International Conference

Audio for Virtual and Augumented Reality

Los Angeles, USA
30th September –
1st October 2016

CONFERENCE REPORT

Platinum Sponsors

DOLBY

oculus

GAUDIO

dts LISTEN

2016 has seen the world of virtual reality finally take its first steps out of the labs of research universities and into the realm of consumer products and support from major corporations. New apps, experiences, and tools are popping up at an ever faster pace. For the creators of video game engines like Unity and Unreal Engine, this has resulted in an escalating arms race to develop new tools for creating VR experiences, while VR headset creators (including HTC's Vive, Facebook's Oculus, and Microsoft's Hololens) prepare to fight for their share of the rapidly expanding market for virtual and augmented reality.

With the billions being poured into creating ever better VR experiences, the market for realistic, immersive, 3D audio experiences is just beginning to heat up. As a result, the Audio Engineering Society hosted the first-ever Audio for Virtual and Augmented Reality conference alongside the 141st AES International Convention at the LA Convention Center. Scheduling technical papers, real-world tutorials, workshops, and a technical showcase into just two days was challenging, but the sold-out crowd showed up ready and eager to learn.

Kicking off the conference, cochairs Andres Mayo and Linda Gedemer thanked everyone involved in the production of the conference, which was organized and executed with a small team in only nine months. Top industry guests from Google, Magic Leap, Linden Labs, and others reached out to be involved in the inaugural conference, leading with keynote presenter Philip Lelyveld of the Virtual Reality/Augmented Reality Initiative at the University of Southern California.

OPENING KEYNOTE

Starting with a brief video of their new interactive VR experience, CLOUDS, Lelyveld walked the audience through the differences between augmented reality (that superimposes virtual entities and environments over your surroundings via glasses like HoloLens and the secretive Magic Leap) and virtual reality, which utilizes a head-mounted display (HMD) to immerse viewers in virtual worlds. Lelyveld defined the core characteristics of strong virtual reality experiences, including providing the user with a sense of presence, a heightened sense of empathy, and varying degrees of agency and ability to interact with virtual

Linda Gedemer and Andres Mayo, conference cochairs, were happy to greet the many delegates to AVAR 2016.

Philip Lelyveld, opening keynote speaker

worlds. Much to the delight of the crowded theater, in his research he has found that at the moment properly executed spatial audio is even more important than the quality of visuals.

Lelyveld also addressed the most common concern of those who are apprehensive of a future where we all live isolated, Matrix-like experiences in our HMD's. After Facebook's acquisition of Oculus, the world seemed to realize that the likely "killer app" for VR will be social VR, enabling users to engage in experiences unachievable in plain-old-reality, simultaneously with friends located anywhere on the planet, or with astronauts on Mars. One of the keys to immersive social VR is not the visual (with good visuals requiring incredibly powerful PCs to render even simple graphical avatars), but the spatial audio that enables the user to hear other users in 3D. It's this that enables up to 10,000 people to attend the same virtual event in apps like AltspaceVR without experiencing it as a cacophony of voices, as only those nearest to your xyz position are both audible and properly positioned in space.

Finally, Lelyveld raised some tough questions for the future development of AR and VR, with challenges like a lack of standards, the pace of consumer adoption, and even questions of who will define emotional intensity ratings for VR experiences. (Your heart has never beat as fast as the first time a horde of zombies gets the drop on you in VR, for example.) Further, as production professionals, it will fall on us to realize that these experiences have the potential to genuinely change a person over time.

And so with that, the stage was set to listen, to question, and to learn about best practices, failed experiments, and the future of audio for virtual and augmented reality (AVAR).

PRESENTATIONS

The first tutorial of the day addressed audio recording and production for VR & 360 applications. Chaired by Jan Plogsties (Fraunhofer Institute), the panelists were Nils Peters (Qualcomm Tech), Chris Pike (BBC R&D), Jamieson Brettle (Google VR), and Frederick Umminger. Starting with a discussion of the trade-offs made in every AVAR project, they outlined questions all audio pros should ask: What's the goal of this audio? Are we shooting for realism or hyperreality? Will this be a scene-based or object-based recording?" Chris Pike went on to discuss how the BBC has had to develop many of its own production tools, even going so far as to introduce an audio definition model for metadata that allows representation of advanced audio formats, and how to embed it into an extension of standard broadcast.wav files.

