

# Field Tests for Immersive and Interactive Broadcast Audio Production using MPEG-H 3D Audio

Christian Simon  
Yannik Grewe  
Nicolas Faecks  
Ulli Scuda

Cite this article:

Simon, Christian, Grewe Yannik, Faecks, Nicolas, Scuda Ulli; 2018. Field Tests for Immersive and Interactive Broadcast Audio Production using MPEG-H 3D Audio. SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi: 10.18580/setijbe.2018.5. Web Link: <http://dx.doi.org/10.18580/setijbe.2018.5>

# Field Tests for Immersive and Interactive Broadcast Audio Production using MPEG-H 3D Audio

Christian Simon, Yannik Grewe, Nicolas Faecks and Ulli Scuda

**Abstract**— Next Generation Audio offers new features for the consumer such as advanced user interactivity, immersive sound and optimized reproduction across different classes of playback devices. All of these strongly depend on additional information that is affixed to the audio in the shape of metadata. Thus, some processes of the final sound reproduction assembly shift from the mix process during production to the rendering stage in the consumers' device. Using the MPEG-H TV Audio System as an example, this paper outlines the workflow for Next Generation Audio content creation with a focus on broadcast production for live and offline applications. Initially, the MPEG-H TV Audio Systems fundamentals will be detailed. It is shown, how new tools can be integrated into well-established production workflows and how they ensure full control of produced mixes in anticipation of the end user's experience. Use cases for immersive and interactive audio broadcasts are presented by example of an international music competition and a major tennis event. Finally, an outlook to future requirements and improvements is given.

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

## I. INTRODUCTION

NEXT Generation Audio (NGA) systems, such as the MPEG-H TV Audio System [1] based on the ISO/IEC 23008-3:2015 MPEG-H 3D Audio Standard [2], provide revolutionary features, such as immersive and interactive audio, supporting channel-based, object-based as well as scene-based audio [3]. MPEG-H meets requirements of increasing types of delivery platforms and infrastructures for broadcast, streaming services, TV on demand or mobile applications by using the same bit stream across different device classes. This feature is called 'universal delivery'. Bit rates of 192 kbit/s to 384 kbit/s, which are commonly known for six-channel transmissions, can be used to broadcast various content such as twelve-channel audio or twelve audio objects, depending on the production scenario.

Since 2017, the MPEG-H TV Audio System is in use for 24/7 broadcasts in South Korea, making it the first region worldwide to use NGA for regular services. In 2018, MPEG-H was selected as transmission codec for the

upcoming UHD TV services in China.

The production, distribution and rendering of immersive and interactive audio poses a number of challenges for content creators and providers:

- How can they use advanced features in an efficient way?
- How is it possible to produce immersive and interactive audio without completely changing well-established production or contribution workflows?
- How can scene related metadata be created, handled, securely transmitted and accessed at any critical point in the chain?

Based on these considerations, specific production tools were developed to offer an efficient workflow for live and post-production using the MPEG-H TV Audio System.

This paper outlines challenges, requirements and solutions for an efficient production workflow for immersive and interactive broadcasting, based on two field tests.

Section II continues by describing fundamental features of the MPEG-H TV Audio System, detailing the need for new production tools. To demonstrate how these tools and concepts can be used in practice, field tests and trial transmissions of MPEG-H Audio content were conducted during the Eurovision Song Contest and the French Tennis Open in 2018 [4] which are described in section III and section IV. The outlined workflow is focused on Serial Digital Interface (SDI) based studio infrastructure, which is most common in today's broadcasting facilities.

Unless explicitly mentioned, MPEG-H refers to the MPEG-H TV Audio System, based on the ISO/IEC MPEG-H 3D Audio standard.

## II. FUNDAMENTALS OF THE MPEG-H TV AUDIO SYSTEM

The MPEG-H TV Audio System enables all NGA features. In the following, most important ones are briefly described.

