Scene-based Audio Implemented with Higher Order Ambisonics

By Nils Peters, Deep Sen, Moo-Young Kim, Oliver Wuebbolt, and S. Merrill Weiss

Abstract

Scene-based audio uses a sound-field technology called "higher-order ambisonics" (HOA) to create holistic descriptions of both live-captured and artistically created sound scenes that are independent of specific loudspeaker layouts. For efficient representation, the audio can be carried as a set of PCM channels that contain predominant sounds and ambience in separate tracks. Standard audio bandwidth-compression techniques can then be applied to the PCM channels. This approach is in contrast to conventional channel-based sound representations in which one signal is used for each loudspeaker of a target reproduction system, with the implication that upmixing or downmixing is required when loudspeaker configurations other than the intended one are used for actual reproduction. This paper examines how, with scene-based audio, there can be satisfactory reproduction of immersive sound at bitrates corresponding to the equivalent of only six channels, whereas an alternative sound-field method that exclusively employs audio objects typically involves much higher bitrates.

In scene based audio, a sound scene is represented by a set of loudspeaker-agnostic HOA coefficient signals. These coefficient signals are the linear weights of spatial orthogonal basis functions and can be rendered to different loudspeaker setups or headphones.

Keywords

Scene-based audio, higher-order ambisonics, HOA, spatial audio, next-generation audio, MPEG-H, ATSC 3.0

Introduction

In this paper, we show how the practical scene-based audio concept, specifically higher-order ambisonics (HOA), can be used to broadcast immersive audio efficiently.

This paper is organized as follows. First, we compare scene-based audio with the other two prominent immersive audio concepts—channel-based audio and object-based audio. Next, we provide further details on the principles of HOA. Then the scene-based compression specified in the recent MPEG-H 3D audio standard is described, including a review of other approaches. The paper ends with an explanation of how HOA is rendered specifically to a consumer's personal reproduction environment.

Three Immersive Audio Concepts

The International Telecommunication Union[1] and the European Broadcast Union (EBU)[2] recognize and support three distinct immersive audio concepts these concepts are entitled channel-based audio, object-based audio, and scene-based audio. Each offers different features and advantages to content creators and audio consumers, while fundamentally differing in the ways in which content is broadcast and is then decoded and processed.

Channel-Based Audio

Channel-based audio has been the *de facto* format in the broadcast industry with the wide deployment of Stereo 2.0 and Surround Sound 5.1 formats. The name originates from the fact that loudspeaker feeds are mixed by content creators for predefined (standardized) loudspeaker setups and delivered in predefined sequences (**Fig. 1**). Because content creators "bake" the content into loudspeaker channels, complex rendering equipment at the decoder is not necessary. To perceive the intended audio image, consumers have to connect their loudspeakers in the correct order and have all loudspeakers placed as dictated by the channel-based format. This has the drawback that a loudspeaker setup that does not match the intended setup will result in a degraded sound image.

Digital Object Identifier 10.5594/JMI.2016.2623398
Date of publication: 16 December 2016

Object-Based Audio

In contrast to channel-based audio, object-based audio delivers unmixed audio components in the form of audio objects in conjunction with object metadata. Based on these metadata, the audio objects are rendered to final loudspeaker feeds at the listening location. Rendering at the listening location has the advantage that each consumer's loudspeaker configuration is taken into account in creating the loudspeaker feeds (**Fig. 2**). Therefore, individual nonstandardized loudspeaker configurations as well as legacy and future immersive loudspeaker configurations are supported. Further, if the system and content allow for it, individual audio objects can be manipulated by users, e.g., by changing the location or the loudness of each object. On the one hand, this enables a personalized audio experience; on the other hand, it reduces the content creator's control over the reproduced mix compared with channel-based content.

An aspect that often is overlooked in object-based audio is the increased cost of transmitting and rendering audio objects: The exclusive use of dynamic object-based audio in the broadcast world is challenging because of the transmission bandwidth needed and the processing power required to decode and render many audio objects simultaneously. To circumvent this challenge, audio objects often are advocated for use in combination with channel-based or scene-based audio.

Scene-Based Audio

A concept that is receiving increasing attention is scene-based audio. Scene-based audio is especially interesting for broadcasting because it combines the benefits known from both channel-based and object-based audio. Similar to channel-based audio, the content creator "bakes" the content into a fixed number of audio signals. In contrast to channel-based audio, and similar to object-based audio, however, scene-based audio signals are agnostic with respect to the loudspeaker reproduction layout. In scene-based audio, content is transmitted in the form of so-called HOA coefficient signals that describe all sounds and their time-varying spatial properties within a sound scene, independent of any particular loudspeaker setup. Based on the actual loudspeaker configuration at the listening location, these HOA coefficient signals are rendered into appropriate loudspeaker signals (**Fig. 3**). Consequently, nonstandard, legacy as well as future immersive loudspeaker configurations can be accommodated—a benefit in common with object-based audio. In contrast to object-based audio, however, in scene-based audio, the number of channels is constant and independent of the number of sound sources. Thus, bandwidth cost and renderer complexity do not scale with scene complexity and are easy to predict for scene-based audio systems.

