1 Scope

This document presents recommended practices regarding the TV 3.0 Audio Coding, defined in ABNT NBR 25604.

2 References

The following documents are cited in the text in such a way that their contents, in whole or in part, constitute requirements for this document. For dated references, only the editions cited apply. For undated references, the most recent editions of that document (including amendments) apply.

ABNT NBR 25604, *TV 3.0 – Audio Coding*

ABNT NBR 15608-2, *Digital terrestrial television – Operational guideline - Part 2: Video coding, audio coding and multiplexing — Guideline for ABNT NBR 15602:2007 implementation*

3 Terms and Definitions

For the purposes of this Document, the following terms and definitions apply.

audio channel
audio signal that is reproduced at a specific nominal loudspeaker position

audio object
audio signal with associated metadata, which includes rendering information (e.g., gain and position) and information about interactivity options (e.g., minimum and maximum allowed gain values) that may change dynamically over time

Audio Preselection
set of Audio Program Components representing a version of a given Audio Program that can be selected by a user for simultaneous decoding

Audio Program
complete collection of all Audio Program Components and a set of accompanying Audio Preselections

NOTE: Not all Audio Program Components of one Audio Program are necessarily meant to be presented simultaneously. An Audio Program can contain Audio Program Components that are always presented and can include optional Audio Program Components.

Audio Program Component
smallest addressable unit of an Audio Program

default Audio Preselection
the Audio Preselection to be selected in cases when no other selection guidance (automatic or user-originated) exists

NOTE: Exactly one Audio Preselection can be identified as the default Audio Preselection, and all of its corresponding Audio Program Components are present in a single elementary stream.

immersive audio
audio system that enables high spatial resolution in sound source localization in azimuth, elevation and distance, and provides an increased sense of sound envelopment

TV 3.0
Digital Terrestrial Television system defined in the suite of standards ABNT NBR 25601 to ABNT NBR 25609 (which include this Document), also known as DTV+

4 Abbreviations

For the purposes of this Document, the following abbreviations apply.

AAC	*Advanced Audio Coding*
AOT	*Audio Object Type*
AD	*Audio Description*
BL	*Baseline*
CICP	*Coding-Independent Code Points*
CMAF	*Common Media Application Format*
DASH	*Dynamic Adaptive Streaming over HTTP*
DE	*Dialog Enhancement*
DRC	*Dynamic Range Control*
EHFR	*Efficient High Frame Rate*
HE	*High Efficiency*
HLS	*HTTP Live Streaming*
IOP	*Interoperability Point*
IP	*Internet Protocol*
IPF	*Immediate Playout Frame*
ISOBMFF	*International Organization for Standardization Base Media File Format*
JOC	*Joint Object Coding*
LC	*Low Complexity*
MAE	*Metadata Audio Element*
MHAS	*MPEG-H Audio Stream*

MPD	*Media Presentation Description*
RAP	*Random Access Point*
SAP	*Stream Access Point*
SBR	*Spectral Band Replication*
TOC	*Table of Contents*

5 Overview

The operational guidelines corresponding to the technologies used in the TV 3.0 Audio Coding are contained in the following Annexes:

- Annex A contains the MPEG-H Audio guidelines;
- Annex B contains the MPEG-4 AAC guidelines;
- Annex C contains the AC-4 guidelines; and
- Annex D contains the E-AC-3 JOC guidelines.

Annex A

MPEG-H Audio

A.1 Scope

This Annex provides Operational Guidelines for the use of MPEG-H Audio (ISO/IEC 23008-3:2022) in TV 3.0. It describes the MPEG-H Audio system features used within TV 3.0 broadcast emissions.

A.2 Personalization and Interactivity

The use of audio objects, usually in combination with channel-based audio, enables the viewers to interact in new ways with the content and create a personalized listening experience. The MPEG-H Audio metadata carries all the information needed to allow viewers to change the properties of audio objects by attenuating or increasing their level, disabling them, or even changing their position in the three-dimensional space. Additionally, the MPEG-H Audio metadata structures empower broadcasters to enable or disable interactivity options and to strictly set the limits to which extent a user can interact with the content.

In addition, with MPEG-H Audio, several versions of the content can be created as “Presets”, which describe how all channels and objects present in the MPEG-H Audio stream are mixed together and presented to the user.

A.2.1 Metadata Structure

The MPEG-H metadata (MAE) is structured in several hierarchy levels. The top-level element is the Audio Scene Information or the "*AudioSceneInfo*" structure as shown in Figure A.1. Sub-structures of the *AudioSceneInfo* contain descriptive information about "*Groups*", "*Switch Groups*", and "*Presets*." An "ID" field uniquely identifies each group, switch group or preset, and is included in each sub-structure.

The group structures ("*mae_GroupDefinition*") contain descriptive information about the audio elements, such as:

- the group type (channels or objects),
- the content type (e.g., dialog, music, effects, etc.),
- the language for dialog objects, or
- the channel layout in case of Channel-based content.

User interactivity can be enabled for the gain level or position of objects, including restrictions on the range of interaction (i.e., setting minimum and maximum values for gain and position offset). The minimum and maximum values can be set differently for each group.

Groups can be combined into switch groups ("*mae_SwitchGroupDefinition*"). All members of one switch group are mutually exclusive, i.e., during playback, only one member of the switch group can be active or selected. As an example, using a switch group for dialog objects ensures that only one out of multiple dialog objects with different languages is played back at the same time. Additionally, one member of the switch group is always marked as default to be used if there is no user preference setting and to make sure that the content is always played back with dialog, for example.

