International Conference
Audio for Virtual and Augmented Reality

August 20–22, 2018

DigiPen Institute of Technology Redmond, WA, USA

CONFERENCE REPORT

The area of virtual and augmented reality technology is rapidly evolving, and with it new and exciting immersive experiences are being created for the consumer that are born out of cutting-edge audio research and development. The VR industry has experienced remarkable growth in the past five years, and now any consumer with a smartphone or commercial VR headset such as Oculus Go or HTC Vive can experience immersive content from first-person VR games to 360-degree videos. While the application possibilities are limitless and include interactive social media experiences, virtual learning environments, and immersive experiences for health and well-being, there remains significant challenges within VR and AR, with many underlying problems falling into perceptual, signal processing, and sound design areas of immersive audio.

Such challenges were tackled head-on at the 2018 AES International Conference in Virtual and Augmented Reality held at the prestigious DigiPen Institute of Technology in Redmond, Washington. The conference was chaired by Matt Klassen with cochairs Linda Gedemer (workshops), **Lawrence Schwedler** (treasurer), and **Edgar Choueiri** (papers chair). The conference brought together 239 registered attendees, most of whom were leading experts in the field of spatial audio, psychoacoustics, and sound design. There was an impressive breadth of keynote talks, workshops, paper presentations, and posters as well as VR and AR technology demonstrations from Facebook, Bacch Labs, Microsoft, and many others.

AES President David Scheirman, left, with Matt Klassen, conference chair, and Linda Gedemer, conference cochair

Jean-Marc Jot, Monday's keynote speaker

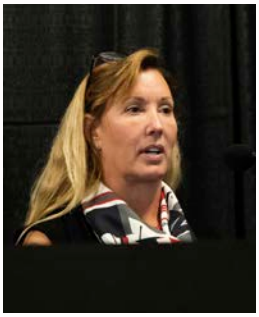
STATE OF THE ART

To set the scene, the conference opened with a compelling keynote from Jean-Marc Jot, Distinguished Fellow of Magic Leap, whose topic was "Wearable 3-D Audio Unleashed." Throughout his talk, Jean-Marc defined the term "unleashed" as meaning that the technology must permeate the consumer level as well as being full six-degrees of freedom, i.e., the end user can move anywhere in the virtual space. He gave a technical overview of immersive audio processing for games and VR with a retrospective look at the last 20 years of research and development including game audio standards VRML, I3DL, OpenAL, and EAX. Jean-Marc touched on the challenges involved in representing the acoustic characteristics inherent in rooms, in particular the representation of early reflections and open environments. This was punctuated with some excellent conceptual examples of "we are there" and "they are here" experiences from Magic Leap. Jean-Marc concluded by giving some final perspectives on the rapidly evolving technology and the convergence between previously disparate media applications and industries. He said that audio plays a critical role in spatial congruence of immersive experiences and a sense of naturalness. To that end he felt that we must evolve the notion of the audio object in immersive audio creation tools to properly include source directivity and orientation.

PERCEPTION

An underlying theme of the conference was the perception of immersive VR and AR spaces within the limitations of the existing state of the art. Leading this theme was a panel discussion chaired by Kedar Shashidhar from Magic Leap that focused on multimodal integration for new realities. Panelists also included Tom Smurdon from Oculus, Chanel Summers from VR Studios, Viktor Phoenix from Headspace, Guy Whitmore from Foxfacerabbitfish, and Robert Rice from Microsoft. The panel discussed the importance of hearing sounds within the build of the game, as the visual percept can have a strong impact on what is heard. This is particularly relevant when procedural audio is used for immersive experiences, where sound effects are synthesized on the fly rather than resorting to triggering of samples during play. The panel discussed many tools for implementing this including Pure Data integration with Wwise. They also posited the question as to whether procedural audio may take away from the artistic intent. Viktor stated that the level of the aesthetic quality may not be where we want it to be yet but having the SFX in real-time and reacting to the user gives a high level of interactive fidelity that goes a long way. The panel felt that there is still a strong need for field recordings and that hybrid approaches give the best of procedural vs. sample-based worlds. The most significant changes within the coming years are likely to be related to physics-based sound generation within the game engines particularly in acoustics and automatic sound propagation. This is of major importance when moving between mixed and virtual reality, with major challenges not only in the accurate generation of the room reflections and reverberation, but the parametric manipulation of the acoustics for artistic intent.