With scene-based audio, content creation in both live-capture and cinema-based workflow situations is possible. Live capture is especially simplified. Using a microphone array, such as the baseball-sized em32 Eigenmike (**Fig. 4**), along with any number of optional spot mics, it is possible to capture live immersive sound scenes realistically with little to no human intervention. As with the workflow for object-based audio, one also can design scene-based content in a digital audio workstation (DAW) by panning audio-objects to desired positions. Captured live or designed in a DAW, scene-based audio also offers a new palette of audio effects to shape immersive content. For instance, it is easy to rotate, mirror, and stretch an entire sound scene[3] or to "Photoshop" sounds originating from a specific direction.

Table 1 provides an overview of the benefits of the three different immersive audio concepts. It is noteworthy that all three concepts can be combined simultaneously to create any desired degree of personalization (such as with multiple language objects) and coding efficiency.

Principles of HOA

HOA[4,5] represents a three-dimensional dynamic sound scene by using a set of orthogonal basis functions. These basis functions are the so-called spherical

FIGURE 1. Channel-based audio concept.

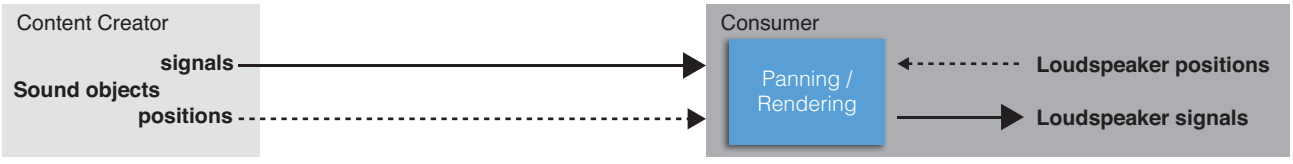

FIGURE 2. Object-based audio concept.

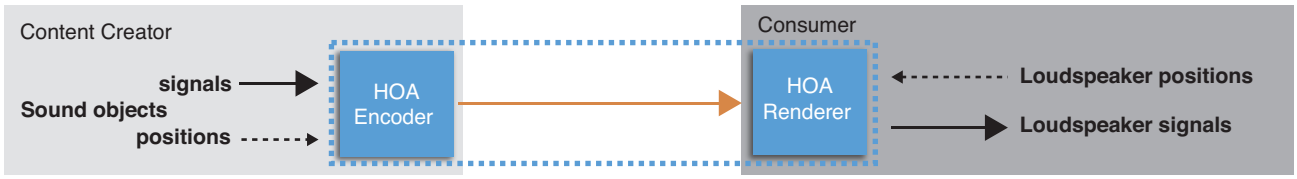

FIGURE 3. Scene-based audio concept.

harmonics, which are hierarchically organized by an increasing order N. **Figure 5** shows the polar pattern of these harmonics from order $N = 0$ (top row) to order $N = 3$ (bottom row). With increasing order, these harmonics become more and more spatially directive and, because higher-order harmonics build up on lower-order harmonics, the spatial resolution of the scene representation increases with each additional order and is (for $N > 3$) approximately $180/N$ degrees.[6]

The order of the spherical harmonics is theoretically unlimited; in practice, a sound scene is described with spherical harmonics up to a given order N. This order N determines the number of audio signals (a.k.a., HOA coefficient signals) required to convey the sound scene, whose number is equal to $(N + 1)^2$. The HOA coefficient signals describe how much weight is given to the associated spherical harmonic basis functions over time. Typically, the order N is between 3 and 6, resulting in 16–49 HOA coefficient signals.

Because HOA is a hierarchical format, it is possible to remove high-order coefficient signals of an HOA signal without loss of audio elements in the scene but with the penalty of reduced spatial resolution.

FIGURE 4. The em32 microphone array (right) and for visual size comparison, a Sennheiser MKH 8070 directional microphone (left).

TABLE 1. Comparison of benefits provided by the three immersive audio concepts

Feature	Channel-Based	Object-Based	Scene-Based
Easy and efficient live capture of immersive sound scenes			x
Easy postprocessing of live capture of immersive sound scenes			x
Creation of immersive content within DAWs	x	x	x
Final creator control over mix through "baking" content into fixed number of audio channels	x		x
Decoder complexity independent from scene complexity	x		x
Transmitted audio signals independent of consumer loudspeaker layout		x	x
Interactive user control of single audio elements		x	
Binaural headphone rendering enabled	x	x	x
Easy deployment in VR applications		x	x

Compression of Scene-Based Audio

Review

Evolving from research on first-order ambisonics, compression of HOA coefficient signals has been investigated in recent years. Hellerud *et al.*[7] and Burnett *et al.*[8] studied whether HOA coefficient signals could be coded with traditional audio coders, such as AAC. It was found that quantization and bit allocation, as a function of the HOA order, shape the perceived reproduction error within the listening area and that the sound quality of the lower orders is more important than the sound quality of higher order HOA coefficients.

