ETSI TS 126 254 V18.2.0 (2026-02)

TECHNICAL SPECIFICATION

LTE;
5G;
Codec for Immersive Voice and Audio Services (IVAS);
Rendering
(3GPP TS 26.254 version 18.2.0 Release 18)

Reference
RTS/TSGS-0426254vi20

Keywords
5G, LTE

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI IPR online database.

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™**, **LTE™** and **5G™** logo are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Legal Notice

This Technical Specification (TS) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found at 3GPP to ETSI numbering cross-referencing.

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

Foreword

This Technical Specification has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

- x the first digit:
  - 1 presented to TSG for information;
  - 2 presented to TSG for approval;
  - 3 or greater indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, modal verbs have the following meanings:

shall indicates a mandatory requirement to do something

shall not indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

should indicates a recommendation to do something

should not indicates a recommendation not to do something

may indicates permission to do something

need not indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

can indicates that something is possible

cannot indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

will indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

will not indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

might indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

might not indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

is (or any other verb in the indicative mood) indicates a statement of fact

is not (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

1 Scope

The present document provides a comprehensive description of the rendering functions of the decoder/renderer for Immersive Voice and Audio Services (IVAS codec).

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 26.250: "Codec for Immersive Voice and Audio Services (IVAS); General overview".

[3] (void)

[4] 3GPP TS 26.253: "Codec for Immersive Voice and Audio Services (IVAS); Detailed Algorithmic Description incl. RTP payload format and SDP parameter definitions".

[5] 3GPP TS 26.258: "Codec for Immersive Voice and Audio Services (IVAS); C code (floating point)".

[6] 3GPP TS 26.249: " Immersive Audio for Split Rendering Scenarios; Detailed Algorithmic Description of Split Rendering Functions".

3 Definitions of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the terms given in TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

rendering: a process of generating digital audio output from the decoded digital audio signal.

3.2 Symbols

Void.

3.3 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in TR 21.905 [1].

BRIR	Binaural Room Impulse Response
CPE	Channel Pair Element
EVS	Enhanced Voice Services
HRIR	Head-Related Impulse Response
HRTF	Head-Related Transfer Function

ISM	Independent Stream with Metadata
IVAS	Immersive Voice and Audio Services
MASA	Metadata Assisted Spatial Audio
MCT	Multi-channel Coding Tool
OMASA	Objects with MASA
OSBA	Objects with SBA
RTP	Real-Time Protocol
SBA	Scene-Based Audio
SCE	Single Channel Element
UE	User Equipment

4 General

4.1 IVAS receiver side processing

The codec for Immersive Voice and Audio Services is part of a framework comprising of an encoder, decoder, and renderer. An overview of the audio processing functions of the receive side of the codec is shown in Figure 4.1-1. This diagram is based on [2], with rendering features highlighted.

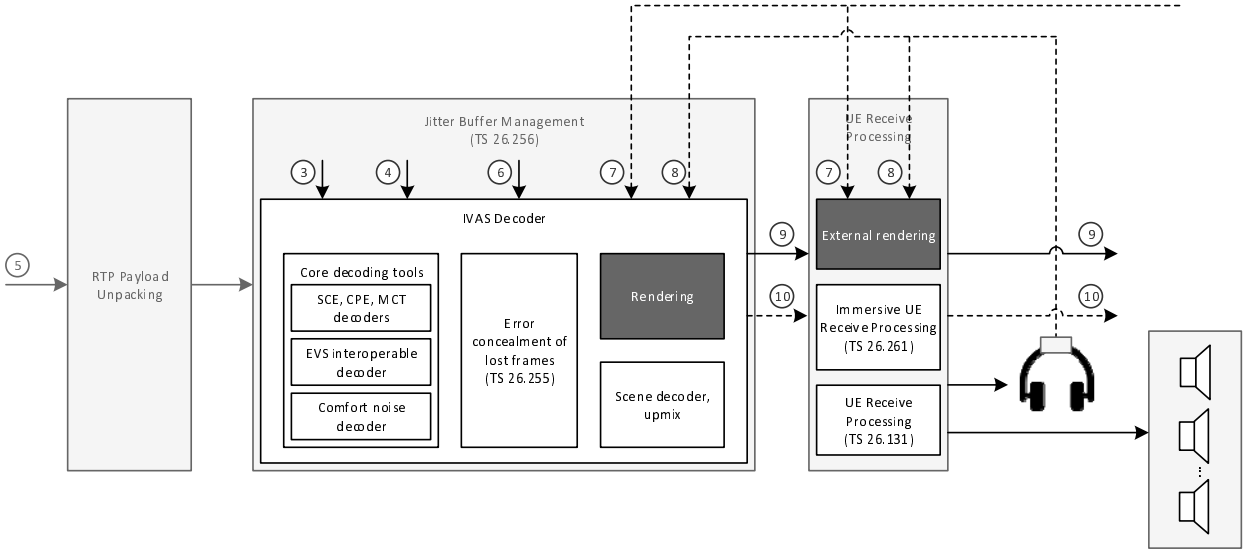

Figure 4.1-1: Overview of IVAS audio processing functions – receiver side.

