

AES International Conference on Audio for Virtual and Augmented Reality 2018

Science, Technology, Design, and
Implementation

Redmond, Washington, USA
20 - 22 August 2018

ISBN: 978-1-5108-7039-0

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2018) by the Audio Engineering Society
All rights reserved.

Printed by Curran Associates, Inc. (2018)

For permission requests, please contact the Audio Engineering Society
at the address below.

Audio Engineering Society
International Headquarters
551 Fifth Ave., Suite 1225
New York, NY 10176
USA

Phone: +1 212 661 8528

www.aes.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Paper Session 1

Production & Tools

Room: Michelangelo

Session Chair: Edgar Choueiri

11:30 - 11:55 Monday, August 20

AUTHORS: **Michael Costagliola**, *Yale University*

TITLE: **Multi-user shared augmented audio spaces using motion capture systems**

1

This paper describes a method for creating multi-user shared augmented reality audio spaces. By using a system of infrared cameras and motion capture software, it is possible to provide accurate low-latency head tracking for many users simultaneously, and stream binaural audio representing a realistic, shared virtual environment to each user. Participants can thus occupy and navigate a shared virtual aural space without the use of head-mounted displays, only headphones (with passive markers affixed) connected to lightweight in-ear monitor belt packs. Potential applications include installation work, classroom use, and museum audio tours.

12:00 - 12:25 Monday, August 20

AUTHORS: **Giordano Jacuzzi**, *Sennheiser Electronic GmbH*

TITLE: **Augmented Audio: An Overview of the Unique Tools and Features Required for Creating AR Audio Experiences**

6

What a user sees in augmented reality is only part of the experience. To create a truly compelling journey, we must augment what a user hears in reality as well. In this paper, we consider Augmented Audio to be the sound of AR, and discuss a technique by which the binaural rendering of virtual sounds (Augmented) is combined with the manipulation of the real-world sound surrounding a listener (Reality). We outline the unique challenges that arise when designing audio experiences for AR, and document the current state-of-the-art for Augmented Audio solutions. Using the Sennheiser AMBEO Smart Headset as a case study, we describe the essential features of an Augmented Audio device and its integration with an AR application.

12:30 - 12:55 Monday, August 20

AUTHORS: **Jukka Holm and Mark Malyshev**, *Tampere University of Applied Sciences, Finland*

TITLE: **Spatial Audio Production for 360-Degree Live Music Videos: Multi-Camera Case Studies**

14

This paper discusses the different aspects of mixing for 360-degree multi-camera live music videos. We describe our two spatial audio production workflows, which were developed and fine-tuned through a series of case studies including rock, pop, and orchestral music. The different genres were chosen to test if the production tools and techniques were equally efficient for mixing different types of music. In our workflows, one of the most important parts of the mixing process is to match the Ambisonics mix with a stereo reference. Among other things, the process includes automation, proximity effects, creating a sense a space, and managing transitions between cameras.

Paper Session 2

Perception & Evaluation

Room: Michelangelo

Session Chair: Rahulram Sridhar

2:00 - 2:25 Monday, August 20

AUTHORS: **Angela McArthur, Mark Sandler and Rebecca Stewart**, *Queen Mary University of London, UK*

TITLE: **Perception of mismatched auditory distance - cinematic VR** 24

This study examines auditory distance discrimination in cinematic virtual reality. Using controlled stimuli with audiovisual distance variations, it determines if mismatched stimuli are detected. It asks if visual conditions - either equally or unequally distanced from the user, and environmental conditions - either a reverberant space as opposed to a free field, impact accuracy in discrimination between congruent and incongruent aural and visual cues. A Repertory Grid Technique-derived design, whereby participant-specific constructs are translated into numerical ratings, is used. Discrimination of auditory event mismatch was improved for stimuli with varied visual-event distances, though not for equidistant visual events. This may demonstrate that visual cues alert users to matches and mismatches, but can lead responses toward both greater and lesser accuracy.

2:30 - 2:55 Monday, August 20

AUTHORS: **Hanne Stenzel, Philip J.B. Jackson**, *University of Surrey, UK*, **Jon Francombe**, *BBC Research and Development, UK*

TITLE: **Reaction times of spatially coherent and incoherent signals in a word recognition test** 34

Using conventional sound design, the audio signal in VR applications is often reduced to a static stereophonic signal that is accompanied by a visual signal that allows for interactive behavior such as looking around. In the current test, the influence of spatial offset between the audio and visual signals is investigated using reaction time measurements in a word recognition task. The audio-visual offset is introduced by a video presented at horizontal offset angles between ± 20 degrees, accompanied with a static central audio. Measurements are compared to reaction times from a test where both audio and visual signal are presented with the same offset. Results show that the spatial offset introduces changes in the reaction times exhibiting greater within-participant differences.