Daniel[9] presented an approach to dynamically threshold HOA coefficient signals based on a psychoacoustic model that accounts for spatial masking and the minimum audible angle.

Pulkki *et al.*[10] proposed an extension to Directional Audio Coding (DirAC) to accommodate HOA. In the conventional DirAC,[11] an energy-based time-frequency analysis is used to estimate the dominant direction and a diffuseness value per frequency bin from a first-order ambisonics signal. The transmitted payload consists of these frequency-dependent directional and diffuseness parameters and the omnidirectional zeroth-order ambisonics coefficient signal (**Fig. 5**), which can be further compressed with a perceptual audio codec. In the proposed HOA extension,[10] an Nth-order HOA signal is subdivided into $M \geq N$ even-sized spatial sectors, and each is processed with conventional DirAC. The transmitted payload increases to M times the frequency-dependent directional and diffuseness parameters and M audio signals.

Somewhat related to DirAC, Cheng *et al.*[12] present a method entitled Spatially Squeezed Surround Audio Coding (S^3AC). Here, the sound at the dominant direction of each time-frequency bin is extracted from the three-dimensional soundfield and downmixed into a two-channel stereo signal in which the original sound direction is monotonically squeezed into a direction in the ±30° stereo field. The downmix can be further compressed with a traditional audio coder to create the compressed bitstream. The S^3AC decoder analyzes the stereo signal and remaps the squeezed directions to their original values.

Efficient Compression of Scene-Based Audio for Broadcasting

The fundamental assumption of the scene-based audio compression presented in this paper is that the sound scene consists of a small and time-variant number of predominant sounds (the foreground) and an ambient soundfield (the background). This concept models the way humans decipher complex soundfields: Humans tend to pay attention to a few important sources at a given time.[13,14] Even if a soundfield consists of many contributing sound sources (e.g., an orchestra or a cocktail-party), sound sources sensed as less important fade away into a diffuse and reverberant background where spatial and timbral properties are less critical.

Figure 6 exemplifies a scene-based encoder that can be used to create MPEG-H 3D Audio[16,17] bitstream. The *HOA Analysis and Decomposition* module analyzes the input HOA coefficient signals over a short time frame and extracts a small number of predominant components of the sound scene during that given

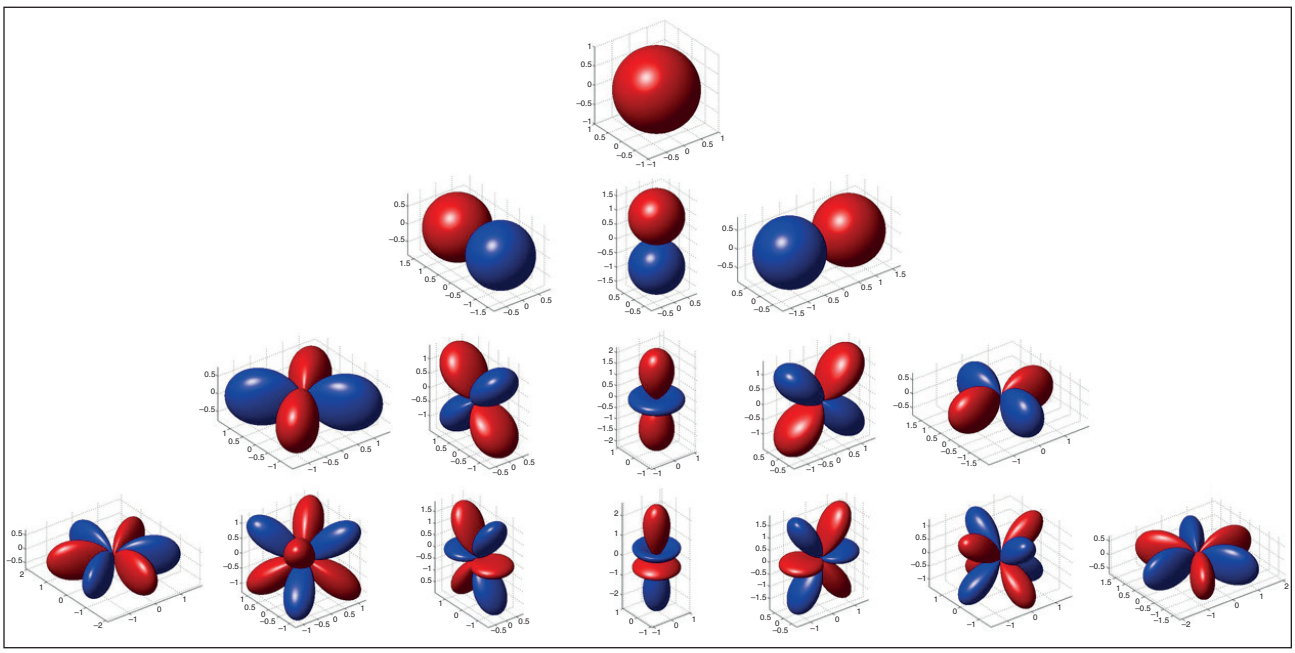

FIGURE 5. Polar pattern of the spherical harmonics from order $N = 0$ (top row) to order $N = 3$ (bottom row).

