Bringing 6DoF to Live Recorded Audio and Video

By
Piotr Szczechowiak
(COO, Zylia)
and Tomasz Żernicki
(CTO, Zylia)

Photo 1: Zylia, applying third-order Ambisonics ZM-1 microphone arrays with Infineon's digital MEMS microphones, has an innovative and unique approach to Six Degrees of Freedom (6DoF) technology in virtual and augmented reality (VR/AR).

Imagine a popular music festival you've always wanted to experience in person—seeing the performers on stage and being surrounded by the music, the people, and the energy of the show. For one reason or another, you can't make it to the festival. But what if you could teleport yourself there? What if you could plug into an immersive experience that allows you to move freely through the whole audio soundscape and visual landscape? This is the promise of Six Degrees of Freedom (6DoF).

True 6DoF experiences enable freedom of movement in 3D space, or along all three movement axes and all three rotational axes. Within this environment a person can move forward/backward (surge), up/down (heave), and left/right (sway), and also rotate—yaw (normal axis), pitch (transverse axis), and roll (longitudinal axis)—to change orientation.

With 6DoF technology in VR/AR, you can experience 3D audio/video content from an unlimited number of listening and vantage points. Wearing an appropriate VR headset and headphones, you could use 6DoF to go anywhere. You could walk along a tropical rainforest path, taking in the sights and sounds of animal life all around you (see **Figure 1**). You could amble among palm trees on a sunny beach, listening to the waves crashing on shore.

While some video games offer virtual worlds that effectively enable 6DoF movement, the real potential of 6DoF lies in exploring the real world and actual places and events within it. The technology and techniques that enable development of real-world 6DoF do exist, and audio and video creatives are just starting to explore the possibilities.

The Advance Toward 6DoF

Until very recently, immersive audio and video were limited to just Three Degrees of Freedom (3DoF). Examples include the 360° videos now available on platforms including YouTube, Facebook, and Vimeo. Sometimes also referred to as immersive videos or spherical videos, these 360°-video offerings incorporate video recordings shot simultaneously using either an omnidirectional camera or a collection of cameras.

Images are stitched together to build a panoramic scene, which can be viewed on a laptop or mobile device, on a head-mounted display, or within a purpose-built space equipped with projectors or displays in a spherical configuration. While people viewing or listening to a 360° or 3DoF scene can rotate to look in different directions or to hear audio from different direction, they themselves remain static—they cannot freely translate their position in space.

The 360° audio is similarly recorded. Algorithms within associated software convert the signals from the recording device into the Ambisonics domain. Ambisonics is the full-sphere surround sound format that has become the standard for 360° filmmakers and video game developers, and it covers not only the horizontal plane, but also sources "above" and "below" the listener (see **Figure 2**).

Unlike other multichannel surround formats, the Ambisonics format carries speaker-independent representations of a sound field called B-format over its transmission channels. The sound field then can be decoded for playback on the target system or device. This approach gives creatives and engineers the flexibility to design sound in terms of source directions rather than loudspeaker positions.

Photo 1 shows Zylia, applying third-order Ambisonics ZM-1 microphone arrays with Infineon's digital MEMS microphones, has an innovative and unique approach to Six Degrees of Freedom (6DoF) technology in virtual and augmented reality (VR/AR).

First-, second-, and third-order Ambisonics, respectively, offer higher spatial resolution. The higher the Ambisonics order, the better the sound source localization in the space. Research shows that it is essential to have at least third-order Ambisonics representation in order to fully experience the 3D audio that is critical to creating a truly immersive experience.

Higher-order Ambisonics allow mixing engineers to create 3D sound tailored to headphones so that the listener experiences panoramic audio. Although this technique—binaural reproduction or binaural recording—is not new, the capture of audio with an Ambisonics microphone and decoding specifically for headphones makes for a listening experience that rivals reality.

In a 360° production, the 3DoF approach to video and audio might allow the viewer to jump between different positions at a concert or other live event. 6DoF takes the experience to a whole new level by enabling the viewer to move freely throughout the scene.

Additionally, with the right software tools, typically plug-ins for digital audio workstations (DAWs), it is possible to process the data captured by the multitrack recording device and perform adaptive isolation of sound sources using a virtual microphone. The virtual microphone is, in fact, a spatial audio filtering technique based on

