DOI: 10.22032/dbt.39945
DOI (proceedings): 10.22032/dbt.39936

Verband Deutscher Tonmeister e.V.

Abstract Reviewed Paper at ICSA 2019

Presented by VDT.

Field Test for Immersive and Interactive Audio Production with the Gewandhausorchester Leipzig using MPEG-H

Jakob Mäsel[1], Christian Simon[1], Jin Choi[2], Andreas Schulz[3], Andreas Silzle[1]

[1] *Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany, Email: jakob.maesel @iis.fraunhofer.de*

[2] *Sempre La Musica Korea, Seoul, Republic of Korea*

[3] *Gewandhaus zu Leipzig, Leipzig, Germany*

Abstract

Next Generation Audio (NGA) codecs offer new features for consumers such as advanced user interactivity, immersive sound and optimized reproduction across different classes of playback devices. Yet, 'real world' experience in the application of NGA workflows in classical music production is limited: The realistic depiction of the original concert hall sound and the treatment of elevated sound sources impose challenges when recording and mixing. Besides producing for immersive reproduction systems, sound engineers face new tasks like the integration of user interactivity, ensuring downmix compatibility and loudness consistency.

This paper expounds our field test for an NGA production workflow found suitable for complex orchestral music. Using MPEG-H 3D Audio as an example, we recorded three full-length concerts with Gewandhausorchester Leipzig. Among them is the prestigious new year's concert with Beethoven's 9th symphony conducted by Gewandhaus-kapellmeister Andris Nelsons. We developed a microphone setup for the immersive recording of classical music in Gewandhaus and achieved convincing results with the recording and mixing strategies presented in this paper. The perceived reverberance in the mixes was found to be realistically relatable to the excellent room sound in the original concert hall. Elevated sources, like the choir and soloists, were clearly depicted as such. For metadata authoring and mastering, MPEG-H production tools were used to regulate loudness, dynamic range and downmix compatibility, targeting playback over loudspeakers and 3D audio soundbars. Additionally, we tested two different approaches for interactivity that are suitable for classical music listeners.

Key words: Next Generation Audio, MPEG-H Audio, immersive music production, user interactivity, classical music.

1 Introduction

Next Generation Audio (NGA) codecs provide the foundation for producing, distributing and playing back immersive and user-interactive content, whether it be channel- or object-based audio. The ISO standard MPEG-H 3D Audio[1] [1] allows for transmitting and reproducing immersive mixes, user interactivity or accessibility features. It is further possible to render content to different reproduction systems with varying channel configuration. This makes it easy to serve different playback scenarios with only one dedicated mix: Whether it be stereo, surround, immersive formats with elevated speakers or binaural reproduction [2].

Although all the necessary tools exist, there are only few immersive music recordings fully exploiting the new features of NGA. There are even less publications on real world applications, especially in classical music. Yet the upcoming of 3D soundbars opens up the mass market towards the reproduction of NGA content.

Recording for immersive playback scenarios with elevated sound sources raises new questions and there is hardly any publication on end to end chains for immersive and user-interactive music productions. Thus, with our field test, we aim to find an NGA production workflow that is suitable for complex orchestral music. This paper documents our immersive recording of the Gewandhausorchester's New Year's concert 2018/2019 and the according preliminary recording sessions. Our main target was the realistic depiction of the listener's perception of the concert hall, maintaining the high standards of today´s classical music production. Besides room accuracy, we focused on the reproduction of elevated sources, such as choir and soloists, as they were placed on a tribune behind the orchestra.

Using the example of our immersive recordings in Gewandhaus (see Fig. 1) and the according mixing process, this paper will further delineate the opportunities of mastering object based audio (OBA) during the authoring process. This includes handling loudness and downmixing in the course of universal delivery. Subsequently, two possible scenarios for user interactivity are exemplified in section 4.3.

Fig. 1: 3D audio recordings were performed in Gewandhaus zu Leipzig, Germany.

