



Theater Room 411

Friday, September 30th

8:30 am

OPENING KEYNOTE

The Journey into Virtual and Augmented Reality
Philip Lelyveld - Virtual Reality / Augmented Reality Initiative Program Manager, USC Entertainment Technology Center

Virtual, Augmented, and Mixed Reality have the potential of delivering interactive experiences that take us to places of emotional resonance, give us agency to form our own experiential memories and become part of the everyday lives we will live in the future. Philip Lelyveld will define what Virtual, Augmented, and Mixed Reality are, present recent developments that will shape how they will potentially impact entertainment, work, learning, social interaction, and life in general, and raise rarely-mentioned but important issues that will impact how VR/AR/MR is adopted. Just as TV programming progressed from live broadcasts of staged performances to today's very complex language of multithread long-form content, so VR/AR/MR will progress from the current 'early days' of projecting existing media language with a few tweaks into a headset experience to a new VR/AR/MR-specific language that both the creatives and the audience understand. Philip's goal is to bring you up to speed on the current state, the potential, and the known barriers to adoption of Virtual, Augmented, and Mixed Reality.

9:45 am

Tutorial 1

Audio Recording and Production for Virtual Reality/360° Applications
 Chair: *Jan Plogsties* - Fraunhofer Institute
 Participants: *Chris Pike* - BBC Research & Development, *Dr. Nils Peters*, Qualcomm Technologies Inc., *Fred Umminger* - Sony Computer Entertainment

At IBC, AES 139th, CES and other events Virtual Reality has been a huge topic. VR producers more and more realize the potential and need for spatial audio processing for VR applications. This workshop will discuss the following topics:

1. How to record audio for 360° video? - Can we use the same techniques as for Movie/TV productions? Does a B-format mic do the trick?
2. How to mix audio for 360/VR? - Channels, objects or ambisonics, a combination, or binaural? Plugins and production tools, How to monitor?
3. How to deliver audio for different VR applications? - What codecs and formats are there?
4. How to render audio for 360/VR? - Headphone rendering for VR glasses - What are resolution and latency requirements?
5. What quality aspects are important? - Accuracy and plausibility, interaction with video?

11:30 am

Tutorial 2

Creating Immersive & Aesthetic Auditory Spaces for Virtual and Augmented Reality
Chanel Summers - University of Southern California / Syndicate 17 LLC

This presentation will discuss the challenges and provide specific solutions for creating audio within interactive virtual and augmented reality experiences. Audio techniques will be revealed that can be used today to advance storytelling and gameplay in virtual environments while creating a cohesive sense of place. Processes and techniques will be demonstrated for use in the creation of soundscapes in shipping products, ranging from immersive mixed reality experiences to multi-participant, multi-site, location based games.

2:00 pm

Tutorial 3

Spatial Audio and Sound Propagation for VR: New Developments, Implementations, and Integration

Chair: *Dr. Dinesh Manocha* - Univ. of North Carolina @ Chapel Hill

Participants: *Dr. Anish Chanda* - Impulsonic Inc.
Neil Wakefield - Linden Lab

In this tutorial, we give an overview of recent research and tools for immersive spatial audio and sound propagation effects for VR. We also discuss sound design and integration aspects of adding these propagation techniques and capabilities to a massive VR platform, Project Sansar from Linden Lab. Spatial audio is important to maintain audio-visual coherence in VR for increased realism and better sense of presence. It refers to 3D audio and environmental effects like sound occlusion, reflection, diffraction, and reverberation. However, quickly simulating spatial audio to update with orientation and positional changes in VR is computationally challenging. We will give an overview of research techniques that have been developed in the last 10 years for efficiently modeling spatial audio for complex VR worlds. There will also be a hands on tutorial using Phonon, that has implemented many of these state of the art spatial audio algorithms. Phonon integrates with a wide variety of game engines and audio engines and we use these applications to demonstrate the performance of these novel spatial audio algorithms. We also demonstrate how 3D audio effect can be applied to environmental effects in spatial audio. Finally, we will discuss various sound design considerations when adding spatial audio for VR as well as practical challenges and considerations when adding spatial audio into a large VR platform, especially with regards to making spatial audio tools accessible for untrained users or content creators.

