First Experiences with the MPEG-H TV Audio System in Broadcast

Stefan Meltzer
Adrian Murtaza

First Experiences with the MPEG-H TV Audio System in Broadcast

Stefan Meltzer, Adrian Murtaza, *Fraunhofer Institute for Integrated Circuits (IIS)*

Abstract— **In May 2017, the worldwide first UHDTV broadcasting service using with the MPEG-H TV Audio System, a Next Generation Audio codec, was launched in South Korea. Now the system is in regular operation for more than one year and in the meantime the leading terrestrial broadcasters have tested the different formats available. It is the perfect time to do an intermediate summary and report about the experience made during the first year. Starting from the necessary modifications to integrate the MPEG-H TV Audio System into the existing infrastructure and operating it with legacy formats (stereo and 5.1 surround), this paper will describe the different phases of the transition to the full usage of the feature set of the MPEG-H TV Audio system.**

*Index Terms—*3D-Audio production, Broadcast Audio, Immersive Audio, Interactive Audio, MPEG-H 3D Audio, Next Generation Audio

I. Introduction

MPEG-H 3D Audio [1] is the latest audio standard developed by MPEG and besides improving the coding efficiency it brings a set of advanced new features. Therefore MPEG-H is more than just an audio codec, it allows to transmit different audio formats, enables user interaction with the audio mix and provides best audio experience for each playback device. Also, often referred to as "Next Generation Audio Codec (NGA)", the MPEG H TV Audio System [2] has been adopted by ATSC [3], DVB [4] and SCTE [5], [6] specifications for usage of audio and video codecs.

II. The MPEG-H TV Audio System

The MPEG-H TV Audio System allows the transport of audio as traditional audio channels, representing a loudspeaker signal in a predefined target setup, as well as audio objects and scene-based audio. While any one of these three formats can be used individually for the delivery of complete audio scenes, the MPEG-H TV Audio System also supports any combination of them. Scene-based audio or Ambisonics, is basically a mathematical representation of the sound-field, where the spatial accuracy of the reproduction increases with the order of the representation. Therefore, a Higher Order Ambisonics (HOA) signal offers a higher spatial accuracy compared to a First Order Ambisonics (FOA) signal. An audio representation using Ambisonics is independent from the final reproduction setup and will be rendered at the receiver side to the available loudspeaker setup. This independency from a loudspeaker configuration makes it also easy to rotate the sound field in all directions.

.

For that reason, the Ambisonics representation of audio is commonly used in virtual and augmented reality applications.

The third option in MPEG-H 3D Audio is the use of audio objects. Here we can distinguish between dynamic and static objects. Dynamic objects are commonly used for describing any audio source which is moving in the sound field. A typical example is a sound effect like a helicopter moving in a scene from the front right to the rear left. Static objects are sound elements which keep their position in a scene, like a commentary in a sports event or the dialogue in a movie. These objects can be used to offer the viewer the opportunity to interact with the audio mix and personalize it according to its own preference. The viewer can change the volume of these objects, for example for enhancing the intelligibility of the dialog, or select between different language versions or commentaries. Audio description (AD) can also be transmitted as an object, which allows the user to position the audio description at a different location than the main dialog during the playback (e.g., position the AD to the right rear loudspeaker, exactly where the visually impaired person sits in the living room). Moreover, the advanced receivers with MPEG-H support can offer audio with audio description over the headphone output, while the speaker system reproduces the audio without the video descriptive service. Audio objects are introducing a lot of flexibility into the mix and create new opportunities for the creatives while offering the viewers an unprecedented flexibility to adjust the audio to their personal preferences.

To avoid that the viewer gets overwhelmed by the new options and flexibility, the MPEG-H TV Audio System includes a user interface with two levels. In the first basic level, the user has the choice to select between a number of presets, which provide different versions of the program. The number of presets and their properties are defined by the broadcaster and offer most commonly used settings, such as: an increased dialogue volume or stadium atmosphere. Viewers who want more control over the settings can access the advanced menu which provides all options for the available objects (e.g., level and position interactivity). The presets as well as the limitations of the user interactions are entirely controlled by the broadcaster, by means of comprehensive metadata carried in the audio stream and used for enabling/disabling specific features and controlling the audio decoding and playback. A reference design of the User Interaction Interface is provided in Figure 1.

