Technical Brief

Installing the MPEG-H Audio Alliance's New Interactive and Immersive TV Audio System in Professional Production and Distribution Facilities

---

Robert Bleidt
Fraunhofer USA Digital Media Technologies
April12, 2015

MPEG-H AUDIO ALLIANCE

The Next-Generation System for Interactive and Immersive Sound

Features and Four-Stage Process

The Alliance's new MPEG-H[1]-based TV audio system will bring three primary new features to television broadcasts:

- Interactivity to enable consumers to choose different audio presentations, such as a home team or away team commentary at a sports event, or to turn up or down particular audio elements in a program – such as dialogue or sound effects – as they prefer.
- Immersive sound to improve the realism of the sound by adding height channels, by using MPEG-H's Higher-Order Ambisonics mode, or by placing statically panned objects above the listener.
- Multi-platform Adaption. Unlike today's TV sound, the MPEG-H system will tailor playback so it sounds best on a range of devices and environments – from quiet home theaters with speakers to the subway or airport with earbuds.

All of these features will be under the control of the broadcaster or content distributor, providing new creative opportunities, such as the ability to efficiently add additional languages, player or official microphones, or, as the Alliance has demonstrated, car-to-pit-crew radios at races.

Since the MPEG-H Audio system is designed to work over unmodified HD-SDI embedded audio channels, stations can begin implementing MPEG-H Audio features as they choose, without changing their internal plant or operating procedures. The MPEG-H Audio Alliance has proposed a four-stage process for broadcasters to consider when adopting MPEG-H:

1. Transmission of stereo and surround programming using MPEG-H Audio: This will allow broadcasters to gain the bitrate efficiency and new mobile audio features of MPEG-H Audio without any operational changes.

2. Addition of audio objects for additional languages or alternate commentary, enabling viewers to Hear Your Home Team™ audio or listen to their favorite race drivers' radios, as well as providing for mandated access features such as visual description.

3. Addition of immersive sound to improve the realism of the sound image by adding height channels, Higher-Order Ambisonics, or statically panned objects above the listener.

4. Addition of dynamic audio objects: In contrast to static objects fixed in position, dynamic objects move over time to track video action or provide creative effects. If sound effects are to be panned, for

[1] In this paper, "MPEG-H" refers to part 3 of the MPEG-H standard, ISO/IEC 23008-3, not to the other parts concerned with MMT transport, HEVC video coding, etc. More specifically, "MPEG-H" refers herein to the new TV Audio System developed by the MPEG-H Audio Alliance based on the MPEG-H Audio codec.

example, a dynamic object can reduce the required bitrate compared to sending a five- or nine-channel static object.

Adapting live production and playout for MPEG-H: two approaches

In today's television plants, live or real-time video signals are transported using the HD-SDI interface, which supports up to 16 channels of embedded audio. The Alliance's system is designed to use these channels directly for the channels, objects, and other audio elements of a program.

For stages 1 to 3 above, the traditional approach of using a fixed channel map or rundown and fixed metadata values may be used. This approach has the advantage of being easy to understand, and requires very little in terms of operational changes if objects are not used or only a few routine objects are used. This approach is termed the "Fixed Mode" in this paper, although presets can be used under external control to change the encoder settings.

For example, the table below shows possible channel assignments for programs in stereo, surround, and immersive formats:

Format	2.0	5.1	5.1 + 4H	7.1 + 4H
Preset	1	2	3	4
1	Left	Left	Left	L
2	Right	Right	Right	R
3		Center	Center	C
4		LFE	LFE	LFE
5		Left Sur.	Left Sur.	Back L.
6		Right Sur.	Right Sur.	Back R.
7			Left Height	Side L.
8			Right Height	Side R.
9			Back L Height	Top Front Left
10			Back R Height	Top Front Right
11				Top Back Left
12				Top Back Right
13				
14				
15				
16				

This approach would easily allow stage 1 or stage 3 programs to air, just by switching presets within the MPEG-H encoder for the various formats. Loudness, DRC, and downmix values could be set in the encoder semi-permanently for each preset, as is common practice today. In these cases, it

would not be strictly necessary to purchase the Audio Monitoring and Authoring Units described later in this paper.

For stage 2 content with interactive objects, this approach potentially becomes more complex, with possible objects shown below in blue:

Format	2.0	5.1	5.1 + 4H	5.1 + 4H
Preset	5	6	7	8
1	Left	Left	Left	Left
2	Right	Right	Right	Right
3		Center	Center	Center
4		LFE	LFE	LFE
5		Left Sur.	Left Sur.	Left Sur.
6		Right Sur.	Right Sur.	Right Sur.
7	Effects L	Effects L	Left Height	Left Height
8	Effects R	Effects R	Right Height	Right Height
9	Effects Ls	Effects Ls	Back L Height	Back L Height
10	Effects Rs	Effects Rs	Back R Height	Back R Height
11	Player 1	Player 1	Player 1	Driver 1
12	Player 2	Player 2	Player 2	Driver 2
13	Visual Description	Visual Description	Visual Description	Visual Description
14	Away Commentary	Away Commentary	Away Commentary	Driver 3
15	Home Commentary	Home Commentary	Home Commentary	Spanish Commentary
16	Primary Dialogue	Primary Dialogue	Primary Dialogue	Primary Dialogue

In these cases, two issues become evident. One is that, if immersive formats are used, as shown for presets 7 and 8, there are basically four to six channels left for interactive objects. Another is that different programs may require different objects, even though the basic bed channels are the same. Thus, for a football game, the layout might be that of preset 7, while for an auto race, preset 8 might be used.