3:45 pm

Tutorial 4

3D Audio Post-Production Workflows for VR

Viktor Phoenix & Scott Gerstin - Technicolor

Overview of solutions to some of the creative and practical challenges encountered in the audio post-production pipeline for 360 videos and Virtual Reality. We discuss methodologies for monitoring, editing, designing, mixing, mastering, and delivering audio for VR and 360 Videos. We discuss how to integrate a 3D audio workflow into existing post-production pipelines by merging best practices from Games, Television, and Feature Film along with new strategies for this emerging medium. The future of content delivery and playback is considered while still respecting current infrastructures for delivering client projects and accommodating the

variety of delivery formats required for various virtual reality and 360 video platforms.

4:45 pm

Tutorial 5

How Can Audiology and Hearing Science Inform AVAR, and Vice Versa?

Chair: *Dr. Chris Stecker* - Vanderbilt University School of Medicine

Participants: *Dr. Erick Gallun* - VA National Center for Rehabilitative Audiological Research, *Dr. Dan Tollin* - University of Colorado, *Dr. Ryan McCreery* - Boys Town National Research Hospital, *Dr Poppy Crum* - Dolby Laboratories

This panel discussion will feature investigators in hearing science and audiology, including experts in binaural hearing, audiological assessment and rehabilitation, next-generation hearing aids, and auditory cognitive neuroscience. Brief presentations will highlight the current and future impacts of hearing science on AVAR—e.g., the evolution of binaural hearing aids as spatially intelligent devices, lessons from auditory scene analysis, and brain-directed signal processing. Applications of AVAR technology to hearing science and the audiology clinic will also be presented, e.g. the use of immersive VR to diagnose and retrain spatial hearing deficits, and the benefits of using binaural devices to study hearing “in the wild.”

6:30 pm

Tutorial 6

VR Audio - The Convergence of Sound Professions

Christopher Hegstrom - Symmetry Audio

At the very least, VR audio is an exciting new paradigm for audio professionals to learn & at most it is a convergence of all the preceding subcategories of audio professions. Built with game engines, using film cinematography on a mobile platform with the presence of theater & live streaming inspired by broadcast, it will take all of our combined knowledge to pull off convincing VR. Audio is the glue that binds all of these sub-genres. This talk will identify what we can apply to VR audio from each of these proficiencies, what we can learn from other VR system technology (such as cameras or haptics) and how audio can inspire other professions with our standardization and collaboration.

Room 409A

Friday, September 30th

9:45 am

Paper Session 1: Sound Localization in 3D Space

1-1 Moving Virtual Source Perception in 3D Space

Sam Hughes and Dr. Gavin Kearney - University of York

This paper investigates the rendering of moving sound sources in the context of real-world loudspeaker arrays and virtual loudspeaker arrays for binaural listening in VR experiences. Near Field compensated Higher Order Ambisonics (HOA) and Vector Base Amplitude Panning (VBAP) are investigated for both spatial accuracy and tonal coloration with moving sound source trajectories. A subjective listening experiment is presented over 6, 26 and 50 channel real and virtual spherical loudspeaker configurations to investigate accuracy of spatial rendering and tonal effects. The results show the applicability of different degrees of VBAP and HOA to moving source rendering and illustrate subjective similarities and differences to real and virtual loudspeaker arrays.

1-2 Disparity in Horizontal Correspondence of Sound and Source Positioning: The Impact on Spatial Presence for Cinematic VR

Angela McArthur - BBC R&D / Queen Mary University London

This study examines the extent to which disparity in azimuth location between a sound cue and image target can be varied in cinematic virtual reality (VR) content, before presence is broken. It applies disparity consistently and inconsistently across five otherwise identical sound-image events. The investigation explores spatial presence, a sub-construct of presence, hypothesising that consistently applied disparity in horizontal audio-visual correspondence elicits higher tolerance before presence is broken, than inconsistently applied disparity. Guidance about the interactions of subjective judgements and spatial presence for sound positioning is needed for non-specialists to leverage VR's spatial sound environment. Although approximate compared to visual localization, auditory localization is paramount for VR: it is lighting condition-independent, omnidirectional, not as subject to occlusion, and creates presence.