Matt Klassen, conference chair,



Linda Gedemer, cochair, workshops,

Lawrence Schwedler, conference treasurer,

Jean-Marc Jot asks a question of a panel group.

Alongside this panel discussion were some very interesting contributions to the foundational understanding of auditory perception and evaluation in virtual environments. Angela McArthur from Queen Mary University of London, UK, discussed auditory distance discrimination in cinematic virtual reality and how mismatched stimuli are perceived in different environmental conditions including free-field to reverberant spaces. Christopher Stecker, from Vanderbilt University presented a paper on objective measures of auditory co-immersion in virtual and augmented reality. He discussed the need for measures to define immersion where real and virtual objects exist in the same multisensory scene. He proposed the use of VR games in combination with psychophysical measurements to assess user sensitivity to differences in room acoustics.

Edgar Choueiri, cochair, papers, opens the Monday morning papers session.

In a similar vein, Olli Rummukainen from Fraunhofer presented his work on audio quality evaluation in virtual reality. This consisted of a classic multiple stimulus test methodology for VR for evaluation of different binaural renderers. Olli described how the method uses a "ranking through elimination" procedure and demonstrated a pilot study that shows the feasibility of the approach for VR in comparison to a nonimmersive offline method. Gregory Reardon from New York University continued the theme of evaluation of binaural rendering by looking at a multidimensional sound quality assessment. He presented a study where subjects evaluated timbral balance, clarity, naturalness, spaciousness, and dialogue intelligibility, and ranked the renderers in terms of preference for a set of music and movie stimuli presented over headphones. He showed that binaural renderer performance was highly dependent on the content, making it difficult to discern an "optimal" renderer for all contexts.

REALISM AND PERSONALIZATION

An underlying question coming from the perception talks was how realistic do immersive audio environments have to be so that they are considered plausible and natural? This was addressed within the keynote of Ivan Tashev from Microsoft who gave a fascinating insight into the capturing, representation, and rendering of audio for virtual and augmented reality. Ivan discussed the importance of individualization and personalization of HRTFs for enhancing realism in VR/AR experiences. Strategies to make personalization a reality include 3-D head scans and depth map measurements from cameras in order to inform models or generate meshes for computational HRTF creation. Ivan mentioned that there is still no real established perceptual measure for showing if two HRTFs match. Other important considerations are ensuring high-resolution HRTF datasets when considering the slight head movements that are required to localize sound and the importance of head-tracking in this regard. He also discussed the possibility of using multichannel loudspeaker arrays to form invisible headphones in real-time motion tracked conditions, for example using Wave Field Synthesis techniques. Ivan finally brought back the discussion to the current limitations of portable mobile technologies for 3D sound for virtual and augmented reality, but he felt that with some effort, we can

Ivan Tashev, Tuesday's keynote speaker

Agnieszka Roginska and Nicolas Tsingos discussing soundfield proopogation.

already create fantastic experiences with current systems and low-cost headtrackers.

Following Ivan's talk was a panel discussion on realistic soundfield propagation for immersive virtual environments, chaired by Edgar Choueiri from Princeton where Ivan was joined by panelists Agnieszka Roginska, from New York University, Ramani Duraswami from the University of Maryland, Hanes Gamper from Microsoft, and Nicolas Tsingos from Dolby. The discussion began by considering the state of the art in measurement methods of HRTFs from loudspeaker-based solutions to numerical mesh solving solutions. Ramani gave an overview of the reciprocity technique, where instead of microphones at the ears and loudspeakers emitting sine sweeps from desired directions, we can reverse the process and have miniature loudspeakers at the ears with a dense grid of miniature microphones distributed about the head. The panel also discussed database methods, where HRTFs are matched to nearest anthropomorphic details of an individual's head and ears. Other methods discussed were HRTF simplification and smoothing processes (how much simplification can be done?) and adaptive HRTFs where a listener can shape the HRTF through some form of system input. Edgar felt that there needs to be some form of metrics established for assessing HRTFs in terms of localization and coloration. Agnieszka made the important point that there is a continuum from simple HRTFs to full Binaural Room Impulse Responses and that the final application, from game audio to cinema, must be considered.