### A. Immersive Audio

By adding elevated sound sources above and below the listeners' position, more detailed spatial reproduction can be achieved [5]. The system supports, but is not limited to, over twenty different loudspeaker configurations including setups such as stereo, 5.1, 7.1 [6] as well as 3D audio setups, namely 5.1+4H, 7.1+4H or 22.2 [7]. Immersive audio can be carried

This manuscript was received on June 30, 2018. Date of current version June 30, 2018.

C. Simon, Y. Grewe, N. Faecks and U. Scuda are with Fraunhofer Institute for Integrated Circuits IIS, 91058 Erlangen, Germany (e-mail:

christian.simon@iis.fraunhofer.de; yannik.grewe@iis.fraunhofer.de; nicolas.faecks@iis.fraunhofer.de; ulli.scuda@iis.fraunhofer.de)

as channels, objects or ambisonics coefficient signals or any combination of the above [2].

The ISO/IEC MPEG-H 3D Audio standard supports up to 128 channels, objects or ambisonics coefficient signals, rendered simultaneously to a maximum of 64 loudspeakers. To constrain implementation complexity, limits were defined and described in an MPEG-H Low Complexity Profile (LC-Profile). This paper refers to Low Complexity Profile Level 3, which is adopted by ATSC 3.0, DVB and SCTE standards for the definition of Next Generation Audio and video coding systems for broadcast and cable applications. Up to 32 audio elements can be transmitted within one bit stream, while 16 of them can be decoded simultaneously. Other broadcasting scenarios going beyond the limitations of the LC-Profile Level 3, for example using a 22.2 channel format, can be covered using the LC-Profile at Level 4 [2].

### B. Interactive and Personalized Audio

Using audio objects and combining them with channel- or scene-based audio enables the listener to interact with the content by using the standard TV remote control. Simple adjustments, such as increasing or decreasing the prominence of dialogue in relation to other audio elements, to more advanced scenarios are possible. Listeners may choose from different languages or commentators or even change position of audio objects.

To achieve grouping of exclusive components, the concept of 'switch groups' was designed. This can be used for switching between different languages or other audio signals, whose semantic content is not meant to be played back simultaneously.

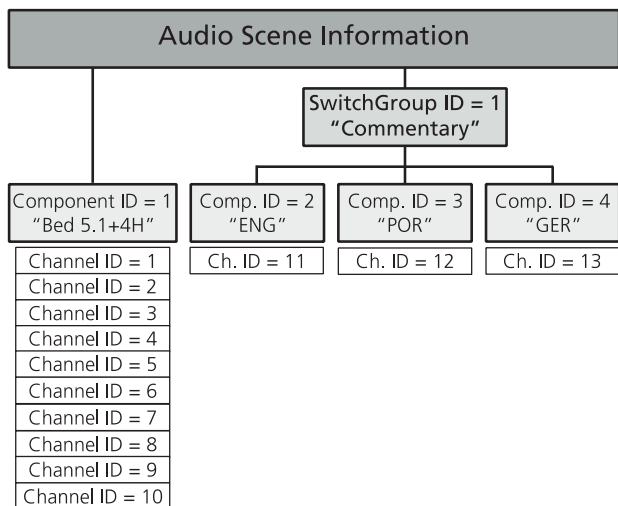


Figure 1: Overview on a MPEG-H scene description. Channel 1-10: ambience bed, channel 11-13: audio objects with different commentators within a switch group.

The MPEG-H TV Audio System describes the sound of a TV program as a set of audio components and its related metadata, combined in an audio scene. Metadata can configure different mix presets offered by the broadcaster, such as a default mix as a first preset and secondly a hearing impaired mix in which the dialogue is boosted. Metadata contain the range limits on the viewers' control of the audio scene, such as dialogue level and object position. Furthermore, metadata include text labels, information about content kind, downmix coefficients and loudness values to

adjust the playback to different device platforms and environments such as a home cinema or a noisy train station. An exemplary illustration of an MPEG-H scene description including a 5.1+4H immersive bed and three commentator tracks compiled within a switch group is presented in Fig. 1.

The creation of MPEG-H related metadata and respective parameters is always under full control of the content provider.