Besides the above-mentioned features, the MPEG-H TV Audio System also offers immersive audio which in combination with a UHD picture leads to the full immersion of the viewer into the scene. The integrated energy preserving
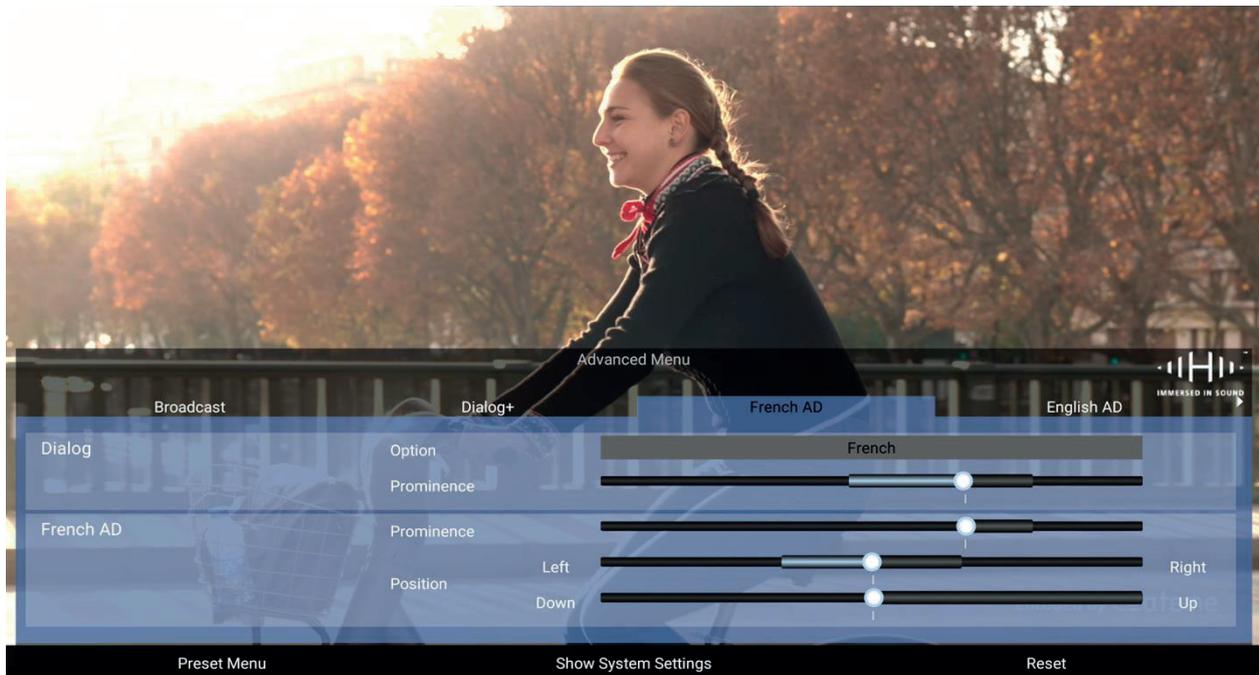
Figure 1. User Interaction Interface Example

format converter adapts the original channel layout to the final reproduction setup using phase alignment methods and avoiding the well-known problems of the passive downmix algorithms used in previous audio coding systems. Advanced loudness and dynamic range compression algorithms allow the use of a single bit stream to be send to different target devices ranging from mobile phone and tablets to TV set and AVRs, while always producing the best sound experience adapted for the target device. The core behind all these advanced functionalities is the carriage of additional information about the single elements of the mix and the overall mix, the so-called metadata.

The metadata describe the properties of each single element of a mix and the structure of the mix. It defines whether the element is an object or a channel or a part of an Ambisonics signal. For example, the metadata associated with an object includes amongst others the position of the object in the sound field, the level for playback, which type of audio and the language ID in case of a dialog. The structure of a mix, also called the audio scene description, defines the relation of the single elements of a mix to each other. It includes the definition of groups, presets and the limitations of user interactions. All these metadata need to be created during the production and then transported through the production facilities to the final emission encoder.