1-3 Lateral Listener Movement on the Horizontal Plane (part 2): Sensing Motion through Binaural Simulation in a Reverberant Environment

Matthew Boerum, Bryan Martin, Richard King, and George Massenburg - McGill University

In a multi-part study, first-person horizontal movement between 2 virtual sound source locations in an auditory

virtual environment (AVE) was investigated by evaluating the sensation of motion as perceived by the listener. A binaural cross-fading technique simulated this movement while real binaural recordings of motion were made as a reference using a motion apparatus and mounted head and torso simulator (HATS). Trained listeners evaluated the sensation of motion among real and simulated conditions in 2 opposite environment-dependent experiments: Part 1 (semi-anechoic), Part 2 (reverberant). Results from Part 2 were proportional to Part 1, despite the presence of reflections. The simulation provided the greatest sensation of motion again, showing that binaural audio recordings present less sensation of motion than the simulation.

11:30 am

Paper Session 2: Real -World Case Studies Part 1

2-1 Object-based 3D audio production for virtual reality using the Audio Definition Model

Chris Pike, Richard Taylor, Tom Parnell and Dr. Frank Melchior - BBC Research & Development

This paper presents a case study of the production of a virtual reality experience with object-based 3D audio rendering using professional tools and workflows. An object-based production was created using a common digital audio workstation with real-time dynamic binaural sound rendering and visual monitoring of the scene on a head-mounted display. The Audio Definition Model is a standardized meta-data model for representing audio content including object-based, channel-based and scene-based 3D audio. Using the Audio Definition Model the object-based audio mix could be exported to a single WAV file. Plug-ins were built for a game engine in which the virtual reality application and the graphics were authored to allow import of the object-based audio mix and custom dynamic binaural rendering.

2-2 Virtually Replacing Reality: Sound Design and Implementation for Large Scale Room Scale VR Experiences

Sally Kellaway - Zero Latency VR

Audio for Virtual Reality (VR) presents a significant array of challenges and augmentations to the traditional requirements of sound designers employed within the video games industry. The change in perspective and embodiment of the player requires the employment of additional tools and consideration of object size, spacing and spatial design as a more significant part of the sound design process. The author presents her approach to these tasks from the perspective of developing audio for the large-scale Room Scale video game developer Zero

Latency. Focusing on the design considerations and processes required in this unique medium, the content of this presentation is designed to give insight in to this large-scale version of VR technology.

2:00 pm

Paper Session 3: Real -World Case Studies Part 2

3-1 Crafting Cinematic High End VR Audio for Etihad Airways

Ola Björling and Eric Thorsell - Media Monks

MediaMonks were approached by Etihad Airways via their ad agency The Barbarian Group to create a Virtual Reality experience taking place aboard their Airbus A380, the worlds largest and most luxurious non-private airplane. Challenges included capturing audio including dialogue aboard the real plane, crafting an experience that encourages repeated viewing and combining a sense of truthful realism with a sense of dream-like luxury without relying on a musical score, all in a head tracked spatialized mix. Artistic conventions around non-diegetic sound and their psychological impact in VR also required consideration.

3-2 Creating an Immersive 360°-A/V Concert Experience at the 50th Montreux Jazz Festival using Real-time Room Simulation

Sönke Pelzer, Dirk Schröder, and Fabian Knauber - AudioBorn GmbH

The Montreux Jazz Festival is the second largest jazz festival in the world. Since the beginning 50 years ago, all concerts have been recorded for the Montreux Jazz Archive, a unique treasure and the largest collection of live music, declared Unesco World Heritage. Following the vision of the deceased founder Claude Nobs, who always pushed the boundaries by applying latest recording technologies, this year's 50th anniversary of the festival introduced capturing of 3D-audio and 360° stereoscopic video. Using a virtual reality camera, ambisonics microphones, as well as multitrack audio recording with 3D post-processing, an immersive capture and reproduction was achieved. This contribution highlights challenges, experiences and solutions of the preparation, recording, post-processing and release of this immersive production.