2 Discussion of the Chosen Production Setup

Understanding the principles behind multichannel recording of classical music for surround reproduction setups is the basis for recording in 3D. Theile [3] presented different approaches, substantiated by ORF (Austrian public service broadcasting) listening tests conducted by Camerer [4]. They compared many different surround microphone arrays and reported outstanding image stability and good evaluation of timbre for at least the Optimized Cardioid Triangle (OCT) and the Decca Tree respectively. Thus, in Gewandhaus we compared OCT 3D and Decca Tree plus wide spaced microphones and the Hamasaki square for surround and height layer according to the findings of Hamasaki et al. [5] [6]. To obtain flexibility for adding more diffuse sound compared to the main microphone system, we set up a more distant wide spaced AB microphone setup. The combination of Decca and AB is well known and has a long history in famous recording studios [7]. Its widely observed preference over the spatially more accurate OCT system for symphonic orchestra content could be explained with the use of omnidirectional microphones in AB/Decca arrays instead of cardioid or super-cardioid microphones in OCT. This provides a more linear bass depiction and is thus said to draw a more realistic picture in terms of timbre [8]. This coincides with the studies of Rumsey et al. [9] on weighted preference: They found that timbral fidelity is a dominant influence on overall sound quality ratings with an impact of 67% for surround sound reproduction. It is followed by front localization with 25% and surround localization (8%). This suggests that localization accuracy might not be the top priority for main microphone array design although it still plays a considerable role. It is surely interesting how the weighted preferences change when introducing envelopment and engulfment as new independent parameters, strongly depending on the use of horizontal and vertical loudspeakers respectively [10].

The principles of spatial and directional depiction in surround recording also apply to 3D production techniques. They are well known but require careful setup adjustment and experimentation [11]. Yet, choosing microphone setups for 3D recording imposes new challenges due to the increasing amount of channels and their according relationships: With more channels, interchannel crosstalk is more likely to happen and more difficult to avoid [12]. This coincides with studies by Scuda [13] and Grewe [14]: For the evaluation of microphone setups, they reported strong dependencies on musical content and recording location. Therefore we used and combined different main microphone setups in order to stay flexible regarding different musical material and instrumentation.

Field tests for recording and transmitting 3D audio content have been publicized rarely but on a regular basis: Nishiguchi et al. [15] from Japanese broadcaster NHK were one of the first ones reporting about it for their 9+10+3[2] (22.2) speaker

[1] Further referred to as MPEG-H

[2] ITU-R BS 2051 notation (Height+Mid+Lower)

configuration. Field tests in sports broadcasting using MPEG-H were conducted by Stenzel et al. [16]. The EU research project ORPHEUS evaluated in detail all aspects of a complete end-to-end OBA chain for audio-only content [17]. In corporation with the European Broadcast Union (EBU) and major European broadcasting stations, they successfully showed that the production and transmission of object-based and user-interactive 3D audio content is possible with the NGA codec MPEG-H, explaining the full chain and workflow [18] including extensive user experience tests [19]. Many ideas for content were created and tested during their experiments. Simon et al. conducted field tests for the immersive and user-interactive live production and transmission of the Eurovision Song Contest 2018 and French Tennis Open 2018 [20]. The field tests outlined in this paper extend the previous described activities to the field of classical music.

3 Production and Distribution of Classical Music with MPEG-H

NGA enables a whole new way to produce and consume audio. Legacy codecs are only able to deliver unchangeable mixes. With OBA, mixes can adapt to the capabilities of their respective playback device and be personalized by consumers. This is realized by the production and delivery of discrete audio components, which are then flexibly rendered in the playback device with the help of affixed metadata. The new process of the production workflow is called authoring, creating a so-called audio scene, including all audio and metadata.

MPEG-H comprises three NGA features, which are all relevant for classical music production:

- Immersive Audio
- Interactivity
- Universal Delivery

Immersive audio refers to three-dimensional sound and enhanced envelopment compared to speaker configurations that are not using the height layer. It is possible to produce for certain 3D audio channel layouts defined in [1] as well as to create layout-independent audio objects defined by their relative position to the listener.
Interactivity allows for changing the mix in the playback device. Producers decide which parameters the listeners can adjust and to what extend. This can be selecting presets with different presentations of the delivered audio components. Even changes in the position of objects or their relative volume levels are conceivably possible.
Universal delivery enables for playback of the same audio content in a wide range of different reproduction scenarios. It comprises a format converter and an elaborate dynamic range control (DRC). The format converter changes the raw multi-channel audio stream to the selected playback format, as can be seen in Fig. 2. This concept allows for producing only in the largest reproduction format, while the adaption for smaller formats happens in the renderer of the playback device. The converter uses an active downmix algorithm to avoid downmix artefacts and ensure artistic integrity. The downmix

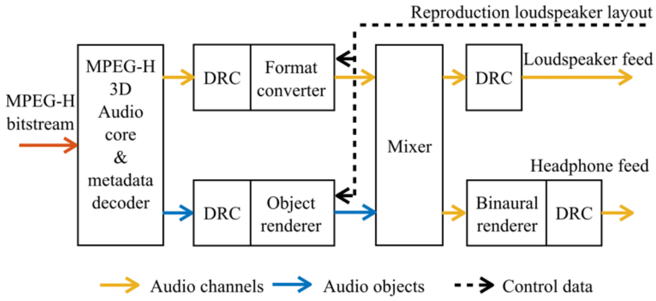

Fig. 2: Top-level block diagram of MPEG-H 3D Audio decoder with dynamic range control (DRC). All depicted components are controlled by metadata.