3:00 - 3:25 Monday, August 20

AUTHORS: **G. Christopher Stecker**, *Vanderbilt University*

TITLE: **Toward objective measures of auditory co-immersion in virtual and augmented reality** 44

“Co-immersion” refers to the perception of real or virtual objects as contained within or belonging to a shared multisensory scene. Environmental features such as lighting and reverberation contribute to the experience of co-immersion even when awareness of those features is not explicit. Objective measures of co-immersion are needed to validate user experience and accessibility in augmented-reality applications, particularly those that aim for “face-to-face” quality. Here, we describe an approach that combines psychophysical measurement with virtual-reality games to assess users’ sensitivity to room-acoustic differences across concurrent talkers in a simulated complex scene. Eliminating the need for explicit judgments, Odd-one-out tasks allow psychophysical thresholds to be measured and compared directly across devices, algorithms, and user populations. Supported by NIH-R41-DC16578.

Paper Session 2 (continued)

Perception & Evaluation

Room: Michelangelo

Session Chair: Rahulram Sridhar

3:30 - 3:55 Monday, August 20

AUTHORS: Olli Rummukainen, Thomas Robotham, Sebastian J. Schlecht, Axel Plinge, Jürgen Herre and Emanuel A. P. Habets, *Fraunhofer, Germany*

TITLE: Audio Quality Evaluation in Virtual Reality: Multiple Stimulus Ranking with Behavior Tracking 50

Virtual reality systems with multimodal stimulation and up to six degrees-of-freedom movement pose novel challenges to audio quality evaluation. This paper adapts classic multiple stimulus test methodology to virtual reality and adds behavioral tracking functionality. The method is based on ranking by elimination while exploring an audiovisual virtual reality. The proposed evaluation method allows immersion in multimodal virtual scenes while enabling comparative evaluation of multiple binaural renderers. A pilot study is conducted to evaluate feasibility of the proposed method and to identify challenges in virtual reality audio quality evaluation. Finally, the results are compared to a non-immersive off-line evaluation method.

4:00 - 4:25 Monday, August 20

AUTHORS: Gregory Reardon, Andrea Genovese, Gabriel Zalles, and Agnieszka Roginska, *New York University*, Patrick Flanagan, *THX Ltd.*

TITLE: Evaluation of Binaural Renderers: Multidimensional Sound Quality Assessment 60

A multi-phase subjective experiment evaluating six commercially available binaural audio renderers was carried out. This paper presents the methodology, evaluation criteria, and main findings of the tests which assessed perceived sound quality of the renderers. Subjects appraised a number of specific sound quality attributes - timbral balance, clarity, naturalness, spaciousness, and dialogue intelligibility - and ranked, in terms of preference, the renderers for a set of music and movie stimuli presented over headphones. Results indicated that differences between the perceived quality and preference for a renderer are discernible. Binaural renderer performance was also found to be highly content-dependent, with significant interactions between renderers and individual stimuli being found, making it difficult to determine an "optimal" renderer for all settings.

Paper Session 3

Posters

Room: Edison

AUTHORS: **Otto Puomio and Tapio Lokki**, *Aalto University, Finland*

TITLE: **ICHO: Immersive Concert for Homes** 74

Concert hall experience at home has been limited to stereo and 5.1 surround sound reproduction. However, these reproduction systems do not convey the spatial properties of the concert hall acoustics in detail, and specifically for headphones the sound tends to be perceived as playing inside the head. The ICHO project introduced in this paper aims to bring an immersive concert hall experience to home listeners. This is realized by using close pick-up of sound sources, spatial room impulse responses, and individualized head related transfer functions; all combined together for spatial sound reproduction with head-tracked headphones. This paper outlines how this goal is going to be achieved and how the quality of the reproduction might be evaluated.