3:15 pm

Paper Session 4: Capture, Rendering and Mixing for VR, Part 1

4-1 Efficient, Compelling and Immersive VR Audio Experience Using Scene Based Audio / Higher Order Ambisonics

Dr. Shankar Shivappa, Dr. Martin Morrell, Dr. Deep Sen, Dr. Nils Peters, and Dr. S.M. Akramus Salehin - Qualcomm Technologies Inc.

Scene-based audio (SBA) also known as Higher Order Ambisonics (HOA) combines the advantages of object-based and traditional channel-based audio schemes. It is particularly suitable for enabling a truly immersive (360,180) VR audio experience. SBA signals can be efficiently rotated and binauralized. This makes realistic VR audio practical on consumer devices. SBA also provides conductive mechanisms for acquiring live soundfields for VR. MPEG-H is a newly adopted compression standard that can efficiently compress HOA for transmission and storage purposes. It is the only known standard that provides compressed HOA end-to-end. Our paper describes a practical end-to-end chain for SBA/HOA based VR audio. Given its advantages over other formats, SBA should be 'the format of choice' for a compelling VR audio experience.

4-2 Interpolation of Multiple-Point Ambisonics Recordings for Virtual Navigation

Joseph Tylka and Dr. Edgar Choueiri - Princeton University

Using multiple Ambisonics recording arrays enables a more accurate estimation of the soundfield at an intermediate position than that achievable with a single array. Two techniques for such interpolation are examined: averaging translated plane-wave signals from each array to the listening position; and direct estimation of the soundfield at the listening position that "best explains" the signals at all recording points. Additionally, a method is presented that triangulates nearfield sources and excludes from the interpolation any arrays for which the listening position is not valid. For each technique, impulse and frequency responses are calculated at various listening positions for simple incident soundfields and localization is predicted with a binaural model. Localization and coloration are also examined through binaural-synthesis-based listening tests.

4-3 Immersive Audio Rendering for Interactive Complex Virtual Architectural Environments

Imran Muhammad and Dr. Jin Yong Jeon - Hanyang University

In this study we investigate methods for sound propagation in virtual complex architectural environments for spatialized audio rendering to use in immersive virtual reality (VR) scenarios. During the last few decades, sound propagation models have been designed and investigated for complex building structures, using geometrical approach (GA) and hybrid techniques. For sound propagation, it is required to design fast simulation tools to incorporate a sufficient number of dynamically moving sound sources, room acoustical properties, and reflections and diffraction from interactively changing surface elements in VR environments. Using physically based models, we achieved a reasonable trade-off between sound quality and system performance. Furthermore, we describe the sound rendering pipeline into a virtual scene to simulate virtual environment.

4-4 Immersive Audio for VR

Joel Susal, Kurt Krauss, Dr. Nicolas Tsingos, and Marcus Altman - Dolby Laboratories

Object based sound creation, packaging and playback of content is now prevalent in the Cinema and Home Theater, delivering immersive audio experiences. This has paved the way for Virtual Reality sound where precision of sound is necessary for complete immersion in a virtual world.

typically relies on mixing of the real acoustic environment of a listener with a virtual one. In this work, we conceptually extend this legacy ARA framework, aiming to the increment of the synthesized acoustic field spatial scale primarily in the real world domain. Such an increment actually acts as a virtual wide-angle sound focuser of wide-area acoustic fields. We demonstrate the above Augmented Reality Audio Network (ARAN) concept in terms of a live electroacoustic music concert. A subjective evaluation derived strong and secure indications that the proposed ARAN framework may represent a strong alternative potential of legacy ARA in the artistic and creative domain.

5:45 pm

Paper Session 5: Streaming Immersive Audio Content

5-1 Streaming Immersive Audio Content

Johannes Kares and Dr. Véronique Larcher - Sennheiser

“Immersion [...] is a perception of being physically present in a non-physical world.” It is critical to think about immersive audio for live music streaming because giving listeners the illusion of being transported to a different acoustic environment makes the experience of streaming much more real. In this paper, we are describing various approaches to enable audio engineers to create immersive audio content for live streaming, whether using existing tools and network infrastructure and delivering static binaural audio, or getting ready for emerging tools and workflows for Virtual Reality streaming.