The preset structures ("*mae_GroupPresetData*") can be used to define different versions of audio elements within the Audio Scene. It is not necessary to include all groups in every preset definition. Groups can be "ON" or "OFF" by default and can have a default gain value. Describing only a subset of groups in a preset is allowed. The audio elements that are packaged into a preset are mixed together in the decoder, based on the metadata associated with the preset, and the group and switch group metadata.

From a user experience perspective, the presets behave as different complete mixes from which users can choose. The presets are based on the same set of audio elements in one Audio Scene and thus can share certain audio objects/elements, like a channel-bed.

Textual descriptions ("*labels*") can be associated with groups, switch groups and presets, for instance "*Commentary*" in the example below for a switch group. Those labels can be used to enable personalization in receiving devices with a user interface.

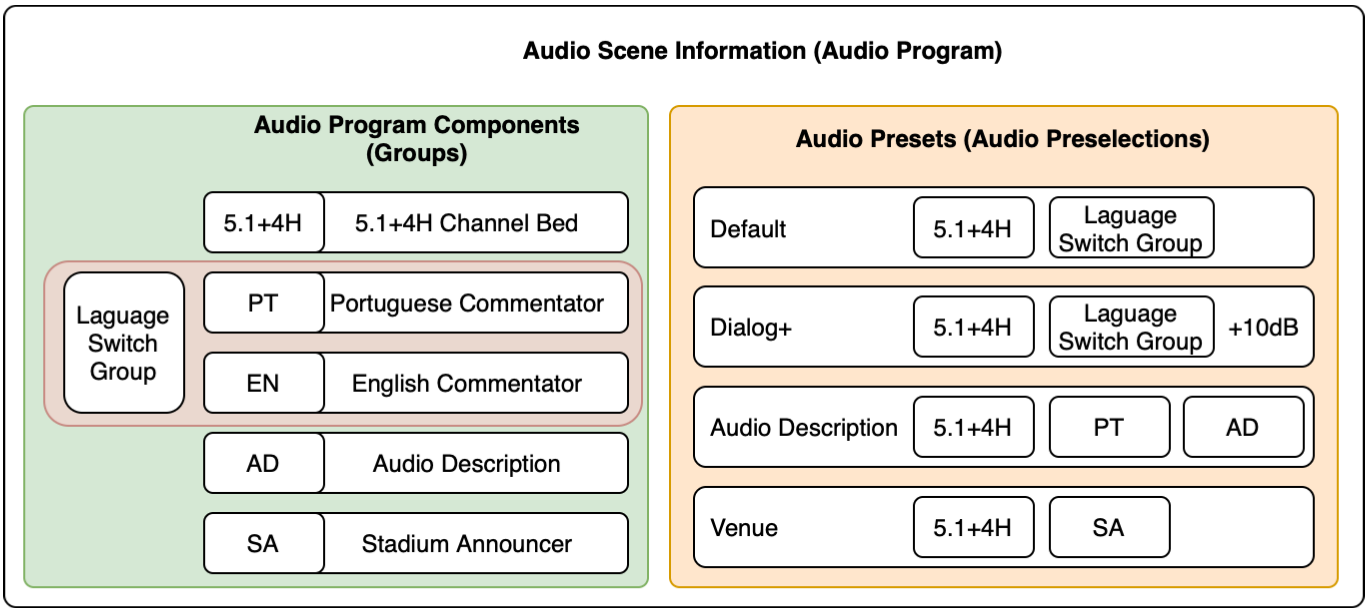

Figure A.1 – Example of an Audio Scene Information

Figure A.1 shows an example configuration of the MPEG-H Audio bitstream which offers the content in two different languages (Portuguese and English) and in the same time enables Audio Description and Dialog Enhancement services. The *Audio Scene information* contains in this example, four different *Presets:*

- "Default" – comprising the Ambience (5.1+4H) audio channels and language *Switch Group* including both languages. Note that only one element of a Switch Group can be active at any instance of time.
- "Dialog+" – comprising the Ambience (5.1+4H) audio channels and language *Switch Group* including both languages and additionally a 10 dB gain for the language switch group.
- "Audio Description" – comprising the Ambience (5.1+4H) channel bed, the Portuguese dialog element and the Audio Description element.
- "Venue" – comprising only the Ambience (5.1+4H) audio channels and an additional audio object carrying the stadium announcer.

In the absence of user interaction or automatic selection based on receiver preferred settings, the "*Default*" Preset will be presented to the user in the Portuguese (default) language.

A.2.2 Accessibility and Personalization

Object-based audio delivery offers the possibility to enable advanced and improved accessibility services and personalization features, especially for:

- Audio Description services (AD),
- Dialog Enhancement (DE), and
- Multi-language services.

This is achieved by carrying each dialog and audio description elements as separate audio objects, each object in one *Group* that can be combined with a channel bed element in different ways and create different Presets, such as:

- a "*Default*" preset without audio description service, and
- an "AD" preset including the audio description element.