AUTHORS: **Sebastian Nagel, Tobias Kabzinski, Stefan Kühn, Christiane Antweiler and Peter Jax**, *Institute of Communication Systems (IKS), RWTH Aachen University, Germany*

TITLE: **Acoustic Head-Tracking for Acquisition of Head-Related Transfer Functions with Unconstrained Subject Movement** 83

Recently proposed Head-Related Transfer Function acquisition methods include head-tracking to allow unconstrained subject movements during the measurement. This enables fast measurements with less equipment than previous fast measurement approaches. In this paper, we propose a novel acoustic head-tracking concept, which is particularly suited for this application. Only a head-mounted microphone array and additional recording channels are required in addition to a regular measurement setup. Unlike other head-tracking systems, the tracking data is inherently synchronized to the acoustic measurements and the angle of sound incidence can be accurately determined without knowledge of loudspeaker positions. The concept is developed and evaluated in simulations and measurements, which show that the proposed acoustic head tracker outperforms a comparison device.

AUTHORS: **Guangzheng Yu and Bosun Xie**, *South China University of Technology*

TITLE: **Multiple sound sources solution for near-field head-related transfer function measurements** 93

Near-field head-related transfer functions (HRTFs) are essential to scientific researches of binaural hearing and practical applications of virtual auditory display. In contrast to far-field HRTFs measurement, the near-field measurement is more difficult, because high efficiency, accuracy and repeatability are required in a near-field HRTF measurement. When multiple sound sources are adopted to accelerate the near-field HRTF measurement, multiple scattering among sound sources could be the major factor to influence the accuracy. In present work, the well-designed multiple sound sources solution is used to accelerate the near-field HRTFs measurement. Results show that the error caused by the multiple scattering among sound source can be controlled within the acceptable values through reasonable design and arrangement for sound sources.

Paper Session 3 (continued)

Posters

Room: Edison

AUTHORS: **Yun-Han Wu, Scott Murakami and Agnieszka Roginska**, *New York Univeristy*

TITLE: **Comparison of Measured and Simulated Room Impulse Response for an Interactive Music Making Environment in Mixed Reality** 98

This paper describes a musical environment created for mixed reality (MR) where virtual sound objects are superimposed on the real environment with the goal to blend seamlessly with the real environment using acoustic simulation techniques. The resulting experience allows a participant to interact with and create music by playing the virtual objects. An informal subjective study is performed where subjects are asked to evaluate the acoustics in the scene, with the different techniques applied, based on preference and the evaluation of the interaction design with the musical objects.

AUTHORS: **Marco Binelli, Daniel Pinardi and Angelo Farina**, *University of Parma, Italy*, **Tiziano Nili**, *ASK Industries, Italy*

TITLE: **Individualized HRTF for playing VR videos with Ambisonics spatial audio on HMDs** 104

Current systems for rendering 360-degrees videos with spatial audio on HMDs rely on a binaural approach combined with Ambisonics technology. These head-tracking systems employ generic HRTFs typically measured with a dummy head in an anechoic room. In this paper, we describe a new solution that has been developed to play 360-degrees video files with spatial audio for desktop and portable platforms, based on existing open source software. The HRTF set can be loaded from a standard WAV file chosen in an existing database or from an ad-hoc measurement or simulation. The capability to switch among multiple HRTF sets while playing has been added.

AUTHORS: **Alan Kan**, *University of Wisconsin-Madison*

TITLE: **On high-frequency interaural time difference sensitivity in complex auditory environments** 114

The Duplex theory of sound localization is a useful principle for guiding trade-offs between realistic production vs implementation complexity for virtual auditory environments. However, there are exceptions to this theory in the psychoacoustic literature that should be noted. One exception is that sensitivity to an interaural time difference (ITD) can be improved when either low-frequency amplitude modulation (AM) or frequency modulation (FM) is introduced into a high frequency tone. This paper presents results from a psychoacoustic experiment that show that having both AM and FM greatly improves sensitivity to high-frequency ITDs compared to AM or FM alone when presented in incoherent broadband noise. The implications of these findings to the generation of virtual auditory environments are discussed.

Paper Session 3 (continued)

Posters

Room: Edison

AUTHORS: **Nail Gumerov, Dmitry Zotkin, Adam O'Donovan and Ramani Duraiswami**, *VisiSonics Corporation*

TITLE: **Spatial Acoustic Field Simulation as a Service** N/A

Many practical applications in VR and AR require the computation of scattered acoustical fields from complex shaped objects. Such computations arise, e.g., in computing head related transfer functions and in designing microphone arrays. We present a parallelized fast-multipole accelerated boundary-element solver which can be used to efficiently compute the solution to such problems in the cloud. This builds on our extensive previous work on developing such methods. Details of the methods, and results from problems relevant to audio in AR and VR will be presented.