coefficients in MPEG-H are highly flexible and can be defined by the producer.
The DRC adapts the overall level to a device-dependent target loudness and reduces the dynamic range accordingly. It is based on the integrated loudness measurement of each component contained in the audio scene. In [1], three target loudness profiles are defined:

- Audio Video Receiver (AVR): -31 LUFS
- TV: -24 LUFS
- Mobile devices: -16 LUFS

Furthermore, the MPEG-H system automatically normalizes program loudness according to common standards (e.g. EBU R-128 [21] [22], ITU-R BS.1770.4 [23], ATSC A/85 [24], etc.) providing loudness consistency between different content. Also, the loudness between different presets is adapted, avoiding leaps in level during user interaction. In certain cases, especially in OBA mastering for music, it can be necessary to replace the automatic loudness adaption with manual loudness settings, for example to preserve the natural or aesthetically motivated loudness relations between single movements. All mentioned features are under full control of producer or broadcaster.

Dedicated production tools enable for authoring MPEG-H content: They provide the mentioned features, allow for monitoring of interactivity and universal delivery and, eventually, export audio and affixed metadata. The export format can be BWF-ADM [25] or the MPEG-H Production Format (MPF), consisting of the audio data and metadata, the latter modulated to an audio file or stream in the so called Control Track [26].

4 Field Test

4.1. Recording and Mixing

Over the course of three different concerts, we experimented with different microphone setups. We focused on symphonic material with elevated sound sources, like choir and vocal soloists, as the third and main recording was supposed to be Beethoven’s 9th Symphony played by the Gewandhaus-orchester and conducted by Gewandhauskapellmeister Andris Nelsons.

As main microphone we compared three different setups (see Fig. 3):

(1) OCT 3D with additional omnidirectional microphones on front left and right to compensate for the restrained bass response of the hyper-cardioid microphones
(2) AB with omnidirectional microphones, spaced 1 meter
(3) Decca Tree with omnidirectional microphones, spaced 2.20 meters, with retracted center (30 cm ahead)[3] and outriggers

The mixing was done on a 4+5+0 loudspeaker setup in the so called "Mozart" studio at Fraunhofer IIS audio laboratories, [27]. During the mixing sessions, we subjectively evaluated the setups and found the OCT to have the most accurate localization and the benefit of very controllable elevation in the 3D mix due to good signal separation. On the other hand, poor low frequency fidelity was perceived due to the use of super cardioids. Adding the signal of the additional omnidirectional microphones helped soften this impression. Still, the OCT's stereo image was perceived more narrow and center-focused compared to all other setups. This was becoming more obvious during playback on a 3D soundbar: the OCT direct signal was found to stay mostly within the soundbar's geometrical limits, while spaced omnidirectional setups appeared broader than the soundbar itself. The spaced omnidirectional setup enhanced the feeling of envelopment.

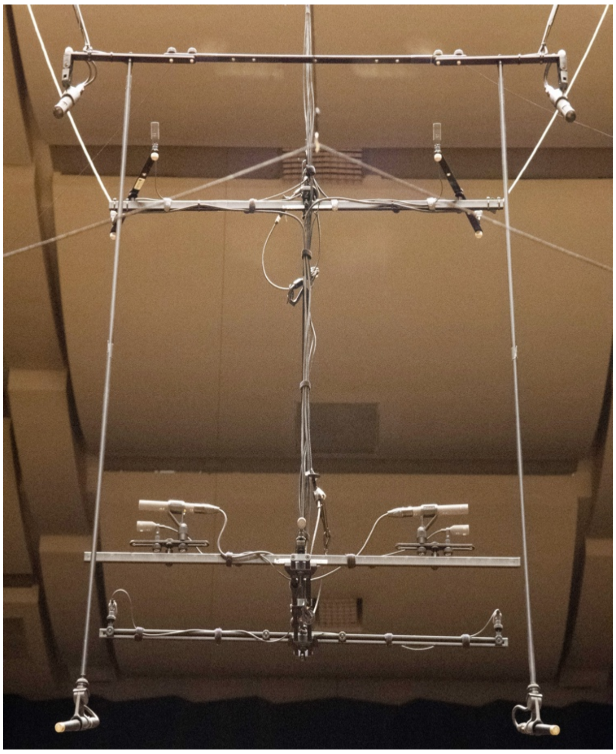

Fig. 3: Experimental main microphone setup in Gewandhaus zu Leipzig with: spaced height AB (Sennheiser MKH 800 Twin); OCT 3D height layer, OCT front with additional omnidirectional microphones on front left / front right, surround; and spaced AB (top to bottom).

[3] This approach is similar to previous surround recordings of Polyhymnia International.

