

MPEG-H Audio For Live Production

Overview for Broadcasters

A Technical Introduction for Broadcast Engineers

Version 1.0.1

Ulli Scuda, Matteo Alessandrini Lupia,
Aimée Moulson, Elias Weißenrieder,
Holger Braun, Christoph Bräuer,
Marcel Remy, Jochen Köhler

Date 2026-03-28

Contents

1	Goal of this document	2
1.1	What is NGA?	3
1.2	Unique Features of MPEG-H Audio	3
2	What You Need to Know to Produce with MPEG-H Audio	3
2.1	What Do You Do During Authoring?	4
2.2	Metadata	4
2.3	Metadata as PCM using the Control Track	4
2.4	Metadata as S-ADM over IP (ST 2110-41)	5
2.5	Metadata Summary	5
2.6	Signal Chain for MPEG-H Audio Production	6
2.7	Audio Monitoring and Authoring Unit (AMAU)	6
3	MPEG-H Audio Live Production	7
3.1	Managing Loudness and Monitoring in MPEG-H Audio	8
3.2	Encoder for MPEG-H Audio Live Production	8
3.3	Other Tools for MPEG-H Audio Production	9
3.4	Further Information	9
3.5	Contact	9
	References	9

1 Goal of this document

This document briefly summarizes what MPEG-H Audio is, what changes are involved for a regular broadcast service in order to make use of MPEG-H Audio and what the resulting benefits are. Technical details are kept to a minimum and only the most important aspects will be explained. The target audience of this paper is broadcast engineers and managers who already have a good understanding of the fundamentals of broadcast production and immersive audio.



1.1 What is NGA?

Next-Generation Audio (NGA) delivers immersive, personalized, and accessible experiences. A single master format contains all elements and metadata, enabling one stream or master file to play reliably across various device classes—from mobile devices and TVs to set-top boxes and AV receivers. Unlike legacy systems, it supports both channel-based and object-based workflows by keeping key signals separate up to the encoder. Interoperable metadata is created during production and carried end to end through encoding and decoding, ensuring consistent behavior across platforms, products, and vendors.

An **audio object** is an audio signal connected with describing metadata. Metadata can include information about the audio objects position in space, loudness, possible allowed interactivity or content kind. Audio objects carry their metadata along the entire production and reproduction path so the metadata can be used by the playback device as rendering instructions.

Authoring is the process of adding descriptive metadata to audio signals. During that process playback and rendering instructions can be configured and multiple presets or representations can be defined.

1.2 Unique Features of MPEG-H Audio

MPEG-H Audio as an NGA system, supports personalization, interactivity, and accessibility features [1]. It allows flexible mixes combining channels and objects, for example, immersive 5.1+4H with five audio objects or pure 9.1.4. It streamlines workflows by carrying multi-audio in a single transmission stream. The standard mix, multiple languages, and immersive program with audio description is not separated across multiple assets any more. Only one stream or file is managed across the entire production, post, and archive, while the Control Track carries all metadata end-to-end as a regular PCM audio track. MPEG-H is an open ISO standard deployed in broadcast in South Korea and Brazil and in Sony's 360RA music service. Certified CE devices are listed at <https://mpeg-h.com/integration/trade-mark/>

Personalization and Interactivity features

- **Presets** can offer multiple mix versions and different combinations of channels and objects. If there are multiple presets available and how they are configured is completely up the producer.
- **Switch Group** is a set of audio elements (so called components) that enables user-controlled, mutually exclusive selection between them. Only one Switch Group member can be active at a time.
- **Gain and Position Interactivity** can be configured to be allowed. What audio components can be changed in volume, azimuth or elevation and where the lower and upper limits of that kind of interactivity are, is controlled by metadata during the production.

As well as personalization and interactivity, MPEG-H Audio offers a series of accessibility features that help to reach a wider audience.

Dialog Boost presets can be configured for hearing impaired users. If wanted, the boost level can be made adjustable by the user with the help of a slider.

Audio Description (AD) objects are specifically marked, so visually impaired users can access them easily. Because the AD component can be just one mono object that is added to an existing preset, the AD mix can be as immersive as the default mix.

Localization refers to the text labels of presets and menu options. Labels of a multi language program are available in different languages, so the preferred language of the playback device can decide in what language interactivity options are presented.

2 What You Need to Know to Produce with MPEG-H Audio

The key for producing MPEG-H Audio is to keep audio signals of interest separate during the mixing process and to create metadata.