AUTHORS: **Hanne Stenzel and Philip J.B. Jackson**, *University of Surrey, UK*

TITLE: **Perceptual thresholds of audio-visual spatial coherence for a variety of audio-visual objects** 120

Audio-visual spatial perception relies on the integration of both auditory and visual spatial information. Depending on auditory and visual features of the stimulus, and the relevance of each sound to the listener, offsets between both signals are more or less acceptable. The current paper investigates to which extent each of these factors influences how critical the perception of spatial coherence is by estimating the psychometric function for seventeen audio-visual stimuli. The results show that the maximum accepted offset angle does not depend on semantic categories but is linked to audio feature classes with harmonic sounds leading to greater acceptable offsets. A regression shows that the perceptual spectral centroid is negatively correlated with the offset angle and the slope of the psychometric spatial-coherence function. This finding, however, is not conclusive and further research is necessary to define all parameters that influence bimodal localization of realistic stimuli.

AUTHORS: **Rishabh Gupta, Rishabh Ranjan, and Woon-Seng Gan**, *Nanyang Technological University, Singapore*, **Jianjun He**, *Maxim Integrated Product Inc*

TITLE: **Investigation of the Effect of VR/AR Headgears on Head-Related Transfer Functions for Natural Listening** 130

With the advent of Virtual/Augmented/Mixed Reality (VR/AR/MR) applications and accompanied head mounted displays (HMD), these devices are becoming increasingly common in daily life use producing immersive virtual and augmented audio-visual content. However, presence of these HMDs can affect the acoustic propagation of the sound waves from the sound sources to the human ears, resulting in changes in head-related transfer functions (HRTFs). In this paper, we conducted the measurements of HRTFs with and without these headgears and investigated the effect of these headgears on HRTFs at different source directions using descriptive analysis and objective metrics. Furthermore, subjective listening tests were conducted to study the perceptual significance of these differences in terms of timbre and spatial performance.

Paper Session 3 (continued)

Posters

Room: Edison

AUTHORS: **Raimundo Gonzalez and Tapio Lokki**, *Aalto University*, **Joshua Pearce**, *Michigan Technological University*

TITLE: **Modular Design for Spherical Microphone Arrays** 140

Spherical microphone arrays are commonly utilized for recording, analyzing and reproducing sound-fields. In the context of higher-order Ambisonics, the spatial resolution depends on the number and distribution of sensors over the surface of a sphere. Commercially available arrays have set configurations that cannot be changed, which limits their usability for experimental and educational spatial audio applications. Therefore, an open-source modular design using MEMS microphones and 3D printing is proposed for selectively capturing frequency-dependent spatial components of sound-fields. Following a modular paradigm, the presented device is low cost and decomposes the array into smaller units (a matrix, connectors and microphones), which can be easily rearranged to capture up to third-order spherical harmonic signals with various physical configurations.

Workshops 1 & 2

Audio Production & Design for VR/AR I

Room: Plato

11:30 - 12:25 Monday, August 20

Chanel Summers, *VRstudios, Inc.*

TITLE: Compete to Win! Creating the Audio for VRcade PowerPlayTM: A Next-Gen Competitive eSport for Location-Based VR N/A

This presentation will discuss specific immersive and aesthetic auditory design techniques and strategies as applied to create VRcade PowerPlayTM, a full-motion, multi-participant virtual reality competitive eSport game experience for live commercial Location Based Entertainment (LBE) installations that allows the participants to actually forget that they are in VR. Challenges and solutions will be revealed for designing the dynamic audioscape for an intensely athletic team eSport played in free-roaming, arena-scale VR.

12:35 - 1:30 Monday, August 20

Jeffery Stone, *Artisyns Audio*

TITLE: Audio Production in a New Reality N/A

The demand for VR/AR content, in the near future, is potentially explosive. We, the creators of 50% of the experience, need to be ready with a solid understanding of what is possible with 3D sound design, as well as its unique limitations, and how to produce it effectively. With this workshop, presented from a sound designer's perspective, I will discuss and demonstrate some tools and platforms I have used, and their potential workflows, for production, post-production, and distribution of 3D cinematic audio for 360 video, as well as production and design techniques for this unfamiliar aesthetic..

