**3GPP SA4 #129-e S4-241598**

**Online, 19-23 August 2024**

**Source: Fraunhofer IIS**

**Title: Draft LSout on IVAS**

**Document for: Discussion & Agreement**

**Agenda Item: 7.4**

## 1. Summary

This contribution proposes a draft LS text to communicate the completion of the IVAS codec work item and the corresponding specifications to relevant groups outside of 3GPP.

Title: LS on the IVAS Codec

Release: Release 18

Work Item: IVAS\_Codec

Source: SA WG4

To: ISO/IEC JTC 1/SC 29, GSMA NG, GSMA TSG, IETF, ETSI STQ, ETSI DECT, CTIA

Cc:

**Contact Person:**

Name: Stefan Döhla

E-mail Address: stefan.doehla AT iis.fraunhofer.de

**Send any reply LS to: 3GPP Liaisons Coordinator,** **mailto:3GPPLiaison@etsi.org**

Attachments: -

**1. Overall Description:**

3GPP SA4 would like to make aware of the finalized IVAS\_Codec work item, which resulted in the set of specifications for the new 3GPP communication codec for Immersive Voice and Audio Services (IVAS).

IVAS is the next generation codec in 3GPP. It is an extension of the 3GPP Enhanced Voice Services (EVS) codec offering:

- Complete bit-exact EVS codec functionality for mono speech/audio signal input

- Support of stereo and binaural audio

- Support of audio formats beyond stereo which include multi-channel audio (5.1, 5.1.2, 5.1.4, 7.1, 7.1.4), scene-based audio (Ambisonics up to 3rd order), metadata-assisted spatial audio (MASA), and object-based audio.

- Support of combined immersive audio formats: object-based audio with scene-based audio (OSBA) and object-based audio with metadata-assisted spatial audio (OMASA)

- VAD/DTX/CNG for rate efficient stereo and immersive conversational voice transmissions

- Error concealment mechanisms to combat the effects of transmission errors and lost packets

- Jitter buffer management

- Binaural rendering functionality for headphone playback including head-tracking and scene orientation control, configurable room acoustic synthesis, as well as loudspeaker rendering functionality for loudspeaker playback

- Binaural split rendering functionality with pre-rendering and transcoding to a head-trackable intermediate representation that can be transmitted to a post-rendering end-device for headphone playback. This functionality is an integral part of the IVAS codec and has also been specified as a codec-agnostic tool in specification TS 26.249.

The codec is optimized for services over 5G mobile networks and implementations on 5G devices with:

- Operation on 20 ms audio frames

- Multi-rate/multi-mode operation at the following discrete bit rates [kbps]: 13.2, 16.4, 24.4, 32, 48, 64, 80, 96, 128, 160, 192, 256, 384, and 512

- Ability to switch bitrate upon command

- Support of sampling frequencies of 8 kHz (only EVS interoperable coding), 16 kHz, 32 kHz and 48 kHz (fullband audio content)

- Low algorithmic delay (≤38 ms)

- Complexity and memory footprint suitable for different device types and application scenarios

The following technical specifications have been developed:

|  |  |
| --- | --- |
| TS [26.249](https://www.3gpp.org/dynareport/26249.htm) | Immersive Audio for Split Rendering Scenarios; Detailed Algorithmic Description of Split Rendering Functions |
| TS [26.250](https://www.3gpp.org/dynareport/26250.htm) | Codec for Immersive Voice and Audio Services (IVAS); General overview |
| TS [26.252](https://www.3gpp.org/dynareport/26252.htm) | Codec for Immersive Voice and Audio Services (IVAS); Test sequences |
| TS [26.253](https://www.3gpp.org/dynareport/26253.htm) | Codec for Immersive Voice and Audio Services (IVAS); Detailed Algorithmic Description including RTP payload format and SDP parameter definitions |
| TS [26.254](https://www.3gpp.org/dynareport/26254.htm) | Codec for Immersive Voice and Audio Services (IVAS); Rendering |
| TS [26.255](https://www.3gpp.org/dynareport/26255.htm) | Codec for Immersive Voice and Audio Services (IVAS); Error concealment of lost packets |
| TS [26.256](https://www.3gpp.org/dynareport/26256.htm) | Codec for Immersive Voice and Audio Services (IVAS); Jitter Buffer Management |
| TS [26.258](https://www.3gpp.org/dynareport/26258.htm) | Codec for Immersive Voice and Audio Services (IVAS); C code (floating-point) |