Following Pike, Brettle brought the audience up to speed on Google's efforts in VR on YouTube, noting that spatial audio is now enabled on Youtube 360 videos. The discussion then turned to future challenges, including a need to build more robust, efficient, and cost-effective middleware, tools, and workflows for audio professionals navigating this new space.

Another tutorial later in the day chaired by Dinesh Manocha (U. of North Carolina) with panelists Anish Chandak (Impulsonic) and Neil Wakefield (Linden Lab) discussed the challenges of creating a cohesive sound environment in VR experiences ranging from storytelling and gaming environments to the complex problem of how things should sound in completely user-generated environments. Various models exist, ranging from the "Box Model" that requires minimal CPU, but requires intensive setup for the game developer (now tasked with creating acoustic "boxes" with known reverb characteristics throughout their game worlds). Other options like the Phonon or physics-based models attempt to use the materials in the virtual environment to influence the acoustics directly, as they would in the real world. However, this can be an incredibly intensive calculation, especially in user-generated worlds like Linden Labs' Second Life, or its upcoming VR version dubbed Project Sansar (where users can build their own environments to live in and explore). Wakefield outlined the methods to reduce latency and CPU usage, including an automated system to simplify the geometry (for audio calculation purposes) of the user's environment, as well as running the impulse response calculations at half sample rate, before recombining with the original dry signal. The audible

Jan Plogsties and panel colleagues discuss audio recording and production techniques for VR.

results were nothing short of amazing. For those interested in these challenges, the afternoon session on spatial audio and sound propagation for VR delved even deeper, with practical challenges and workflows addressed.

VisiSonics followed with the presentation, "Creating Scientifically Valid Spatial Audio for VR and AR: Theory, Tools, and Workflows." Like many of the presentations by manufacturers, VisiSonics' Ramani Duraiswami and Adam O'Donovan provided a thorough overview of their research into spatial audio systems, and their creation of a workflow designed to capture, measure, model, render, and personalize the sound for each participant in a VR experience. Incredible creations like their 5/64 Audio/Visual Camera (a fascinating spherical creation with 5 cameras and 64 microphones) facilitated incredibly realistic capture and reproduction.

Viktor Phoenix and Scott Gershin of Technicolor discussed postproduction workflows for 3D audio and VR. They started with a premise that is familiar in the audio postproduction world, namely that while the visual has traditionally provided the viewer with critical elements like direction and pacing, in the world of 360 video (where the audience can look in any direction), it is up to the audio pro to fill in the gaps by directing the viewer's attention, providing pacing, and of course, supporting the emotional tone of the experience. Or as Gershin so aptly put it, "It's all about finding the balance. What do you put on screen, versus what do you fill in with sound?"

Faced with that question, the audio professional's mind likely fills with more questions. "What's the best way to mic for a VR experience when the camera is able to see in all directions? Should you hide close mics? Should you put up a few ambisonic/B-format mics? Does one have to mic with the new higher-order ambisonic mics to create an immersive experience?" Their response to this conundrum was simple, "Yes, use them all. Art and physics don't always match…it's our job to create sounds that people believe, the job is creating what feels right." The consensus was that 360 VR shoots required more advanced planning, both for preproduction and post. As far as the miking of the shoot, they stressed that the engineer should scout for mic placement from the angle of the 360 camera, knowing that they need to be completely out of view of the camera and mics during recording.

When it comes to postproduction and mixing, educating composers and sound designers was also emphasized. Few composers have yet written for VR and its individually immersive environments. One challenge noted was that viewers will turn their heads quickly toward higher-pitched sounds. If the sound professional isn't careful, this can result in users whipping their heads back and forth, creating fatigue, or worse. "For example, the sound designer needs to know that our ears, and then our eyes, naturally turn toward the higher frequencies, so avoiding rapid pans of high-pitched sounds

will prevent the viewer from getting whiplash."

Phoenix and Gershin stressed that, due to the fledgling size and nascent status of 360 audio, extensive collaboration and information sharing is needed among all practitioners. Despite the decades of audio and postproduction experience of the panelists, each agreed that even the most experienced had worked on no more than 25 VR projects. Truly, the sky is the limit in terms of what techniques and tools remain to be discovered. "With this community, we're through the desert, now we have to climb the mountain."