Using fixed presets certainly is possible, particularly for initial deployment or use only on premier events or programs. For example, a network's normal programming might be set to use preset 2, and preset 4 might be used for a special music awards show mixed in 7.1 + 4H immersive sound, while preset 7 is used for a weekly football game.

If immersive or interactive programming becomes more common, then the strategy of using a fixed channel map or encoder presets becomes more

complex. What about commercial breaks? They can be carried during a program with dialogue mixed in the channel bed or using the same dialogue objects as the program – if they are ingested or prepared that way.

Eventually, a more flexible scheme will be needed to identify content and assign channels. One approach is to associate an XML file with each content item or program in an automation or asset management system, which then is used to set the encoder configuration. This is a sensible approach and perhaps practical in the all file and IP-based TV plant of the future.

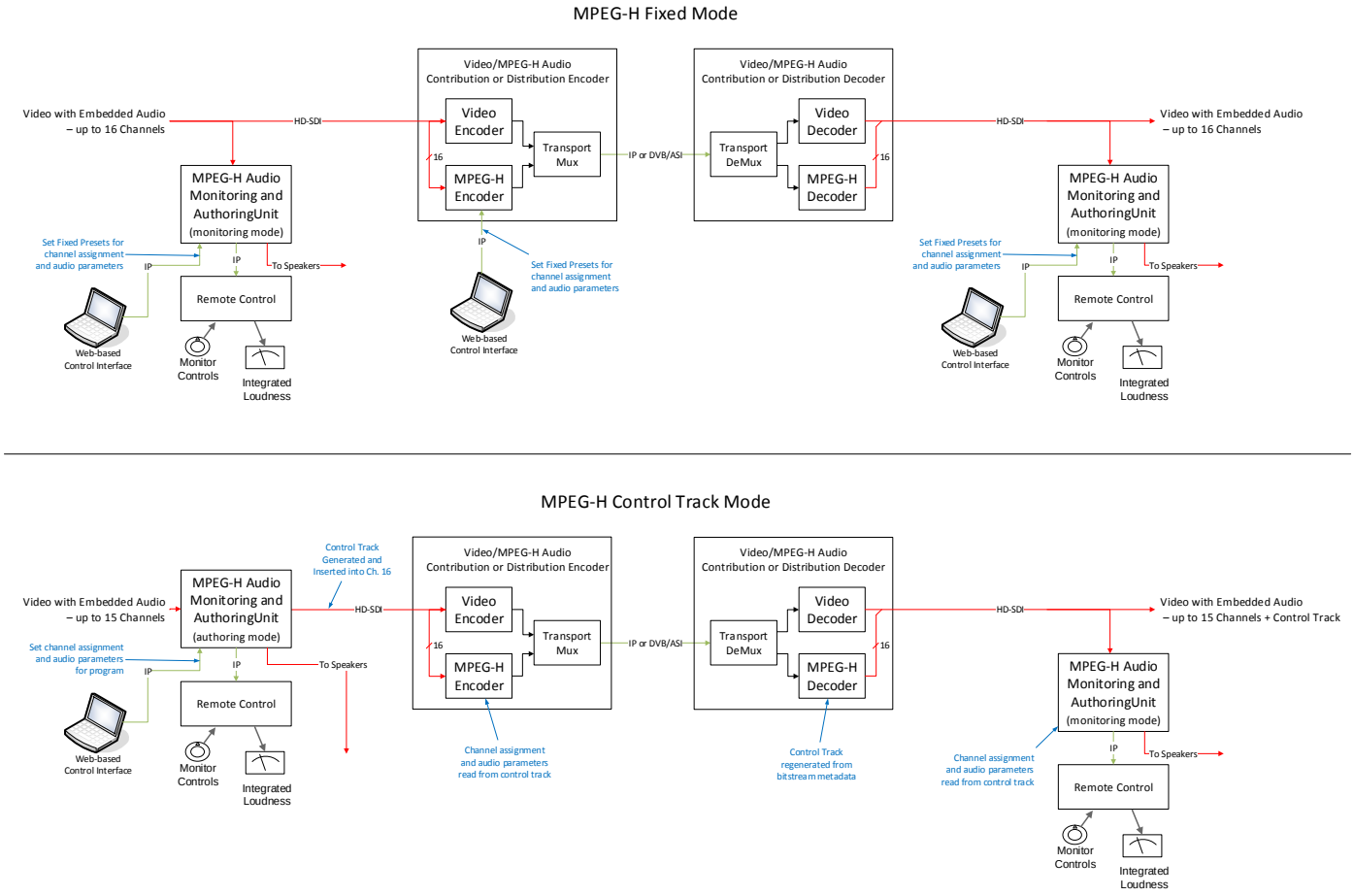

Figure 1. Fixed Mode and Control Track Mode

For today's plants, the Alliance has developed an alternative approach – the Control Track Mode, which uses a Control Track placed on one of the audio channels, usually channel 16. The Control Track:

- contains all the configuration information needed by the encoder, including
    - channel map or rundown
    - object names or labels
    - object groups and control limits

    - program reference level ("dialnorm" in MPEG terminology), downmix gains, and DRC profiles
    - position information for dynamic objects
- may be switched in routing, production, or master control switchers
- will pass through frame synchronizers and other terminal equipment
- may be edited with the other audio tracks in a video editor or audio workstation
- will pass through an audio console with the other audio tracks
- provides frame-accurate transitions of the encoded or monitored audio to match video program switches or edits[2]
- does not require configuring equipment for "data mode" or "non-audio mode" treatment of the control track channel[3]

The Control Track, since it is carried in an audio channel with the content, provides automatic setting of all parameters of the MPEG-H Audio Encoder without any manual programming or need to modify other equipment in the plant. The Encoder translates the Control Track information into MPEG-H audio metadata that is transmitted in the encoded bitstream to the MPEG-H Audio Decoder. This mode of operation is termed the "Control Track Mode."