5-2 An Augmented Reality Audio Live Network for Live Electroacoustic Music Concerts

Nikos Moustakas, and Dr. Andreas Floros - Ionian University

Dr. Bill Kapralos - University of Ontario Institute of Technology

Augmented reality audio (ARA) represents a well-established and widely investigated concept that

Theater Room 411

Saturday, October 1st

8:30 am

10:30 am

Workshop 1

End-to-End VR Audio Solution for Fully Immersive/Interactive Experience

Presenter: *Dr. Henney Oh* - GAUDIO Lab, Inc.

GAUDIO LAB VR develops audio technologies, one of which was adopted as the MPEG-H 3D Audio international standard. Based on these, we provide end-to-end solution for whole VR ecosystem as producer intended. In this workshop, audience will be guided the basics of VR audio solution when creating, delivering, and playing back 360/VR contents. All levels including beginner could enjoy the workshop.

This workshop will cover:

- Immersive and interactive binaural rendering: a key technology for real immersive VR audio
- Object-based audio vs. scene-based audio
- VR audio distribution: app-based or codec-based
- Considerations on VR audio creation workflow
- VR sound recording: mono, Ambisonics and binaural
- Mixing and mastering with GWorks, AAX plugin for Pro Tools HD
- Quality matters. But how to guarantee the same sound quality at the end-user stage?
- GPlayer, a reference quality 360/VR audio player (API)

9:30 am

Workshop 2

Object Based Audio Mixing for AR/VR Applications

Jurgen Scharpf - Dolby Laboratories

Dolby has been a pioneer in developing world's leading object-based audio technologies and mixing tools for filmmakers and sound engineers around the world. In the last two years, we have also been working closely with a number of VR pioneers in the content community to develop the tools and playback technologies for enabling high quality linear VR experiences. This workshop will cover the unique advantages of using object-based audio mixing for cinematic and experiential VR experiences. The AES audience can walk away with an understanding the power and flexibility of object-based audio mixing for creating more precise and convincing sound to match the visual — giving viewers a strong sense of presence.

Workshop 3

Facebook 360 Spatial Workstation: Tools, Workflows and Best Practices

Joel Douek and Benedict Green - Ecco,
Varun Nair - 2 Big Ears

This workshop will guide participants in the use of the FB360 Spatial Workstation toolset for 360 spatialized audio work in VR, from both the technology and content creation perspectives. It will cover: Session configuration for VST (Reaper/Nuendo) and AAX (Pro Tools HD). A deep exploration of the tools and feature set for a variety of project types, highlighting examples and case studies. Step-by-step workflows to stay organized and achieve best-in-class VR audio outcomes. Maintaining differentiated spatialized vs non-spatialized experiences across delivery formats. Encoding, levels and ingestion considerations across different platforms (mobile iOS/Android, Oculus Rift etc). The role and future developments of FB360 Spatial Workstation for audio in VR and AR.

11:30 am

Workshop 4

Immersive Sound Capture for Cinematic Virtual Reality

Chair: *Sofia Brazzola* - Sennheisen

Participants: *Henrik Oppermann* - Visualise, *Jean-Pascal Beaudoin* - Headspace Studio (Felix & Paul Studios), *Benedict Green* - ECCO VR, *Mikkel Nymand* - Sennheiser

The workshop will explore the requirements and best practices of on-location sound capture for Cinematic Virtual Reality. The panel will sketch out the differences for the on-site sound engineer when working on a Cinematic VR shoot compared to a traditional cinema shoot. In this context, it will examine the benefits and draw backs of different spatial audio capture solutions – such as binaural and Ambisonics – and look at their best practices. As well, the panel will look at when and how to capture non-spatial sources on set and discuss the still unresolved pain points facing location engineers today.

2:00 pm

Workshop 5

Real-Time Production Chain for immersive 3D Audio
Sönke Pelzer and Dr. Dirk Schröder - Audioborn

Using the novel 3D audio engine, Auratorium, this tutorial discusses a new approach to design compelling audio environments for virtual reality applications such as cinematic VR, 360 degree videos and immersive sound productions. A light-footed production chain for the object-based creation of 3D audio content is presented including a fully fledged room acoustic simulation and reproduction, targeting binaural headphone reproduction as well as arbitrary loudspeaker layouts. Participants experience how to setup an easy workflow within their existing design tools (e.g. Unity and Pro Tools) and create realistic 3D audio by placing sound sources, receivers and adjusting material properties, room shape and directivities while listening to real-time rendered 3D audio.