A.2.2.1 Personalization options

For enabling advanced personalization options, the metadata in the *Audio Scene Information* (see mae_AudioSceneInfo() in ISO/IEC 23008-3, section 15) should be set accordingly:

- For each *Group* (e.g., dialog element) the following bit field elements in the *mae_GroupDefinition() structure* should be used:
    - *mae_allowGainInteractivity* should be set to "1" for enabling gain interactivity. The minimum and maximum gains allowed for gain interactivity should be given by the *mae_interactivityMinGain* and *mae_interactivityMaxGain* fields as described in ISO/IEC 23008-3, section 15*.*
    - *mae_allowPositionInteractivity* should be set to "1" for enabling position interactivity. The minimum and maximum allowed values for azimuth and elevation should be given by the *mae_interactivityMinAzOffset, mae_interactivityMaxAzOffset, mae_interactivityMinElOffset* and*mae_interactivityMaxElOffset* fields *as described in* ISO/IEC 23008-3, section 15*.*
- For each Preset the bit field elements in the *mae_GroupPresetDefinition()* should be used for enabling or disabling individual features, such as gain interactivity or position interactivity.
- For each Group, Switch Group and Preset part of the Audio Scene Information, the *mae_Description()* structure should be used to carry a dedicated textual label in the corresponding *mae_descriptionData* element. The textual labels can be defined in multiple languages using the *mae_bsNumDescLanguages* element.

A.2.2.2 Audio Description

The MPEG-H Audio system allows the delivery of Audio Description (AD) in multiple languages. AD services can be enabled by automatic device selection (prioritization) as well as by manual user selection. For each AD audio object, all advanced personalization options described in A.2.2.1 are available and can be enabled by the broadcaster during production. The level of the Audio Description can be adjusted independently, and moreover, the AD audio object can be spatially moved f (e.g., to the left or right) or better spatial separation from the main dialog element.

For using the Audio Description features, the metadata in the *Audio Scene Information* (see mae_AudioSceneInfo() in ISO/IEC 23008-3, section 15) should be set accordingly:

- For each *Group* containing an Audio Description object the *mae_contentKind element* in the *mae_ContentData() structure* should be set to "9" (audio description / visually impaired) as described in ISO/IEC 23008-3, section 15. The *mae_Description()* structure should be used to carry an appropriate textual label in the corresponding *mae_descriptionData* element (e.g., “Audio Description”). The textual labels can be defined in multiple languages using the *mae_bsNumDescLanguages* element.
- For each *Preset* including an Audio Description object, the *mae_groupPresetKind* element in the *mae_GroupPresetDefinition()* should be set to "7" (audio description / visually impaired) as described in ISO/IEC 23008-3, section 15. The *mae_Description()* structure should be used to carry an appropriate textual label in the corresponding *mae_descriptionData* element (e.g., “Audio Description”). The textual labels can be defined in multiple languages using the *mae_bsNumDescLanguages* element.

A.2.2.3 Dialog Enhancement

MPEG-H Audio includes Dialog Enhancement (DE) options that enable automatic device selection (prioritization) as well as user manipulation. For ease of user selection or for automatic device selection (e.g., enabling TV "Hard of Hearing" TV setting), a DE preset can be created, as illustrated in Figure A.2 using a broadcaster defined enhancement level for the dialog element (e.g., 10dB).

Moreover, if the broadcaster allows personalization of the enhancement level, MPEG-H Audio supports advanced DE which enables direct adjustment of the enhancement level via the user interface. The enhancement limitations (i.e., maximum level) are defined by the broadcaster/content creator and carried in the metadata. This maximum value for the lower and upper end of the scale can be set differently for different elements as well as for different content.

For using the Dialog Enhancement features, the metadata in the *Audio Scene Information* (see mae_AudioSceneInfo() in ISO/IEC 23008-3, section 15) should be set accordingly:

- For each *Group* containing a dialog object the *mae_contentKind element* in the *mae_ContentData() structure* should be set to "2" (dialog) or "7" (commentary) as described in ISO/IEC 23008-3, section 15. The *mae_Description()* structure should be used to carry an appropriate textual label in the corresponding *mae_descriptionData* element (e.g., “Dialog”). The textual labels can be defined in multiple languages using the *mae_bsNumDescLanguages* element.

- A Dialog Enhancement *Preset* can be defined using the *mae_groupPresetKind* element in the *mae_GroupPresetDefinition()* set to "5" (hearing impaired - light) or "6" (hearing impaired - heavy) as described in ISO/IEC 23008-3, section 15. For this Preset, the *mae_groupPresetGain* should be set to value higher than 0 dB (e.g., 10 dB).

- The *mae_Description()* structure corresponding to the Dialog Enhancement *Preset* should be used to carry an appropriate textual label in the corresponding *mae_descriptionData* element (e.g., “Dialog+”). The textual labels can be defined in multiple languages using the *mae_bsNumDescLanguages* element.

For each dialog audio object, all advanced personalization options described in A.2.2.1 are available and can be enabled by the broadcaster during production.

The loudness management tools of the MPEG-H Audio system automatically compensate for loudness changes that result from user interaction (e.g., switching presets or enhancement of dialogue) to keep the overall loudness on the same level, as illustrated in Figure A.2. This ensures constant loudness level not only across programs but also after user interactions.