III. Phased introduction of the MPEG-H TV Audio System

Looking at its advanced features it may sound complicated to use the MPEG-H TV Audio System in today's broadcast facilities and that this would require large investments in new equipment. But there is no requirement to use all features with the introduction of the MPEG-H TV Audio System. Broadcasters can gradually enable features of the MPEG-H TV Audio System in their production workflow based on their needs.

In a first stage a simple replacement for their current audio coding system will allow usage of the MPEG-H TV Audio System for producing only stereo or 5.1 surround sound with a fixed set of metadata. In this scenario the MPEG-H audio

encoders are just a drop-in replacement of the legacy encoders and no metadata authoring and transport is required in the production and broadcast facilities. The immediate advantage is the increased bit rate efficiency of MPEG-H compared to the legacy codec and of course that a future proof audio coding system is now introduced which can be exploited at any time in the future.

In a second step the interactivity feature using audio objects in combination with a stereo or 5.1 channel bed for music and effects can be introduced. Even without immersive sound, the interactivity features greatly improve the attractiveness of the offering, efficiently enabling for example multiple languages or video description services. For this step, the generation and transport of metadata within the broadcast facilities is required. The necessary changes will be described in the following section.

The third phase would be the addition of the immersive sound experience. This makes the user experience complete and opens a new dimension in the entertainment. With the introduction of immersive soundbars, it is also possible to reproduce the experience into the viewers home environment without the need to install a large number of speakers.

IV. Required Changes To the Broadcast Infrastrucutre

While during the first step of the introduction of the MPEG-H TV Audio System no changes to the basic infrastructure and production workflow are necessary, the following phase requires additional equipment and editing work. The second step introduces metadata authoring and monitoring into the production process and requires a secure and reliable mechanism to transport the metadata within the broadcast plant. The metadata can be either authored live or during post production. In the file-based workflow of the post production these metadata elements are stored in an XML structure and later on inserted in the workflow by a playout server. For the generation of the metadata plugins for DAWs and standalone metadata generation tools are available (see Figure 2). These plugins and tools also allow the monitoring of already existing productions with metadata. In the live
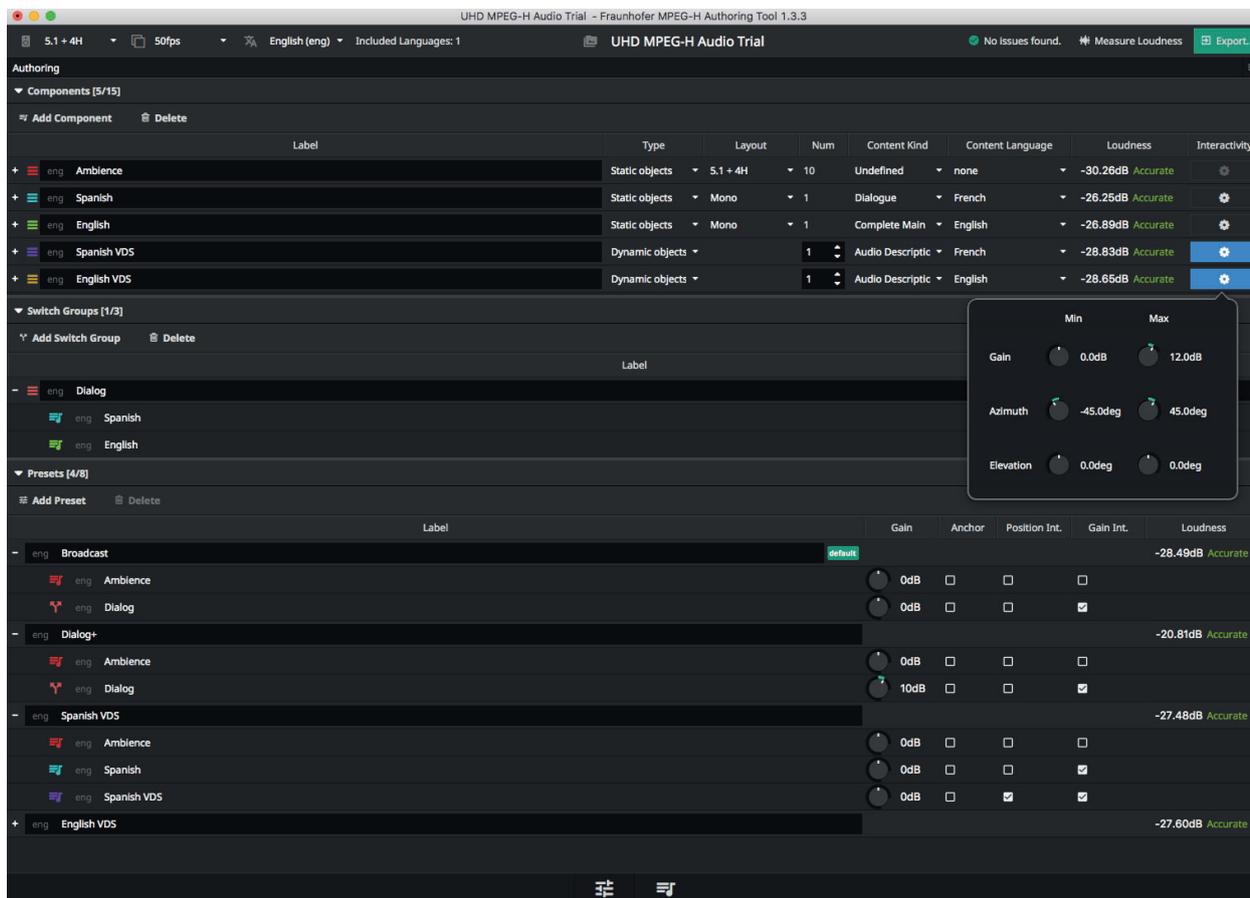
Figure 2. Fraunhofer MPEG-H Authoring Tool