Professional Decoders may be operated in a contribution or transmission mode, in which they recreate the Control Track signal from the received metadata, or in an emission mode, in which they render the audio channels just as a consumer decoder would.

The Control Track may be generated by an Audio Monitoring and Authoring Unit used by an audio operator for a live program. For ingest of recorded content, either the HD-SDI signal may be passed through an Audio Monitoring and Authoring Unit for adding the control track during real-time dubbing, or file-based utilities may be used to insert the control track into common file formats such as QuickTime/MP4FF or MXF. Of course, the Audio Monitoring and Authoring Unit also uses the Control Track during monitoring to simulate the actions of an MPEG-H Audio Decoder.

Since the control track may be edited just like any other audio channel, programming with different channel assignments or different objects can be combined in an editor just by dropping items on the editing timeline.

Use of the Control Track means one audio channel is no longer available for objects or channels, but it also opens the possibility of using dynamic objects.

---

[2] The Control Track is synchronized to vertical sync to allow easy video editing and switching.

[3] The Control Track is designed to operate just like a longitudinal time code signal. It will survive normal processing of a PCM audio channel, but it cannot be successfully transmitted over a compressed audio channel such as a Layer II contribution codec. When audio compression is required, an MPEG-H Audio contribution encoder may be used, which compresses the audio channels for transmission and converts the control track into metadata carried in the MPEG-H Audio bitstream.

For panned sounds, such as sound effects, several channels of static objects could be required to create the effect that may be produced with a single-channel dynamic object.

The Control Track approach allows full flexibility in the MPEG-H audio modes used during a broadcast day. As shown in our NAB 2015 demonstration, it is easily possible for content having a stereo bed and two dialogue objects to be interrupted by program inserts in full immersive 7.1 + 4H sound, or even Higher-Order Ambisonics, interspersed with commercial breaks in stereo or 5.1 surround.

One new possibility shown is the ability to broaden the reach of commercials to include demographics that are more comfortable listening to advertisements in their primary languages. Local spots intended to reach the broadest possible audience could have voiceovers or dialog in several languages selected by the advertiser. The Preferred Language feature of the Alliance's system will present a commercial in a viewer's preferred language, if broadcast, and automatically switch back to a default language for other programming or commercials that do not have the preferred language present.

With certain restrictions on content transitions, primarily during network break and join operations, it is possible to have a mixture of new content with the Control Track signal and legacy content without it. For example, when no Control Track data is present, an MPEG-H Audio Encoder and MPEG-H Audio Monitoring and Authoring Unit can be set to switch to 5.1 surround mode, with a fixed loudness of -24 LKFS and standard downmix gains and DRC profiles, as a facility typically uses today. In this manner, legacy content would be encoded as it is today, and new content with immersive or interactive features automatically would be encoded with the correct settings.

Hybrid Approaches for Multi-Tier Distribution

Networks that include an affiliate model, or that operate multiple distribution facilities, may contemplate operating MPEG-H in either the Fixed Mode or the Control Track Mode in both network and affiliate plants. A professional MPEG-H Audio Decoder may be configured to operate in Fixed Mode, outputting 16 channels of audio, or in Control Track mode, outputting 15 channels of audio and the Control Track signal.

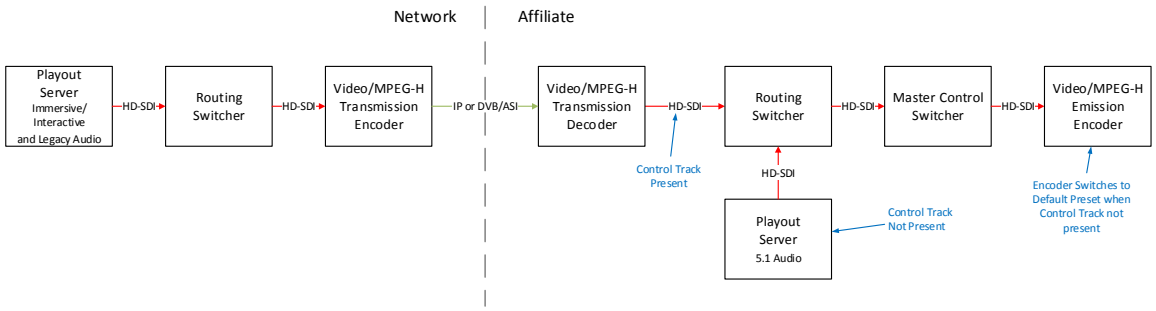

Figure 2. Multi-tier Distribution

Of particular interest is the possibility to operate network distribution in the Control Track mode and offer affiliates the possibility to continue to operate in Fixed Mode until they begin producing complex content, as illustrated in the diagram above. Network programming may be distributed using all the features of MPEG-H, but the affiliate can continue to operate in Stage 1 with legacy 5.1 local programming. The emission encoder will switch to a 5.1 (or even stereo) preset when its input signal does not have the control track present.