2:45 pm

Workshop 6

Positioning Sounds in VR Post Production
Dr. Nuno Fonseca - Sound Particles

Although many VR projects only use the sound of the original venue, other projects require additional audio content, especially on high budget productions, incorporating sound design content and other sound material. The problem with adding additional audio content to VR shots, is the need to positioning sound in a very accurate way, which can be a nightmare when handling moving sound sources. This workshop will present an easy and very accurate way of positioning sounds on 360° videos, by dragging sounds on top of the image and track them around with key frame animation, a well-known CGI technique. Using a 3D CGI-like software for audio applications, with a virtual microphone approach, the same video can then be exported to different output formats, without the need to redo all sound positioning.

3:45 pm

Workshop 7

OZO Audio Workflow
Hannu Pulakka - Nokia

Discussion of how audio production for live VR events compares to a "traditional" live production. We'll review the signal chain, from capture through to distribution & payout, discuss the unique challenges live VR presents,

and pose best practices for producing high quality immersive audio experiences that teleport the listener to the stadium or concert hall.

4:30 pm

Workshop 8

Using MEMS Microphones for Ambisonic Audio in a Live Streaming Spherical Video Camera
Lucas McCauley - Videostitch

A few months ago, VideoStitch announced the Orah brand and its Orah 4i, a live streaming spherical camera, complete with immersive audio. The 4i uses four lenses and realtime stitching algorithms to produce 4K spherical video, and four on-camera MEMS microphones to capture first order ambisonics. The output can be streamed directly to any platform that accepts live feeds of immersive content. This workshop explores the design of the Orah 4i camera's integrated ambisonic audio capture, from selection and placement of the microphones, capsule correction EQ, level matching, omnidirectional to cardioid polar pattern processing, and conversion to first order Bformat. The final design decisions and several measurements are shown.

5:30 pm

Workshop 9

A Lightweight & Versatile 3D-Audio Codec for Storing & Transmitting Immersive Audio and its Application to Live VR 360 Events
François Becker, Clément Carron, Longcat Audio Technologies and Benjamin Bernard - Medialab Consulting SNP and Longcat Audio Technologies

Many producers are currently adopting 360 content creation. For image production, multiple hardware and software solutions are now available off the shelf. However, although audio is a key point to immersion, this is not paralleled on the audio side. An adequate, versatile codec for 3D audio is introduced in this workshop that adapts seamlessly to the needs of the VR industry, from storage to production, distribution/delivery, and rendering. Besides VR, the presented technology suite also finds uses for 2D movies, music, telepresence & teleconferencing. Other related topics will be addressed: hybrid sound capture for live events, production formats, and integrated rendering techniques.

6:30 pm

CLOSING KEYNOTE**Future Nostalgia, Here and Now: Let's look back on today from 20 years hence.***George Sanger - Magic Leap*

Two decades of progress can change how we live and think in ways that boggle the mind. Audio is a small piece of that, but it's our piece. 20 years ago, the PC got rudimentary sound cards; now the entire "multitrack recording studio" lives on our computers. Some of us saw that development as inevitable, but in 1996 those smart people sounded fairly edgy, to say the least. And of those smart people, who among them saw that texting would be the "killer app" for smart phones, in many ways trumping audio communication? Let's take our accumulated wisdom from the past 20 years of growth and non-growth of audio and computing, and see if we can't get some feel for what it will be like in this room 20 years from now, looking back. With luck, we will be nodding our heads sagely, saying, "Yep, we saw that one coming way back in 2016!"

Room 409A

Saturday, October 1st

8:30 am

Workshop 10

Audio content creation for VR with standard DAWs

Tom Ammermann - New Audio Technology

More than 50% of the emotional impact of a movie is audio. So creation of audio content becomes a new important topic in VR applications. But using common digital audio workstations (DAWs) for that is currently difficult. Strategies to connect DAWs with VR video devices, game/VR engines and how to use the new space and headphone virtualizations are questions the presentation will answer and show opportunities and tools.