### C. Universal Delivery

The MPEG-H TV Audio System offers flexibility by delivering the same bit stream through different distribution platforms (e.g. satellite, broadband or mobile network) to different devices (e.g. TV set, AVR, soundbar, tablet, virtual reality devices) in different environments (e.g. living room, home theater, noisy public).

Therefore, multiple technologies have been implemented to ensure that the consumer experience always is complying with the content producers' intention. The core element for the object-based audio is a high quality renderer, while a format converter handles channel based audio. In this way, the main intention of a mix can still be transmitted even though it is played back on different reproduction platforms. Furthermore, the MPEG-H 3D Audio standard features binaural rendering technology to directly output signal for stereo headphones, creating immersive audio experiences.

Loudness-, dynamic range- and peak control, as well as ducking for voice over applications are conducted by advanced Dynamic Range Control (DRC) mechanisms [8].

### D. MPEG-H Metadata for SDI-based Infrastructure

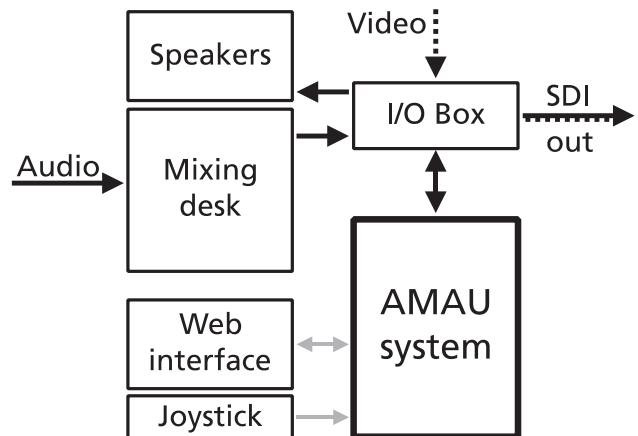


Figure 2: Signal flow of an Audio Monitoring and Authoring Unit system (AMAU). Solid lines represent audio, dotted lines represent video. A solid and dotted line represent audio and video. Grey lines indicate control data.

As mentioned, metadata are essential to control audio objects and interactivity within NGA. The most fundamental information being channel count and layout, types and labels of audio objects, interactivity control limits, loudness information and position data for dynamic objects. In general, metadata need to be synchronized and attached to the corresponding audio, in order to be processed together while encoding. During decoding, metadata control the rendering process.

For live productions using MPEG-H, metadata are usually created as a so-called control track (CT) [9] with the help of an Audio Monitoring and Authoring Unit (AMAU) [10]. The

AMAU modulates metadata into the CT and feeds this additional channel on a MADI or SDI link back to the OB van's or studio's audio equipment, as shown in Fig. 2. Later the encoder uses this CT to encode audio accordingly to metadata created in the AMAU.

The CT is a timecode like audio signal and can be handled as a regular audio channel. Typically, the CT is carried on channel 16 within an SDI framework (see Fig. 3). This tightly coupled transport of the CT together with the audio channels carrying the audio essence ensures integrity of the transmitted audio scene. In a future IP based production workflow, video, audio and metadata parts will be transmitted as separate IP streams according to SMPTE ST2110, where all streams are synchronized by global time stamps and recipients are able to extract only data portions relevant for a certain application [11].

Handling the CT as an audio channel ensures that it is always synchronized with its corresponding audio. It is robust enough to survive A/D and D/A conversions, level changes, sample rate conversions or frame-wise editing. The CT does not force audio equipment to be put into data mode or non-audio mode in order to pass through.

For post-production scenarios, the CT can be created using standalone applications, such as the Fraunhofer MPEG-H Monitoring and Authoring Tool (MHAT) as well as plugin solutions for digital audio workstations (DAW) [12].

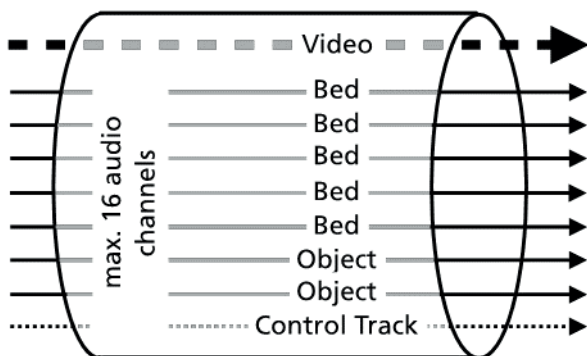


Figure 3: SDI signal for MPEG-H distribution. It contains max. 15 channels of audio (Bed and/or audio object tracks) and a control track. The control track includes all scene related metadata on channel number 16.