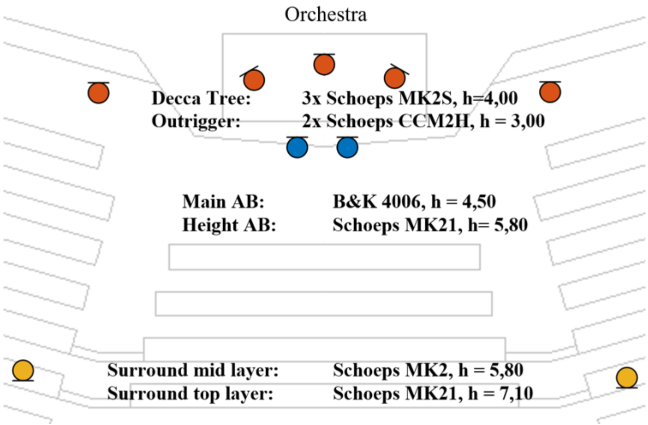

Fig. 4: Final main microphone setup for the 3D recording of Beethoven's 9th Symphony in Gewandhaus zu Leipzig: Decca Tree with Outriggers; spaced AB and height AB; mid and top layer surround microphones (from top to bottom). All designations in meter.

Furthermore, it produced a distinct string sound that could neither be achieved with the OCT nor its combination with omnidirectional microphones. As we were recording string-focused symphonic content, we eventually decided on a combination of Decca Tree and a more diffuse AB system to gain flexibility in the mix (see Fig. 4). It is important to point out that this is not a general recommendation, as the choice of main microphone setups remains a rather subjective decision depending on the recording location, material and artistic taste. This being said, in the very same concert hall we would have opted for the OCT setup for music with many individually localizable elements. Our decision was based on the musical content not least, suggesting the blending of the orchestral instruments and a very broad envelopment to support the composer's intention: the fusion "of the worldly and the sacred" [28].

Using omnidirectional microphones for the height layer bears the risk of overly prominent direct sound in the height channels [29], which could lead to vertical blur in localization and tone coloration after downmixing. Thus, in vertically spaced arrays Lee et al. [30] suggest directional polar patterns, angled outwards to minimize interchannel crosstalk. For experimentation purposes we therefore used MKH 800 twin microphones to stay flexible regarding the degree of directivity as well as a combination of omnidirectional and cardioid microphones. In the final setup we used two Schoeps MK21, placed 1,30 meters above the main AB with same base width. They were angled up- and outwards to provide sufficient separation from the orchestra's direct sound and still catch enough direct sound from the elevated sources to situate them in the height layer. We used the Hamasaki Square as additional room system feeding the height channels.

In the surround microphones we were facing disturbingly hard reflections from the concrete balustrade. The choice of microphone type, position and angle overcomes this challenge. The surround's height layer is designed in the exact same way as the front height: 1,30 meters above the according mid layer microphone, angled up- and outwards. They are widely spaced (distance L-R: 12 meters) in order to increase interaural fluctuations in the low frequencies and thus envelopment [11].

The concerts were monitored and recorded in a special 3D audio OB Van from B&R Media, allowing loudspeaker playback of immersive content. It was also possible to listen to the active MPEG-H downmixes and a binauralized version of the recording in order to evaluate downmix compatibility in an early stage.

In our mix, the main height system played a distinct role, not only increasing vertical spread and the perceived size of room and orchestra but also providing subjectively satisfying musical balance and fusing of voices in the choir. Yet, regarding the apparent elevation of sources, their spot microphones were the decisive contributors. The very narrow polar pattern of Microtech Gefell's KEM microphones, used as spots for the elevated sources, was useful to isolate the choir from the instruments and make sure that the orchestra sound does not bleed into the height layer when enhancing the choir's direct sound. We delayed these spots in accordance to their measured distance from the main system to keep its sonic blending and spectral balance and filtered them to reduce the apparent proximity. Spots used to elevate sources were not panned to intermediate positions between loudspeakers. So they were panned directly to the height L and R speaker position. In order to ensure downmix compatibility, they were distributed to one respective channel. Further, we used two surround reverberators; one for the height and one for the mid layer to create reflection patterns for the spot microphones used.

Subtle denoising was applied in the pre-mastering process, yet we did not remove coughs and audience noise because, especially during silent parts, they give a good impression about both, the size and quality of the concert hall and the playback system.

4.2. Mastering and Authoring in MPEG-H

The necessary steps in the MPEG-H mastering and authoring process for the production described in this paper were:

(1) Configuration of channel layout
(2) Configuration of contained objects
(3) Configuration of presets
(4) Configuration of user-interactivity
(5) Configuration of loudness values
(6) Modification of downmix coefficients
(7) Labeling in different languages
(8) Audio and metadata export

We used the MPEG-H Authoring Plugin (MHAPI) [31] for authoring the final mix in the DAW mastering session. In MHAPI, audio is organized in two layers: Components and presets. A component represents an incoming mono or multichannel audio track which is affixed with metadata like labels and interactivity ranges. Presets contain one or more components in different compilations and/or with different possibilities of interactivity. They allow the user to easily switch between different versions of the audio scene, but it is also possible to only author *one* preset.