9:30 am

Paper Session 6:

Perceptual Consideration for VR / AR

6-1 Spatial Auditory Feedback in Response to Tracked Eye Position

Dr. Durand Begault - NASA ARC

Fixation of eye gaze toward one or more specific positions or regions of visual space is a desirable feature within several types of high-stress human interfaces, including vehicular operation, flight deck control, target acquisition, etc. It is therefore desirable to have a means to give spatial auditory feedback to a human in such a system about whether or not the gaze is specifically directed towards a desired position. Alternatively, it is desirable to use eye position as a means of controlling a device that provides auditory feedback so that there is a correspondence between eye position and control voltages that manipulate aspects of an auditory cue that includes spatial position, pitch and/or timbre.

6-2 Perceptual Weighting of Binaural Information: Toward an Auditory Perceptual “Spatial Codec” for Auditory Augmented Reality

Dr. Chris Stecker and Anna Diedesch - Vanderbilt University School of Medicine

Auditory augmented reality (AR) requires accurate estimation of spatial information conveyed in the natural scene, coupled with accurate spatial synthesis of virtual sounds to be integrated within it. Solutions to both problems should consider the capabilities and limitations of the human binaural system, in order to maximize relevant over distracting acoustic information and enhance perceptual integration across AR layers. Recent studies have measured how human listeners integrate spatial information across multiple conflicting cues,

revealing patterns of “perceptual weighting” that sample the auditory scene in a robust but spectrotemporally sparse manner. Such patterns can be exploited for binaural analysis and synthesis, much as time-frequency masking patterns are exploited by perceptual audio codecs, to improve efficiency and enhance perceptual integration.

6-3 DeepEarNet: Individualizing Spatial Audio with Photography, Ear Shape Modeling, and Neural Networks

Shoken Kaneko, Tsukasa Suenaga, Satoshi Sekine - Yamaha Corporation

Individualizing spatial audio is of crucial importance for high-quality virtual and augmented reality audio. In this paper, we propose a method for individualizing spatial audio, by combining the recently proposed ear shape modeling technique with computer vision and machine learning. We use a convolutional neural network to obtain estimates of the ear shape model parameters from stereo photographs of the user ear. The individualized ear shape and its associated individualized head-related transfer function (HRTF) can be calculated from the obtained parameters, based on the ear shape model and numerical acoustic simulations. Preliminary experiments, evaluating the shapes of the estimated individual ears, proved the effect of individualization.

6-4 Adjustment of the Direct-to-Reverberant-Energy-Ratio to Reach Externalization within a Binaural Synthesis System

Dr. Thomas Sporer - Fraunhofer Institute for Digital Media Technology, Stephan Werner and Florian Klein - Technische Universität Ilmenau, Electronic Media Technology Group

The contribution presents a study which investigates the perception of spatial audio reproduced by a binaural synthesis system. The quality features externalization and room congruence are measured within a listening test. Former studies imply that especially externalization is decreased if acoustic divergence between the synthesized and listening room exists. Other studies show that the adjustment of the Direct-to-Reverberant-Energy-Ratio (DRR) can increase the perceived congruence between synthesized and listening room. Within this experiment test persons are able to adjust the DRR of the synthesis until perceptual congruence between the synthesis and the internal reference concerning the listening room occurs. The ratings show that the test persons are able to adjust DRR of the listening room and therefore externalization increases.

2:00 pm

Paper Session 7: Music for VR / AR Projects

7-1 Spatial Music, Virtual Reality and 360 Media

Dr. Enda Bates and Dr. Francis Boland - Trinity College Dublin

The following paper documents the composition, recording and post-production of a number of works of instrumental spatial music for a 360 video and audio presentation. The filming and recording of an orchestral work of spatial music is described with particular reference to the various ambisonic microphones used in the recordings, post production techniques, and the delivery of 360 video with matching 360 audio. The recording and production of a second performance of a newly composed spatial work for an acoustic quartet is also presented and the relationship between spatial music and 360 content is discussed. Finally, an exploration of the creative possibilities of VR in terms of soundscape and acousmatic composition is presented.

