Considerations for the MPEG-H Audio Standard

This article discusses the plan for the new MPEG-H Audio standard, a real-time 3-D audio encoder system with applications for broadcast and streaming distribution, developed in large part by the Fraunhofer Institute in Germany, and recently finalized and submitted as an ISO/MPEG standard.

By
João Martins
(Editor-in-Chief)

MPEG-H is a group of standards under development by the ISO/IEC Moving Picture Experts Group (MPEG) for a digital container standard, a video compression standard, an audio compression standard, and two conformance-testing standards. The group of standards is formally known as ISO/IEC 23008—high-efficiency coding and media delivery in heterogeneous environments. MPEG-H Audio is just the third part of 13 document chapters and basically covers the codec for 3-D audio, which supports: an object-audio codec for "interactive sound mixing;" a realistic audio experience with additional front- and rear-height speaker channels to traditional surround sound setups; and higher order ambisonics (HOA), a technique for reproducing a complete spherical soundfield to provide a fully immersive live sound experience.

The Applications

The basic MPEG-H audio technology, which is under discussion for publication as an ISO/IEC international standard, results basically from cooperation between the Fraunhofer Institute for Integrated Circuits IIS, Technicolor and Qualcomm.

Much of the research work followed demonstrations of new emerging immersive formats (e.g., Dolby Atmos and Auro-3D), and the direct commitment from several organizations including the European Broadcast Union (EBU/UER) and the Society of Motion Picture and Television Engineers (SMPTE). The organizations are promoting new "enhanced" audio experiences and technologies, including live rendering of audio objects for multichannel sound and binaural reproduction.

Since 2011, those efforts have also involved several prestigious European institutions, from the British Broadcasting Corp. (BBC) to the Fraunhofer IIS in Germany, which is also the provider of the HE-AAC multichannel audio codec powering half the world's TV surround sound today.

While Dolby Atmos is a format directly promoted for cinema production and exhibition, that later became also used in home-theater application, the before mentioned institutions are looking at consumer delivery of immersive audio in broadcast and streaming media on any type of distribution network—from direct Internet streaming to cable/IPTV operator platforms. To promote the MPEG-H standard, Fraunhofer IIS, Qualcomm Technologies, and Technicolor created the MPEG-H Audio Alliance.

In the MPEG-H Audio Alliance perspective, the MPEG-H Audio standard enables live broadcasts with object-based 3-D audio across all devices, providing viewers the ability to tailor the audio to suit their personal listening preferences. The system encodes elements of the audio as interactive objects so viewers at home can adjust the sound to their preference while also enhancing today's traditional multichannel surround sound broadcasts to create a more realistic audio experience.

From Object-Based to Speakers

In this context, it is important to understand how to differentiate the various channel-based

Pictured from left to right are Robert Bleidt, Division General Manager at Fraunhofer USA Digital Media Technologies and Claude Gagnon, Technicolor's SVP, Content Solutions & Industry Relations.

audio delivery solutions from the object-based audio concept. According to Nuno Fonseca (Instituto Politécnico de Leiria, Portugal), one of the leading researchers in the area of object-based audio, "Immersive sound is the designation that was given to the existence of sound in a 3 dimensional world, including a height dimension." Immersive audio can follow the channel-based and object-based approaches. In channel-based audio "all sound sources are mixed to output channels, where each output channel corresponds to a speaker (or an array of speakers) that will exist at a predefined position. Object-based audio uses a different approach, using metadata. Instead of using channels with predefined positions, each audio object will have the desired audio stream and metadata with time varying information on how that audio stream should be played (position in space, apparent size, etc.)."

Thanks mostly to Dolby's efforts in the promotion of its Dolby Atmos format for cinemas and the home, new delivery solutions with support for the

Delivering Multi-Platform Immersive Audio with the New MPEG-H Audio Standard

By
Robert Bleidt
(Fraunhofer USA)

New standards for TV audio and new content offerings from streaming media providers will soon provide consumers with immersive sound in their living rooms. Immersive sound is a step beyond today's surround sound, offering sound from above to complete the auditory illusion of being in the audience or at the performance instead of at home.

Immersive sound is being deployed in cinemas today, with Dolby's Atmos and Auro's Auro-3-D systems the most popular, while DTS has indicated it will re-enter the cinema market with the DTS:X system. Derivatives of these systems are also finding their way into Blu-ray releases as extensions to existing codecs in the disc's specification.

However, even the Blu-ray versions of these codecs require several megabits of data to deliver immersive audio. This is not practical for TV broadcast or Internet streaming, and for these purposes next-generation audio codecs must be used.