### III. IMMERSIVE AND INTERACTIVE AUDIO PRODUCTION FOR THE EUROVISION SONG CONTEST

#### A. Scope of the Event

The Eurovision Song Contest (ESC) is an international song competition produced by the European Broadcasting Union (EBU) and one of the most viewed non-sporting-events worldwide. In 2018, it took place in the Altice Arena Lisbon, Portugal and was watched live by 186 million viewers [13]. It was supported by the national Portuguese broadcaster Rádio e Televisão de Portugal (RTP). The EBU and Fraunhofer Institute for Integrated Circuits IIS collaborated to conduct a field test for immersive and interactive live audio production, based on the MPEG-H TV Audio System. The field test was done to evaluate future production and reproduction scenarios. The authoring and monitoring of the immersive mixes was done live for demonstration purposes as well as executed in post-production after the event. The immersive mix was not

broadcasted but played back during the event in Lisbon.

Producing immersive audio raises several general questions to content creators, such as:

- Which audio production and reproduction formats need to be used?
- How can the immersive part of the content be captured in an efficient way?
- What are changes regarding the editing and mixing workflow?

Firstly, production and reproduction scenarios need to be clarified. Both scenarios comprise of different requirements for live broadcast and offline recording, sound quality and transmission. In the case of the ESC field test, an offline audio production was targeted towards reproduction over loudspeakers, soundbars and headphones. The focus of the following considerations lies on the immersive aspect with a production format of 5.1+4H, by adding four height channels to an ITU-R BS.775 surround configuration as defined in [2].

#### B. Production Scenario

The legacy audio production format for the international feed at ESC was a 5.1 surround sound mix and an additional 2.0 stereo mix. Both were created by the host broadcaster's OB van on location. The signals used were a mix of microphone feeds and pre-produced material, such as special sound effects (SFX) and trailers for the participating competitors.

In total, 232 microphones were employed by the broadcaster to capture sound from the main stage, hosts, interview partners and arena audience. Based on these signals, international feeds were mixed. 26 ambience microphones were put into different zones throughout the audience, both in front of the stage and on the rear floor area, aiming to the upper tiers and downwards from the PA rigs. Positioned this way, a versatile mix of audience reactions can be created, surrounding the listener by blending more diffuse and more direct ambience signals and giving the mixer the possibility to select the most appropriate sounds. Microphone signals representing the diffuse upper part of the arena, preferably with minimum direct PA sound, were missing in the conventional production workflow to create an immersive experience. For that, a Hamasaki Square [14] was additionally placed underneath the roof in the center of the arena about 25 meters above the floor. Minimizing direct sound from PA and audience by its polar patterns and direction, it provided the upper layer ambience signal by adding just four additional microphones.

A 3D audio mixing room was set up, receiving all OB van's signals including sub-mixes and additional commentator feeds from a second service provider, as shown in Fig. 4. All signal transmissions were realized by Multichannel Audio Digital Interface (MADI) connections.

For creating an immersive, three-dimensional sound impression, the Hamasaki square signals were panned 100% to the loudspeakers' upper layer, which created the intended 'dome' effect. Microphone signals from the upper part of the arena that are panned to the middle layer for regular 5.1 audio production, have been used differently for the immersive mix. Here, they were panned slightly upwards to bring them to their natural position in the sound field, about 15 degrees

upwards from the arena floor. It is also possible to use the existing 5.1 ambient mix from the OB van, adding only the additional height layer microphone setup. The latter case has a significant drawback. Microphone signals that carry spatial information of the upper part of the sound field cannot be elevated by panning, which leads to a loss of spatial information. Therefore, in case of the ESC field test, a new ambience mix was done.