Wrapping up the first day, Christoper Hegstrom from Symmetry Audio discussed how AVAR represents both a new paradigm for audio professionals, while also bringing many areas of audio together.

Exploring how AVAR pulls from areas as diverse as game engines, film conventions, mobile and even live streaming, the day ended with participants of diverse audio backgrounds expanding their relationships and opportunities through discovery, sharing ideas, and looking toward the future.

Day Two saw no lull in the enthusiasm of the audience filling the theater. Henney Oh from GAUDIO Lab reviewed the various approaches and challenges of 360 audio. He demonstrated the company's "End-To-End VR Audio Solution" and its unique approach to capturing panning automation within the DAW. Using their GWorks plugin, he was able to live pan the selected track around on the floating 360 Video window, following the sound of the captured characters, and recording the move for later processing.

Simultaneously, Tom Ammermann of New Audio Technology entertained an adjacent crowd with his insights and techniques encountered while producing 360 audio for industrial, professional, and end-user applications. One tip that will be familiar to any audio engineer: help your clients appreciate the value of a 360 mix. By starting the client listening to the 360 mix, he noted that many clients, upon being switched to the standard stereo mix, asked if the headphones were broken.

Nathaniel Kunkel and Jurgen Scharpf of Dolby Labs, Tim Gedemer of Source Sound and Glenn Kiser discussed the unique attributes of the Dolby Atmos object-based approach to surround mixing, instead of relying on channel-based surround (i.e., 5.1, 7.1 etc.) or scene-based binaural audio. Noting several high-profile VR projects they had recently mixed in Atmos (including the stellar Bohemian Rhapsody VR experience available from Google, and new VR TV experience "The Invisible Project"), the conversation explored everything from workflow changes to how we should utilize "objects" in mixing for VR. Kunkel further reframed the challenge, "We're used to getting so much of our audio information from rectangles, a smartphone, a tv, a film etc…but now in VR we've got a sphere, so which sounds do you track with the users head, and which ones do you not?"

From left, Glenn Kiser, Jurgen Scharpf, Tim Gedemer, and Nathaniel Kunkel discuss audio object-based mixing techniques for immersive production.

The consensus from all was to let the scene always dictate the sound. If the scene has a band playing as characters, one would tend to track those sounds with the users movement, retaining immersion by having the band move about in space. Alternatively, for many VR experiences where the music creates the emotional support for the experience, it should not be tracked, but instead emanate from a nonmoving "center of the listeners head." It was a fascinating insight into the expertise and experiences of pioneers in surround audio.

While most engineers would think of a company like Dolby as a pioneer in 360 audio, the next presentation would surprise anyone unfamiliar with the recent history of Facebook, and its $2.1 billion acquisition of VR headset-maker Oculus. Presenters Joel Douek and Benedict Green of Ecco and Varun Nair (2 Big Ears) discussed the workflow for the newly rebranded Facebook 360 Spatial Workstation Tools. Of the many challenges discussed, one of the most perplexing is the dichotomy presented by the supported workstations. Offering workflows for only two DAWs, Reaper (the $60 indie) and Pro Tools HD (previously only available with $5,000+ hardware purchase), this hopefully temporary limitation had many audience members holding out for other tools with a broader range of support.

After a morning of show and tell with these incredible new production tools, Sofia Brazzola and Mikkel Nymand of Sennheiser, Henrik Oppermann of Visualise, Jean-Pascal Beaudoin of Felix & Paul Studios, and Benedict Green of ECCO VR came together for a practical panel on Immersive Sound Capture for Cinematic VR. Keeping the audience alternating between laughter and furious note-taking would have been a daunting task for most panels, but these experts weighed in with good humor and useful advice on the challenges of miking for VR, when to record wild takes of lines and foley, how to combine point sources with B-format ambisonics, and the influence of postproduction on their recording workflow. Specifics like "always bring a mid-side, stereo, and B-format mic, as well as a mix of omni and directional lavaliers" provided the crowd with useful insights into a properly outfitted gear kit, and the reminder that it's always better to be overprepared.