Interfaces:

3: Encoded audio frames (50 frames/s), number of bits depending on IVAS codec mode
4: Encoded Silence Insertion Descriptor (SID) frames
5: RTP Payload packets
6: Lost Frame Indicator (BFI)
7: Renderer config data
8: Head-tracker pose information and scene orientation control data
9: Audio output channels (16-bit linear PCM, sampled at 8 (only EVS), 16, 32, or 48 kHz)
10: Metadata associated with output audio

Please note that the interface numbering is consistent with IVAS General Overview [2].

4.2 IVAS rendering

Rendering is the process of generating digital audio output from the decoded digital audio signal. Rendering is used when output format is different than input format. In case output format is the same as input format, the decoded audio

channels are simply passed through to the output channels. Binaural rendering is a special case, where binaural output channels are prepared for headphone reproduction. This process includes head-tracking and scene orientation control, head-related transfer function processing, and room acoustic synthesis. Rendering for loudspeaker reproduction is also supported for preset or custom loudspeaker configurations.

IVAS rendering is available as an integral component of the IVAS decoder (internal renderer) or can be operated standalone as external rendering. The external renderer can be applied e.g., in the case of rendering outputs originating from multiple sources, such as decoders or audio streams.

IVAS rendering include:

- support for provisioning of HRIR/BRIR filter sets as control data for binaural rendering. The format of HRIR/BRIR data is provided in clause 5.10 of [5],
- support for default HRIR/BRIR sets for binaural rendering,
- support for rendering control features:
  - head-tracking data as control data and external orientation data for binaural audio rendering in quaternions (and for head-tracking also in Euler notation). The format of orientation data is provided in clause 5.11 of [5],
  - binaural reverb and early reflections controlled by reverb parameters, the format of reverb parameters is provided in clause 5.14.1, and in Annex B of [5],
  - object editing of decoded ISM, OMASA, and OSBA bitstreams in internal IVAS renderer (clause 5) according to interface provided in clause 5.18 of [5].

A special feature of the renderer is that it supports split operation with pre-rendering and transcoding to a head-trackable intermediate representation that can be transmitted to a post-rendering end-device. This enables moving a large part of the processing load and memory requirements for IVAS decoding and rendering to a (more) capable node/UE while offloading the final rendering end-device. The IVAS specific split rendering functionality is mostly described in TS 26.253 [4] whereas more generic split rendering functionality is specified in TS 26.249 [6].

This document provides a high-level specification of the internal (clause 5) and external renderer (clause 6). Furthermore, the rendering library interface is provided (clause 7). Split rendering is described on high level in clause 8. Specific rendering algorithms and processing paths are out of scope of this specification and are provided in TS 26.253 [4].

5 Internal renderer

5.1 Overview

The internal IVAS renderer is integrated into the IVAS decoder. In case of specific operating points, this integration allows for combining decoding and rendering processes, resulting in efficient processing. The internal renderer supports rendering for loudspeaker and headphone reproduction. In the case of loudspeaker rendering, the audio output is mapped to the loudspeaker positions of the loudspeaker setup. In the case of headphone reproduction, binaural rendering is applied. The following binaural output modes are supported:

- Binaural output without room acoustic synthesis (no room), command line option BINAURAL,
- Binaural output with room acoustics synthesized using impulse responses (room with IR), command line option BINAURAL_ROOM_IR,
- Binaural output with room acoustics synthesized using parametric reverb, with or without early-reflections (room with reverb), command line option BINAURAL_ROOM_REVERB.

There are four binaural renderer implementations available in the IVAS codebase: parametric binaural renderer, FastConv renderer, Crend convolution renderer, and time-domain object renderer. The application of these renderers depends on IVAS input format, bitrate, IVAS encoding mode, and binaural rendering output mode. These dependencies are summarized in Table 5.1-1.

