-based Audio, 3 Object requirements**

The following table illustrates corresponding detailed performance requirements for Object-based audio content with 3 objects:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| BW | Bitrate (kbit/s) | FER/ Delay Loss Profile | DTX(\* | Requirements(\*\* | Objectives |
| SWB  FB | 13.2 | All | On/Off |  |  |
| 16.4 |  |  |
| 24.4 | **NWT EVS @ 3x 8.0 kbps (WB)** |  |
| 32 | **NWT EVS @ 3x 9.6 kbps (SWB)** |  |
| 48 | **NWT EVS @ 3x 16.4 kbps (SWB/FB)** |  |
| 64 | **NWT EVS @ 3x 24.4 kbps (SWB/FB)** |  |
| 80 | **NWT EVS @ 3x 24.4 kbps (SWB/FB)** |  |
| 96 | Off | **NWT EVS @ 3x 32 kbps (SWB/FB)** |  |
| 128 | **NWT EVS @ 3x 48 kbps (SWB/FB)** |  |
| 160 | **NWT EVS @ 3x 48 kbps (SWB/FB)** |  |
| 192 | **NWT EVS @ 3x 64 kbps (SWB/FB)** |  |
| 256 | **NWT EVS @ 3x 96 kbps (SWB/FB)** |  |
| 384 | **NWT EVS @ 3x 128 kbps (SWB/FB)** |  |

(\* DTX will be tested for rates up to 80 kbit/s where mandatory DTX operation is available for the multi-mono EVS reference.

(\*\* The EVS reference shall be produced by 3 individual EVS codings of the object signals using unquantized object metadata.

The requirements will be tested with binaural rendering over headphones. The requirements may also be tested with rendering over a room loudspeaker system in experienced listener tests.

Note: No performance requirements are defined for WB content.

**6.5.4 Object-based Audio, 4 Objects:**

**High-level definition of Object-based Audio, 4 Objects requirements**

The general requirement is that IVAS operated at rate X shall

* be no worse than the EVS multi-mono system, where each EVS instance is operated at the closest bit rate to X/4.
* For the EVS multi-mono system, the object metadata is fed without any quantization directly to the renderer.
* The identical set of minimal IVAS object metadata (as defined in IVAS-4, Annex C) is used for both, IVAS and the EVS multi-mono system.

**Detailed Object-based Audio, 4 Object requirements**

The following table illustrates corresponding detailed performance requirements for Object-based audio content with 4 objects:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| BW | Bitrate (kbit/s) | FER/ Delay Loss Profile | DTX(\* | Requirements(\*\* | Objectives |
| SWB  FB | 13.2 | All | On/Off |  |  |
| 16.4 |  |  |
| 24.4 |  | **NWT EVS @ 4x 8.0 kbps (WB)** |
| 32 | **NWT EVS @ 4x 8.0 kbps (WB)** |  |
| 48 | **NWT EVS @ 4x 13.2 kbps (SWB)** |  |
| 64 | **NWT EVS @ 4x 16.4 kbps (SWB/FB)** |  |
| 80 | **NWT EVS @ 4x 16.4 kbps (SWB/FB)** |  |
| 96 | **NWT EVS @ 4x 24.4 kbps (SWB/FB)** |  |
| 128 | Off | **NWT EVS @ 4x 32 kbps (SWB/FB)** |  |
| 160 | **NWT EVS @ 4x 48 kbps (SWB/FB)** |  |
| 192 | **NWT EVS @ 4x 48 kbps (SWB/FB)** |  |
| 256 | **NWT EVS @ 4x 64 kbps (SWB/FB)** |  |
| 384 | **NWT EVS @ 4x 96 kbps (SWB/FB)** |  |
| 512 | **NWT EVS @ 4x 128 kbps (SWB/FB)** |  |

(\* DTX will be tested for rates up to 96 kbit/s where mandatory DTX operation is available for the multi-mono EVS reference.

(\*\* The EVS reference shall be produced by 4 individual EVS codings of the object signals using unquantized object metadata.

The requirements will be tested with binaural rendering over headphones. The requirements may also be tested with rendering over a room loudspeaker system in experienced listener tests.

Note: No performance requirements are defined for WB content.

# **9. EVS Objective Performance Requirements and Objectives**

The IVAS codec shall/should meet the following objective requirements/objectives.