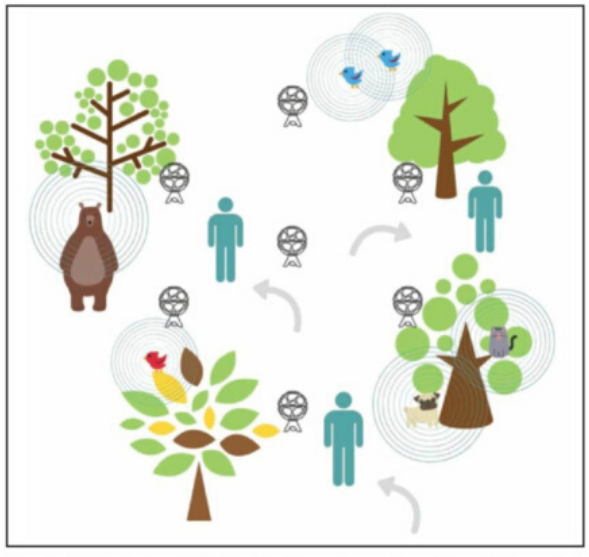

Figure 1: 6DoF technology in VR/AR enables the user to experience 3D audio/video content in the virtual space by introducing an unlimited number of listening and vantage points. Wearing an appropriate VR headset and headphones, you could walk along a tropical rainforest path, taking in the sights and sounds of animal life all around you.

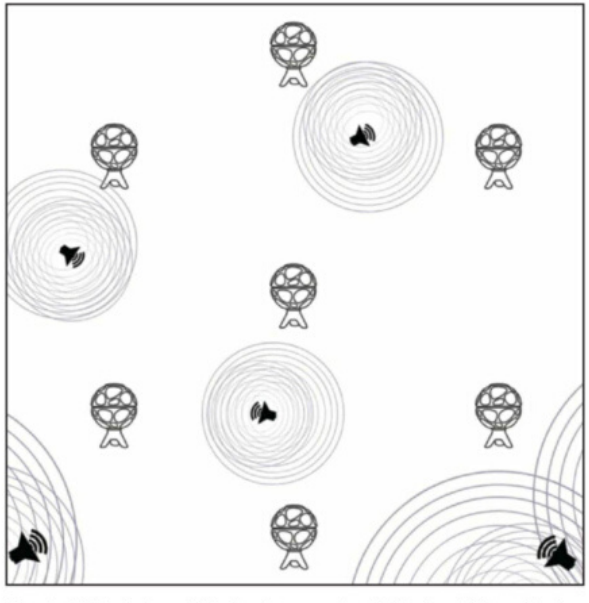

Figure 2: 6DoF technology with Ambisonics covers not only the horizontal plane, but also sources "above" and "below" the listener.

advanced beamforming and sound source separation algorithms, and it accounts for both phase and amplitude differences between signals captured by microphone arrays on the recording device. Each virtual microphone can be configured to with unique spatial characteristics, width, or shape, and each can be used as a single-spot microphone with super-narrow directivity response or with other virtual microphones in a multichannel surround recording set.

For the creative user or engineer, the process of isolating sound can be as simple as pointing the virtual microphone at a particular voice or instrument. Software algorithms interpolate any given position in the scene, allowing desired sounds to be emphasized and unwanted sounds suppressed. Adjustments such as balancing direct and reverberant sound can be fine-tuned during mixing. With the ability to change a spatial characteristic of virtual microphone in postproduction, musicians and audio engineers can avoid the need to repeat pre-recording and recording steps. The high spatial resolution of third-order Ambisonics audio ensures there are no gaps in sound and helps creatives and engineers bring a natural character to content.

Photo 2: 6DoF audio environments deliver on the promise of lifelike, or live, immersive experiences with complete freedom of movement.

About the Authors

Piotr Szczechowiak is the Co-founder and COO of Zylia. He manages the day-to-day operations. Piotr has a PhD in embedded systems and makes sure that ZYLIA hardware and software fits together.

Tomasz Zernicki is the Co-founder and CTO. He manages the technical operations. Tomasz received a PhD in digital audio processing and is an audio expert in the ISO/IEC MPEG standardization group.

Effective 6DoF Workflow for Live Recordings

Implementing a simulated immersive audiovisual scene that supports 6DoF in 3D video games has become fairly common. The same cannot be said for cinematic audio-video or live recordings. Of the two current approaches to 6DoF production, only one opens the door to live recorded navigable 6DoF sound from an actual physical space (see **Photo 2**).

The first of today's techniques for 6DoF content is synthetic object-based 6DoF, which is much like the approach used to build audio in post editing for video games. Each sound source recorded is assigned to a channel, and then rendering engines and specialized DSP algorithms are used to mix thousands of tracks to try to mimic a real environment. Mono or stereo sound is artificially spatialized by software processing that attempts to replicate reverberation, diffusion, and other acoustic behaviors. This is a complicated and time-consuming process that can take months of work in postproduction.

The second approach to creating 6DoF audio environments is much simpler, and it delivers on the promise of lifelike, or live, immersive experiences with complete freedom of movement. Like the aforementioned 3DoF workflow, it leverages a recording solution equipped with a microphone array to capture the entire sound scene in a real environment, as well as software capable of delivering third-order Ambisonics signals.