One of these codecs is the recently approved MPEG-H Audio Standard from the MPEG audio group that developed the MP3 and the AAC family of codecs. An audio system built around this codec is being developed by the MPEG-H Audio Alliance of Fraunhofer, Technicolor, and Qualcomm. It is now being proposed for new streaming services as well as being evaluated for the new ATSC 3.0 TV broadcast standard for the US.

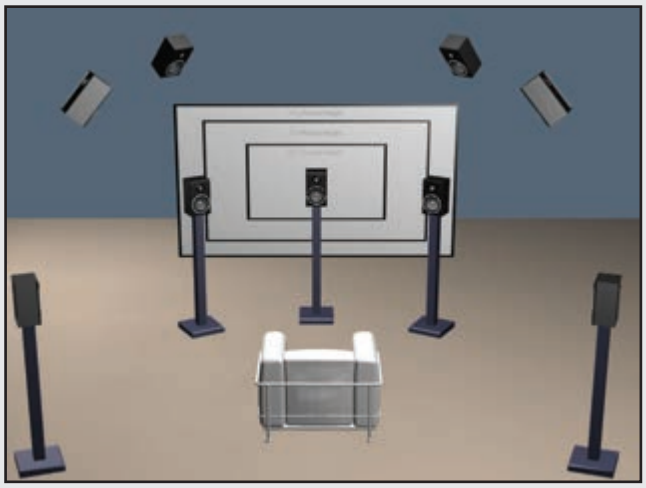


Photo 1: The 5.1 + 4 H speaker configuration is shown with four overhead speakers. A 7.1 + 4 H configuration is also possible, and for limited hardware only two front channels may be considered.

Although the standard is capable of carrying up to 128 channels of audio, and loudspeaker configurations are defined up to 22.2 channels, in practical use in the Alliance system, the

use of today's 5.1 or 7.1 surround speaker layouts is proposed, with four overhead corner speakers. Listening tests at Fraunhofer have shown the 5.1 + 4 overhead speaker arrangement can make a substantial improvement in the perceived realism of sound, almost as great as the step from stereo to 5.1 surround (see **Photo 1**).

Systems for playback of Blu-ray immersive audio were shown at CES 2015 and have begun to enter the market. Most of them use additional drivers in the speaker enclosure aimed to reflect sound for the height channels off the ceiling of the listening room. Add-on bounce speakers for existing 5.1 or 7.1 installations have also appeared. These systems will provide immersive, though diffuse, overhead sound in many rooms, but they represent a compromise that still requires additional wiring and setup.

While ceiling-mounted loudspeakers offer the ultimate immersive sound quality, the consumer trend toward flat-screen displays have led them to abandon discrete speakers for soundbars. This has actually reduced the possible sound quality, as almost all soundbars are stereo devices. Fraunhofer has shown a potential solution to this problem with 3-D soundbars that have additional speakers and acoustic processing to enable an array of speakers surrounding the TV to produce realistic immersive sound (see **Photo 2**). This offers the mainstream consumer a way experience good immersive sound without any additional wiring or configuration—a décor-friendly, hang it and hear it experience.

Another trend is toward consumption on tablets and other mobile devices, even within the home. This represents a difficult challenge for audio reproduction given the typical 16-mm speakers in tablets, likely with a resonant frequency in the 600-to-800-Hz range, and often only exposed to the listener through rear-firing grilles or even internal ducts for mechanical convenience. Fraunhofer has had experience in addressing these challenges through psychoacoustic processing, perhaps applied in combination with specialized speaker driver circuits that extract the maximum possible excursion from the speaker.

One example is the Fraunhofer Cingo technology that appears in Google's Nexus tablets and now in the Gear VR virtual reality headset adapter for Samsung phones. Fraunhofer has extended this technology, originally developed for rendering surround sound on tablet speakers and headphones, to be able to recreate a realistic sound image for immersive audio as well. Another technology is the binaural rendering function built into the MPEG-H codec itself for headphone reproduction.

Thus, with these delivery modes—speakers, soundbars, and tablet software—MPEG-H can deliver immersive sound to the universe of devices consumers may use to experience new TV and video programming. Additionally, we have included functions in the system to enable the dynamic range to be tailored to the listening environment, moving beyond the "night mode" setting of today's codecs, to multiple dynamic range settings with different target levels for different use cases.

These different target levels are needed to account for the fact that, on mobile devices, the same volume control is used to set the loudness for all content. At home in the living room, a consumer may decide to play music from his phone,

Photo 2: This is an early concept prototype of a 3-D soundbar. Commercial products would be smaller and perhaps shaped more like a conventional soundbar.