2.1 What Do You Do During Authoring?

During the authoring process, the MPEG-H scene is configured (see p. 2.7). For this, the **Reference Layout**, **Presets**, and **Switch Groups** are established using an AMAU (see p. 2.7). Text labels in different languages, specific downmix parameters and slider for the limits of gain or position interactivity are also created during authoring—if wanted. All of that can be prepared before the production and can be loaded into an AMAU as a template.

An **Audio Monitoring and Authoring Unit (AMAU)** is a device to create, edit and author MPEG-H scenes in real-time for live applications. It can be hardware or software. It also offers rendering options for monitoring allowing engineers to preview interactivity options and audio playback.

The **Reference Layout** is the target speaker reproduction format for the program. For immersive productions, this is typically 5.1+4H, though 5.1 or 2.0 are also possible. If you are unsure which Reference Layout to choose, use the layout of your largest component—usually the ambience bed.

2.2 Metadata

Metadata are essential for NGA. Created during production, they can be monitored and edited during the entire production chain. Depending on the type of signal distribution, metadata are transported in different forms, all types being fully cross-compatible and losslessly interoperable with post production workflows.

2.3 Metadata as PCM using the Control Track

Technically, the Control Track is a true linear audio signal. The CT itself cannot be encoded, it always needs to stay PCM audio. It is usually carried as the last audio channel of an SDI signal (see fig. 1). This way, audio and video are always transported together and aligned. The CT works within 50 fps and 59.94 fps environments.

The CT as a PCM signal can be carried over an 2110-30 connection (see 2.5) [2]. Because an 2110-30 stream can only carry eight audio signals, two parallel streams are need for a full 16-channel MPF including CT. This way of transportation can be considered IP transport, but it is important to note that this is not the same as S-ADM!

The Control Track: All necessary MPEG-H Audio metadata can be modulated into a timecode-like audio signal, which is typically located on the last channel of a 16 ch linear PCM signal. The so called Control Track (CT), due to its robust nature, does not require to put audio equipment into data mode or non-audio mode to pass through. The CT contains all metadata as well as all dynamic metadata, such as object positions. The CT can be edited and cut in the frame grid in synch with the video. It is safe against level changes and resampling.

MPF: A 16-channel PCM signal, that carries the MPEG-H Control Track with all its metadata on channel 16, is called an **MPEG-H Production Format (MPF)**. It can be created and transmitted in real time, intercut for configuration changes (e.g., seamless ad insertion), multiplexed with video, and passed through legacy audio chains (routers, AD/DA, sample-rate converters) without data/non-audio mode. It is 16-channel-based so it can be used in SDI based environments. MPF can also be saved as a multichannel WAV file or inside a MOV container.

2.4 Metadata as S-ADM over IP (ST 2110-41)

Within an IP based workflow, MPEG-H metadata can be transported using S-ADM over IP ST 2110-41 [3, 4]. The serialized frame based data packages are fully interoperable with the CT or file based ADM. S-ADM cannot be transported using SDI! It is always bound to an IP network. An ST-2110-41 connection only carries the metadata but no audio. For the audio signals, additional ST-2110-30 streams are needed.

Serialized Audio Definition Model (S-ADM) is defined in ITU-R BS.2125, which extends the Audio Definition Model (ADM, ITU-R BS.2076) [5, 6]. S-ADM is a frame-based metadata format designed to describe object-based audio scenes for real time applications. It enables the synchronized transmission of audio and metadata within IP-based production and distribution systems. Its key features include a time-segmented metadata structure, synchronization with audio frames, and real-time as well as streaming-capable operation, making it suitable for live production [7].

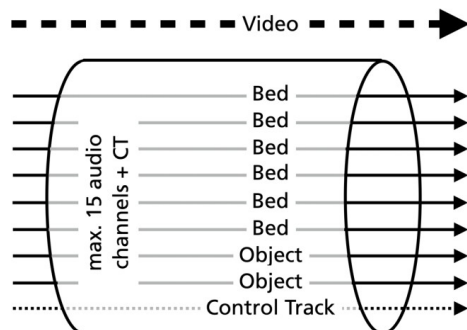


Figure 1: Metadata in an SDI signal as used with coax cables. The first 15 channels are used for the bed and audio object signals, the last channel is used for the Control Track.