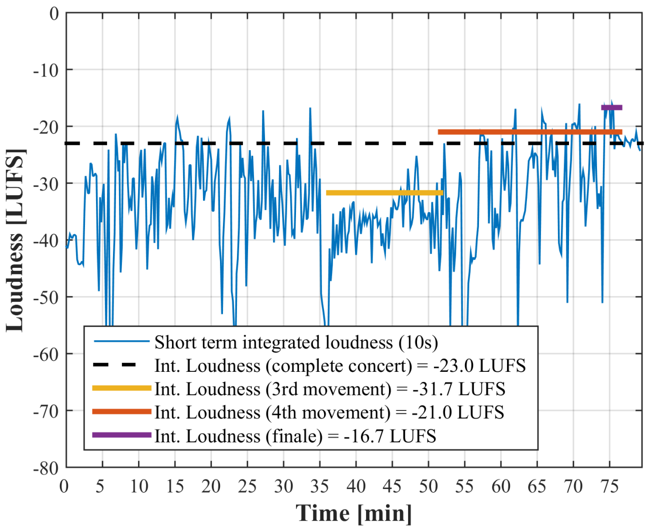

Fig. 5: Short term integrated loudness over the whole concert recorded. The overall program loudness was adjusted to -23.0 LUFS (dotted line), the loudness values of further parts of the symphony are shown.

For simplicity's sake, the following delineates the procedure for authoring audio content with only one preset without interactivity: First, the components have to be created. We chose a 4+5+0 channel format as target layout and assigned the respective input channels in the right order. No further components needed to be defined. Next, we created a preset and included the 4+5+0 audio component. As this is the only preset, it acts as default preset, active whenever the item is played back.

As explained in chapter 3, MPEG-H comprises a program loudness measurement and a loudness normalization based on it. Different movements or even parts of movements, like the finale marked purple in Fig. 5, have different loudness values. In order to preserve the artistic intention, these relative loudness differences must be maintained when exporting singular items from one cohesive program. After exporting e.g. the quiet third movement (yellow) solely, decoders would raise the item's level to match their respective target level, e.g -23 LUFS for TV reproduction. The much louder fourth movement, depicted in red, would be attenuated to the same target level. Thus the relative loudness difference between the movements would be lost, resulting in a flattening of the symphony´s dynamic that diametrically opposes the musical intent. To overcome this challenge, *album loudness* values can be set [32]. In this example, for each item being exported, its measured loudness was replaced with the measured overall program loudness, so decoder level adaptions do not influence the relative loudness differences within a compilation of individual items.

To ensure compatibility to reproduction systems with fewer channels, downmix rules can be defined by producers. In the case of the production described here, it was important to preserve the direct sound of the elevated sources, choir and soloists, which were exclusively panned to the top front speakers in the 4+5+0 mix. Thus, those channels were only slightly attenuated for the 0+5+0 and 0+2+0 downmix configurations. On the other hand, surround and top surround objects were attenuated more, in order to avoid overly diffuse downmixes. The downmixes can be monitored in MHAPI, allowing control over all channel layouts available in MPEG-H smaller than the target layout.

Before exporting audio and metadata to the MPEG-H production format, labeling presets and components in different languages provides good service for the multilingual dissemination of audiovisual content. Depending on the preferred language settings on their target device, users will see the labeling in their respective language accordingly.

4.3. Suggestions for Interactivity in Classical Music

We implemented two possible user interactivity scenarios in this production in order to show possibilities and encourage content producers to invent further scenarios for recipients:

(1) Interactive choice of seats
(2) Multilingual music commentary

For the interactive choice of seats, users could choose between presets called *Conductor* for a close-up orchestral sound, *Queen's seat* for the default mix and *Balcony* for a predominantly spatial experience enhancing the diffuse, enveloping sound as can be seen in Fig. 6. To implement this, we split the mixed signals in four components, separating the basic mix from diffuse surround and height signals:

1. Two mono objects panned to ± 30° comprising the front left and right signal respectively (front main system, spot microphones, reverberation).
2. A 5-channel object comprising the front center signal (front main system, spot microphones), surround reverberation and the height layer front signal
3. A 2-channel object panned to ± 110° comprising the mid layer diffuse surround microphones
4. A 4-channel object panned to ± 30° and ± 110° with elevation at 35° (height speaker positions) comprising the height layer surround microphones, Hamasaki square and the 4 channel upper layer reverberation.