Program or Commercial Insertion in the Compressed Domain

Today, many MVPDs and at least one broadcast network insert commercials in an already-encoded audio and video bitstream, typically using SCTE 35 messages to signal insertion points in an MPEG-2 transport stream. Such "stream splicing" insertion of programming or commercials is easily possible with the Alliance system, with frame accuracy and seamless audio transitions, due to special features incorporated in the system design.

Next-generation TV standards currently under development may replace MPEG-2 transport with other technologies such as MPEG-DASH, ROUTE/DASH, or MPEG-H MMT. The Alliance system will operate seamlessly using MPEG-DASH transport, and the Alliance, of course, will work in the standards process to ensure that splicing is possible with the specific technologies standardized.

Extending the Limits of Audio Transport

As productions become more sophisticated in the future, the limit of 15 or 16 active audio channels with HD-SDI transport may become an issue. Several approaches are available to extend the number of channels carried. One is to use two HD-SDI channels, or to use MADI or AES 67 audio transport. Of course, plants upgrading to 1080p or 4K video will need to use at least 3G-SDI signals, which provide 32 channels of audio. Eventually, as plants migrate to all-IP, file-based transport and playout, only the -channel or 128-object limits of MPEG-H will constrain programs.

It is important to remember, however, that each channel and object in the Alliance's system will require 20-40 kb/s of bitrate in emission. For superior audio quality, a program using the full HD-SDI 16-channel layout might need 600 kb/s of emission bitrate.

Another approach is to encode the audio signals with an MPEG-H Audio Encoder in contribution mode and to transport the encoder output over a pair of PCM channels operating in "data mode" or "non-audio mode," similar to the manner in which Dolby E signals are carried today. Doing so is theoretically possible but creates the same operational issues as Dolby E transmission does. In such cases, all equipment in the video signal path has to be programmed for data mode on the pair of channels used. Devices such as audio embedders or frame synchronizers must have sample rate conversion or audio resampling disabled. Any master control switchers or router clean switches must have audio crossfading or "V" fading disabled. Any monitoring equipment, even for casual monitoring, must be equipped with an MPEG-H decoder. Routing the signal through an audio console is almost impossible, and the audio must be decoded for any audio editing. For these reasons, the Alliance system currently does not provide this mode of

operation, given the advantages of the Control Track approach or of simply moving to alternative audio transport standards.

Adapting audio consoles and live production equipment for MPEG-H

Very few consoles and digital audio workstations are equipped today to handle immersive audio production directly. They have a maximum bus size of 5.1 or 7.1 channels, and their panners are designed for two-dimensional surround panning. In immersive film production, the trend has been to use the busses of the console for a two-dimensional "bed" adapted from surround mixing, and to use many dynamically panned objects for any sources with positions above the horizontal plane. In broadcasting, the bandwidth for dynamic objects is more costly, and the preferred approach is to mix a three-dimensional "bed" and use objects primarily for interactivity and dynamically panned sounds.

Because consoles have only 5.1- or 7.1-channel busses, they only have 5.1- or 7.1-channel monitoring capability as well. Static and dynamic objects also must be rendered in three-dimensional space and converted to loudspeaker signals. Thus, outboard equipment is needed to adapt the console for panning, object rendering, and monitor control. In the MPEG-H Audio system for live production, this is done with the MPEG-H Audio Monitoring and Authoring Unit. The unit also handles loudness monitoring and allows the audio mixer or operator to author appropriate Control Data for the program.

As shown in the diagram below, a surround broadcast can take the first step towards testing MPEG-H's new features by routing interactive sources such as dialogue (either as individual channels or as sub-mixed busses) over direct outputs or effects sends to an MPEG-H Audio Monitoring and Authoring Unit. The 5.1-channel bed then is created by making a "mix-minus" bus containing the entire program except the interactive sources, the outputs of which also are sent to the monitoring unit. Typically, the monitoring unit is simply connected to the console's MADI output and appropriate routing is set up in the console's routing matrix.

The monitoring unit also receives an HD-SDI video or black signal input to synchronize the control track, if used, to vertical sync. The monitoring unit then combines audio, control track, and video in an HD-SDI output signal that feeds a contribution video encoder for remote production.

For production in a studio or control room, an HD-SDI or MADI output containing only the audio and control track or an output with the video included can be used, depending on the facility architecture. The combined HD-SDI output is suitable for direct connection to a routing switcher, video server, or other plant equipment.

Monitor speakers are connected to the monitoring unit's speaker outputs. This is necessary to render the interactive objects back into the mix. Of course, the A1 or audio mixer also can use the console in this case to mix in the isolated channels for IFB feeds or other purposes. The monitoring unit will allow the mixer to listen to each presentation authored for the program, such as the main broadcast or network dialogue, secondary language dialog, or home and away team commentary.