Workshops 3 & 4

Audio Production & Design for VR/AR II

Room: Plato

2:30 - 3:25 Monday, August 20

Robert Rice, *Microsoft*, **Kedar Shashidhar**, *Magic Leap*

TITLE: The Role of Audio and Multimodal Integration for New Realities N/A

An effective VR/AR/MR experience is dependent on content and technology to work together in harmony to suspend disbelief for the participant. These new realities demand an understanding of the connection between audio design and technologies that vary in immersiveness and deliver to variable end-user environments. On this panel, experts of interactive audio will discuss the role that immersive audio plays when interacting with multiple modalities of perception in order to create top-notch virtual and augmented reality experiences.

MODERATOR: Kedar Shashidhar

PANELISTS: Tom Smurdon, Chanel Summers, Viktor Phoenix, Guy Whitmore, and Robert Rice

3:35 - 4:30 Monday, August 20

Martin Rieger, *VRTONUNG, Germany*

TITLE: Sound for challenging 360° productions N/A

The talk shows various examples of 360° productions under challenging conditions for sound on location and in post-production. Working on virtual reality films requires a whole new understanding for the entire sound department. Not only that, 360° videos tend to be more adventurous than the average film production since they start feeling like pioneers again and want to shot footage that simply hasn't been there before. This is not only crucial for the camera but also for the sound department. Sometimes even more. How to approach onboard recordings on moving motorcycle with dialogue to ski-stunts in snowy mountains -Martin describes 360° recordings on set and fixing them in the post.

Workshop 5

Immersive Audio & Education

Room: Plato

5:00 - 6:30 Monday, August 20

John Merchant, *Middle Tennessee State University*

TITLE: **Immersive Audio + Education** N/A

This will be a panel discussion of topics regarding higher education for immersive audio, with consideration of skills and knowledge required for long-term student successes as well as ways to balance the needs of industry with calls for academic research. The selected panelists represent a cross-section of industry professionals and academics to assess essential concepts and teaching methodologies. The panel will include significant time for questions and answers.

Panelists:

Agnieszka Roginska, Assoc. Director of Music Technology at New York University

Tom Smurdon, Audio Content Lead at Oculus VR

Jean-Marc Jot, Distinguished Fellow at Magic Leap

Sally-anne Kellaway, Senior Audio Designer at Microsoft

(Moderator) John Merchant, Assoc. Professor and Director of the Immersive Story Lab, Middle Tennessee State University

Paper Session 4

HRTF Personalization

Room: Michelangelo

Session Chair: Matt Klassen

11:30 - 11:55 Tuesday, August 21

AUTHORS: David Poirier-Quinot and Brian Katz, *Sorbonne Université, CNRS, Institut Jean Le Rond d'Alembert, France*

TITLE: Impact of HRTF individualization on player performance in a VR shooter game II 147

We present the extended results of a previous experiment to assess the impact of individualized binaural rendering on player performance in the context of a VR “shooter game”. Participants played a game in which they were faced with successive enemy targets approaching from random directions on a sphere. Audio-visual cues allowed for target localization. Participants were equipped with an Oculus CV1-HMD, headphones, and two Oculus Touch hand tracked devices as targeting mechanisms. Participants performed six sessions alternatively using their best and worst-match HRTFs from a “perceptually orthogonal” optimized set of 7 HRTFs [Katz2012]. Results suggest that the impact of the HRTF on participant performance (speed and movement efficiency) depends both on participant sensitivity and HRTF presentation order.

12:00 - 12:25 Tuesday, August 21

AUTHORS: Rishi Shukla, Rebecca Stewart, Mark Sandler, *Queen Mary University of London, UK*, Agnieszka Roginska, *New York Univeristy*

TITLE: User selection of optimal HRTF sets via holistic comparative evaluation 155

If well-matched to a given listener, head-related transfer functions (HRTFs) that have not been individually measured can still present relatively effective auditory scenes compared to renderings from individualised HRTF sets. We present and assess a system for HRTF selection that relies on holistic judgements of users to identify their optimal match through a series of pairwise adversarial comparisons. The mechanism resulted in clear preference for a single HRTF set in a majority of cases. Where this did not occur, randomised selection between equally judged HRTFs did not significantly impact user performance in a subsequent listening task. This approach is shown to be equally effective for both novice and expert listeners in selecting their preferred HRTF set.