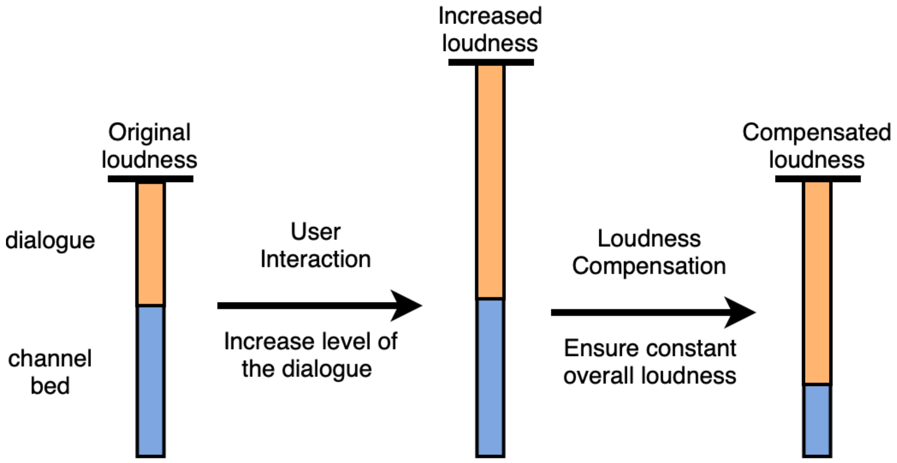

Figure A.2 – Loudness compensation after user interaction

A.2.2.3 Multi-language

If multiple languages are available, e.g., Portuguese, Spanish and English, the audio objects corresponding to the available languages can be included in a *Switch Group.* This will ensure that two dialog audio objects corresponding to two different languages will never be played back at the same time. Only one audio language can be active at any instance of time, based on the user selection and/or device preferred settings.

For using the multi-language features of the MPEG-H Audio system, all dialog elements carrying the different languages:

- should be signaled as part of the same switch group (i.e., using the same *mae_switchGroupID*) and
- should accordingly set the *mae_contentLanguage* part of the *mae_ContentData()* structure to the corresponding ISO 639-2 language code as described in ISO/IEC 23008-3, section 15. For example:
    - *mae_contentLanguage* = 0x70 0x6F 0x72 (POR) for Portuguese language
    - *mae_contentLanguage* = 0x73 0x70 0x61 (SPA) for Spanish language
    - *mae_contentLanguag*e = 0x65 0x6E 0x67 (ENG) for English language

For each switch group, all advanced personalization options described in A.2.2.1 are available and can be enabled by the broadcaster during production.

A.3 MPEG-H Audio Authoring

A.3.1 Audio Definition Model (ADM)

The Audio Definition Model (ADM) has been defined as a codec agnostic audio metadata standard and has been published in Recommendation ITU-R BS.2076. It is an XML-based model describing many types of audio content including channel- and object-based representations for immersive and interactive audio experiences. It is intended for production, exchange and archiving of audio content in file-based workflows. A serial representation of the Audio Definition Model (S-ADM) is specified in Recommendation ITU-R BS.2125 and defines a segmentation of the original ADM for use in linear workflows such as real-time production for broadcasting and streaming applications.

To ensure maximum interoperability according to the specific requirements for production, distribution and emission, applications adopting the ADM format should be able to convert the native metadata formats to ADM metadata and vice versa, such that artistic intent is preserved in a transparent way. This is achieved through the specification of ADM profiles.

The MPEG-H ADM Profile [MPEG-H ADM] defines a set of constraints on the Recommendations ITU-R BS.2076 and ITU-R BS.2125 that enable interoperability with established content production and distribution systems for MPEG-H Audio, as defined in ISO/IEC 23008-3. The profile allows the definition of ADM configurations providing full control over all features and parameters of the MPEG-H Audio system.

With the publication of the MPEG-H ADM Profile, the BWF/ADM file format has been established as a primary format for MPEG-H Audio content masters and is now supported natively by MPEG-H production tools. A BWF/ADM file is a Broadcast Wave File according to Recommendation ITU-R BS.2088, where ADM metadata is included in separate chunks in addition to the PCM audio data chunk.

In addition to the BWF/ADM format, SMPTE has published the standards SMPTE ST 2127-1 and ST 2127-10, which define a frame-based mapping of audio and time-synchronized S-ADM metadata into the Material

Exchange Format (MXF). The frame-based structure of this format makes it an ideal fit for workflows involving video frame aligned processing such as cutting and concatenating of content and for real-time processes with linear access to the frames in a file. The frame-based structure of the format is illustrated in Figure A.3.

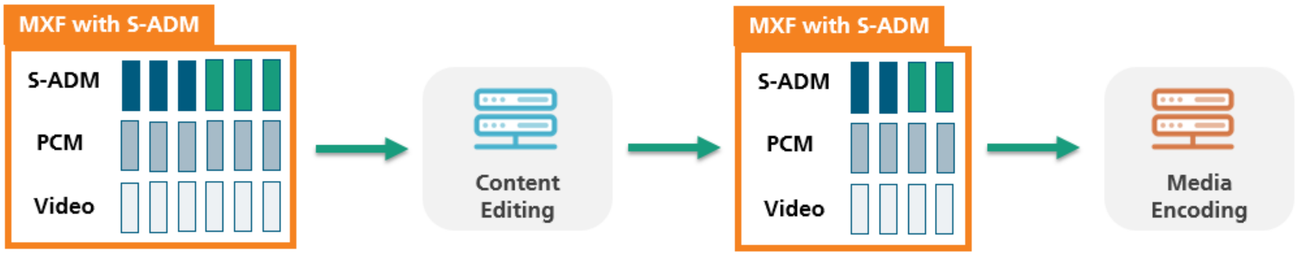

Figure A.3 – Frame-based structure of the MXF format

A.3.2 MPEG-H Master

In file-based workflows, the authoring of MPEG-H Audio scenes and the metadata is handled either by stand-alone tools or by plug-ins for Digital Audio Workstations (DAWs). In those tools, the scene is authored based on the audio mix, all user interactivity options are set, and loudness is measured. After completion of the authoring, the metadata is exported together with the audio data as an MPEG-H Master file. Using the tools of the MPEG-H Authoring Suite [MPEG-H MAS] or other post-production tools, an MPEG-H Master can be exported as a bundle of metadata and audio content. MPEG-H Audio metadata contains all control information for user interactivity and also all necessary information that the playback device needs for reproduction and rendering to ensure the best audio experience on any platform. The MPEG-H Master can be exported in the following formats:

- MPEG-H BWF/ADM: An MPEG-H BWF/ADM file (short for Broadcast Wave Format with embedded ADM metadata) is a multi-channel wave--file which contains all the audio and metadata of the MPEG-H Audio scene. The exported BWF/ADM file is compliant to the MPEG-H ADM Profile [MPEG-H ADM].