Internet radio, or media server, but with a different volume setting for each. On a phone or tablet, the operating system does not distinguish between TV programming recorded with normal headroom, and music which is now uniformly compressed, limited, and clipped to maximize loudness, typically 10 to 15 dB louder. A compromise solution to this difference is to boost the level of video programming on mobile devices to match that of music.

Both of these issues—reducing dynamic range and matching video to music loudness—are addressed in the MPEG-H system through additional sets of dynamic range control coefficients sent in the bitstream. The system also offers special processing of the dialog to increase the intelligibility in noisy environments.

MPEG-H also offers three ways to carry immersive sound—traditional channels, audio objects, and higher-order ambisonics. More information on the MPEG-H Audio Alliance system is available at www.MPEGHAA.com.

About the Author

Robert Bleidt is Division General Manager of Fraunhofer USA Digital Media Technologies. He is the inventor of the award-winning Sonnox/Fraunhofer codec plug-in, widely used in music mastering, led the extension of Fraunhofer's codec business to an open-source model by inclusion in Android, and developed Fraunhofer's Symphoria automotive audio business. Before joining Fraunhofer, he was president of Streamcrest Associates, a product and business strategy consulting firm in new media technologies. Previously, he was Director of Product Management and Business Strategy for the MPEG-4 business of Philips Digital Networks and managed the development of Philips' Emmy-winning asset management system for television broadcasting. He also worked for Sarnoff Real Time Corp. and was president of Image Circuits, a consulting engineering firm and manufacturer of HDTV research equipment.

Fraunhofer's 3-D TV Audio System enables viewers to personalize sound broadcasts to suit their personal preferences. With a click of the remote, they can create their favorite mix of sports broadcasts. (©Getty Images/Fraunhofer USA)

MPEG-H audio format will be able to translate the additional height element, using overhead speakers or up-firing additional drivers. Of course, Dolby Atmos is more than simply adding ceiling speakers. It effectively combines complete authoring, distribution, and playback tools, including improved audio quality and timbre matching, as well as greater spatial control and resolution with which, the immersive or "3-D" perception of audio can benefit from developments in the area of real-time encoding of audio objects.

Previously, several other sound reproduction systems explored additional dimensions in sound reproductions, from ambisonics to 3-D Vector Base Amplitude Panning (VBAP) and Auro-3D. The MPEG-H 3-D audio standard could effectively have a lot to gain from the consumer's adoption of Dolby Atmos solutions (or the new DTS:X format) in helping to introduce elevated loudspeakers to create a real 3-D audio impression, since those loudspeaker setups are able to deliver higher spatial fidelity than the established 5.1 setup.

Of course, the MPEG-H audio project is also intended to solve additional challenges, which were not addressed by Dolby Atmos or other "immersive audio" solutions. Considerations such as production compatibility between different immersive 3-D audio formats, or efficient distribution to existing channels and delivery systems, including a single "3D" loudspeaker setup.

The MPEG-H Audio Effort

From earlier work dating back to 2011, with initial presentations on 3-D audio at the MPEG meetings, discussions around 3-D audio lead to a "Call for Proposals" for such new 3-D audio technologies in January 2013.

The requirements and application scenarios specified for the new technology included bit rate parameters and compatibility of rendered audio on various loudspeaker setups from 22.2 down to 5.1, plus binauralized rendering for virtualized headphone playback. The document also specified that submissions would be independently conducted for "channel and object (CO) based input" and "Higher Order Ambisonics (HOA) input content types." At the 105th MPEG meeting in July/August 2013, the Reference Model technology was selected from the received submissions (four for CO and three for HOA) based on their technical merits, to serve as the baseline for further collaborative technical refinement of the specification. The winning technology came from Fraunhofer IIS (the CO part) and Technicolor/Orange Labs (the HOA part). In a subsequent step, both companies agreed to merge the proposals into a single harmonized system and work together in areas such as binaural rendering. The final stage of the specification, which is the submission of a document for approval as an international ISO/IEC standard, was issued at the 111th MPEG meeting in Geneva, Switzerland, in February 2015. The documents have been available for public discussion on The MPEG website, http://mpeg.chiariglione.org/standards/mpeg-h.

Trials

With the MPEG-H standard complete and undergoing field tests and deployment, there is a practical solution to translate object-based audio productions into existing delivery multichannel systems, from broadcast chains to home-theater setups.