With binaural audio such an integral part of VR and AR rendering, there were many excellent paper contributions to fast HRTF measurement processes. Traditionally, HRTFs are measured in an anechoic environment, with a person constrained in the center of an array of loudspeakers that play measurement stimuli picked up by microphones at the ears. Such measurements can be quite tedious to undertake and the measurement phase can be long if a high-resolution dataset is required. Some alternative and more user-friendly approaches were proposed at the conference. Sebastian Nagel from Aachen University presented a novel method for using head-tracking to measure HRTFs with unconstrained subject movement. Rishi Shukla presented a system for HRTF selection that relies on holistic judgments of users to identify their optimal match through a series of pairwise adversarial comparisons, resulting in a clear preference toward a single HRTF set for most listeners. Ramani Duraiswami presented a paper that discussed different methods of HRTF individualization, including reciprocity and cloud-based mesh calculation. This work was also showcased by Michel Henein from VisiSonics who discussed the basics of HRTF personalization. Faiyadh Shahid (EmbodyVR) discussed HRTF prediction using machine learning and image recognition techniques, focusing on prediction and optimization by using an artificial intelligence method based on an image of the pinnae.

Ramani Duraiswami discusses HRTF individualization.

AMBISONICS

Another underlying technology within VR/AR audio is ambisonics, which is a 3D-capture and rendering concept that has its mathematical basis in spherical harmonic representation of the soundfield. Originally developed by Michael Gerzon in the 1970s it has found a perfect home in the VR/AR world, not only due to the elegant representation of sound scenes but also the ease at which an ambisonic soundfield can be rotated, tilted, and tumbled to compensate for head-movements. Despite the age of the technology, there is still much research and development on the topic, particularly in the area of higher-order ambisonics and its place in the VR/AR space.

A good summary of developments in the field was presented by Markus Zaunschirm from IEM in Graz, Austria. He discussed different approaches to binaural ambisonic rendering over headphones from classic methods that often suffer from localization blur and timbral coloration to newer techniques that allow for higher order performance from low-order signals. Calum Armstrong from the University of York's AudioLab presented a novel approach to ambisonic binaural decoding that considers two independently rendered soundfields at each ear. The method utilizes new datasets of HRTFs with measurements taken from the perspective of the ear canal rather than the center of the head. Also from York's

Aaron Mcleran, Sally Kellaway, and Nathan Harris on the content creation tools panel.

Brian Schmidt, left, and Jean-Marc Jot

AudioLab was Thomas McKenzie who showed how to improve low-order ambisonic rendering in desired directions by employing directional bias equalization. The method utilizes a two-stage approach of diffuse-field equalization followed by a secondary direction-dependent filtering stage. Olli Rummukainen from Fraunhofer IIS proposed a workflow for facilitating full six degrees of freedom (6DOF) VR rendering over headphones. The model utilizes a parametric approach to first-order ambisonic rendering that exploits knowledge of the distances to each audio source in the scene. Fabian Brinkmann (Technical University of Berlin) presented a paper on spherical harmonic HRTF representation for ambisonics. He used psychoacoustic localization and timbre models to assess different methods of decomposition of spherical harmonic transform with order truncation, and results showed how time alignment produces good rendering results.

The DigiPen Institute of Technology Campus made an excellent venue.

On the soundfield capture side, Fernando Lopez-Lezcano from Stanford University gave an entertaining presentation on the creation of high-quality and low-cost ambisonic microphones, from first-order tetrahedral designs to second-order eight-channel arrays. Michael Goodwin from DTS also presented a hybrid ambisonic beamforming approach that uses directional analysis and spatial synthesis in the frequency domain.