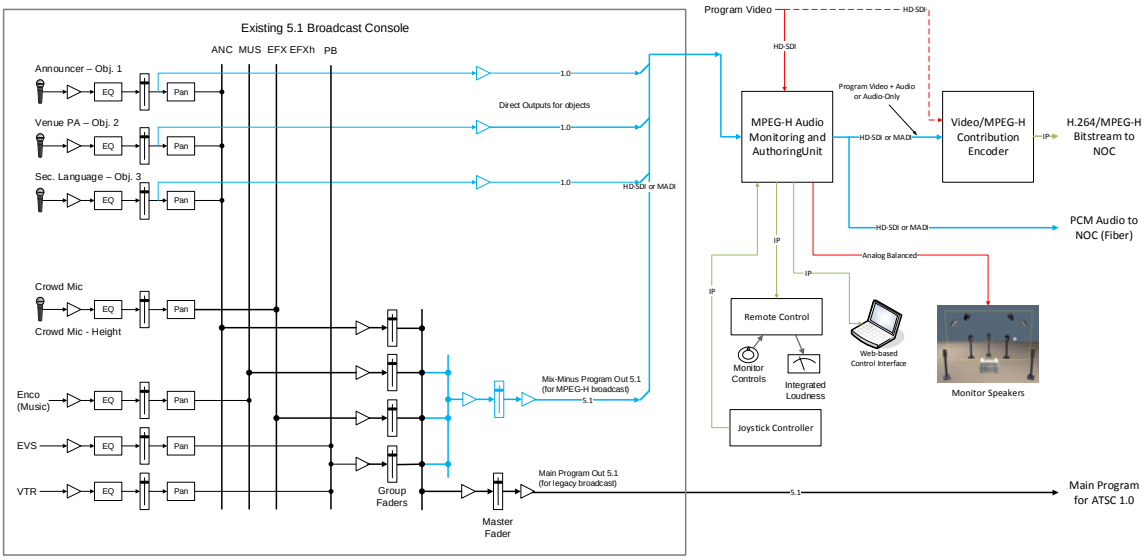

Figure 3. Extending 5.1 Console for Live Interactive Production

In the photo below, common commands of the monitoring unit have been routed to an intelligent button panel for rapid use. These are divided into the following categories:

Monitor level:

- Mute (disabling monitor speaker output)
- Dim (reducing speaker output level by an adjustable amount)
- Mix (adjusting speaker volume to a stored level, to normalize output to a specific listening level)

Downmix Control:

- 2.0 (stereo downmix)
- 5.1 (surround downmix)
- 5.1 + 4 (no downmix)

DRC Profile Auditioning

- Loudness (target loudness normalization, default -24dbFS target level, user adjustable)
- AVR (no dynamic range control applied, -31 dbFS target level)
- TV (light dynamic range control applied, -24 dBFS target level)
- Mobile (strongly compressed dynamic range, -16 dBFS target level)
- Noisy (extremely compressed dynamic range, with dialog accentuated, -16dBFS target level)

Presentation Auditioning

- Network
- Venue
- Norwegian

(These are presets authored for the particular show portrayed in the photograph. In normal use, names displayed will be those set up in the monitoring unit's web interface.)

Mode indication

Mix: Authoring Mode, Listen: Monitoring Mode, Pass: Bypass Mode

5.1 + 4H + 4o (5.1 surround mid-layer speakers, 4 top-layer speakers, 4 objects)

Figure 4. Junger MPEG-H Audio Monitoring and Authoring Unit Remote Control

The monitoring unit remote control also offers a volume knob for monitor level and comprehensive loudness metering facilities. As configured in the photo, the center numeric displays show the integrated and short-term loudness of the channel bed and that of each isolated object, as well as the loudness of each authored presentation, from left to right respectively[4]. Pressing the buttons below will activate a Solo mode for listening to each metered item individually. The bar graph displays to the left of the unit provide quick visual feedback of the short term loudness of the corresponding numeric displays. This allows the operator to quickly see if presentations are above or below the desired target level, which is indicated by the center green segment of the bar. The "---" button allows the integration period of the integrated loudness measurements to be reset.

To mix immersive broadcasts, the previous diagram is extended as shown below.

[4] Loudness regulations for immersive or interactive TV audio have not been established yet, although the ITU is considering BS. 1770-3. Other measurement techniques could be used for loudness measurement, such as measuring the loudness of a 5.1-channel downmix using only the default presentation.