- MPEG-H Production Format (MPF): An MPF file is a multi-channel wave--file which contains all the audio and metadata of the MPEG-H Audio scene. The metadata is stored in the "Control Track", which is one of the audio tracks in the multichannel audio file and contains a "time-code like" signal that is robust against sample rate conversions or level changes. The Control Track is used for live productions using SDI-based workflows.

- MPEG-H MXF/S-ADM: An MPEG-H MXF/S-ADM file is an MXF file with embedded S-ADM metadata which contains all the audio and metadata of the MPEG-H Audio scene. The audio and S-ADM data are carried in a frame-based structure, and the file is compliant to the MPEG--H ADM Profile [MPEG-H ADM]. The MPEG-H MXF/S-ADM file format is especially designed for workflows involving video frame aligned processing and for real-time processes requiring linear access to the frames in a file.

A.3.3 MPEG-H Audio Authoring Suite

The Fraunhofer MPEG-H Authoring Suite [MPEG-H MAS] is a software suite for content generation, quality control, content validation, encoding and playback of MPEG-H Audio, and contains the following components:

- The MPEG-H Authoring Plugin enables all the steps of creating object- and/or channel-based MPEG-H productions inside a VST3 or AAX enabled Digital Audio Workstation (DAW).

- The MPEG-H Authoring Tool, a standalone application to create MPEG-H metadata based-on existing audio material. The MPEG-H Authoring Tool allows for easy MPEG-H metadata authoring and monitoring without the need for a DAW.
- The MPEG-H Production Format player, an audio and video player for quality control of already authored MPEG-H metadata and audio mix, with or without an accompanying video.
- The MPEG-H Conversion Tool, a software tool providing file format conversion for MPEG-H compliant content masters. The MPEG-H Conversion Tool serves as an interface to the MPEG-H Audio ecosystem and supports the import and export of MPEG-H Production Format (MPF), BWF/ADM and MPEG-H MXF/S-ADM files.
- The MPEG-H Info Tool, a standalone application for visualization of MPEG-H compliant content masters. The tool also analyzes ADM based content regarding its compatibility to various ADM profiles.
- The MPEG-H Encoding and Muxing Tool, a standalone application for the encoding of MPEG-H Master files and muxing with video.
- The MPEG-H VVPlayer, a standalone video player for the playback of encoded MPEG-H content together with video. The player includes the personalization user interface, binaural monitoring and flexible loudspeaker playback.

Figure A.4 provides a high-level overview of an end-to-end workflow for creating MPEG-H Audio content in post-production, either by authoring new projects or by ingesting existing material. The ingest step can convert an existing BWD/ADM file to the MPEG-H ADM Profile and enhance the content with advanced personalization and interactivity options.

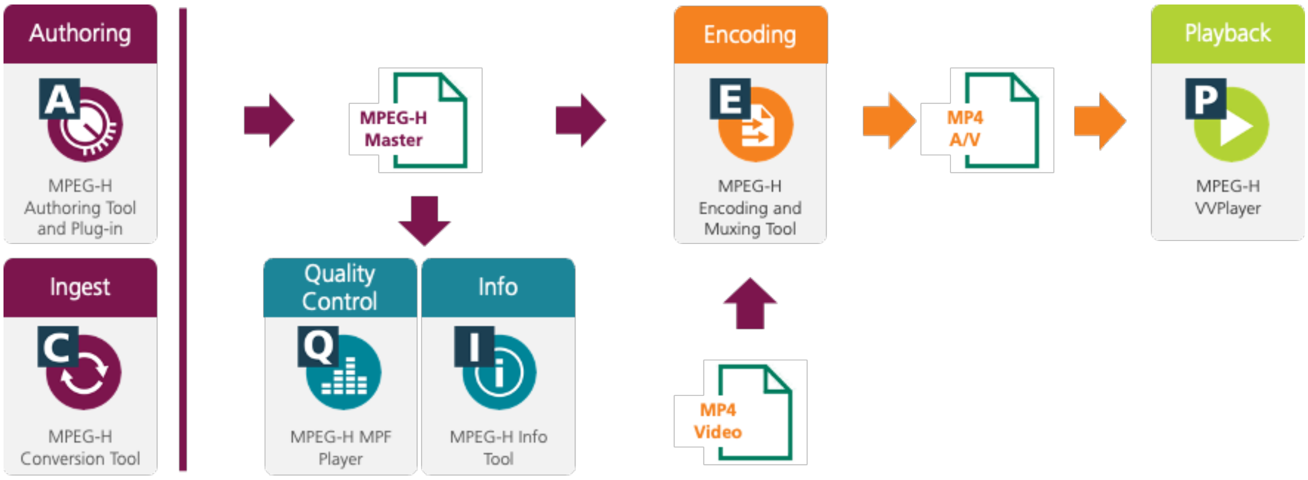

Figure A.4 – MPEG-H Audio post-production workflow.