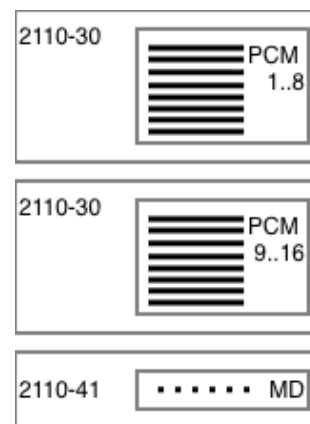


Figure 2: MPEG-H as an AoIP stream. To transport an MPF signal over IP, you can use two 2110-30 streams and one 2110-41 stream for metadata.

2.5 Metadata Summary

All different representations of metadata contain the full set of MPEG-H Audio information. All flavors can be converted into each other losslessly.

ST 2110-30 is a subcategory in the Standard for IP-based transmission of media signals and data SMPTE ST 2110. The '-30' part contains the transmission of uncompressed audio signals with common audio sampling frequencies and quantization. Typically an ST2110-30 Stream includes eight audio channels [2].

ST 2110-41 is a subcategory in ST 2110 and contains the transmission of 'fast meta-data' which are needed for NGA-Codecs [3].

BWF ADM is an xml-based, formalized description of audio content for file-based export formats, used in post production workflows. It is standardized in ITU-R BS.2076 [6]. The a-xml data chunk can be embedded as an ADM header in a BW64-RIFF-container, commonly known as a WAV-file [8]. The audio data is saved as linear PCM.

2.6 Signal Chain for MPEG-H Audio Production

A typical chain for an MPEG-H Audio production consists of recording, mixing, authoring, monitoring, encoding, and emission (see fig. 3). Authoring needs to happen after mixing and before encoding. Monitoring and QC checkpoints can vary by operation, but can be at any stage. Advertisements that are not authored can be injected into an AMAU where a Control Track can be generated using an .xml authoring template. Feeding ads directly into the encoder without a Control Track is possible but not recommended. This fallback mode in the encoder would engage to ensure a constant MPEG-H stream but this has been designed for a last resort scenario and not for regular operational use.

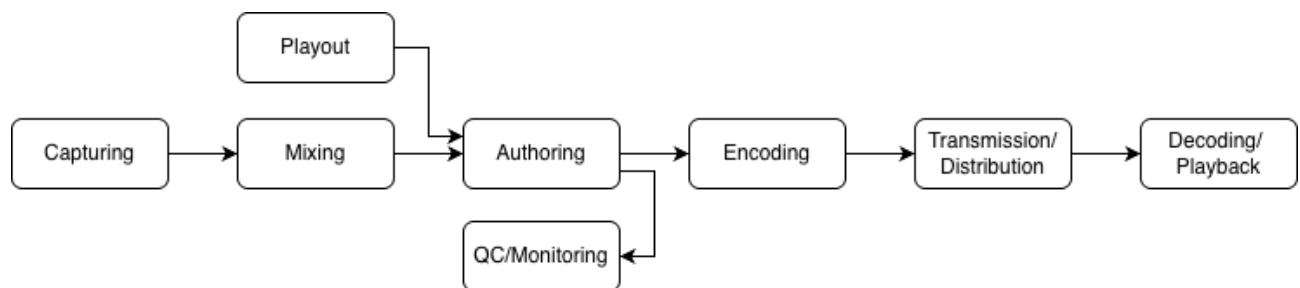


Figure 3: A simplified production chain for MPEG-H Audio broadcast.

2.7 Audio Monitoring and Authoring Unit (AMAU)

The AMAU (see 2.7) is used during a live production to create MPEG-H Audio scenes or to load pre-configured standardized templates, as well as to define dynamic parameters for personalization and accessibility. During production, the AMAU generates metadata in real-time and allows flexible monitoring.

Authoring in the AMAU includes defining components, languages, dynamic objects, and presets. Interactivity and personalization options are also set, such as for gain and position. Templates can be loaded or saved. Incoming metadata can be captured and edited, as well.

In the AMAU you can also configure a 'stereo legacy component' or create a legacy rendering output for parallel delivery.

The **MPEG-H Audio Scene** Information is a set of meta-data which defines different presentations of audio elements contained in an MPEG-H Audio bit stream as well as available interactivity features for personalized playback.

What AMAU products are available?