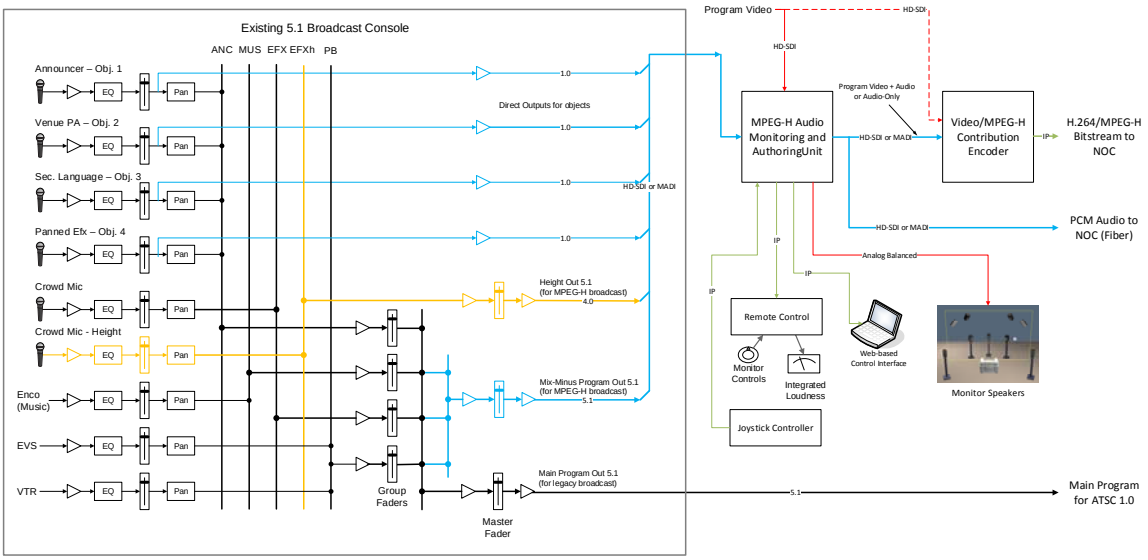

Figure 5. Extending 5.1 Console for Live Interactive and Immersive Production

In addition to ambience microphones routed to the "EFX" bus in the console, additional microphones placed to deliver immersive ambience are routed to an "EFXh" bus. The EFHh bus is then sent separately as a 4.0 or 5.1 bus to the Audio Monitoring and Authoring Unit, where it becomes the height channels of the MPEG-H broadcast signal. It is possible in an indirect way to pan sources between the middle and top channel layers by using a traditional 5.1 panner to control the horizontal position, and an effects send to control the level of the signal sent to the EFX and EFXh buses, for example.

It is also possible in any channel configuration – stereo, surround, or immersive – to use dynamic objects panned to follow action or for creative effect. In the diagram, object 4 is used to send a mono object to the monitoring unit, where panning data from the joystick controller is added to the Control Track signal.

If a show is simulcast in both ATSC 1.0 and 3.0, the monitoring unit may be set to Bypass Mode so that the main console outputs are sent directly to the speakers for monitoring the 5.1 ATSC 1.0 program output.

Creating Derivative Mixes or Programs Using MPEG-H

Often a broadcaster needs to create more than one audio mix from an event. Perhaps a stereo version is needed for legacy internet streaming, or music used in the show is not licensed for an international feed and must be replaced, or a "clean feed" version of the show must be distributed.

Producing a show in MPEG-H makes this process much easier. In a 16 channel SDI feed, at most two 5.1 surround programs can be transmitted today. Other feeds must be stereo given the limited number of channels available. The table below shows an example typical of a complex broadcast. With traditional channel-based audio, only the complete main program can be transmitted in surround. The other feeds must be sent as stereo. Using

MPEG-H and sending the effects as objects would allow all feeds to be available as surround programs.

Format	Traditional Audio	MPEG-H Audio with Objects
1	Complete Main L	Channel Bed Left
2	Complete Main R	Channel Bed Right
3	Complete Main C	Channel Bed Center
4	Complete Main LFE	Channel Bed LFE
5	Complete Main Ls	Channel Bed Left Sur.
6	Complete Main Rs	Channel Bed Right Sur.
7	International Feed L	Effects L
8	International Feed R	Effects R
9	Alt. Main with Different Music L	Effects Ls
10	Alt. Main with Different Music R	Effects Rs
11	Natural Sound Only L	Main Music L
12	Natural Sound Only R	Main Music R
13	Complete Main Stereo L	Alt Music L
14	Complete Main Stereo R	Alt Music R
15		Primary Dialogue
16		Control Track

MPEG-H in File-Based Workflows and Post Production

As the Alliance's system is designed to operate using standard PCM audio tracks, embedded in HD-SDI or carried over MADI, AES 67, or other audio transport standards, it can be recorded on audio and video servers directly, without any modifications. Since the audio is maintained as uncompressed PCM samples, there is never any generational loss as would be encountered with mezzanine compression encoding. Programs produced in the live scenarios discussed above can be recorded on standard video servers and archived or distributed in the video servers' file formats.

In a similar manner, content can be edited in standard video editors, such as Avid Media Composer or Adobe Premiere Pro, using up to 16 audio tracks, and output in common file formats such as MOV/MP4FF or MXF. Each audio track can be auditioned, scrubbed, or modified in the editor directly, unlike the situation with mezzanine encoded audio (Dolby E, for example). For

content using the Control Track, it can be edited along with the audio and video and in the same way, being placed on one of the audio tracks.[5]

In the immediate term, the output of audio edits in a video editor may be auditioned by feeding the output of the editor over HD-SDI or MADI to an MPEG-H Audio Monitoring and Authoring Unit. Similarly, the monitoring unit may be used in authoring mode to create a Control Track with presentations, object labels, default object gains, ranges, and positions, downmix gains, dynamic range profiles, loudness levels, and other metadata. The Control Track then can be recorded in the editor and dropped onto the timeline for the content being edited. In the near future, the Alliance intends to make available plug-ins for generating the Control Track directly in popular video editors and audio workstations.

Unlike in video editing or in live production using existing 5.1-channel consoles, audio mixing in a studio, post-production, or re-recording environment needs a full suite of capable tools to manipulate tens or potentially hundreds of simultaneous sources at once, under automation control. While it is possible to extend hardware devices such as an MPEG-H Audio Monitoring and Authoring Unit to connect to post-production hardware consoles, or interface with digital audio workstation software, this has not been the approach used by the Alliance for production of its demonstration content. Most of this content has been created or modified using standard DAW software such as Pro Tools or Nuendo, by adding plug-ins for immersive panning and reverb developed internally or with close partners.

Four cases exist for panning signals during a mix session:

1. Panning sources to channels. This is the traditional panning operation, where a microphone, instrument, or other signal source is mixed to several channels with a "pan pot." Since immersive sound requires panning in three dimensions, a 3D panning plug-in is needed to perform this function.
2. Panning sources to HOA. This is a similar operation to (1), but the source is panned to HOA components instead of channels. This only requires the panner plug-in to use HOA spherical harmonic basis functions instead of traditional panning law functions.
3. Panning objects statically. As objects are maintained as separate channels, strictly speaking, a panner is not required, as the positioning is done by metadata authoring. As a practical matter, the panning plug-in has to render the object in its desired position for monitoring during mixing.
4. Panning objects dynamically. This is similar to (3), except the object position moves over time, usually controlled by an external joystick or by adjusting automation track or lane values. In this case, the panned position must be sampled periodically and used to generate metadata stored in the Control Track.