Ambisonics is of course one of several potential formats for VR, the others being channel- and object-based audio. To tackle the delivery of these different formats Patrick Flanagan from THX outlined how the MPEG-H codec could be beneficial to the AR/VR industry in this regard and gave a rundown of the standard's capabilities as well as introducing THX's new compatible plugins.


Scott Selfon, panel moderator on VR content-creation tool.s


Sally-Anne Kellaway takes part in a panel.

CONTENT CREATION

VR technologies are only as good as the content that is consumed with them, and it was no surprise that a large part of the conference was dedicated to the tools and techniques used to create compelling immersive audio experiences. Recording, sound design strategies, and mix workflows were discussed alongside content creator perspectives on the future of VR and AR production.

Scott Selfon from Facebook Reality Labs led a panel discussion on the state of the art in VR audio content creation tools and workflows. He was joined by Jean Marc Jot, Nathan Harris from Audio-Kinetic, Sally Kellaway from Microsoft Mixed Reality, Aaron McLeran from Epic Software, and Brian Schmidt from DigiPen. The panel discussed how VR tools have evolved from games, but that people who are used to linear workflows struggle with such tools due to their many shortcomings. Conversely linear DAWs are not going to cut it in the VR world even though they give good creative flow to sound designers. The panel felt that overall more DAWS could embrace better workflows and UX design for VR audio content creation. They also discussed the difficulties in the creative workflow when considering end-user proximity to sound sources resulting in dynamic range issues. A good example of this is where we may not want a gunshot close to the player to be as loud as it would be experienced in real life or a human voice shouting from afar versus the same dialogue presented very near the listener. We therefore need to balance realism, artistic intent, and user expectations. These points were echoed in the workshop of Jelle van Mourik, Simon Gumbleton, and Nick Ward-Foxton from Sony Interactive Entertainment Europe, who discussed recording and implementing dialogue in VR applications. The addition of three (and further six) degrees of freedom introduces far more complications to audio rendering than might first be thought. One proposal was that by educating nonaudio specialists, such as script writers, we may be able to avoid some of these issues entirely, for example by not having to accommodate characters whispering to each other.

There were also several workshops dedicated to audio production and design in VR. Chanel Summers from VR Studios Inc. gave a compelling workshop on next-generation competitive e-Sports for location-based VR. She demonstrated how participants can actually forget they are in the VR space in such experiences and discussed the sound design for free-roaming arena scale VR. Jeffery Stone from Artisyns Audio and Martin Rieger from VRTONUNG both gave workshops on their own unique perspectives of the workflow, production, and design techniques for 360 videos with many intriguing examples of 360 soundfield recording for VR in challenging conditions. Magic Leap's Anastasia Devena presented what she believes to be the four domains of spatial audio: frequency, time, space, and context. Within context, she highlighted the extra complexity added by the ability of a user to move freely within a given reality and the need to adjust audio scenes accordingly.

Tim Gedemer and Francois LaFleur from Source Sound, Inc. presented three case studies of sound design for complete VR experiences: Pixar's Coco VR, Dreamscape's Alien Zoo, and CNN's VR news portal. They talked about both the good and stressful times they had on these projects and encouraged the audience to keep pushing for the prioritization of audio in VR projects.

A final panel discussion from Facebook saw Varun Nair, Jon Ojeda, and Andrew Boyd discuss the blurry lines between AR, VR, and 360 content sound design. The panel discussed the new requirements and design choices that need to be made and said that good sound design should go unnoticed, but that extra dimensions requires more sound and (yet again) more complexity.