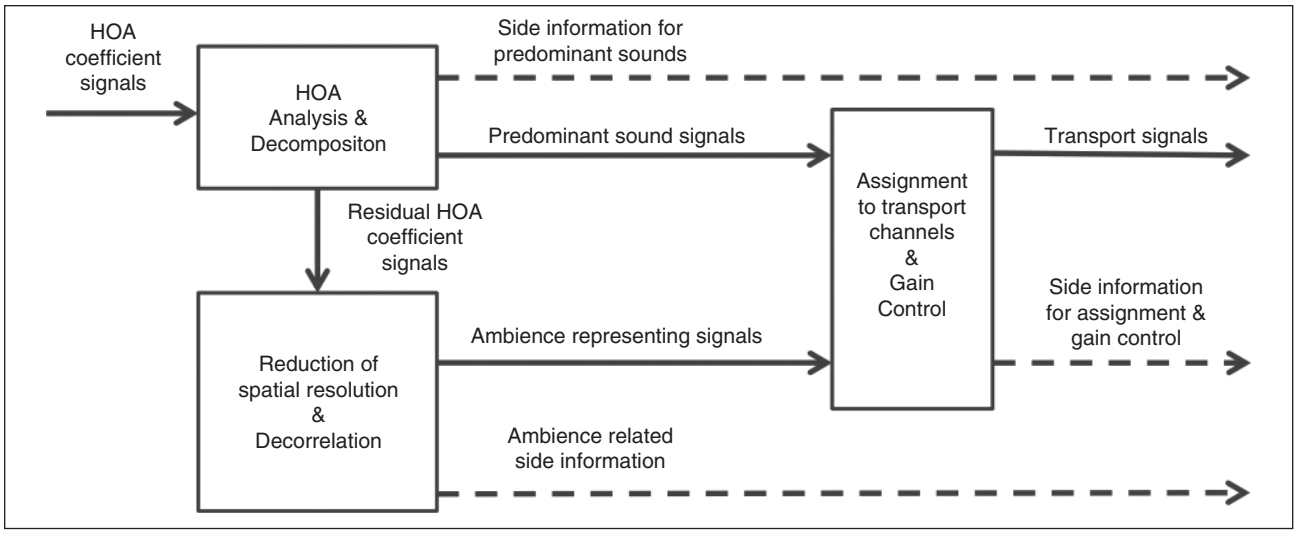

FIGURE 6. HOA spatial encoder.[15]

period. The length of the time frame determines the temporal resolution of the soundfield analysis, and the number of preserved predominant sound elements may depend on the content as well as on the desired spatial compression ratio.

Each predominant sound element is represented by monaural PCM audio data plus parametric side information describing its spatial properties (i.e., spatial distribution, source width). The residual HOA coefficient signal describes the ambient background of the sound scene with all remaining (subsidiary) sounds in the HOA domain. Because the spatial resolution is less critical for this ambient soundfield, its HOA order can be reduced without significant perceptual impact on the overall sound scene, resulting in fewer coefficients required to describe the ambience. To prevent audible artifacts due to spatial unmasking of quantization noise, the low-order HOA ambience coefficients can go through a decorrelation process before further bandwidth-compression.

In summary, due to decomposition of HOA input content into predominant sound elements and order-reduced ambient background elements, HOA content can be reduced to a lower number of PCM transport signals plus associated side information. Satisfactory results have been achieved with as few as six transport signals. Notably, the number of transport signals is independent of the HOA order of the original content but, rather, depends on the desired spatial compression ratio. The compression ratio is approximately $(N + 1)^2/T$, with N being the order of the HOA input content and T being the number of resulting PCM transport signals.

Subsequently, the resulting PCM transport signals are coded using the MPEG-H core coder and are multiplexed with the HOA side information to create the final bitstream (**Fig. 7**). To prepare the PCM signals for the *Core Encoder*, an automatic gain control ensures that all PCM signals have optimal amplitude (**Fig. 6**).

In combination with the MPEG-H core encoder, the HOA content can be coded with a rate-distortion curve, as illustrated in **Fig. 8**. Excellent broadcast quality (≥80 MUSHRA points) is currently achieved at about 300 kbits/sec, independent from the HOA order of the input content. Because this coding technology is at the beginning of its optimization phase, the encoder efficiency will likely improve further over time.