The Fraunhofer MPEG-H Authoring Suite is freely available and can be downloaded according to the terms and conditions provided in [MPEG-H MAS].

As mentioned in A.2.1, BWF/ADM and MXF/S-ADM files conforming to the MPEG-H ADM Profile are considered as native file formats for MPEG-H Audio content as the profile ensures a transparent mapping of all MPEG-H Audio features and parameters. In addition to providing file format conversions between native MPEG-H audio content formats, the MPEG-H Conversion Tool, which is part of the MPEG-H Authoring Suite, also allows the import of ADM-based content from tools outside of the MPEG-H content ecosystem.

The MPEG-H Info Tool, also included in the MPEG-H Authoring Suite, provides an in-depth conformance analysis functionality. The tool runs a conformance framework equipped with exhaustive sets of checks derived from Recommendations ITU-R BS.2076, ITU-R BS.2088, and ITU-R BS.2125 as well as from ADM profiles related to the MPEG-H Audio system. After analysis, the tool summarizes the conformance checks' findings and indicates whether the tested content is supported by the MPEG-H Audio system. It also compiles a detailed report of all conformance issues with the ADM specifications and the MPEG-H ADM Profile that were encountered and provides information on how they can be resolved.

A.3.4 MPEG-H Audio Live Production workflows

The MPEG-H Audio system is designed to work with today's streaming and broadcast equipment using SDI-based workflows as well as with future IP-based infrastructure.

The most important step during a live production is the metadata authoring. At this stage all information about the MPEG-H Audio scene is created (e.g., labels for various objects and presets, interactivity settings for each preset, etc.) and at the same time each audio preset is monitored for quality control. This is described in the following by the functional block called "Authoring and Monitoring Unit" (AMAU), which typically is a hardware unit. The AMAU exports the metadata in realtime, tightly coupled with the audio signals and synchronized with the video signal on any of the connections that are commonly used in linear productions, such as SDI, MADI, or AoIP.

A.3.4.1 SDI-based Live Production workflow

A simplified SDI-based workflow for live production and distribution is illustrated in Figure A.5. To ensure the integrity of metadata in an SDI environment in any production step, the metadata is delivered over SDI using the so-called "Control Track", normally carried on the 16th SDI audio channel.

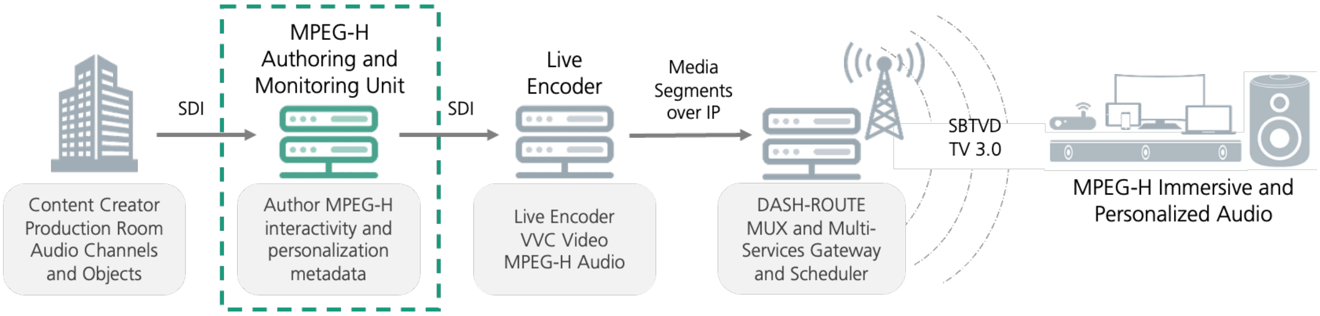

Figure A.5 – MPEG-H Audio live-production SDI-based workflow.

The MPEG-H Control Track is a "time-code like" audio signal and can be treated as a regular audio channel. This ensures the synchronization of metadata with its corresponding audio and video signal. The MPEG-H Control Track is robust enough to survive A/D and D/A conversions, level changes, sample rate conversions or frame-wise editing. The MPEG-H Control Track does not force audio equipment to be put into data mode or non-audio mode in order to passthrough.

The metadata for the audio signal is collected into packets synchronized with the video signal and is modulated with analog-channel modem techniques into a Control Track signal that fits in the audio channel bandwidth. This signal is unaffected by typical filtering, resampling, or scaling operations in the audio sections of broadcast equipment.

The use of AMAU systems allows productions to enable all MPEG-H features without changing the entire workflow. The monitoring stage during a production is extremely important. Many different speaker layouts from stereo to 7.1+4H can be connected for 3D Audio playback and used for monitoring in an AMAU. Additionally, all interactivity options and the audio quality can be monitored during production using an emulation of end-user receivers with different reproduction configurations.

AMAU systems measure the loudness and true peak values of all channels, objects, output busses and formats, as well as every created audio preset in real-time. With the resulting data, correction values are added to the metadata stream compliant with the applicable loudness regulation. The measurement of all generated DRC profiles and real-time loudness correction are also included. Additionally, AMAU production tools support the user with visualizations of all crucial measurement values.

