**Source: Editor[[1]](#footnote-0)**

**Title: IVAS Design Constraints (IVAS-4)**

**Version: 0.2.1**

**Agenda Item:**

1. **Scope**

This document presents the Design Constraints of the EVS Codec Extension for Immersive Voice and Audio Services (IVAS). The development of IVAS was initiated at SA4 #94, approved at SA#77 in September 2017 and the Work Item is described in SP-170611. The target for the standardisation is to complete codec specifications for Release 16.

The remaining sections of this document describe the design constraints for the IVAS codec in detail.

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 including the VR use cases envisioned in 3GPP TR 26.918 (See SP-170611).

1. **IVAS Codec Design Constraints**

|  |  |
| --- | --- |
| **Sampling Frequency** **and Audio Bandwidth** | The encoder shall support 16, 32, and 48 kHz sampling rates in all operation modes.  The decoder shall support 16, 32, and 48 kHz sampling rates in all operation modes.  Note: Functions that are necessary for achieving the arbitrary selection of encoder / decoder input / output sampling rate are envisioned to be a part of the IVAS candidate.  The encoder shall support input signals with different input signal bandwidth (NB, WB, SWB, and FB) with frequency masks as defined for EVS.  The encoder and decoder shall support 8kHz sampling when EVS bit-exact operation is used (See Backward Interoperability) |
| **Diegetic and Non-diegetic audio** | The IVAS codec shall support diegetic and non-diegetic input audio.  Editor’s note: Switching of diegetic/non-diegetic audio type is TBD |
| **Encoder Input Formats** | The encoder shall support the following input formats:  Channel-based audio, including mono (1.0), stereo (2.0), surround (5.1 and 7.1), [surround + height (5.1+4 and 7.1+4), TBD]  Binaural audio  Scene-based audio, first-order (FOA) and up to [N]-order ambisonics.  Note: ACN component ordering and SN3D normalization.  [Spatial audio, [N] channels and spatial metadata defined by [TBD].]  [Editor’s Note FFS: Spatial metadata definition for the spatial audio format will require further input.]  Object-based audio, with support for at least [TBD] individual [mono] object streams. Each audio object shall be defined by [TBD metadata parameters].  [In addition, the IVAS codec shall support combinations of the above, totalling to no more than [TBD] audio streams.  Note: It will be necessary to specify how capture/presentations could be achieved in mobile communications.]  Editor’s note: Input is invited for defining the combinations of input formats. The inclusion of Binaural Audio for the combinations may matter to the rendered output formats. |
| **IVAS renderer** | Proponents shall provide a renderer solution as part of their IVAS candidate.  The renderer will be specified in IVAS series. |
| **Rendered Output Formats** | The IVAS codec shall support the following rendered output formats for the corresponding input format:   |  |  | | --- | --- | | **Encoder Input Format** | **Rendered Output Format** | | Multi-channel 7.1 | Multi-channel 7.1, Binaural Audio, Stereo, Mono.  Multi-channel on arbitrary loudspeaker configurations of up to [K] speakers. | | Multi-channel 5.1 | Multi-channel 5.1, Binaural Audio, Stereo, Mono.  Multi-channel on arbitrary loudspeaker configurations of up to [K] speakers. | | Binaural Audio | Binaural Audio, [Stereo, Mono]  [Binaural Audio output assumes listening over headphones while Stereo output assumes listening over two channel Stereo loudspeaker configuration.  Editor’s note: Mono and Stereo output will not be tested in the selection phase] | | Stereo | Stereo, Mono | | Mono | Mono | | Scene-based audio | Binaural audio, Stereo, Mono  Multi-channel on arbitrary loudspeaker configurations of up to [K] speakers.  Editor’s note: at least one multi-channel configuration will be tested in the TBD phase | | Object-based audio | Binaural audio, Stereo, Mono  Multi-channel on arbitrary loudspeaker configurations of up to [K] speakers.  Editor’s note: at least one multi-channel configuration will be tested in the TBD phase | |  |  | |  |  |   Editor’s note: Specification of rendered output formats for the remaining input formats is needed.  Editor’s note: the term “arbitrary loudspeaker configuration” needs to be defined. One proposed definition is: rendered up to [K] loudspeaker positions on a 3D sphere. Potential further definition of minimum number of loudspeakers in an arbitrary configuration could be considered. More input is invited. |
| **Pass-Through Operation** | The IVAS codec shall support pass-through operation, where the output format is identical to the encoder input format, for the following format categories (as specified in Encoder Input Formats):   * Channel-based audio * Binaural audio * Scene-based audio * [Spatial audio] * Object-based audio   Editor’s Note: The exact codec configurations (bitrates etc.) for which pass-through operation is required is TDB, e.g., to be specified in IVAS-3 (Performance Requirements). |
| **Interface to external rendering** | Candidates shall provide interface specification to external renderer.  Requirements on the interface are TBD.  Note: Performance requirements on the external renderers are to be defined, that is outside of the scope of this document.  [Fraunhofer proposal: The IVAS codec shall support the following interface formats for optional external rendering solutions:   * Channel-based audio, including mono (1.0), stereo (2.0), surround (5.1 and 7.1), [surround + height (5.1+4 and 7.1+4), TBD] * Scene-based audio, first-order (FOA) and up to [N]-order ambisonics. * Object-based audio, with support for at least [TBD] individual [mono] object streams. Each audio object shall be defined by [TBD metadata parameters]. ]   [Dolby proposal: An external renderer is a renderer that is connected to the IVAS decoder via the External Renderer API. The IVAS codec candidate shall offer the possibility to connect an external renderer via the external renderer API.  The API shall support render of any received and decoded input audio.] |