Layered Coding with HOA

A coding strategy especially known within the video broadcasting field is to divide media content into hierarchical coding layers. In a layered (a.k.a., scalable) coding method, a base layer delivers media content in a basic audio quality that can be stepwise raised by having access to one or more hierarchical enhancement layers. This layering makes it possible, for example, to apply a priority-based error protection scheme to the different bitstream layers or to broadcast the same content to devices with different decoding capabilities.

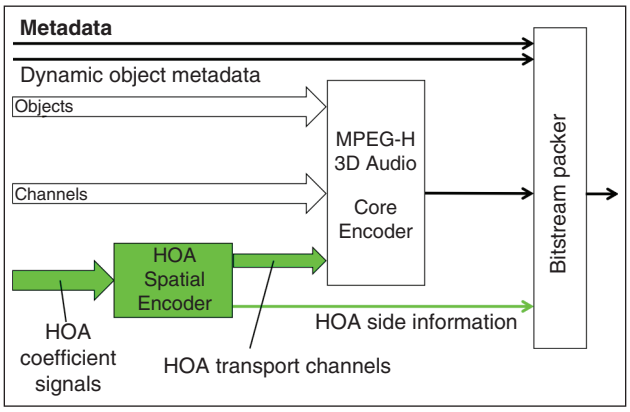

FIGURE 7. MPEG-H reference encoder.[15]

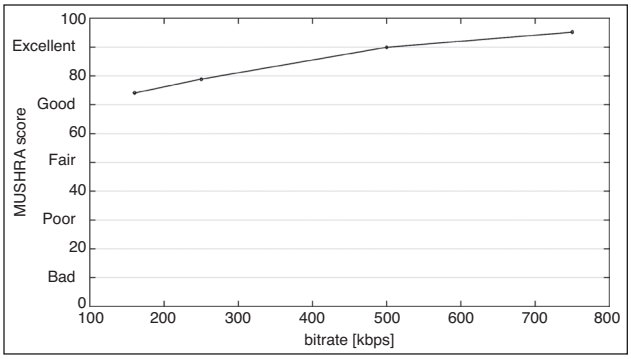

FIGURE 8. Rate-distortion curve for HOA coding with MPEG-H

Because HOA is a hierarchical format (see Section Principles of HOA), it is perfectly suited for layered coding. The layered coding method in HOA is different from other multichannel layered coding approaches in which, for instance, the frequency spectrum is divided (e.g., no high frequencies in the base layer) or in which scene elements are prioritized, resulting in a base layer with just a subset of all audio elements.

In HOA, a base layer could be created by encoding just the few low-order HOA coefficient signals of the HOA input content. From this base layer, it still is possible to reproduce the entire immersive sound scene but with a reduced spatial accuracy. When the remaining higher-order HOA coefficient signals are compressed into enhancement layer(s), decoders can stepwise regain the spatial resolution of the original HOA content.

Decoding of Scene-Based Audio

An overview of the MPEG-H 3D audio decoder is shown in **Fig. 9**.

The *Bitstream unpacker* module takes the received bitstream and extracts the HOA side information together with other metadata. The *Core Decoder* module decodes the remaining portion of the bitstream into PCM signals. These PCM signals can be loudspeaker feeds (channel-based), audio objects (object-based), and/or HOA transport signals. In the case of the latter, the HOA decoder first restores the amplitude of the transport signals and sorts the transport channels into

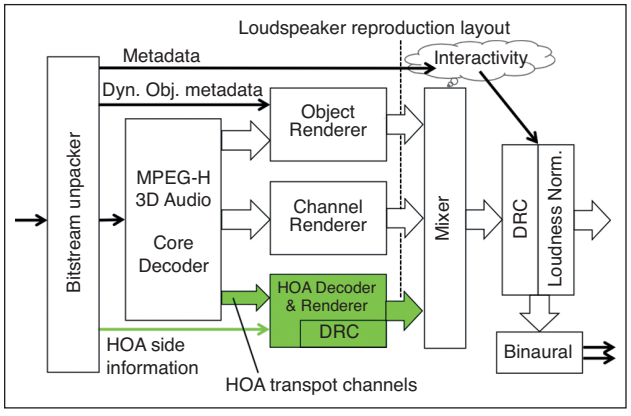

FIGURE 9. MPEG-H reference decoder

predominant sounds elements and ambient background elements (**Fig. 10**). By utilizing the side information, the predominant sound elements are spatially restored into the predominant soundfield. By multiplexing this predominant soundfield with the ambient background soundfield, the final HOA coefficient signals are reconstructed in the *HOA Composition* module.

Listener-Centric Postprocessing

For generating the final audio output, the decoded HOA content is further processed, taking into account the actual reproduction environment. Device-specific or listener-defined settings for dynamic range control (DRC) can be applied, too. Loudness-normalization can be applied to conform to the CALM act or to EBU R128.