- Jünger Audio: as hardware called AIXpressor or in the cloud in the form of flexAI
- Linear Acoustics: AMS
- New Audio Technology: SAD-P
- Salsa Sound: MIXaiR

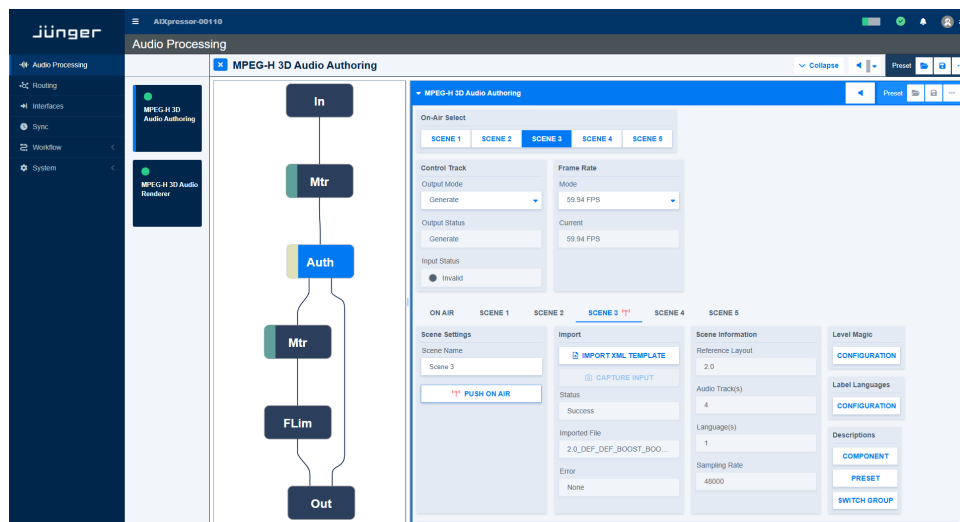


Figure 4: Screenshot of an Authoring Unit.

3 MPEG-H Audio Live Production

However, the key difference is the authoring stage, whereby operators define their MPEG-H scene. This is the stage where the metadata is generated and the scene can be monitored in real time alongside regular audio operations.

When producing in MPEG-H, it is important to first consider which audio elements the audio engineer would like to be interactive (e.g. a commentator). All audio elements which the audio engineer desires to give interactivity to, should be defined as an object component in the authoring stage and should therefore be kept separate from the main channel-based bed.

Components are a way of organizing audio tracks in MPEG-H Audio productions. A channel component can form the bed, or exist alongside it as separate channel-based element. For channel components, you can use any channel layout such as 7.1+4H, 5.1, 2.0, 1.0, and many more, or even 5.0 or 4.0+4H. Object components have more flexibility, e.g. they can have additional dynamic panning metadata. Both component types have metadata for personalization and interactivity.

For the bed component of the mix, many audio signals can be part of the bed and they are mixed together before the authoring stage. All components must remain separate during the mixing stage and have to be fed separately to the Authoring Unit. The channel layout for 5.1+4H follows the SMPTE-order and is L, R, C, LFE, Ls, Rs, UL, UR, ULs, URs.

Audio output needed to advance further in an MPEG-H Audio chain, transmission can be analog, AES3, MADI, DANTE, Ravenna, Livewire, ST2110-30, and SDI.

Depending on the interfaces available on the AMAU, it can be necessary to use signal converters, such as Analog <-> AES3, AES3 <-> MADI, MADI <-> DANTE, DANTE <-> Ravenna, Ravenna <-> Livewire, Livewire <-> ST2110, ST2110 <-> SDI and up-down-cross video conversion, plus frame rate conversion (50 and 59.94 fps).

3.1 Managing Loudness and Monitoring in MPEG-H Audio

All components are kept separate which has to be considered during mixing and monitoring. Loudness values and information for all components and presets are written into the metadata. Target loudness values for the components are entered by the audio engineer into the Authoring Unit. It is essential that integrated loudness values of components entering the Authoring Unit are conform and consistent. Correct loudness values of all components ensure consistent and R-128 conform loudness during rendering and playback, even when switching between presets.

When the loudness of the input component signals is correct, the loudness of all possible rendering variations (presets, interactivity) will be accurate as well. Especially members of a Switch Group need to be equal in loudness (e.g. -23 dBFS). A mismatch of loudness within members of a Switch Group for example can cause a dialog loudness jumps when switching languages during playback.

Rendering refers to the process of reproducing object-based audio scenes. It translates audio signals and their object-based metadata into the final audio signals for the target reproduction system (headphones, loudspeakers, etc.) in real time [9]. Metadata for dynamic adjustment of presets, level, and panorama are also taken into account. The MPEG-H renderer enables consistent playback of object-based audio across different output systems.

Important: During the mixing stage, integrated component loudness must meet predefined targets.

When it comes to monitoring it is recommended to focus on the most important preset, which is usually the first 'Default' preset, occasionally checking the other presets and interactivity options.