IVAS is the first 3GPP communication codec which natively supports stereo and immersive audio. This enables completely new use-cases beyond traditional telephony: In a stereo or immersive telephony use-case, for example, a participant can capture and convey an immersive scene to a remote participant, e.g., to share the full immersive experience of an event. For spatial conferencing applications, the flexibility of the IVAS codec will provide multiple options for:

- Ad-hoc conferencing calls with the transmission of the physical immersive scene picked up by a UE, e.g., placed on a table. Rendering of the immersive scene makes it easier to distinguish the talkers’ voice, clearly separated from ambient sounds, leading to more natural and effort-less conferencing.

- More complex scenarios with multiple participants, transmitted as individual streams and spatially rendered on the receiving UE to match the video scene, for example.

- Scenarios where an intermediate call server combines multiple participants into an immersive scene.

Moreover, the flexibility of the IVAS codec with respect to supported audio formats, bitrates, rendering options and suitability for different device types might enable completely new usage scenarios for communication. In addition, the IVAS codec will support content distribution use-cases including streaming of stereo/immersive content and advanced VR/AR applications.

3GPP's rigorous and transparent standardization process involved the definition of demanding terms of reference (ToRs). During the selection phase, the IVAS codec was tested against in total 144 ToRs in 46 subjective tests performed in six languages by four contracted external test labs, where each experiment was conducted in two different labs. During these tests, the IVAS codec has met the requirements in 98.3% of the cases with no systematic fails in both labs. The full performance characterization is available in the following technical report:

|  |  |
| --- | --- |
| TR [26.997](https://www.3gpp.org/dynareport/26997.htm) | Codec for Immersive Voice and Audio Services (IVAS); Performance characterization |

The split rendering feature was characterized separately, with the results documented in the following technical report:

|  |  |
| --- | --- |
| TR [26.996](https://www.3gpp.org/dynareport/26996.htm) | Immersive Audio for Split Rendering Scenarios; Performance characterization |

IVAS is the mandatory codec for immersive audio communication for 3GPP's multimedia telephony services over IMS (MTSI); immersive audio communication support is recommended in 3GPP Release 18. To enable conversational and other immersive media services with IVAS, the following technical specifications have been extended:

|  |  |
| --- | --- |
| TS [26.114](https://www.3gpp.org/dynareport/26114.htm) | IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction |
| TS [26.117](https://www.3gpp.org/dynareport/26117.htm) | 5G Media Streaming (5GMS); Speech and audio profiles |
| TS [26.119](https://www.3gpp.org/dynareport/26119.htm) | Media Capabilities for Augmented Reality |
| TS [26.140](https://www.3gpp.org/dynareport/26140.htm) | Multimedia Messaging Service (MMS); Media formats and codecs |
| TS [26.141](https://www.3gpp.org/dynareport/26141.htm) | IP Multimedia System (IMS) Messaging and Presence; Media formats and codecs |
| TS [26.511](https://www.3gpp.org/dynareport/26511.htm) | 5G Media Streaming (5GMS); Profiles, codecs and formats |

In addition, the specifications related to the required support of IVAS in the network have been updated in Rel. 18.

Also, a set of UE testing methods and performance requirements based on IVAS has been defined in Rel. 18 specifications:

|  |  |
| --- | --- |
| TS [26.260](https://www.3gpp.org/dynareport/26260.htm) | Objective test methodologies for the evaluation of immersive audio systems |
| TS [26.261](https://www.3gpp.org/dynareport/26261.htm) | Terminal audio quality performance requirements for immersive audio services |

**2. Actions:**

**3. Date of Next SA4 Meetings:**

SA4#130 18th Nov – 22nd Nov 2024 Orlando, US