The MPEG-H standard enables high-efficiency coding and media delivery of high-quality 3-D audio content, independently of the number of loudspeaker channels used for reproduction using an appropriate coding and rendering algorithm. Based on the AAC codec family and new technologies, this system enables an easy upgrade path of existing infrastructures when adopting MPEG-H without major equipment changes.

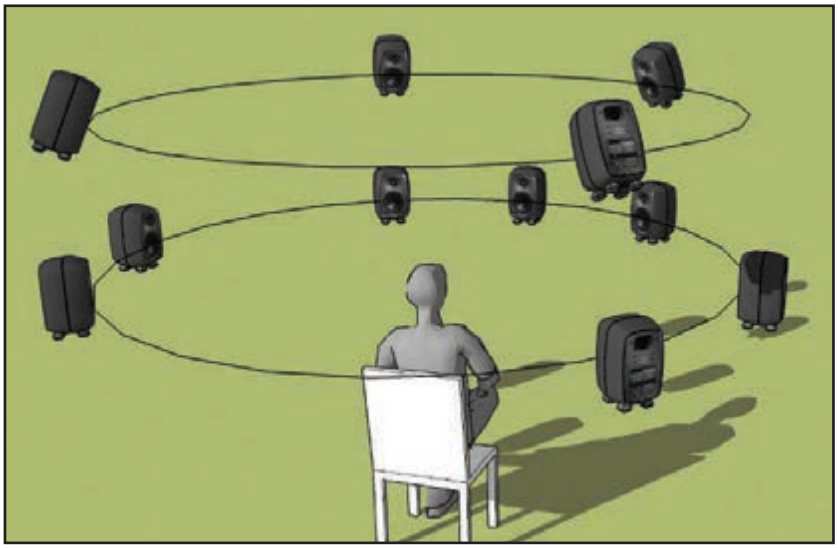
At IBC 2014, the Fraunhofer/Technicolor demonstration showed the world's first real-time encoder for interactive and immersive TV audio. The real-time hardware featured the ability to encode audio for live broadcasts from stereo up to 3-D

sound in 7.1+4H format with additional tracks for interactive objects including commentary in several languages, ambient sound, or sound effects.

The solution includes a professional decoder to recover the uncompressed audio for further editing and mixing in the studio; a real-time encoder for delivery to consumers—over the Internet for new media use or for upcoming over-the-air broadcast systems (such as ATSC 3.0); and a professional decoder used to monitor the transmitted encoder's output.

In all the IBC 2014 and the National Association of Broadcasters (NAB) 2015 demonstrations were supported by Qualcomm, while Technicolor supervised the production services for content creation. Fraunhofer also introduced a new "surround soundbar frame" for television sets, a concept that delivers 3-D audio without using multiple external speakers.

As Robert Bleidt, General Manager of the Audio and Multimedia Division for Fraunhofer USA Digital Media Technologies, explained, "Future UHDTVs might build similar technology into the TV itself, offering consumers an 'un-box, plug-in, enjoy' experience with immersive sound. That will be much better than the soundbars of today, with no wires or external components at all. While high-end home theater enthusiasts will likely still prefer nine or 11 separate speakers for the ultimate 3-D sound quality, this new soundbar concept will greatly enhance the audio experience for a broad consumer base without complex installation and setup," he added.



MPEG-H has defined speaker configurations from 1.0 (mono) to 22.2, including 7.1 and 5.1 plus 4 height speakers above corner as shown. (Image courtesy Fraunhofer IIS)

Resources

Dolby, "Dolby® Atmos® Next-Generation Audio for Cinema," White Paper, 2014, www.dolby.com/uploadedFiles/Assets/US/Doc/Professional/Dolby-Atmos-Next-Generation-Audio-for-Cinema.pdf.

J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H Audio—The New Standard for Universal Spatial/3D Audio Coding," Audio Engineering Society (AES) paper, December 2014.

Moving Picture Experts Group (MPEG), http://mpeg.chiariglione.org/standards/mpeg-h.

N. Fonseca, "Immersive Sound Design Using Particle Systems," Convention e-Brief 193, presented at the 138th Audio Engineering Society (AES) Convention, 2015.

———, "3D Particle Systems for Audio Applications," 16th International Conference on Digital Audio Effects, 2013.

V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *Journal of the Audio Engineering Society*, June 1997

T. Sporer, J. Liebetrau, S. Werner, S. Kepplinger, T. Gabb, and T Sieder, "Localization Of Audio Objects In Multi-Channel Reproduction Systems," Audio Engineering Society (AES) paper, 57th AES International Conference, 2015.