Loudspeaker monitoring is recommended, but binaural monitoring (e.g. with an additional binaural rendering output) also works well for double-checking, especially when producing for a headphone-targeted audience.

3.2 Encoder for MPEG-H Audio Live Production

The encoder converts the audio components including metadata, that have been created in the AMAU, into a compressed MPEG-H Audio bitstream for transmission to broadcast or streaming. The bitstream can be encoded with an AVC, HEVC, or VVC video codec, with video resolutions of HD, UHD, 4K, or 8K. You can use DASH or HLS as well as MPEG-TS or RF modulation (ATSC 3.0/ISDB-T/SBTVD) for transmitting the encoded signal.

What **Encoder** products are available?

- Main Concept: Live Encoder
- ATEME: Titan Live
- ATEME: Kyrion CM5000e
- ATEME: Kyrion DR5000 (monitoring section, decoder)
- SpinDigital: SpinEnclive
- Sapec: Laguna
- KORG: Live Encoder
- Harmonic: XOS

Contribution Mode is an optional feature that encodes audio components and metadata into a data-compressed, perceptually transparent MPEG-H contribution bitstream. It enables secure, reliable, and robust transport of multichannel signals—such as from an OB van to the playout center—while reducing the number of contribution channels and bandwidth compared to PCM. A contribution decoder is required to reconstruct the audio components and metadata for downstream processing. A contribution bitstream can carry up to 16 audio signals, including the Control Track. Supported transport protocols include SRT, NDI, ST 2110, Zixi, and SDI.

3.3 Other Tools for MPEG-H Audio Production

- MPEG-H Authoring Suite (MAS): Software applications can be used to create templates for AMAUs. Furthermore, can be used in postproduction for authoring, encoding, quality control, and playback.
- Marquise MIST: a Software application for both live and file based QC and conformance checks.
- Spin Digital Spin Player: Software player for MPEG-H bitstreams.
- Wohler iAM-12G-SDI: Monitor for audio signals and metadata.
- ffmpeg: Software for setting up automated workflows.

3.4 Further Information

- <https://github.com/Fraunhofer-IIS/mpegdec/wiki/MPEG-H-UI-manager-XML-format>
- <https://github.com/Fraunhofer-IIS/mpeg-h-audio/wiki>
- <https://mpeg-h.com>

3.5 Contact

More infos and contact: yannik.grewe@iis.fraunhofer.de

References

- [1] ISO/IEC. *ISO/IEC 23091-3 Information technology - Coding-independent code points. Part 3: Audio*. ISO/IEC, Geneva, 2018.
- [2] SMPTE. *SMPTE ST 2110-30:2017: Professional Media Over Managed IP Networks: PCM Digital Audio*. 2017.
- [3] SMPTE. *SMPTE ST 2110-41:2024: Professional Media over Managed IP Networks: Fast Metadata Framework*. 2024.
- [4] SMPTE. *SMPTE ST 2116:2019: Format for Non-PCM Audio and Data in AES3 - Carriage of Metadata of Serial ADM (Audio Definition Model)*. 2019.
- [5] ITU. *ITU-R BS.2125-1: A serial representation of the Audio Definition Model*. 2022.
- [6] ITU. *ITU-R BS.2076-3 Audio definition model*. 2025.
- [7] ITU. *ITU-R BS.2168-0: Audio definition model and serial representation of audio definition model profile for advanced sound systems emission*. 2025.
- [8] ITU. *ITU BS.2088: Long-form file format for the international exchange of audio programme materials with metadata*, volume 0. 2019.
- [9] ITU. *ITU-R BS.2127-1: Audio Definition Model renderer for advanced sound systems*. 2023.
- [10] Yannik Grewe, Christian Simon, and Ulli Scuda. Producing Next Generation Audio Using the MPEG-H TV Audio System. In *Broadcast Engineering and Information Technology Conference*, Las Vegas, 2018. NAB.
- [11] Aimée Moulson, Max Walley, Yannik Grewe, Rob Oldfield, Ben Shirley, and Ulli Scuda. *Object-Based Audio Workflows in Live Sports Using AI-Based Mixing*. 2023.
- [12] Aimée Moulson, Max Walley, Yannik Grewe, Rob Oldfield, Ben Shirley, and Ulli Scuda. Practical Implementation of Automated Next Generation Audio Production for Live Sports. *Journal of the Audio Engineering Society*, 72(7/8):517–525, July 2024.