A.3.4.2 IP-based Live Production workflow using S-ADM

The SMPTE ST 2110 suite of standards defines a system for the transport of media essences such as video, audio and metadata in professional broadcast and studio infrastructures based on managed IP (Internet Protocol) networks. The standard suite specifies the carriage, synchronization, and description of separate elementary essence streams over IP, making it possible to route all essences individually over Real-Time Transport Protocol connections (RTP, IETF RFC 3550) and accurately synchronize them using the Precision Time Protocol (PTP, IEEE 1588, SMPTE ST 2059).

The SMPTE ST 2110 family comprises multiple parts describing separately synchronization and system aspects as well as video, audio and metadata transport:

- SMPTE ST 2110-10 covers the system as a whole, the timing model, and common requirements across all essence types.
- SMPTE ST 2110-20 specifies the real-time, RTP-based transport of uncompressed active video essence over IP networks.
- SMPTE ST 2110-30 specifies the real-time, RTP-based transport of PCM digital audio streams over IP networks by reference to AES67.

Hybrid systems based on combining SMPTE ST 2110-based and SDI-based media transport are enabled by SDI/IP converter devices. The requirements of such hybrid systems have been carefully taken into account in the development of the SMPTE ST 2110 standard suite.

The MPEG-H Audio metadata can be transported in SMPTE ST 2110-based production environments, by including the Control Track alongside the PCM audio, the MPEG-H Production Format, in a SMPTE ST 2110-30 link, similarly to the approach for SDI infrastructures.

However, under the RTP-based transport paradigm of SMPTE ST 2110, a more straightforward and future-proof solution is to convey audio metadata via a separate RTP stream for synchronized data essence (i.e., not modulated as PCM). A suitable metadata format for mapping MPEG-H Audio metadata into RTP streams is the Serialized Audio Definition Model (S-ADM, Recommendation ITU-R BS.2125). The frame-based structure of the S-ADM format makes it a perfect fit for real-time workflows with time-variant metadata. MPEG-H metadata can be defined in the form of S-ADM frames according to the MPEG-H ADM Profile.

SMPTE is currently developing new standards that specify the RTP transport of data packages that contain S-ADM and other metadata. The new standards are integrated with all common system aspects of SMPTE ST 2110 such as identification and synchronization to video and audio essence streams carried according to SMPTE ST 2110-20 and ST 2110-30. The standardization for the transport of metadata including S-ADM metadata (SMPTE ST 2110-41) is still ongoing at the time of writing.

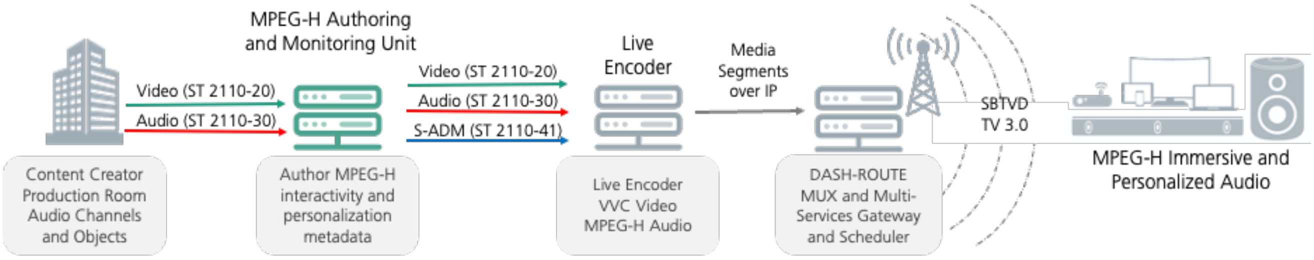
MPEG-H Authoring and Monitoring Unit
Video (ST 2110-20)
Audio (ST 2110-30)
Video (ST 2110-20)
Audio (ST 2110-30)
S-ADM (ST 2110-41)
Live Encoder
Media Segments over IP
SBTVD TV 3.0
MPEG-H Immersive and Personalized Audio
Content Creator Production Room Audio Channels and Objects
Author MPEG-H interactivity and personalization metadata
Live Encoder VVC Video MPEG-H Audio
DASH-ROUTE MUX and Multi-Services Gateway and Scheduler

Figure A.6 – MPEG-H Audio live-production IP-based workflow.

Figure A.6 illustrates the IP-based workflow using the ST 2110 suite of standards for an MPEG-H Audio live production and distribution.

A.4 Audio bitrate recommendations

Because of the flexibility of MPEG-H Audio when it comes to signal configurations, the bitrate depends on the number of channel signals or object signals. With an increasing number of signals in a configuration, the efficiency of the codec increases, and the resulting total bitrate is smaller than the sum of single-encoded signals.

Table A.1 indicates bitrates for some common channel configurations resp. a combination of channel and object signals, starting with stereo and 5.1 surround to several immersive configurations (indicated by “H” for the height channels) and combinations with different numbers of object signals.

Table A.1 - Recommended Bit Rates for Excellent Audio Quality in Broadcast Applications

Bitrates in kbit/s for	Good	Excellent	Transparent
2.0 Stereo	48	64	96
2.0 Stereo + 2 Audio Objects	128	160	192
5.1 Multi-channel surround	128	192	256
5.1+2H Immersive sound	160	256	320
5.1+4H Immersive sound	192	320	448
5.1+4H Immersive sound+ 2 Audio Objects	256 – 288	384 – 420	512 – 576
5.1+4H Immersive sound + 5 Audio Objects	352 – 384	480 – 576	640 – 768

A.5 Hybrid Delivery

The multi-stream-enabled MPEG-H Audio system is capable of handling streams delivered in a hybrid environment (e.g., one stream, containing one complete preselection, delivered over Broadcast and one or more additional streams, containing different dialog elements or Audio Description, delivered over Broadband), as specified in ISO/IEC 23008-3, subclause 14.6.