Over lunch, the presentations took at deeper look at the challenges of live VR audio. Panelists Tim Gedemer (Source Sound Inc), Martin Walsh (DTS), Robert Dalton (Dysonics) and Matti Hamalainen (Nokia Technologies) explored the unique challenges associated with producing immersive surround audio for both game-based and cinematic VR, as well as 360 audio for live VR broadcasts (including events like Jack White Live and Sir Paul McCartney Live in Jaunt VR). Creating great-sounding mixes

that also translate to the VR environment not only created a range of challenges, but they also offered unique opportunities for engineers to reestablish the value of great sound in the age of the MP3. Given that there are still so many standards to establish, as well as techniques to develop and perfect, the panel encouraged all participants to maintain the standard for high-quality audio, and excellence in practice. Gedemer summarized it well, "With the advent of VR, we have an opportunity to show people that great sound matters again."

Demonstrating a deftness in their scheduling, the conference organizers filled the afternoon with remarkable presentations on bleeding-edge technology. Expert research demonstrated by companies like Sound Particles explored unique approaches to positioning sounds in 3D using virtual microphones. Nokia demonstrated its DAW workflow for capturing and manipulating spatial sound with its cinematic VR camera the Ozo (which looks more like a UFO than any blurry photographs your crazy uncle has ever taken). VideoStitch followed that with a demonstration of its VR camera with built in MEMS microphones, capable of capturing first-order ambisonics with capsule correction, EQ, level matching, and multiple polar patterns built right in. Finally presenters Francois Becker, Clement Carron, and Benjamin Bernard (Longcat Audio Technologies) demonstrated their own 3D Audio for VR codec and tools they've developed for binaural and surround audio production. Truly, for a room full of audio geeks immersed in the nascent world of VR, it was an afternoon spent peering into the future.

In addition to all of the presentations, both days offered a valuable collection of research papers covering every aspect of sound localization, binaural simulation, object-based audio, 360 capture, mixing, and delivery, as well as new approaches to sound design and composition. Real-world case studies examined the challenges and solutions discovered in specific projects, ranging from luxury airline advertisements, to live streaming a jazz concert in 360 cinematic VR.

CLOSING KEYNOTE

Finally, the pièce de resistance was the closing keynote "Future Nostalgia, Here and Now: Looking Back on Today from 20 Years Hence," presented by George Sanger of Magic Leap. For those behind on the rumor mill, Magic Leap is developing a next-gener-

The attentive audience at sold-out AVAR 2016.

for the average consumer. Twenty years ago, the PC was just getting its first professional sound cards. Now, the entire multitrack studio can live in anyone's computer, and it only gets better from here. While he noted that audio will only be a piece of the next technical revolution, Sanger said "It's our piece...and with any luck, in 20 years, we'll be nodding our heads sagely, saying, 'Yep, we saw that one coming back in 2016.'"

Thought-provoking sessions leave delegates concentrating hard.

ation augmented reality headset, purportedly capable of superimposing photorealistic visuals and audio on the world around you. Having already taken over $1.5 billion in investments from the tech industry, the crowd was buzzing to hear what he had to say about the future of audio for AR/VR.

Sanger began with a newsreel clip of himself demonstrating cutting-edge technology to enrapt crowds in 1984. Showing off software that we now consider comically ancient (including gasps from 1984 onlookers at his ability to draw with MS Paint in 8-bit color), he noted that the computer revolution took place not because people loved technology but because it allowed them to expand their creativity. After inviting the crowd to join him on a journey to the future, he outlined a number of examples from major technology advances in twenty- to forty- year periods, and how each explosion was propelled by consumers drive to expand their creativity. Forty years ago, nearly all audio computing was priced unattainably high

AES THANKS AVAR COMMITTEE

Cochairs: Andres Mayo and Linda Gedemer

Business & Finance Chair/Treasurer:
Richard Wollrich

Event Logistics Chair:
Greg Kadorian

Secretary: Brian Seagrave

Social Media Coordinator:
Pablo "Tango" Formica

Website Coordinator: Gary Epstein

Convention Volunteer Coordinator: Steven O'Hara

Session Chair: Sarah Stevenson

Editor's note: the papers presented at this conference can be obtained from http://www.aes.org/publications/conferences/