[5] It is desirable to lock the audio track containing the Control Track when applying dissolves or fades to the audio. There are settings available in the system to allow the Control Track to be faded at transitions.

An audio mixing session intended for MPEG-H broadcast may be exported from Pro Tools or Nuendo as follows:

1. Mixdown to a set of tracks representing the channels and objects in the program.
2. Read the automation data for dynamic object panning in a plug-in and generate Control Track
3. Combine the Control Track with the other tracks, either in the DAW or with external tools.
4. Generate a multichannel WAV or BWF file for import into a video editor, or multiplexing with a finished video track.

The Alliance plans to make its internal plug-ins available through commercial partners in the near future.

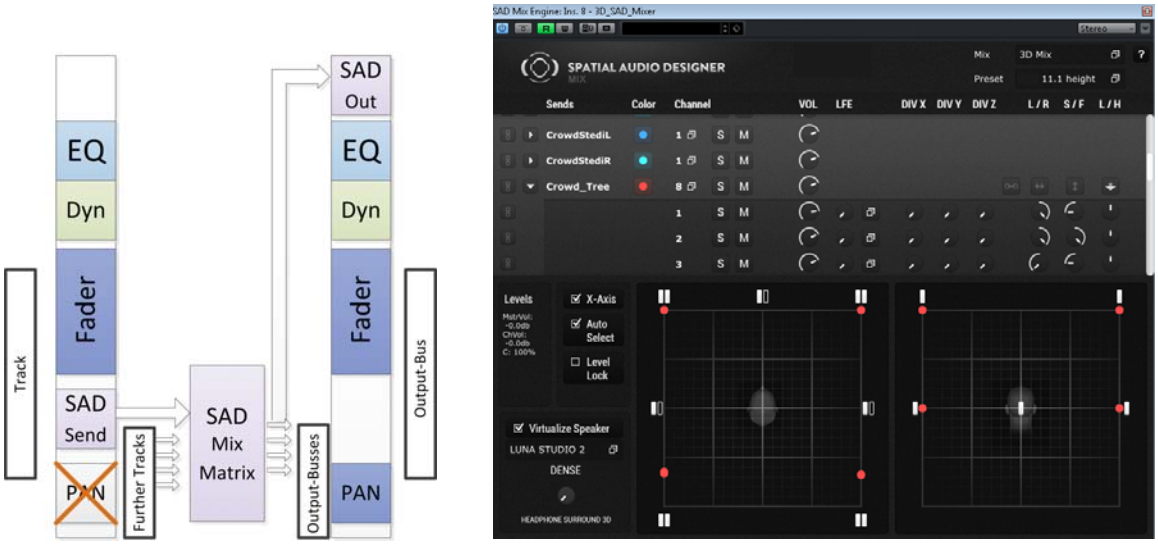

Figure 6. Spatial Audio Designer Plug-In from New Audio Technology, GmbH

Finally, archiving or interchange of content between facilities must be considered. MPEG-H bitstreams – streamed live or stored in the MPEG-4 File Format, contain all the audio essence and metadata information needed for a program and are an international standard. In MP4FF files, the essence may be compressed using the MPEG-H audio codec or stored as uncompressed PCM samples.

MP4FF files are an attractive way to store audio mixing sessions or projects, with or without video. Another developing alternative is the EBU / ITU BWF format, extended by its new ADM object model. MPEG-H Audio's internal object metadata is very similar to the ADM model and translation to and from BWF-ADM is easily possible.

Of course, a practical method of archiving and exchange is to use the project or session files generated by DAW tools themselves, as long as similar tools are used at the receiving end.

Example System: Alliance NAB 2015 Live Broadcast Demonstration

A working example of how MPEG-H may be integrated into facilities is the demonstration shown by the MPEG-H Audio Alliance at the 2015 NAB Show. The demonstration is a complete simulation of a live broadcast, including mixing the audio live in a remote truck, combining the remote content with recorded programming from video servers at a network operations center, distribution to affiliate stations, insertion of local commercials, and emission to viewers' living rooms.

The truck, network, and affiliate systems all operate in the MPEG-H Control Track Mode. MPEG-H dynamic audio objects are used live on the air to carry sound effects while static objects are used for carrying English, foreign language, and venue PA commentary. The audio bed is mixed in the 5.1 surround plus 4 height speakers (5.1+4H) immersive format.

The audio and video from the event pickup is sent from the remote truck to the network operations center, where the live audio and video are combined with other programming stored on standard video servers. This programming includes audio formats ranging from stereo to 5.1 surround, to 7.1 + 4H and Higher-Order Ambisonics immersive sound.

Programming from the network is sent to a local affiliate station where local advertisements are added in formats ranging from stereo to immersive sound. The affiliate's TV signal is then transmitted to a Technicolor MPEG-H Audio-enabled set-top box in a consumer's living room for playback in the 7.1 + 4H speaker configuration.

In the remote truck, pre-recorded microphone signals from an extreme sports event are mixed live on an unmodified Calrec Artemis console adapted for interactive and immersive sound using a Jünger Audio MPEG-H Audio Monitoring and Authoring Unit. The remote truck output feeds a Fraunhofer prototype MPEG-H Audio and H.264 video contribution encoder.