Table 5.1-1: Input format to renderer mapping

IVAS Input Format	Bitrate Range [kbps]	IVAS Mode (if applicable)	Binaural rendering output mode (if applicable)	Renderer Used
SBA	13.2 – 80	-	-	Parametric Binaural Renderer
SBA	96 – 512	-	-	FastConv Binaural Renderer
MASA	13.2 – 512	-	-	Parametric Binaural Renderer
ISM (3 or 4 objects)	24.4 – 32	ParamISM	-	Parametric Binaural Renderer
ISM	13.2 – 512	DiscISM	No room or room with reverb	Time Domain Object Renderer
ISM	13.2 – 512	DiscISM	Room with IR	Crend Binaural Renderer
MC	See Table 5.1-2	McMASA	-	Parametric Binaural Renderer
MC	See Table 5.1-2	ParamMC	-	FastConv Binaural Renderer
MC	See Table 5.1-2	ParamUpmix	-	FastConv Binaural Renderer
MC	See Table 5.1-2	DiscMC	All except below	Crend Binaural Renderer
MC Planar Layouts (5.1 and 7.1)	See Table 5.1-2	DiscMC	Head tracking enabled for either no room or room with reverb	Time Domain Object Renderer
OMASA	See text below	-	-	Same as non-combined format
OSBA	See text below	-	-	Same as non-combined format

The IVAS modes applicable for multi-channel input formats are summarized in Table 5.1-2. More details regarding multi-channel operation are provided in TS 26.253 [4] clause 5.7.

Table 5.1-2: Multi-channel format and bitrate mapping to IVAS coding modes

Bitrate [kbps]	MC layout				
	5.1	7.1	5.1.2	5.1.4	7.1.4
13.2 – 32	McMASA	McMASA	McMASA	McMASA	McMASA
48 – 80	ParamMC	ParamMC	ParamMC	McMASA	McMASA
96	DiscMC	ParamMC	ParamMC	ParamMC	McMASA
128	DiscMC	DiscMC	DiscMC	ParamMC	ParamMC
160	DiscMC	DiscMC	DiscMC	DiscMC	ParamUpmix
192 – 512	DiscMC	DiscMC	DiscMC	DiscMC	DiscMC

For the OMASA and OSBA cases, the IVAS coding modes depend on the number of objects and the total IVAS bitrate. For details refer to TS 26.253 [4] clause 5.9.2 and 6.9.7 for OMASA, and clause 5.8.1 and 6.8 for OSBA.

The details on binaural rendering algorithms are provided in TS 26.253 [4] clause 7.2.2.

5.2 Time-Domain Renderer

The time domain (TD) renderer operates on signals in time domain. In the IVAS internal renderer it is used for binaural rendering of discrete ISM, where each audio signal is encoded and decoded with a dedicated SCE module. This covers all ISM bit rates, except 3-4 objects for bit rates 24.4 kbps and 32 kbps. Further it is used in the decoder for binaural rendering of 5.1 and 7.1 signals when headtracking is enabled. An overview of the TD binaural renderer is found in Figure 7-2.1 in TS.26.253 [4]. An HRIR model accepts the object position metadata along with the headtracking data and generates an HRIR filter pair. The ITD may be modelled as a part of the HRIR, or it may be modelled as a separate parameter. In case an ITD parameter is output, the ITD is synthesis is performed in the ITD synthesis stage. The time aligned signals are then convolved with the HRIR filter pair to form a binauralized signal. Details are described in 7.2.2.2 in [4].

5.3 Parametric Binauralizer and Parametric Stereo Renderer

The parametric binauralizer and stereo renderer operates on the following IVAS formats and operations: MASA, OMASA, multi-channel (in McMASA mode), SBA, OSBA, and ISM, i.e., the input to the encoder has been audio signals (and potentially spatial metadata) in one of these formats, and it is now being rendered to binaural or stereo output. Details are described in 7.2.2.3 in [4].

5.4 Fast Convolution Binaural Renderer

The fast convolution binaural renderer operates on signals in the CLDFB domain. It is used for binaural rendering for the following IVAS formats and operating points (cf. Table 5.1-1): SBA (96 kbps upwards), OSBA and Multi-channel (ParamMC and ParamUpmix modes). Details are described in 7.2.2.4 in [4].

5.5 Crend Binaural Renderer

The Crend binaural renderer operates on signals in time domain. In the IVAS decoder, it is used for binaural rendering of multichannel signals, where each audio signal is encoded and decoded using discrete multi-channel mode or for discrete ISM with binaural output with room acoustics synthesized using impulse responses. The convolver uses a zero-delay block DFT implementation. DFT/IDFT is implanted using MDFT/IMDFT allowing buffer size being equal to the decoder frame size. Details are described in 7.2.2.5 in [4].

6 External renderer

6.1 Overview

The external IVAS renderer offers a standalone rendering capability employing the same rendering algorithms as the internal IVAS renderer. It is intended to receive the outputs of the IVAS decoder and further render them to other output formats. In addition, the IVAS external renderer is able to receive multiple different input streams that are rendered into a single output format. This provides a mixing functionality to use with multiple IVAS decoder outputs and a pre-renderer functionality for use before IVAS encoding. More details on pre-rendering algorithms are provided in TS 26.253 [4] clause 7.5.