Once the immersive ambience was created, remaining stems like music and hosts were mixed in a legacy way, panning them in the mid-front line of the immersive loudspeaker configuration. The stereo SFX stem was panned slightly upwards to produce an immersive effect. For future events, sound effects and interstitials should already be pre-produced in an immersive channel layout to be more impressive. Additionally, a selection of the 20 available commentary feeds were added to the audio scene, whereupon interactivity was configured and monitored.

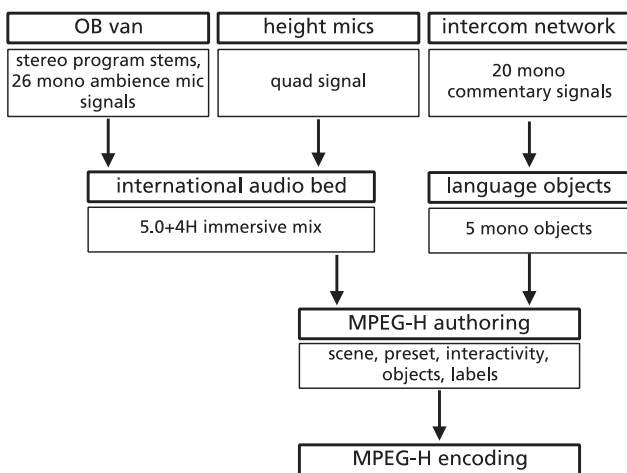


Figure 4: Schematic audio production and authoring workflow of the ESC MPEG-H field test.

The resulting 5.1+4H mix can be monitored using an AMAU or DAW with MPEG-H enabled plugins [12] to monitor downmixes from 5.1+4H to 5.1 and stereo as well as the binauralized playback. In a live scenario, the 5.1 and stereo mixes from the OB van should also be monitored frequently to prevent major differences in the level balance of the stems.

After finishing the mix, a CT including all metadata was generated by the MPEG-H production tools and used for controlling the MPEG-H Audio encoder. An exemplary GUI on the receiver side, based on the authored audio scene and preset creation, is shown in Fig. 5.

The field test showed that the already existing infrastructure for sound capturing and processing only needs few additional changes to be able to produce immersive and interactive sound. In the described case, four additional microphones and loudspeakers plus an MPEG-H enabled authoring tool were required. The resulting immersive and interactive audio mix received positive feedback compared to the legacy stereo or 5.1 production, as stated by involved producers and audio engineers during on-site demonstrations.

#### IV. IMMERSIVE AND INTERACTIVE AUDIO FOR THE FRENCH TENNIS OPEN

##### A. Scope of the Event

During the major tennis tournament held in Paris in May/June 2018, the host broadcaster, France Télévisions, provided a test channel for UHD broadcast including MPEG-H Audio supported by the Fraunhofer Institute for Integrated Circuits IIS. The aim was to broadcast appropriate NGA for UHD video including immersive and interactive features as well as High Dynamic Range with Hybrid Log Gamma (HLG-HDR) [15] to preserve more details in the darkest and brightest areas of the picture. Along with the video, immersive sound and interactive audio objects were transmitted over satellite and terrestrial broadcast. The tournament was the first event for an immersive and object-based production using the MPEG-H TV Audio System to go on air in Europe. The test channel covered all games taking place on the center court. The transmission was receivable via DVB-T2 [16] in the area of Paris and via DVB-S2 [17] all over France and was on air until the end of the tournament.

##### B. Production Scenario

Immersive audio is well suited for live sports events broadcast because the ambience of the event location including audience reactions can be reproduced much more detailed, resulting in a stronger emotional experience for the viewers. Using audio objects allows for level interaction on commentary to support better understanding, especially if ambient and audience noise is high. Furthermore, a 'Venue preset' with stadium sound only can deliver a more realistic live experience on the consumer device. In addition to French commentary, an English commentary was provided for several matches. Both commentaries were delivered as separate audio tracks and configured as static dialogue objects.