However, while 3DoF production uses one Ambisonics microphone to recreate the spatial audio surrounding a particular recording device, the 6DoF approach uses many such microphones, distributed throughout the space, to deliver the extensive audio data needed to enable freedom of movement throughout the space. Software can process data in real time and calculate the sound experience for any given point within that space.

Because this model enables 6DoF audio production using sound captured in a natural environment, it features the actual acoustics within the recording venue. As a result, this approach also reduces the need for added processing and specialization in creating believable true-to-life sound. The higher quality of captured sound ensures a smooth listening experience as a person moves around the larger soundscape and gives them the natural feeling of actually being there.

Practical Applications and Benefits

Of course, the feeling of full immersion in a virtual world can only be achieved if realistic sound follows high-quality images. 6DoF has been the subject of a great deal of work in the video realm, with leading use cases being in sports. In this area, 6DoF technology is enabling the creation of replays using a simulated camera—not entirely unlike the virtual microphone in the use of extensive interpolation from data acquired by high-resolution recording devices.

6DoF audio and video are both progressing, though at different rates. Eventually, they will meet at a point in their evolution where they together enable the fully immersive VR experience. In the meantime, demonstrations of 6DoF audio within an artificial visual environment simulated by a Unity game engine are offering glimpses at the future of immersive content.

For artists, the possibilities extended by 6DoF are incredible. The drop in revenues from traditional music sales makes it more important than ever to find ways to connect with audiences. Live shows are a tried-and-true way to engage with fans and to generate revenues. With 6DoF technology, musicians gain much greater flexibility in addressing fans' desires for unique and up-close experiences of the live shows. Musicians also can offer much broader access to their performances.

Fans anywhere can experience the show, with the ability to move throughout the soundscape and to listen from different perspectives. They enjoy control over the experience, moving at will toward the stage or further into the crowd. While it is exciting to get closer to the performer, it's also exciting to be able to watch and hear the show from any vantage point. That exceptional quality of 6DoF sound with today's sophisticated recording and processing solutions also benefits musicians and their fans.

With respect to production, an efficient 6DoF workflow for live recordings can significantly reduce the time, talent, and costs associated with building a 3D audio space for film or TV. In a conventional scenario, experienced sound engineers would be hired to record different actors throughout a scene. Using multiple Ambisonics-capable recording systems with multiple microphone arrays, however, it becomes possible simply to place devices across the 3D space, trigger recording, and then use one or more virtual microphones in postproduction to isolate and manipulate sounds and speakers as desired.

Future Work and Development

To support ongoing progress and, eventually, widespread adoption of immersive media, MPEG has been carrying out explorations on technologies that enable 6DoF. Zylia was the first company to offer practical demonstrations of 6DoF audio enabling reproduction of the listening point from any given location in the scene, and the company's recordings have been adopted by MPEG as reference material for its standards work. Current work includes describing the stream that will carry the elements of 6DoF content.

As standards development progresses, scientists, researchers, and artists are exploring real-world use cases for 6DoF—and they have the tools to experiment, thanks in part to Zylia's 6DoF VR/AR Development Kit. It is extremely challenging to record high-quality sound from many sources present in the sound scene at the same time. The new development makes it easy for users to experiment independently using a multiple-microphone array system in various spatial arrangements around the recorded sound scene. Users can experience not only the rotation of the sound scene (360° audio), but also freely change their position with respect to the sound objects.

The technology for 6DoF audio is not something buried deep in a research laboratory. It's a solution that's already on the market, with top companies and universities investing in exploring how 6DoF content can drive their future applications. Ultimately, the key to success will lie in being able to create immersive live recorded content using 6DoF and bring that experience into the home. The advance of 6DoF audio using multiple Ambisonics microphone systems is laying the foundation for this exciting future.

INNOVATIONS IN AUDIO • AUDIO ELECTRONICS • THE BEST IN DIY AUDIO

www.audioxpress.com

audioxpress

ADVANCING THE EVOLUTION OF AUDIO TECHNOLOGY

Audio Installation Market Update
Integrated, Networked, and Secure

Show Report
2019 Midwest Audiofest Speaker Design Competition
By Thomas Perazella

Audio Electronics
Rethink Your Class D Audio Design and Measurement Practices
By Jens Tybo Jensen (Infineon)

You Can DIY!
The Bantams Mini-Speaker System
By Thomas Zarbo, Jr.

Fresh From the Bench
The DPA d:vice MMA-A Mobile Recording Interface
By Stuart Yaniger

Sound Control
Auralization—Building Virtual Acoustic Models
By Richard Honeycutt

Book Review
Bob Cordell's *Designing Audio Power Amplifiers*, 2nd Edition
By Jan Didden

Audio Praxis
Bringing 6DoF to Live Recorded Audio and Video
By Piotr Szczechowiak and Tomasz Żernicki (Zylia)

OCTOBER 2019