scenario the so-called audio monitoring and authoring units generate the metadata. For the transport of the metadata within the broadcast facilities based on SDI infrastructure a so-called Control Track was developed. This Control Track is basically a PCM signal which carries the metadata as modulation. So, the 16th PCM channel of 16 channel SDI signal is used for the Control Track.

This solution ensures that the metadata is always carried together with the associated audio signals and no additional complexity is introduced since the audio signals are carried in PCM format. This format is designed to be robust against any processing usual in a production chain (e.g., re-sampling, attenuation, cutting to video frame boundaries) and is also used for the connection between the playout server and the emission encoder. In summary, the required changes are limited to adding authoring and monitoring units to the production chain and to install the necessary plugins in DAWs or use the standalone tools for the authoring and monitoring the metadata.

When adding the immersive sound component, it is necessary to upgrade the audio monitoring system in the control rooms to the immersive target format. In many cases the target loudspeaker layout is a 5.1 or 7.1 surround setup with four additional height speakers in the corners. In addition, the production environment needs to upgraded to handle immersive sound. There are several 3D audio production plugins available in the market, which will support the generation of immersive sound. The MPEG-H TV Audio System can be used with any available 3D audio production plugins.

V. First real deployment of the MPEG-H TV audio system

In May 2017, Korea was the first country worldwide to launch a regular terrestrial UHD service. The Korean standard for this service mandates the MPEG-H TV Audio System as the only audio codec [7]. The Korean broadcasters adopted the phased approach described above. During the first year of operation they transmitted the content in the legacy stereo and 5.1 format. This allowed them focus on solving the usual problems which occur when you start a completely new system with new modulation, new transport format, new video resolution and new audio and video codecs.

Now, they start to move to the next phase, which will combine the above mentioned second and third stages. At the end of May 2018, the first immersive and interactive transmissions went on air. The selected target loudspeaker setup was 5.1+4H and included two additional objects and three presets [8].