For the 2018 event, the host broadcaster added an ORTF 3D microphone array to their legacy microphone setup on the center court. This microphone array captures 3D audio with a compact setup and provides a 4+4H audio signal [18]. The array was placed behind the umpire, above the lower terrace about one meter in front of the gallery. The legacy field microphone signals as well as the lower layer of the 3D array were used to create the international stereo and international 5.1 mixes. Broadcasters all over the world used these mixes and the UHD or a down-scaled HD picture to serve their customers with a legacy broadcast. The upper layer output of the 3D microphone array created the upper layer in the 5.1+4H immersive bed. Since 3D audio monitoring was not enabled in the used OB van, it took some iteration to tweak the immersive bed mix before broadcasting. Recordings from the encoded signal were checked in an ITU-R BS.1116 listening room [19] off-site and the observations and suggestions were provided to the responsible audio engineer in the OB van. Both mixes, international stereo and 5.1, as well as the signal of the upper layer and commentaries, went through a MADI link into an AMAU.

To generate metadata, the channel layout was first defined in the AMAU by selecting related channels from the MADI link and grouping them. In this case, the ambience bed was defined in the production layout configuration of 5.1+4H.

Two additional commentary tracks were added for French and English, respectively. Furthermore, the AMAU measured loudness for every audio component. Loudness measurement is an important step in the production to ensure adequate loudness metadata, which control the renderer during the decoding process.

To allow more interaction with commentary beyond selecting the language, corresponding metadata were created in the AMAU. Gain and position interaction for both commentary audio objects were configured. On the users' TV on screen display, interactivity options are shown and can be selected via the remote. An example which shows user interactivity for commentary language and its prominence level is shown in Fig. 5. The prominence of an audio object describes the level relation of this object to all other audio components in the scene. If for example the commentary level is increased in reproduction to improve intelligibility, the level of the bed channels is attenuated. This way the overall loudness remains consistent.

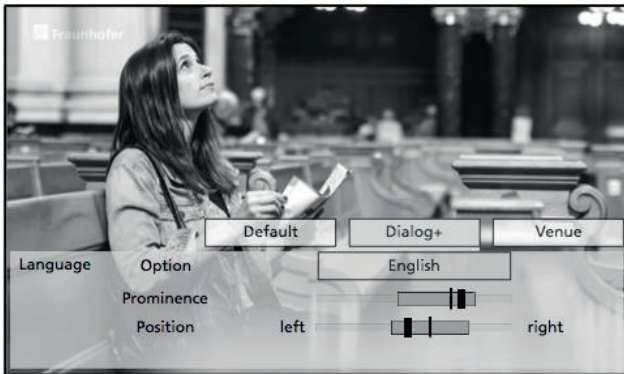


Figure 5: Schematic on screen display on the user's TV set to show available interactivity options. The top line (Default, Dialog+, Venue) represents pre-configured mix presets to choose from. The bottom grey part shows options for the language audio object.

In addition, the user may manipulate the position of the commentary audio object to left and right and—on an immersive reproduction setup—to the upper layer. During production, the host broadcaster emphasized this as an important feature, due to its benefit for users who use audio description services. Users are able to separate the audio description service from the regular dialog (e.g. by positioning the audio description service to a rear speaker in a 5.1 system).

To ease the use of MPEG-H features, user presets were created in the AMAU. In addition to a default one, which represents an immersive broadcast mix, a preset with increased dialogue prominence and a preset without commentary were configured.

As the last processing step, the AMAU modulated the metadata into the CT and fed this together with all audio tracks back to the OB van. The metadata creation in the AMAU is the only additional processing in the workflow for an MPEG-H based production. The basic signal flow for this production is shown in Fig. 6.

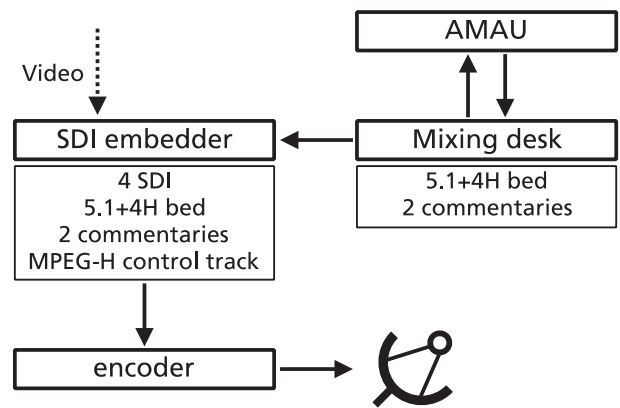


Figure 6: Signal flow used in the sports event production. The AMAU is the only device added especially for the NGA production.