Dynamic Range Control

DRC allows for dynamic loudness adaptation to the user's listening situation. For instance, a mobile home cinema is capable of delivering a high dynamic range, whereas the loudspeakers on a tablet are not designed for that. Similarly, based on DRC gain sequences in the bitstream and a user-specified DRC preset, the dynamic range of the HOA content can be modified, e.g., to reduce the dynamic range for late-night TV watching. Algorithms for DRC modify the dynamic range of loudspeaker feeds. To apply those DRC algorithms in the HOA context, the HOA content is rendered to a predefined virtual loudspeaker setup with $(N + 1)^2$ nearly equidistant loudspeakers. For these $(N + 1)^2$ predefined virtual loudspeakers, the (multiband) DRC gain sequences are stored as gain parameters in the bitstream and the desired DRC effect is applied accordingly, similar to the processing for channel-based content. Finally, all $(N + 1)^2$ DRC-processed virtual loudspeaker signals are converted back to the HOA domain. In case of direction-independent DRC gain sequences the DRC effect can be efficiently realized by processing the T transport channels before HOA decoding rather than $(N + 1)^2$ virtual loudspeaker channels.

Rendering

To listen to the scene-based audio content over loudspeakers, scene-based audio content is rendered into loudspeaker feeds. This process is efficiently carried out by multiplying the decoded HOA content with a rendering matrix. This matrix can be generated based on the number of decoded HOA coefficients and the locations of the loudspeakers in the reproduction environment. Several approaches for generating an optimal rendering matrix can be found.[16,18,19]

In contrast to the situation with channel-based audio, the infamous workaround of upmixing and then downmixing to accommodate loudspeaker configurations that differ from the configuration for which content was intended is not necessary in scene-based audio.

The HOA order N (and therefore the number of $[N + 1]^2$ HOA coefficient signals) does not dictate

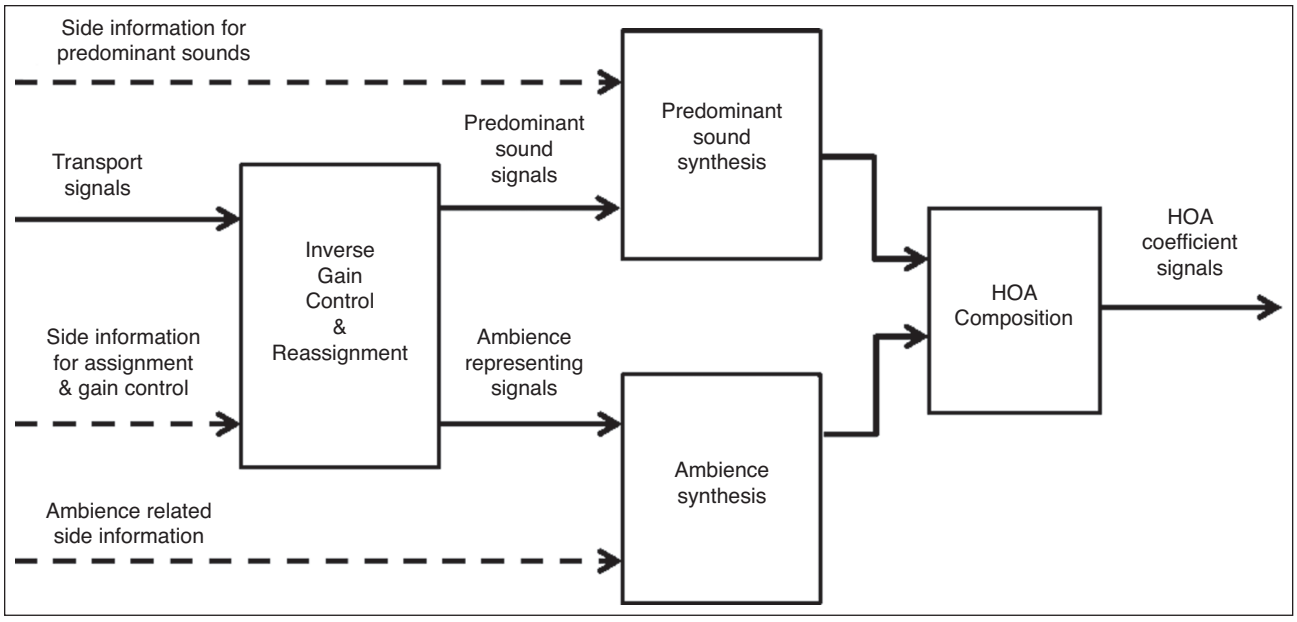

FIGURE 10. HOA spatial decoder.

a certain number of loudspeakers or a certain loudspeaker arrangement. Even HOA content with a small order (e.g., $N = 2$, $[N + 1]^2 = 9$ reconstructed HOA coefficient signals) can be rendered to a large loudspeaker array; and reconstructed HOA content with a higher order (e.g., $N = 6$, 49 HOA coefficient signals) can be rendered for stereo. This all depends only on the rendering matrix, which adapts to the conditions in the listening environment.