The remote truck signal is received in the Network Operations Center for the "The MPEG Network" using a Fraunhofer prototype MPEG-H Audio and H.264 contribution decoder. The output of the decoder flows through a standard, unmodified Lawo frame synchronizer and an Evertz HD-SDI router[6] to the NOC's Master Control position, where the truck signal is switched into recorded programming from an Abekas video server under Aveco automation control. The switched program is then input to a Fraunhofer prototype distribution encoder for transmission to affiliate stations.

An MPEG Network affiliate, WMPG-TV, receives the transport stream from the network with a Fraunhofer prototype distribution decoder. The HD-SDI signal is then fed through a Lawo frame synchronizer and input to an Evertz HD-SDI router. Stored local commercials are inserted under automation control into the signal and fed to the Fraunhofer emission encoder.

The emission encoder output is then fed to a Technicolor set-top box in a simulated consumer living room. The signal from the local affiliate also may

[6] Master Control switchers could have been employed in the demonstration for "clean switch" switching instead of the abrupt switch of a routing switcher. The Alliance system is designed to work in both cases.

be received over an Internet connection by a tablet computer. On-screen displays on both devices allow the viewer to select audio presentations prepared by the broadcaster, or directly control the audio elements within limits set by the broadcaster. Additionally, content may be played on the Fraunhofer prototype 3D Soundbar or a prototype AVR based on the Texas Instruments DA830 DSP.

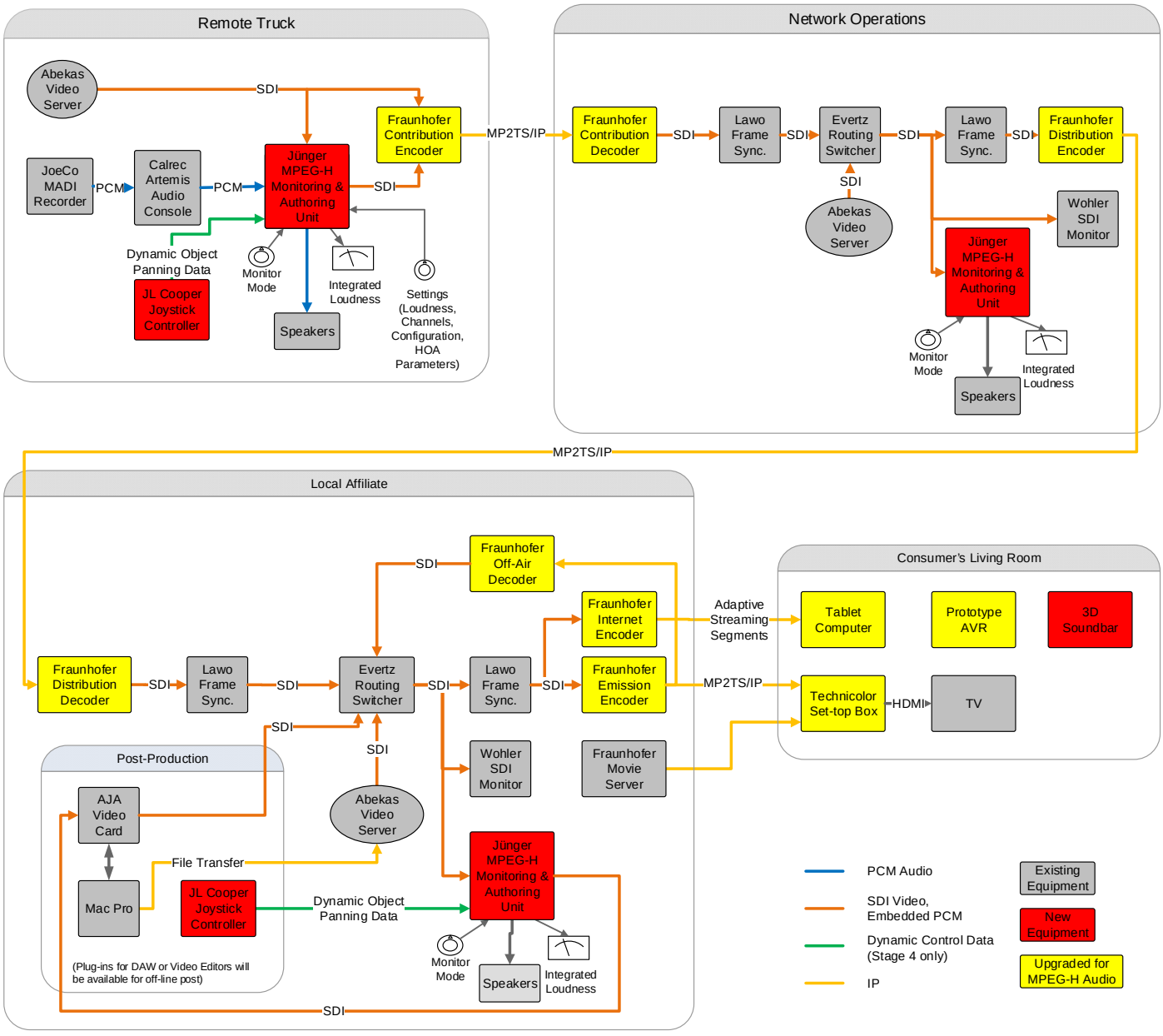

Figure 7. Alliance NAB 2015 Live Broadcast Demonstration