Changing the relative level between the components conveys the impression of sitting closer or further away from the orchestra. When choosing the *Conductor* preset, left and right objects from the first component are spread in panorama using dynamic position metadata. This creates the impression of sitting in the middle of the orchestra.

As a second interactive use case, we produced a music commentary that users can switch on and off, choosing between two different languages. The commentators shed light on compositional and content-related aspects of the pieces. Theoretical backgrounds of the compositions are thus made comprehensible to a broad audience. The commentary offers an additional informative level that combines entertainment and cultural education. Users are further allowed to change volume and position within restrictions that we set in the MPEG-H Authoring Plugin. For the multilingual music commentary, we created a 4+5+0 channel bed

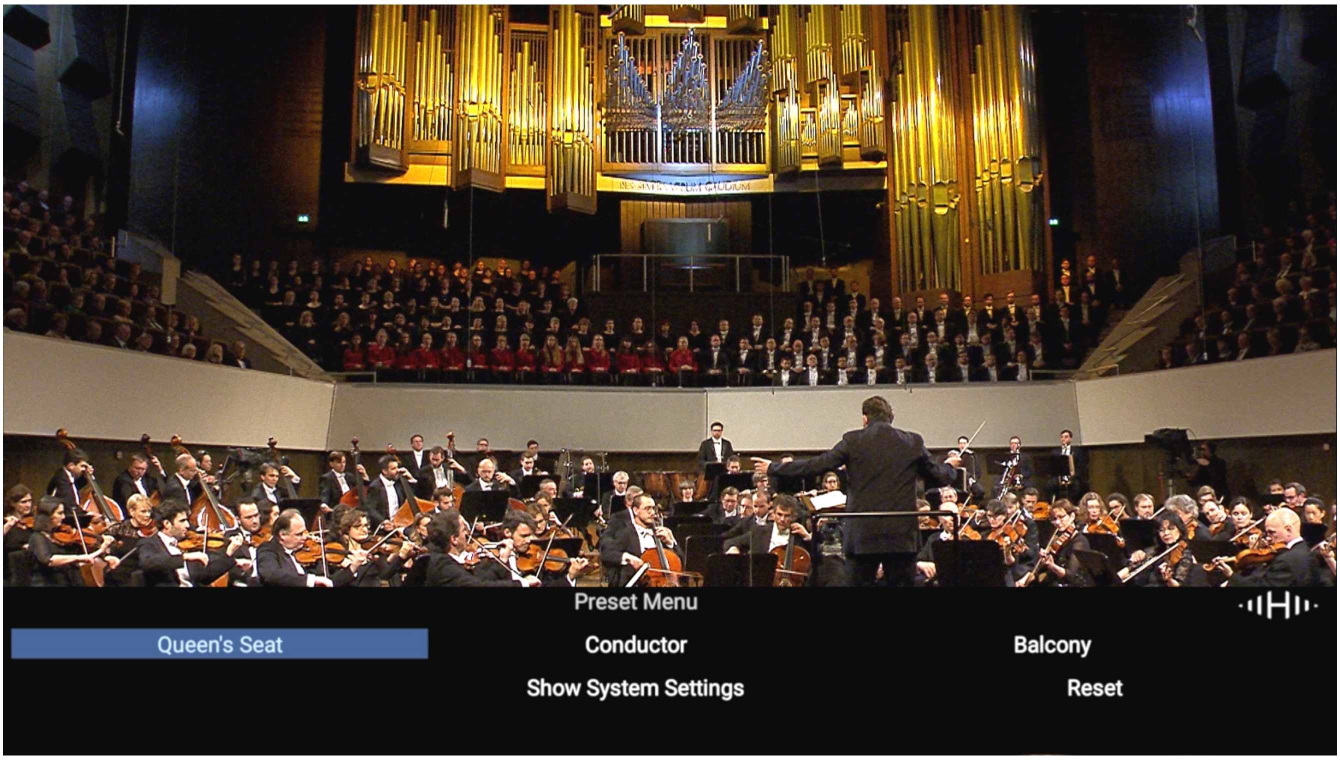

Fig. 6: On-screen display on the user´s TV set to show available interactivity options. The top line (Queen's Seat, Conductor, Balcony) represents pre-configured mix presets to choose from.

comprising the music mix (as described in 4.2) and two mono objects containing the commentary. User interaction restraints were set, so volume and position can be modified within certain limitations: We allowed commentary gain shifts between -2 dB and +4 dB and position changes in the horizontal layer of ±110° and between 0° and +35° in the vertical layer. The default position was set to 0° in both, vertical and horizontal layer equivalent to the M0 or center speaker in a 4+5+0 speaker layout.