The Korean government gave the broadcasters strict rules and timeline for the transition from the current HD service to the future UHD service. From the start in 2017 on the Korean broadcasters were required to produce 5% of the content for the UHD service in native UHD resolution. To fulfill this quote and to stay on the safe side, the Korean broadcasters decided to broadcast in UHD the news programs. This allowed them to make the first UHD productions in a very controlled environment with little risk.
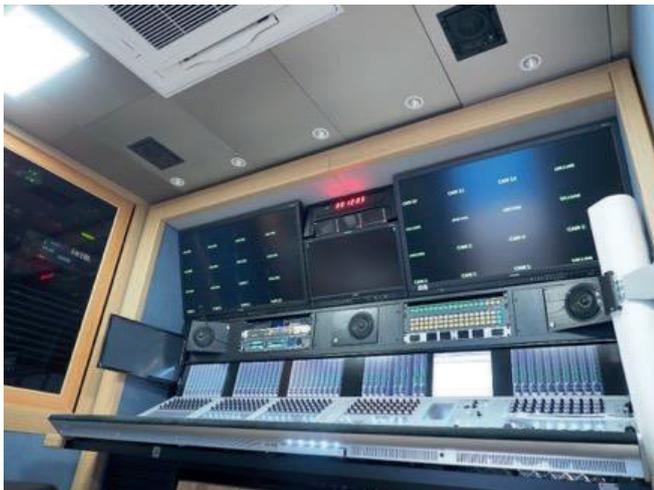
Figure 3. MBC's UHD OB van with immersive audio monitoring setup

The Koran government timeline requires the percentage of native UHD content to increase in three steps from 5% now to 25% in 2020, to 50% in 2025 and finally to 100% in 2027. In 2027 the HD service, currently operated in a simulcast, will be switched off. Although, the three major broadcasters in Korea might have different strategies how to fulfill the necessary quote of native UHD content, it is safe to assume that they will do a more linear increase of the native UHD content then the government requirement. And using the advanced audio features is part of their strategy. In the meantime, they have already started to implement immersive monitoring setups in control rooms and OB vans. Professional equipment from encoders, metadata authoring and monitoring units as well as test receivers are currently available. On the consumer side TV sets from Samsung and LG are on the market supporting the full feature set of the MPEG-H TV Audio System. Immersive soundbars supporting MPEG-H Audio from Samsung and Sennheiser were announced during the Consumer Electronics Show in January 2018 in Las Vegas. Thus, the complete end-to-end chain is available to allow broadcasters to make full use of the advanced features offered by the MPEG-H TV Audio System.

Currently the Korean broadcasters, CE manufacturers and government agencies are discussing the next system, a mobile TV service based on the ATSC 3.0 standard to replace the existing mobile TV service based on T-DMB. The audio side of this service will also be using the MPEG-H TV Audio System.

VI. FURTHER TRIALS IN EUROPE

Besides Korea, the MPEG-H TV Audio System was also successfully tested during the French Tennis Open in Paris, France, end of May 2018 [9]. During the tournament, France Télévisions provided a live UHD signal from the center court, which was transmitted using DVB-T2 for the terrestrial distribution and DVB-S2 for a satellite signal. The live signal consisted of UHD video and a 5.1+4H immersive sound bed for the court atmosphere and two objects for commentaries in French and English with three presets. The signal was on air 24h per day during the tournament with a live coverage during the matches and a replay loop with the matches from the previous day in between. The trial demonstrated how easy it was to enable MPEG H into a common remote production

Besides Korea, the MPEG-H TV Audio System was also successfully tested during the French Tennis Open in Paris, France, end of May 2018 [9]. During the tournament, France Télévisions provided a live UHD signal from the center court, which was transmitted using DVB-T2 for the terrestrial distribution and DVB-S2 for a satellite signal. The live signal consisted of UHD video and a 5.1+4H immersive sound bed for the court atmosphere and two objects for commentaries in French and English with three presets. The signal was on air 24h per day during the tournament with a live coverage during the matches and a replay loop with the matches from the previous day in between. The trial demonstrated how easy it was to enable MPEG H into a common remote production

VII. CONCLUSION

The MPEG-H TV Audio System offers revolutionary new features for the broadcasters and the viewers. It can be introduced in several phases into the broadcasting process and should therefore be considered for the first changing opportunity in the broadcast infrastructure, such as the upgrade from HD to UHD.