Headphone Rendering—Binauralization

An optional processing step for headphone listening is binauralization of the audio content.

The aim is to provide the headphone listener with the sensation of listening to the rendered loudspeaker signals in a desired room. The auditory character of this room is captured by a set of so-called binaural room impulse responses (BRIRs).

For creating binaural headphone signals, each rendered loudspeaker signal is filtered with its associated BRIR and summed together to create a left ear and a right ear headphone signal.

An advantage of scene-based audio is that it can be binauralized, including dynamic head movements. Dynamic head movements are important for creating a convincing illusion, including out-of-head sensation and accurate localization without front-back confusion.[20] This makes scene-based audio a viable option for emerging virtual reality or augmented reality applications. For this purpose, the HOA content is inversely rotated to compensate for head movements sensed by a head tracker. The time-varying soundfield rotation is performed by multiplying the $(N + 1)^2$ HOA coefficient signals with a $(N + 1)^2 \times (N + 1)^2$ rotation matrix before binauralizing.[21]

Conclusion

In this paper, we discussed why scene-based audio, in particular HOA, is a practical, elegant solution for creating and transmitting immersive content for next-generation audio services.

We outlined the advantages of HOA and compared it to other immersive-audio concepts. We further described how, in MPEG-H, scene-based audio can be coded and transmitted with high efficiency and rendered specifically to a consumer's personal reproduction environment.

More information on scene-based audio can be found in a white paper.[22]

References

1. ITU-R BS.2076-0 Audio Definition Model, International Telecommunication Union, Geneva, Switzerland, June 2015.
2. EBU Tech 3364, Audio Definition Model, European Broadcasting Union, Geneva, Switzerland, 2014.
3. M. Frank, F. Zotter, and A. Sontacchi "Producing 3D audio in ambisonics," in *Proc. 57th Int. AES Conf., The Future of Audio Entertainment Technology – Cinema, Television and the Internet*, p. 1–8, 2015.
4. M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, 53(11): 1004–25, Nov. 2005.
5. J. Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," Ph.D. dissertation, University of Paris VI, France, 2000.
6. B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *J. Acoust. Soc. Am.*, 116(4): 2149–57, Oct. 2004.
7. E. Hellerud, A. Solvang, and U. P. Svensson, "Spatial redundancy in higher order ambisonics and its use for lowdelay lossless compression, *IEEE Int." Conf. Acoust., Speech Signal Process. (ICASSP)*, pp. 269–72, Apr. 2009.
8. I. Burnett et al., "Encoding higher order ambisonics with AAC," in *Proc. 124th AES Conv.*, p. 1–8, May 2008.

9. A. Daniel, R. Nicol, and S. McAdams, "Multichannel audio coding based on minimum audible angles," *Proc. 40th Int. Conf., Spatial Audio, Sense Sound Space*, p. 1–10, 2010.
10. V. Pulkki et al., "Parametric spatial audio reproduction with higher-order B-format microphone input," *Proc 134th AES Conv.*, p. 1–10, 2013.
11. V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc.*, 55(6): 503–16, June 2007.
12. B. Cheng et al., "A general compression approach to multi-channel three-dimensional audio," *IEEE Trans. Audio, Speech, Lang. Process.*, 21(8): 1676–88, 2013.
13. M. Kashino and T. Hirahara, "How many concurrent talkers can we hear out," *Proc. Acoust. Soc. Japan*, 467–8, 1995.
14. D. Huron, "Voice denumerability in polyphonic music of homogeneous timbres," *Music Perception*, 6(4): 361–82, 1989.
15. A. Krueger, "Higher order ambisonics relevant architecture," *ISO/MPEG-H Audio – The New Standard for Universal Spatial/3D Audio Coding, 138th AES Convention*, Warsaw, PL, 2015.
16. ISO/IEC 23008-3:2015, "Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio," International Organization for Standardization, 2015.
17. J. Herre et al., "MPEG-H 3D Audio – the new standard for coding of immersive spatial audio," *IEEE J. Selected Topics Signal Process.*, 9(5): 770–9, Aug 2015.
18. F. Zotter, M. Frank, and H. Pomberger, "Comparison of energy-preserving and all-round ambisonic decoders," *Proc. German Ann. Con. Acoust.* (DAGA), 2013.
19. A. J. Heller, E. Benjamin, and R. Lee, "A toolkit for the design of ambisonic decoders," *Proc. Linux Audio Conf.*, p. 1–12, 2012.
20. D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *J. Audio Eng. Soc.*, 49(10): 904–16, 2001.
21. M. Noisternig et al., "3D binaural sound reproduction using a virtual ambisonic approach," *Proc. IEEE Int. Symp. Virtual Environments, Human-Comput. Interfaces Meas. Syst. (VECIMS'03)*, pp. 174–8, Lugano, Switzerland, July 2003.
22. "Scene based audio: A novel paradigm for immersive and interactive audio user experience," Qualcomm Technologies, Inc., accessed November, 2016, https://www.qualcomm.com/documents/whitepaper-scene-based-audio-mpeg-h.