| **Bit Rates** | When EVS bit-exact operation is used (see Backward interoperability), the IVAS codec shall operate at bit rates of EVS (including all EVS Primary and AMR-WB IO modes)  Note: As implemented in EVS, the bit rates up to 24.4 kb/s are gross bit rates and they are net bit rates above 24.4 kb/s.  In other cases than EVS bit-exact operation: the IVAS codec shall operate at bit rates of [5.9VBR, 7.2, 8, 9.6, 13.2, 16.4], 24.4, 32, 48, 64, 96, 128, 160, 192, 256, [384, 512] kb/s [min and max TBD].  Note: The gross bit rate supported in the DTX/CNG/SID operation is [TBD]. |
| **Algorithmic Delay** | TBD  [Editor’s Note: The EVS Algorithmic delay is 32ms] |
| **Complexity** | TBD |
| **Backward Interoperability** | Having interoperability with the EVS is an important feature.  The full EVS codec algorithm shall be part of the IVAS candidate codec solution. EVS bit-exact processing shall be used when the input to the IVAS codec is a simple mono signal without spatial metadata and should also be applied whenever possible. Possible exceptions for particular operation modes are [tbd]. When multiple mono audio channels without spatial metadata are negotiated they shall all be bit-exact with EVS. Possible exceptions are [tbd].  The IVAS Codec shall support certain stereo modes of operation which include an embedded bit-exact EVS mono downmix bitstream at the bit-rates from 9.6 kbit/s to 24.4 kbit/s SWB (9.6/13.2/16.4/24.4 kbit/s).  Editor’s Note: Definition of bit-exactness for embedded stereo modes is FFS.  Note 1: When the IVAS codec uses EVS bit-exact operation then features such as AMR-WB I/O mode, EVS SID update rates and 8kHz sample rate support shall be supported.  [Note 2: TS 26.445 supports multiple mono channels in the same way as described in IETF RFC 4867, which supports multi-channel audio content by means of multiple mono channels. Use of the EVS multiple mono interoperability for encoding multi-channel audio is FFS. Being bit-exact with EVS for the multiple mono case would not be appropriate because channels are not synchronized in the EVS encoding of multiple mono channels which may result in different VAD decisions for different channels as noted in the section of A.2.5.1 of TS 26.445. Contributions are invited if some synchronization is necessary for such use case.] |
| **Frame length** | The candidate codecs shall operate with a frame size of 20 ms. |
| **Jitter Buffer Management (JBM)** | A JBM solution conforming to the requirements in TS 26.114, except for the functional requirement in sub-clause 8.2.2 of TS 26.114: “Speech JBM used in MTSI shall support all the codecs as defined in clause 5.2.1”, shall be provided with the candidate codecs.  [Note: The JBM defined in TS 26.448 may form the basis of the JBM provided with the candidate codecs.] |
| **Rate switching** | [The candidate codecs shall perform rate switching upon command to the encoder throughout the entire bit rate range at arbitrary frame boundaries. The rate switching may imply switching between different bandwidths and between mono coding modes and coding modes for multiple audio streams.] |
| **Packet loss concealment (PLC)** | A PLC solution shall be provided by the IVAS candidate codecs. |
| **RTP payload format** | Candidate codecs shall provide an RTP payload format specification supporting the full set of features and functionality of the IVAS candidate codecs. |
| **DTX** | The candidate codecs shall provide a complete VAD/DTX/CNG framework. It shall be possible to operate the codec with DTX on or DTX off.  SID update frames shall be sent with a frequency not exceeding once per 8 frames. |
| **[Encoder Interface]** | [Ericsson proposal: The encoder may provide an interface for optional indication of the expected playback format.  The encoder shall provide an interface for activation/deactivation of pass-through operation for each audio stream individually (e.g. for an object or for an HOA stream).]  Editor’s Note: Definition of a common interface is FFS. |
| **[Interface for Rendered Output]** | [Ericsson proposal: The decoder/renderer shall provide an interface for specification of the output audio format to be rendered. All Rendered Output Formats shall be supported.] |
| **[Interface for Non-Rendered Output]** | [Ericsson proposal: The decoder shall provide an interface for external rendering supporting all Non-rendered Output Formats.  Editor’s Note: Definition of a common interface is FFS.] |
| **Interface for binaural rendering** | The IVAS decoder/renderer shall provide an interface to provide [HRTF/BRIR] data for binaural rendering. The interface is [tbd].  [Note: There was some support for this interface to follow the SOFA SimpleFreeFieldHRIR convention - See AES69-2015].  The IVAS decoder/renderer shall provide an API to provide [TBD scene displacement data].  [The IVAS decoder/renderer shall support direct headphone presentation.] |
| **Control Data For Binaural Audio Rendering** | The IVAS decoder/renderer shall support the following control data for binaural audio rendering:  [HRTF/BRIR] data for binaural rendering on command line interface. The format for [HRTF/BRIR] data is [tbd].  [Note: There was some support for this interface to follow the SOFA SimpleFreeFieldHRIR convention - See AES69-2015].  [TBD scene displacement data]. |
| **Direct headphone presentation** | The IVAS codec shall support direct headphone presentation for one-channel non-diegetic audio (with application of associated panning gain provided at the decoder/renderer) and two-channel (stereo or binaural) non-diegetic audio. |
| **[Decoder/Renderer Motion to Sound Algorithmic Delay]** | [The maximum algorithmic delay from a detected change in head roll, azimuth & elevation to a binaural sound rendered within +/- [Y] degree(s) of the detected change shall be [20 ms].] |
| **Output gain limitation** | TBD |