The MPEG-H TV Audio System has demonstrated its strength and maturity in regular operation in Korea for more than a year now. Successful DVB-T2/S2 transmission trial in France and streaming trial Norway demonstrated that the MPEG-H TV Audio System can also be used with other transmission standard then ATSC3.0. The necessary equipment on the professional side as well as on the consumer side is already available and the system can be used immediately.

REFERENCES

[1] "*Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio*," International Organization for Standardization (ISO), Geneva, Standard ISO/IEC 23008-3:2015, 2015.

[2] R. Bleidt et al. "Development of the MPEG-H TV Audio System for ATSC 3.0," in IEEE Transactions on Broadcasting, vol. 63, no. 1, March 2017.

[3] A/342 Part 3: ATSC Standard – MPEG-H System, March 2017.

[4] TS 101 154 v2.3.1, "Digital Video Broadcasting (DVB) – Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream". Available: http://www.etsi.org/deliver/etsi_ts/101100_101199/101154/02.03.01_60/ts_101154v020301p.pdf

[5] SCTE 243-2:2017, "Next Generation Audio Coding Constraints for Cable Systems: Part 3 –MPEG-H Audio Coding Constraints".

[6] A. Murtaza et al. "MPEG-H TV Audio System for Cable Applications," in SCTE Journal of Digital Video Subcommittee, Volume 2, Number 1, July 2017.

[7] TTAK-KO-07.0127R1, "TTA - Transmission and Reception for Terrestrial UHDTV Broadcasting Service", Revision 1, December 2016. http://www.tta.or.kr/include/Download.jsp?filename=stnfile/T-TAK.KO-07.0127_R1.pdf

[8] SBS Conducts Worldwide First Regular Broadcast with Immersive and Interactive Sound Powered by MPEG-H Audio: https://www.iis.fraunhofer.de/en/pr/2018/20180614_AME_Worldcup.html

[9] Successful Terrestrial And Satellite Reception Of MPEG-H Audio During The Roland Garros French Open: http://idfrancetv.fr/successful-terrestrial-and-satellite-reception-of-mpeg-h-audio-during-the-roland-garros-french-open.

Stefan Meltzer studied electrical engineering at the Friedrich-Alexander University in Erlangen, Germany. In 1990 he joined the Fraunhofer Institute for Integrated Circuits (IIS) in Erlangen, Germany. After working in the field of IC design for several years, Stefan became the project leader for the development of the WorldSpace Satellite Broadcasting system in 1995 and in 1998 of the XM Satellite Radio broadcasting system. His team was responsible for the system design, chip set design, field trials and development of a reference signal generator.

In 2000 he joined Coding Technologies in Nuremberg as Vice President for business development, Germany. His responsibilities included broadcasting and consumer electronics. During his time at Coding Technologies HE AAC was accepted in numerous broadcasting standards and applications.

After Coding Technologies was acquired by Dolby Labs, Stefan joined Iosono as CTO in April 2008.

Since January 2010, he works as independent technology consultant with the main focus on audio and multimedia. In this role he supported Fraunhofer IIS in the business development and marketing activities within the TV broadcast market. In 2018 he rejoined Fraunhofer IIS.

Adrian Murtaza received his M.Sc. degree in Communication Systems from the École Polytechnique Fédérale de Lausanne, Switzerland in 2012 with a thesis on Backward Compatible Smart and Interactive Audio Transmission. Upon graduation he joined Fraunhofer IIS, where he works as a researcher on Semantic Audio Coding, parametric multi-object and multi-channel audio coding and 3D Audio.

Adrian Murtaza actively participates in several standardization organizations, including MPEG, DVB, ATSC, CTA and SCTE, and contributed as main editor to several standards in those groups. His recent activity is focused on deployment of MPEG-H Audio in different broadcast and streaming ecosystems.