The external IVAS renderer supports inputs of Ambisonics, ISM, multi-channel, and MASA format streams. The available output formats are binaural (with head-tracking and room effect options), Ambisonics, multi-channel, and MASA format (limited to pre-renderer mixing).

Supported input and output format mapping is provided in Table 6.1-1.

Table 6.1-1: Supported pre-rendering input/output mapping

Input format	Output Format				
		Channel based	**SBA**	**MASA**	**Binaural**
	Channel based	●	●	●	●
	SBA	●	● (mixing)	●	●
	MASA	●	●	● (mixing)	●
	ISM	●	●	●	●

In the case of rendering to binaural formats, the renderer implementations as discussed in clause 5.1 are used. Similarly to the case of internal rendering, binaural output modes with and without room acoustics are supported.

6.2 Time-Domain Renderer

In the external renderer the TD renderer is used for all ISM configurations, custom loudspeaker configurations, and multichannel formats 5.1 and 7.1 with headtracking enabled. Details are described in 7.2.2.2 in [4].

6.3 Parametric Binauralizer and Parametric Stereo Renderer

The parametric binauralizer and stereo renderer is used for the MASA input format in the external renderer. That is, if the output format is any form of binaural output or stereo output, then this renderer is used. Details of the renderer are described in 7.2.2.3 in [4].

6.4 Fast Convolution Binaural Renderer

Fast convolution rendering is currently not applicable for external rendering.

6.5 Crend Binaural Renderer

Crend binaural renderer (7.2.2.5 in [4]) is used when input format are SBA and MC except for rendering of MC formats 5.1 and 7.1 with headtracking enabled see 6.2.

7 Rendering interface

7.1 High-level rendering interface description

IVAS renderer and its interface provide support to IVAS codec design constraints. The rendering modes and rendering control mechanisms are discussed in clause of TS 26.253 [4].

The details of the rendering library API are provided in TS 26.258 [5] for the floating-point code. The API functions of the IVAS rendering library provide access to the following groups of functionalities:

- Initialization,
- Configuration (input/output),
- Metadata (input/output),
- Audio (input/output),
- Head tracking and orientation tracking (input/output),
- Object editing

8 Rendering interface

IVAS supports split rendering wherein the process of binaural rendering and headtracking is split between a main device (pre-renderer) and a light-weight head-worn device (post-renderer). The split-rendering architecture in IVAS is such that the complexity at the post-renderer is substantially less than the complexity of the IVAS decoder and renderer.

There are two architectures of split rendering supported in IVAS. The first architecture extends IVAS decoder and internal renderer to perform the pre-rendering part of split renderer, whereas the post rendering is done using a separate post-renderer. In the second architecture, the IVAS decoder runs in pass-through mode and the IVAS external renderer is extended to perform the pre-rendering part of split renderer. The respective pre- and post-rendering functions are specified in [4] clauses 7.6.2 and 7.6.6.

Note that the latter architecture also applies to the case where a custom external renderer is connected to the IVAS decoder through the rendering interface. In that case, the external renderer can connect to pre-renderer functions according to [4], clause 7.6.8.

Annex A (normative): Renderer control metadata processing tools

[A placeholder for the renderer control metadata processing scripts, including custom HRIR/BRIR conversion to binary format, etc. The actual scripts to be provided as Tdoc attachment.]

Annex B (informative): Change history

Change history							
Date	**Meeting**	**TDoc**	**CR**	**Rev**	**Cat**	**Subject/Comment**	**New version**
2023-11	SA4#126	S4-231845				Initial version presented to SA4 for information	0.0.1
2023-11	SA4#126	S4-231969				Submitted to SA4 plenary for information	0.1.0
2023-12	SA#102	SP-231299				Version created by MCC	1.0.0
2024-01	SA4#127	S4-240230				Version presented to SA4 for approval	1.0.1
2024-02	SA4#127	S4-240326				Submitted to SA4 plenary for approval	1.1.0
2024-03	SA#103	SP-240026				Version 2.0.0 created by MCC	2.0.0
2024-03						Version 18.0.0 created by MCC	18.0.0
2024-06	SA#104	SP-240693	0001	-	B	Adding ISAR track-a split rendering feature	18.1.0
2024-06	SA#104					Change of TS title as approved by TSG SA in SP-240917	18.1.0
2025-12	SA#110	SP-251431	0003	2	F	Correction of IVAS rendering specification	18.2.0

History

Version	Date	Status
V18.0.0	May 2024	Publication
V18.1.0	July 2024	Publication
V18.2.0	February 2026	Publication