IMMERSIVE AUDIO EDUCATION

John Merchant from Middle Tennessee State University led a workshop focused not only on how to use VR and AR to meet educational needs but also on the best ways for the industry and academia to work together to mentor and inspire the next generation of audio professionals and lift immersive audio to its highest potential. To help navigate this landscape, John was joined by Agnieszka Roginska from New York University, Tom Smurdon from Oculus VR, Jean-Marc Jot from Magic Leap, and Sally-Anne Kellaway from Microsoft. The panel discussed the many ways in which we can inspire students including conferences, papers, talks, and practical demonstrations and emphasized the importance of mentors: high-school teachers, parents, academic supervisors, and leading engineers in the field, who can all be major sources of inspiration. Agnieszka felt that the biggest motivator is involvement and finding out what aspect of immersive audio you don't want to do is just as important as what you do want to do. The panel also felt that the most significant and important skill to give a student in this field is to learn, relearn, and evolve and that students must move forward and teach themselves in order to push the boundaries of the VR and AR landscape. Moreover, students should allow themselves to think creatively and make mistakes, even if it fails. Jean-Marc said that education cannot only provide a protective space for this, but allow for students to acquire the skills on how to fail correctly through hypothesis testing, and employing the correct productive approach.

Ravish Mehra, Wednesday's keynote speaker

Winners of the best papers awards were recognized during the Wednesday morning Plenary Session.

THE FUTURE OF AUDIO FOR VR AND AR

The future of VR and AR was widely discussed at the conference, largely in terms of emerging technologies and workflow requirements. A clear goal of VR is to be perceptually indistinguishable from real life, but AR could go above and beyond improving and enhancing our daily experience. This was a strong theme in the third and final keynote of the conference by Ravish Mehra from Facebook Reality Labs. Ravish discussed the steps needed to take us from the current state of the art to "where we want to be" and talked about his work concerning the accurate modeling of sound propagation in virtual environments, a not unfamiliar topic of this conference. He commented on the limited computing power made available to most audio systems and the importance of coherency between real and virtually generated sounds. The keynote opened up a wide discussion among both content creators and engineers on whether audio in games or 360 video should be driven by an aim to be more artistic, more plausible, or more authentic and realistic.

The keynote was followed by a panel discussion moderated by Scott Selfon that looked at realistic sound propagation in VR environments. Scott was joined by Ravish as well as Nikunj Raghuvanshi from Microsoft, Lakulish Antani from Valve Sofware, and Ethan Geller from Epic Games. Despite the inevitable increase in computing power that will come over the next couple of decades, the panel addressed the very valid point that products are required to ship today. This means we must provide solutions using the computing power made available to us now. The question of design over accuracy was also raised. Is it correct to sacrifice story line if a key game character walks around a corner mid-dialog and is no longer within an auditory line of sight? Would a user be more confused by a spaceship that doesn't make an entirely unrealistic "fast moving object" sound as it flies though the silent depths of space? Once again, the artistic intent must be considered.

CONFERENCE BANQUET

The conference banquet was held at the Hollywood Schoolhouse, a beautiful 1912 venue in the heart of the Woodinville Wine Valley. Attendees were welcomed with a wine and cheese reception followed by a wonderful buffet dinner. The banquet was the perfect opportunity to continue the engaging discussion of the conference in a more informal setting and for conference chair Matt Klassen to thank the hard work of the committee and volunteers as well as the support of the sponsors. The best peer-reviewed paper was also announced at the banquet and went to Olli Rummukainen, Thomas Robotham, Sebastian J. Schlecht, Axel Plinge, Jürgen Herre, and Emanuël A. P. Habets for their paper entitled "Audio Quality Evaluation in Virtual Reality: Multiple Stimulus Ranking with Behavior Tracking."

CONCLUDING REMARKS

Virtual and Augmented Reality technologies have huge potential to go beyond mere entertainment value, touching our lives in previously unthought of ways. Audio for VR and AR is intrinsic to these experiences, and its importance in creating plausible immersive environments cannot be overstated. With current audio technologies, compelling AR and VR content can already be created but the future relies on strong collaboration between psychoacousticians, sound designers, and audio engineers to take it to a level where the experiences are indistinguishable from reality. To this end, the 2018 AES Audio for Virtual and Augmented Reality Conference was a resounding success, not only for pushing the boundaries of audio technologies for VR and AR, but also for acting as a forum where immersive audio content creators and technologists could synergize their thoughts to inspire new and innovative research in the field.

Editor's note: AES Members, view the conference papers via the AES E-library at http://www.aes.org/publications/conferences/