To ensure speech intelligibility, the channel bed is lowered by the decoder, as soon as commentary signal is present, comparable to a normal voice-over mix. This is attained by gain metadata, so called *gain sequences*, that are attached to the regarding preset and are applied in the decoder to a pre-assigned audio component. Gain sequences are comparable to volume automation. They are encoded and transported within the regular MPEG-H stream. This allows for sending only one, unattenuated bed mix and the related objects. Depending on the user preset selected, the respective objects are activated and their gain sequences process the mix accordingly.

5 Conclusions

This paper describes our field test for immersive and interactive audio production in classical music at Gewandhaus zu Leipzig. We recorded three different concerts, i.a. Beethoven's 9th symphony conducted by Gewandhauskapellmeister Andris Nelsons. We found convincing microphone setups for the immersive recording of classical symphonic music in Gewandhaus. Elevated sources like choir and soloists were captured by our height microphone system and supported by their corresponding spots in the mix. Elevation and spatial fidelity were rated credibly authentic in informal expert listening evaluations. As the main microphone system, a combination of Decca Tree and a more diffuse spaced AB was used for the mid layer. Broad cardioids were used in a spaced AB system for the front height channels. Surround microphones were widely spaced with omnidirectional microphones as mid-layer surrounds and broad cardioids in the height layer. We used a Hamasaki square as room microphone system for the height layer.

Beyond that, we specified the common process for mastering and authoring classical music using MPEG-H. Downmix compatibility as well as loudness consistency are ensured by MPEG-H active downmix and loudness adaption algorithms respectively. Challenges can occur when exporting individual items with differing loudness and compiling them to one program. We suggested a solution using *album loudness* and detailed the further procedure up until the encoding process, using MPEG-H specific software.

Furthermore, we suggested two possible scenarios for user-interactivity: Choice of seats within the concert hall and a commentary which explains the theoretical background to the composition within an edutainment context. From the concerts recorded, we created AV- and audio-only demos, which show that MPEG-H is suitable for the presupposed requirements.

We recommend to put further work into the microphone setup research and optimization. Also, NGA music mastering and authoring processes should be further investigated and optimized. The evaluation of the described interactivity scenarios is another topic to investigate further.

6 References

[1] ISO/IEC 23008-3, *High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D Audio,* Geneva, Switzerland: International Standard, 2019.

[2] J. Herre, J. Hilpert, A. Kuntz et al. "MPEG-H Audio - The New Standard for Universal Spatial / 3D Audio Coding," in *137th AES Convention* , Los Angeles, USA, 2014.

[3] G. Theile, "Multichannel Natural Music Recording Based on Psychoacoustic Principles," Bolkesjø, Norway, 2001.

[4] F. Camerer and C. Sodl, "Classical Music in Radio and TV - a multichannel challenge," *The ORF Surround Listening Test 2001,* 2001.

[5] K. Hamasaki, K. Hiyama and R. Okumura, "The 22.2 Multichannel Sound System and Its Application," in *118th AES Convention*, Barcelona, Spain, 2005.

[6] K. Hamasaki and W. V. Baelen, "Natural Sound Recording of an Orchestra with Three-dimensional Sound," in *138th AES Convention*, Warsaw, Poland, 2015.

[7] A. Gernemann-Paulsen, "Decca-Tree - Gestern und Heute (The Decca-Tree - Past and Present)," www.uni-koeln.de/phil-fak/muwi/ag/tec/deccatree.pdf, Universität Köln, 2002.

[8] S. Weinzierl, Handbuch der Audiotechnik, Berlin - Heidelberg: Springer Verlag, 2008.

[9] F. Rumsey, S. Zielinski and R. Kassier, "On the relative importance of spatial and timbral fidelities in judgements of degraded multichannel audio quality," *J. Acoust. Soc. Am.,* vol. 118, no. 2, pp. p. 968-976, 2005.

[10] R. Sazdov, "The effect of elevated loudspeakers on the perception of engulfment and the effect of horizontal loudspeakers on the perception of envelopment," in *International Conference on Spatial Audio (ICSA)*, Detmold, Germany, 2011.

[11] D. Griesinger, "General Overview of Spatial Impression, Envelopment, Localisation and Externalisation," in *AES 15th International Conference*, Copenhagen, Denmark, 1998.

[12] G. Theile and H. Wittek, "Die dritte Dimension für Lautsprecher-Stereofonie (The Third Dimension for

Loudspeaker Stereophony)," *VDT Magazin,* pp. p. 31-37, 02 2011.

[13] U. Scuda, "Comparison of Main Microphone Systems for 3D-Audio Recording," in *28th Tonmeistertagung - VDT International Convention*, Cologne, Germany, 2014.