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

# Definitions

Binaural audio: Binaural audio is defined as a two-channel spatial representation of a soundfield as typically captured at the entrance of the ear canals and intended for direct presentation to the left and right ears over headphones. In terms of spatial representation, binaural audio may be natural (truly recorded with microphones) or artificial (e.g. using HRTFs). No additional spatialization (e.g. by an additional HRTF/BRIR convolution) should be carried out before direct presentation over headphones.

Diegetic audio: Audio intended to be presented such that it is perceived to be fixed in relation to the listening environment.

Non-diegetic audio: Audio intended to be presented such that it is perceived to be fixed in relation to the listener’s head.



# **4. Revision history**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Meeting** | **Subject/Comment** | **Old** | **New** |
| 2017-10-13 | SA4#95 | Agreement of Initial Skeleton of Design Constraints (IVAS-4)  S4-171036 | N/A | 0.0.1 |
| 2017-11-17 | SA4#96 | Agreement of updating Initial Skeleton of Design Constraints (IVAS-4) S4-171353 | 0.0.1 | 0.0.2 |
| 2018-02-09 | SA4#97 | Agreement of updating Initial Skeleton of Design Constraints (IVAS-4) S4-180265 | 0.0.2 | 0.0.3 |
| 2018-04-13 | SA4#98 | Agreement of updating Initial Skeleton of Design Constraints (IVAS-4) S4-180605 | 0.0.3 | 0.0.4 |
| 2018-07-13 | SA4#100 | Agreement of updating Design Constraints (IVAS-4)  S4-181218 | 0.0.5 | 0.0.6 |
| 2019-02-01 | SA4#102 | Agreement of updating Design Constraints (IVAS-4)  S4-190248 | 0.0.6 | 0.0.7 |
| 2019-02-18 | AHEVS#56 | Changing editorship from Wang Bin to Huan-yu Su AHEVS-443 | 0.0.7 | 0.0.8 |
| 2019-02-20 | Post AHEVS#56 | Incorporating suggested changes from AHEVS#56 | 0.0.8 | 0.0.9 |
| 2019-04-11 | SA4#103 | Incorporating agreed changes during SA4#103 in Newport Beach S4-190450 | 0.0.9 | 0.0.10 |
| 2019-07-5 | SA4#104 | Incorporating agreed changes during SA4#104 in Cork | 0.0.10 | 0.0.11 |
| 2019-07-5 | SA4#104 | Update the header information | 0.0.11 | 0.0.12 |
| 2019-07-5 | SA4#104 | Implement decisions from SA4 Plenary | 0.0.12 | 0.1.0 |
| 2019-10-24 | SA4#106 | Incorporating agreed changes during SA4#106 in Busan | 0.1.0 | 0.2.0 |
| 2021-11-xx | SA4#116-e | Incorporating agreed changes during SA4#116-e | 0.2.0 | 0.3.0 |

1. Huan-yu SU – Huawei Technologies Co Ltd [↑](#footnote-ref-0)