The OB van provided an SDI feed including the UHD video signal and all needed audio channels including the CT (see Fig. 3). The live match or a rerun of a recorded SDI stream was fed into two redundant encoders for a 24/7 live encoding broadcast. The encoders processed HLG HDR video and MPEG-H audio simultaneously in one service. A satellite carrier and a terrestrial carrier received the encoder output on a fiber-based IP connection. They mixed the signal into their regular DVB-S2 and DVB-T2 multiplex, respectively. An Over-The-Air receiver allowed playback of these signals (DVB-S2 and T2) on different playback devices such as TV sets, soundbars and binauralized on headphones.

## V. CONCLUSION

This paper outlines the workflow for Next Generation Audio content creation with a focus on broadcast production for live and offline applications. Initially, MPEG-H TV Audio Systems fundamentals have been detailed. On the basis of two major events, an international music contest and a tennis tournament, production workflows have been described. It has been shown, that capturing, mixing and authoring interactive and immersive content can easily be conducted without major changes in the existing legacy live and offline production chain. All needed tools and devices are available on the market. Practical advice and experience has been described to optimize the consumer experience regarding immersion and interactivity.

To simplify future NGA productions, next needed steps are to upgrade production facilities and to educate broadcast engineers and producers about the required workflow and about the additional features and experiences offered by NGA.

## ACKNOWLEDGMENT

The authors would like to express gratitude to Dr. A. Kuntz, M. Kratschmer and A. Murtaza for sharing their expertise in audio for broadcasting as well as T. Robotham for his valuable feedback. Furthermore, the authors would like to thank all partners who enabled the MPEG-H field tests.