About the Authors

Nils Peters is a senior staff research engineer and manager at the Multimedia R&D labs at Qualcomm Technologies. Before joining Qualcomm, he was a postdoctoral fellow at the International Computer Science Institute, the Center for New Music and Audio Technologies, and the Parallel Computing Lab at UC Berkeley, CA. Peters earned the PhD degree in music technology from McGill University, Montreal, Canada and an MSc degree in electrical and audio engineering from the University of Technology in Graz, Austria.

Peters has been a researcher in the field of spatial audio for more than a decade and works on technical and perceptual aspects of spatial audio, including soundfield analysis and compression, acoustic sensing, spatial sound reproduction, auditory perception, and sound quality evaluation. He has published more than 40 peer-reviewed papers. Previously, as an audio engineer, he has worked in the fields of recording and postproduction. Currently, Peters serves as co-chair of the Technical Committee for spatial audio at the Audio Engineering Society.

Deep Sen is a senior director in the Multimedia R&D labs at Qualcomm Technologies and leads the 3D-Audio research team in San Diego, CA. From 1994 to 2003, he was at Speech and Audio labs at AT&T Bell Laboratories, NJ, and from 2003 to 2011, he was a faculty member at the School of Electrical Engineering, University of New South Wales, Australia. He has more than 60 peer-reviewed publications. His past experience includes speech and audio coding, perception, biomechanical modeling of the cochlea, blind source separation, beamforming, and hearing prosthetics. He is a member of the Acoustical Society of America, Audio Engineering Society, a senior member of IEEE, and an elected member of the IEEE Speech & Language Technical Committee.

Moo Young Kim is a principal engineer in the Multimedia R&D lab at Qualcomm Technologies. He received an MSc degree in electrical engineering from Yonsei University, Seoul, South Korea, in 1995, and the PhD degree in electrical engineering from the Royal Institute of Technology, Stockholm, Sweden, in 2004. From 1995 to 2000, he worked as a member of the research staff of the Human-Computer Interaction Laboratory at Samsung Advanced Institute of Technology. From 2005 to 2006, he was a senior research engineer in the Dept. Multimedia Technologies at Ericsson Research. From 2006 to 2014, he was an associate professor in the Department of Information and Communication Engineering, Sejong University, Seoul, South Korea. His research interests include speech and audio coding, based on information theory, biometrics, including speaker recognition, speech enhancement, joint source and channel coding, and music information retrieval.

Oliver Wuebbolt, received Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from the University of Hannover, Hannover, Germany, in 1998 and 2006, respectively. From 1999 to 2005, he was a research engineer and teaching assistant with the Institut für Informationsverarbeitung, University of Hannover. From 2005 to 2016, he worked in the Audio Department

of at Technicolor Research & Innovation. As a principal scientist, he was responsible for Technicolor's audio standardization activities. In this role, he has been an active contributor to the Motion Picture Experts Group of ISO/IEC SC29, especially to the recently published MPEG-H 3D Audio standard. In 2016, he joined the Fraunhofer Institute for Integrated Circuits, Erlangen, Germany, pursuing his activities in audio coding, spatial audio, and signal processing.

S. Merrill Weiss is a consultant in electronic media technology and electronic media technology management. His career spans nearly 50 years in electronic media and related fields, with 41 years in management and consulting. He began his career in radio in 1967 and in 1968 developed a method for voltage distribution of analog audio, as opposed to the power distribution scheme that was in wide use at the time. He spent much of his early career designing and building radio and television facilities.

Weiss has been involved in development of SMPTE standards for almost 40 years, participating in work on nearly all of the television technologies with which the Society has been involved. He has been a working group or technology committee chair for the past 35 years and a member of the Standards Committee for 29 years. He served as producer for the 1981 First Demonstrations of Component Coded Digital Video and designed the microprocessor-based machine control system that became the foundation of the physical layer of the SMPTE-standard ESbus (RS-422) machine control protocol. He also served four years as engineering director for television and was a co-chair of the joint SMPTE/EBU Task Force for Harmonized Standards for the Exchange of Program Material as Bit Streams. He currently chairs development of the Archive eXchange Format (AXF). Weiss was elevated to SMPTE Fellow membership status in 1987. He received the David Sarnoff Gold Medal Award in 1995 and the Progress Medal in 2005. He received the NAB Television Engineering Achievement Award in 2006 and the ATSC Bernard J. Lechner Outstanding Contributor Award in 2012. He won the IEEE BTS Matti S. Siukola Memorial Award in both 2012 and 2013. Weiss holds four issued US patents and two international patents. He is a graduate of the Wharton School of the University of Pennsylvania.