In order to use the multi-stream delivery mechanism, the audio streams have to be created using the same encoder since all the streams are part of the same audio scene. The encoder has to be aware of which audio elements are encoded into which audio streams. This information is delivered to the MPEG-H Audio Encoder via the MPEG-H metadata in an MPF file. A multi stream configurator tool can be used to decide which audio elements are part of the main stream and which are part of the side streams.

The MAE information contained in each stream allows the MPEG-H Audio decoder to correctly merge the streams into one stream containing several sub-streams:

- The main stream is identified via the *mae_isMainStream* set to value “1” and may use an *MHASPacketLabel* value in the range from 1 to 16, as specified in ISO/IEC 23008-3, subclause 14.6.
- The auxiliary streams are identified via the *mae_isMainStream* set to value “0” and each side stream will use a different range of values for the *MHASPacketLabel*, as specified in ISO/IEC 23008-3, subclause 14.6 and shown in Table A.2.

Table A.2 - Meaning of MHASPacketLabel in multi-stream environments

MHASPacketLabel value	Meaning
1 - 16 (0x01 - 0x10)	Main stream
17 - 32 (0x11 - 0x20)	First auxiliary stream
33 - 48 (0x21 - 0x30)	Second auxiliary stream

It should be noted that the main stream delivered via broadcast always contains a complete preselection, meaning that in the absence of an internet connection for delivery of the side streams, the viewer at home would still receive a complete presentation.

Annex B

MPEG-4 AAC

B.1 Scope

This Annex provides Operational Guidelines for the use of MPEG-4 AAC audio codec (ISO/IEC 14496-3) in TV 3.0.

B.2 General Guidelines

The Common media application format specification (ISO/IEC 23000-19) contains important considerations with respect to delay compensation, signaling and more.

It is recommended to follow the “Considerations for AAC audio encoding” specified in ISO/IEC 23000-19, section 10.3.3 and Annex G.

Annex C

AC-4

C.1 Scope

This Annex provides Operational Guidelines for the use of AC-4 audio codec (see ETSI TS 103 190-1 V1.3.1 (2018-02) for channel-based audio and ETSI TS 103 190-2 V1.2.1 (2018-02) for immersive and personalized audio) in TV 3.0.

C.2 Transitioning from AC-3 to AC-4

For implementers with existing AC-3 workflows, the Dolby Audio Handbook [12] provides valuable information on how to transition to AC-4. While aimed primarily at implementers of ATSC 3.0, many of the recommendations apply to SBTVD TV 3.0 Digital Television System as well.

Annex D

E-AC-3 JOC

D.1 Scope

This Annex provides Operational Guidelines for the use of Enhanced AC-3 (E-AC-3) Joint Object Coding (JOC) audio codec in TV 3.0. E-AC-3 is defined in ETSI TS 102 366 V1.4.1 (2017-09). Joint Object Coding (JOC) is a parametric coding technique that inserts an object-based audio representation into a multichannel signal for transmission and subsequent playback. JOC is defined in ETSI TS 103 420 V1.2.1 (2018-10).

For the current version of this document, no further information is provided regarding the usage of E-AC-3 JOC on TV 3.0.

Bibliography

[1] ISO/IEC 23008-3:2022, Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio

[2] ITU Recommendation, ITU-R BS.2076-2, Recommendation ITU-R BS.2076-2, Audio definition model, Geneva 10/2019

[3] ITU Recommendation, ITU-R BS.2088-1 (10/2019), Long-form file format for the international exchange of audio programme materials with metadata

[4] ITU Recommendation, ITU-R BS.2125, A serial representation of the Audio Definition Model, Geneva 01/2019, available at https://www.itu.int/rec/R-REC-BS.2125-0-201901-I

[4] MPEG-H ADM The MPEG-H ADM Profile, available at https://www.iis.fraunhofer.de/en/ff/amm/dl/whitepapers/adm-profile.html

[5] MPEG-H MAS The MPEG-H Authoring Suite, available at https://www.iis.fraunhofer.de/en/ff/amm/dl/software/mas.html

[6] SMPTE ST 2127-1:2022, SMPTE Standard - Mapping Metadata-Guided Audio (MGA) signals into the MXF Constrained Generic Container

[7] SMPTE ST 2127-10:2022, SMPTE Standard - Mapping Metadata-Guided Audio (MGA) signals with S-ADM Metadata into the MXF Constrained Generic Container

[8] SMPTE ST 2110-10:2022, SMPTE Standard - Professional Media over Managed IP Networks: System Timing and Definitions

[9] SMPTE ST 2110-20:2022, SMPTE Standard - Professional Media over Managed IP Networks: Uncompressed Active Video

[10] SMPTE ST 2110-30:2017, SMPTE Standard - Professional Media Over Managed IP Networks: PCM Digital Audio

[11] SMPTE ST 2110-41, SMPTE WD Standard - Fast Metadata Framework

[12] Dolby Audio Handbook, Dolby Laboratories, available at https://professional.dolby.com/siteassets/tv/home/dolby-vision/dolby_atsc3_hdbk_digi_v04_share.pdf

[13] White Paper: Dolby AC-4 Audio Delivery for Next Generation Entertainment Services, Dolby Laboratories, available at https://professional.dolby.com/siteassets/technologies/dolbt_atmos_ac-4_whitepaper.pdf