7-2 Positioning of Musical Foreground Parts in Surrounding Sound Stages

Christoph Hold, Lukas Nagel, Dr. Hagen Wierstorf and Dr. Alexander Raake - Technische Universität Berlin

Object-based audio offers several new possibilities during the sound mixing process. While stereophonic mixing techniques are highly developed, not all of them generate promising results in an object-based audio environment. An outstanding feature is the new approach of positioning sound objects in the musical sound scene, providing the opportunity of stable localization throughout the whole listening area. Previous studies have shown that even if object-based audio reproduction systems can enhance the playback situation, the critical and guiding attributes of the mix are still uncertain. This study investigates the impact on listening preference evoked by different spatial arrangements of sound objects, with a strong emphasis on the high-attention foreground objects of the presented music track.

7-3 The Soundfield as Sound Object: Virtual Reality Environments as a Three-dimensional Canvas for Music Composition

Dr. Richard Graham and Dr. Seth Cluett - Stevens Institute of Technology

Our paper presents ideas raised by recent projects exploring the embellishment, augmentation, and extension of environmental cues, spatial mapping, and immersive potential of scalable multi-channel audio systems for virtual and augmented reality. Moving beyond issues of reproductive veracity raised by merely recreating the soundscape of the physical world, these works exploit characteristics of the natural world to

accomplish creative goals that include the development of models for interactive composition, composing with physical and abstract spatial gestures, and linking sound and image. We are presenting a novel system that allows the user to treat the soundfield as a fundamental building block for spatial music composition and sound design.

4:00 pm

Paper Session 8: Capture, Rendering and Mixing for VR Part 2

8-1 XY-Stereo Capture and Up-conversion for Virtual Reality

Dr. Nicolas Tsingos, Pradeep Govindaraju, Cong Zhou and Abhay Nadkarni - Dolby Laboratories

We propose a perceptually-based approach to creating immersive soundscapes for VR applications. We leverage stereophonic content obtained from XY microphones as a basic building block that can be easily recorded, edited and combined to provide a more compelling experience than can be obtained from recording at a single location. Central to our approach is a novel up-conversion algorithm that derives a nearly full-spherical parametric soundfield, including height information, from an XY recording. This approach enables a simpler, improved capture, when compared to alternative soundfield recording techniques. It can also take advantage of new object-based delivery formats for flexible delivery and playback.

8-2 Augmented Reality Headphone Environment Rendering

Dr. Jean-Marc Jot and Dr. Keun Sup Lee - DTS, Inc.

In headphone-based augmented reality audio applications, computer-generated audio-visual objects are rendered over headphones or earbuds and blended into a natural audio environment. This requires binaural artificial reverberation processing to match local environment acoustics, so that synthetic audio objects are not distinguishable from sounds occurring naturally or reproduced over loudspeakers. Solutions involving the measurement or calculation of binaural room impulse responses in a consumer environment are limited by practical obstacles and complexity. We propose an approach exploiting a statistical reverberation model, enabling practical acoustical environment characterization and computationally efficient reflection and reverberation rendering for multiple virtual sound sources. The method applies equally to headphone-based “audio-augmented reality” – enabling natural-sounding, externalized virtual 3-D audio reproduction of music, movie or game soundtracks.

8-3 Capturing and Rendering 360° VR Audio Using Cardioid Microphones

Dr. Hyunkook Lee - University of Huddersfield

Recording and listening experiments were carried out to evaluate the horizontal localisation performances of quadraphonic near-coincident and coincident (B-format) microphone configurations in both loudspeaker and binaural reproductions with simulated head rotations. The design philosophy for the near-coincident arrays was the 'equal segment microphone array (ESMA)' concept by Williams. Three microphone spacings of 50cm, 30cm and 24cm, which were determined based on three different stereophonic localisation models, were compared. Results show that the localisation performances of the near-coincident configurations were considerably better than that of the coincident one overall. The 50cm spacing achieved the intended stereophonic recording angle of 90° more accurately and consistently than the 30cm and 24cm. However, the differences among these spacings in response distribution were relatively small.