[14] Y. Grewe and U. Scuda, "Comparison of Main Microphone Systems for 3D-Audio Recordings," in *29th Tonmeistertagung - VDT International Convention*, Cologne, Germany, 2016.

[15] T. Nishiguchi, R. Okumura and Y. Nakayama, "Production and Live Transmission of 22.2 Multichannel Sound with Ultrahigh-definition TV," in *122nd AES Convention*, Vienna, Austria, 2007.

[16] H. Stenzel and U. Scuda, "Producing Interactive Immersive Sound for MPEG-H: A Field Test for Sports Broadcasting," in *137th AES Convention*, Los Angeles, USA, 2014.

[17] A. Silzle, "The EU Project ORPHEUS: Object-based Broadcasting - for Next Generation Audio Experiences," *VDT Magazin,* vol. 1, no. ISSN: 2509-5927, pp. 24-27, 2017.

[18] EBU-TR-042, "Example of an End-to-end OBA Broadcast Architecture and Workflow," European Broadcasting Union, Geneva, Switzerland, 2018.

[19] A. Silzle, R. Schmidt, W. Bleisteiner et al. "Quality of Experience Tests of an Object-Based Radio Reproduction App on a Mobile Device," *J. Audio Eng. Soc.,* vol. 67, no. 7/8, pp. 568-583, 2019.

[20] C. Simon, Y. Grewe, N. Faecks et al. "Field Tests for Immersive and Interactive Broadcast Audio Production using MPEG-H.," *Set International Journal of Broadcast Engineering,* vol. 4, pp. 40-46, 2018.

[21] "EBU Recommendation R 128: Loudness Normalisation and Permitted Maximum Level of Audio Signals," European Broadcasting Union, Geneva, Switzerland, 2014.

[22] "EBU Tech 3343 v3, Guidlines for Production of Programmes in Accordance with EBU R 128," European Broadcast Union, Geneva, Switzerland, 2016.

[23] "ITU-R Recommendation BS.1770-4, Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level," Intern. Telecom Union, Geneva, Switzerland, 2015.

[24] "ATSC Doc. A/85, ATSC Recommended Practice: Technique for Establishing and Maintaining Audio Loudness for Digital Television," 2013.

[25] ITU-BS-2076, "Audio Definition Model," in *ITU Recommendation*, Geneva, Switzerland, 2017.

[26] R. Bleidt et al. "Development of the MPEG-H TV Audio System for ATSC 3.0," *IEEE Transactions on broadcasting,* vol. Vol. 63, no. No. 1, pp. 202 - 236, 2017.

[27] A. Silzle, S. Geyersberger et al. "Vision and Technique Behind the New Studios and Listening Rooms of the Fraunhofer IIS Audio Laboratory," in *126th AES Convention*, Munich, Germany, 2009.

[28] D. B. Levy, Beethoven: The Ninth Symphony (Revised Edition), New Haven and London: Yale University Press, 2005.

[29] W. Howie and R. King, "Exploratory Microphone Techniques for Three-dimensional Classical Music Recording," in *138th AES Convention*, Warsaw, 2015.

[30] H. Lee and C. Gribben, "Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array," *J. Audio Eng. Soc.,* vol. 62, no. 12, pp. 870-848, 2014.

[31] M. Rose, "MPEG-H Authoring Plug-in 2.0," Fraunhofer IIS, 2019. [Online]. Available: https://www.iis.fraunhofer.de/de/ff/amm/dl/software/mhapi.html. [Accessed 29 07 2019].

[32] ISO/IEC 23003-4, *MPEG Audio Technologies Part 4: Dynamic Range Control,* Geneva, Switzerland, 2015.

[33] W. Howie, R. King and M. Boerum, "Listener Preference for Height Channel Microphone Polar Patterns in Threedimensional Recording," in *139th AES Convention*, New York, USA, 2015.

[34] N. Zacharov, C. Pike and F. Melchior, "Next Generation Audio System Assessement Using the Multiple Stimulus Ideal Profile Method," in *8th International Conference on Quality of Multimedia Experience (QoMEX)*, Lisbon, Portugal, 2016.

[35] T. H. Pedersen, and N. Zacharov, "The Development of a Sound Wheel for Reproduced Sound," in *138th AES Convention*, Warsaw, Poland, 2015.

[36] F. Kuech, M. Kratschmer, B. Neugebauer et al. "Dynamic Range and Loudness Control in MPEG-H 3D Audio," in *139th AES Convention*, New York, USA, 2015.

[37] Y. Grewe, C. Simon and U. Scuda, "Producing Next Generation Audio using the MPEG-H TV Audio System," in *NAB*, Las Vegas, 2018.

[38] ISO/IEC/JTC1/SC29/WG11, "MPEG2014/N14462 Active Downmix Control," València, Spain, 2014.