## REFERENCES

- [1] "ATSC Standard: A/342 Part 3, MPEG-H System," in *Doc. A/342-3:2017, 3rd March 2017*, Washington D.C., 2017.
- [2] "High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio, MPEG-H 3D Audio Phase 2," in *ISO/IEC 23008-3*, Geneva.
- [3] N. Peters, D. Sen, M.-Y. Kim, O. Wuebbolt and S. M. Weiss, "Scene- based Audio Implemented with Higher Order Ambisonics (HOA)," in *Proc. SMPTE Annu. Tech. Conf. Exhibit.*, Hollywood, 2015.
- [4] Fraunhofer IIS, "EBU and Fraunhofer IIS conducted live MPEG-H Audio Production Trial at Eurovision Song Contest 2018," 05 2018. [Online]. Available: <http://www.audioblog.iis.fraunhofer.com/ebu-fraunhofer-mpeg-h-eurovision/>. [Accessed 28 06 2018].
- [5] A. Silzle, S. George, E. A. P. Habets and T. Bachmann, "Investigation on the Quality of 3D Sound Reproduction," in *Proc. International Conference on Spatial Audio*, Detmold, 2011.
- [6] ITU-R BS.775-3: Multichannel stereophonic sound system with and without accompanying picture, Geneva, 2012.
- [7] K. Hamasaki, "The 22.2 Multichannel Sounds and its Reproduction at Home and Personal Environment," in *Proc. AES 43rd Conference*, Pohang, 2011.
- [8] F. Kuech, M. Kratschmer, B. Neugebauer, M. Meier and F. Baumgarte, "Dynamic Range and Loudness Control in MPEG-H 3D Audio," in *Proc. AES 139th Convention*, New York City, 2015.
- [9] R. L. Bleidt, D. Sen, A. Niedermeier, B. Czelhan, S. Fug, S. Disch, J. Herre, J. Hilpert, M. Neuendorf, H. Fuchs, J. Issing, A. Murtaza, A. Kuntz, M. Kratschmer, F. Kuch, R. Fug, B. Schubert, S. Dick, G. Fuchs, F. Schuh, E. Burdiel, N. Peters and M.-Y. Kim, "Development of the MPEG-H TV Audio System for ATSC 3.0," *IEEE Transactions on broadcasting*, vol. 63, no. 1, 2017.
- [10] P. Poers, "Monitoring and Authoring of 3D Immersive Next Generation Audio formats," in *Proc. AES 139th Convention*, New York City, 2015.
- [11] SMPTE ST 2110-21:2017. Professional Media Over Managed IP Networks: Traffic Shaping and Delivery Timing for Video, New York City, 2017.
- [12] Y. Grewe, C. Simon and U. Scuda, "Producing Next Generation Audio using the MPEG-H TV Audio System," in *Proc. BEITC Conference*, Las Vegas, 2018.
- [13] Eurovision, "186 Million Viewers for the 2018 Eurovision Song Contest," 05 2018. [Online]. Available: <https://eurovision.tv/story/186-million-viewers-2018-eurovision-song-contest>. [Accessed 28 06 2018].
- [14] K. Hamasaki, "Multichannel Recording Techniques for Reproducing adequate Spatial Impression," in *Proc. AES 24th Conference*, Banff, 2003.
- [15] ITU-R BT.2100-1: Image Parameter Values for High Dynamic Range Television for Use in Production and International Programme Exchange, Geneva, 2017.
- [16] ETSI EN 302 755 V1.4.1 Digital Video Broadcasting (DVB): Frame Structure Channel Coding and Modulation for a Second Generation Digital Terrestrial Television Broadcasting System (DVB-T2), 2015.
- [17] ETSI EN 302 307-1 V1.4.1 Digital Video Broadcasting (DVB): Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and other broadband Satellite Applications. Part 1: DVB-S2, 2014.
- [18] H. Wittek and G. Theile, "Development and application of a stereophonic multichannel recording technique for 3D Audio and VR," in *Proc. AES 143th Convention*, New York City, 2017.
- [19] ITU-R BS.1116-3: Methods for the Subjective Assessment of Small Impairments in Audio Systems, Geneva, 2015.



**Christian Simon** was born in Düsseldorf, Germany, in 1976. He received his Dipl.-Tonmeister in Audiovisual Media from the Film University in Babelsberg, Germany.

He has over 15 years of experience in audio recording and post-production with a focus on mixing and dialogue editing. With his award-winning startup *Easy Listen*, he was the first developer to realize a service for optimization of speech intelligibility for AV media in Germany. At present, Christian is working as a scientist and member of the SoundLab group at Fraunhofer IIS with a key focus on Next Generation Audio and accessibility. Furthermore, he is a visiting lecturer at the Ansbach University of Applied Sciences.



**Yannik Grewe** was born in Böblingen, Germany, in 1991. He received the B.Sc. degree in audiovisual media engineering from university of applied sciences Offenburg, Germany. He joined the Fraunhofer Institute for Integrated Circuits IIS in 2013 as a scientist and field application engineer

for MPEG-H 3D Audio. Y. Grewe is mainly enrolled in developments of production tools for Next Generation Audio, field tests for MPEG-H 3D Audio, 3D sound for broadcasting and virtual reality.



**Nicolas Faecks** was born in Hamburg, Germany in 1986. He received his B.Sc. in Media Technologies and his M.A. in Time-based Media (Sound/Vision) from the University of Applied Science in Hamburg. In 2014, he joined the Fraunhofer Institute for Integrated Circuits (IIS) as a research and

application engineer in the department of Media Systems and Application. His recent activities are focused on the rollout of MPEG-H 3D Audio.



**Ulli Scuda** was born in Berlin, Germany in 1979. He received a Dipl.-Tonmeister degree at the Film University Babelsberg in Germany. His experience covers sound recording, sound design and mixing for various film and music formats. Currently, he works as a Tonmeister for Fraunhofer IIS. As head

of the SoundLab group in the audio and multimedia department, U. Scuda researches 3D audio production and reproduction technologies. His main expertise is 3D audio content production for Next Generation Audio.

Received in 2018-06-30 | Approved in 2018-08-01