## 9.1. IVAS Objective Performance [Requirements/Objectives] for DTX operation

**Definition of AFR (Active Frame Ratio) and CADR (Correct Active Detection Ratio):**

*AFR for IVAS Codec and reference multi-mono EVS are defined as follows:*

AFRIVAS (in percentage) = 100 x (# frames that are neither SID nor NO DATA) / (# frames in DTX-off operation)

AFRNxEVS (in percentage) = 100 x (# super-frames that are neither SID nor NO DATA) / (# super-frames in DTX-off operation)

For AFRNxEVS the following definition applies:

A super-frame for a multi-mono EVS system with N EVS instances at frame index k is the union of EVS frames of the N instances for frame index k. A super-frame in an AFRNxEVS measurement is classified as SID or NO DATA if all frames of the super-frame are, respectively, SID or NO DATA. N is set to number of EVS instances of the multi-mono EVS reference system used in a corresponding subjective performance requirement.

Note: AFR computation does not include the preamble used in the test.

*CADR for IVAS Codec and reference multi-mono EVS are defined as follows:*

CADRIVAS (in percentage) = 100 x (# frames that are neither SID nor NO DATA)/ (# frames in DTX-off operation)

CADRNxEVS (in percentage) = 100 x (# super-frames that are neither SID nor NO DATA) / (# super-frames in DTX-off operation)

For CADRNxEVS the following definition applies:

A super-frame for a multi-mono EVS system with N EVS instances at frame index k is the union of EVS frames of the N instances for frame index k. A super-frame in an CADRNxEVS measurement is classified as SID or NO DATA if any frame of the super-frame is, respectively, SID or NO DATA. N is set to number of EVS instances of the multi-mono EVS reference system used in a corresponding subjective performance requirement.

Note: # frames only take into account true active frames. True active frames are defined as frames where the corresponding super-frame of a multi-mono system with N instances comprises at least one frame with energy larger than [-56dBov].

CADR is only used for generic audio content excluding predominant voice, for which an ideally no DTX should be used.

**[Requirements/Objectives] for objective evaluation:**

The AFR performance evaluation is based on a large database of FB speech and noisy speech of length (approximately 10 to 30 min) with an AFRNxEVS of approximately 40%.

The CADR performance evaluation is based on a large database of FB general audio content (mixed and music) of length (approximately 10 to 30 min).

The reporting of objective evaluation is to be detailed in the processing test plan.

The following objective [requirement/objective] applies for all IVAS operation modes / input audio formats for which subjective performance requirements with DTX on are defined:

* the AFR performance on the speech and noisy speech database of the IVAS system at a given bitrate [shall/should] be not worse compared to the respective multi-mono EVS reference system of any of the applicable subjective performance requirements. The [requirement/objective] shall be met within a relative tolerance of [10%].
* the CADR performance on general audio content database of the IVAS system at a given bitrate [shall/should] be not worse compared to the respective multi-mono EVS reference system of any of the applicable subjective performance requirements. The [requirement/objective] shall be met within a relative tolerance of [10%].

Note: the [requirement/objective] shall be evaluated using the largest available audio bandwidth operation mode of the IVAS system under test and the multi-mono EVS reference system. A full evaluation should be done during characterization.

This [requirement/objective] is exemplified for IVAS scene-based audio operation at 48 kbps.

The corresponding subjective requirement is

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| BW | Bitrate (kbit/s) | FER | DTX(\* | Requirements(\*\* | Objectives |
| SWB  FB | 48 |  | On | **NWT EVS @ 4x 16.4 kbps (SWB/FB) OR**  **BT EVS @ 4x 13.2 kbps (SWB)** |  |

The objective [requirement] is met if

AFRIVAS@48kbps < AFR4xEVS@16.4kbps \* (1 + [10%]) or AFRIVAS@48kbps < AFR4xEVS@13.2kbps \* (1 + [10%])

CADRIVAS@48kbps > CADR4xEVS@16.4kbps \* (1 - [10%]) or CADRIVAS@48kbps < CADR4xEVS@13.2kbps \* (1 - [10%])

IVAS is operated with FB audio bandwidth, EVS@16.4 is operated with FB audio bandwidth, EVS@13.2 is operated with SWB audio bandwidth.

## 9.3. JBM objective performance requirements for delay, time scaling and jitter-induced concealment operations

JBM behaviour in terms of delay, time scaling and jitter-induced concealment operations shall not be substantially worse than for a corresponding EVS reference system.

This requirement will be verified after the selection.

# **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 |
| 2022-11-16 | SA4#121 | General updates and updates to/inclusion of requirements for stereo, scene-based, MASA and multi-channel audio formats (output of Audio SWG editing session) | 0.2.0 | 0.3.0 |

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