12:30 - 12:55 Tuesday, August 21

AUTHORS: Ramani Duraiswami, Nail Gumerov, Adam O'Donovan and Dmitry Zotkin, *VisiSonics Corporation*, Justin Shen, Matthias Zwicker, *University of Maryland*

TITLE: Large Scale HRTF personalization N/A

Many applications in creating realistic audio for augmented and virtual reality require individual head-related transfer functions. Typically individual HRTFs are measured typically via a slow and relatively tedious procedure, making them unusable in practical applications. We discuss two approaches that allow the creation of HRTFs at large scale for applications. The first is a fast reciprocal measurement approach which allows fast measurement. We have applied the technique to measure the HRTF of several individuals. The second approach is based upon computation of the HRTF in the cloud using meshes in a few minutes. We have shown this approach works with 3D scans, and are now extending it to photograph based processing.

Paper Session 5

Ambisonics

Room: Michelangelo

Session Chair: Joseph Tylka

2:00 - 2:25 Tuesday, August 21

AUTHORS: **Fernando Lopez-Lezcano**, *CCRMA/Stanford University*

TITLE: **The *SpHEAR project update: the TinySpHEAR and Octathingy soundfield microphones** 165

This paper is an update of the *SpHEAR (Spherical Harmonics Ear) project, created with the goal of using low cost 3D printers to fabricate Ambisonics microphones. The initial four-capsule prototypes reported in 2016 have evolved into a family of full-featured high quality microphones that include the traditional tetrahedral design and a more advanced eight capsule microphone that can capture second-order soundfields. The project includes all mechanical 3d models and electrical designs, as well as all the procedures and software needed to calibrate the microphones for best performance. A fully-automated robotic arm measurement rig is also described. Everything in the project is shared through GPL/CC licenses, uses Free Software components, and is available on a public GIT repository (<https://cm-gitlab.stanford.edu/ambisonics/SpHEAR/>)

2:30 - 2:55 Tuesday, August 21

AUTHORS: **Michael Goodwin**, *Xperi, Inc / DTS*

TITLE: **A hybrid beamforming framework for B-format encoding with arbitrary microphone arrays** 177

In this paper we consider the problem of B-format encoding of live audio scenes captured using practical compact microphone arrays with unconstrained microphone locations. We formulate the problem and provide a mathematical framework for a hybrid adaptive beamformer which combines active beamforming based on frequency-domain spatial analysis and synthesis for estimated directional audio sources and least-squares passive beamforming for residual audio scene components. The necessary calibration measurements are described and algorithmic details for efficient implementation are given.

3:00 - 3:25 Tuesday, August 21

AUTHORS: **Calum Armstrong, Damian Murphy and Gavin Kearney**, *University of York, UK*

TITLE: **A Bi-RADIAL Approach to Ambisonics** 184

This paper introduces Binaural Rendering of Audio by Duplex Independent Auralised Listening (Bi-RADIAL), a new scheme for the reproduction of 3D sound over headphones. Principles, considerations and methods of Bi-RADIAL decoding and their application within the binaural rendering of Virtual Ambisonics is discussed. Three methods of delivering Bi-RADIAL Ambisonics are compared and the advantages of exploiting a Bi-RADIAL scheme over traditional binaural-based Ambisonic playback are highlighted. Simulation results for standard and Bi-RADIAL Ambisonic decoders are given using a perceptual based comparison of their frequency spectra. Analysis is made for 1st, 3rd and 5th order decoders considering both basic and maxRE weighting schemes. Results show a clear improvement in spectral reconstruction when using a Bi-RADIAL decoder

Paper Session 6

Binaural Rendering of 3D Sound Fields

Room: Michelangelo

Session Chair: Agnieszka Roginska

4:00 - 4:25 Tuesday, August 21

AUTHORS: César Salvador, Shuichi Sakamoto, Jorge Trevino and Yôiti Suzuki, *Tohoku University, Japan*

TITLE: **Enhancing binaural reconstruction from rigid circular microphone array recordings by using virtual microphones** 194

Spatially accurate binaural reconstruction from rigid circular arrays requires a large number of microphones. However, physically adding microphones to available arrays is not always feasible. In environments such as conference rooms or concert halls, prior knowledge regarding source positions allows for the prediction of pressure signals at positions without microphones. Prediction is performed by relying on a physical model for the acoustically rigid sphere. Recently, we used this model to formulate a surface pressure interpolation method for virtual microphone generation. In this study, we use virtual microphones to enhance the high-frequency spatial accuracy of binaural reconstruction. Numerical experiments in anechoic and reverberant conditions demonstrate that adding virtual microphones extends the frequency range of operation and attenuates the time-domain artifacts.

4:30 - 4:55 Tuesday, August 21

AUTHORS: Axel Plinge, *Fraunhofer IIS, Germany*, Sebastian Schlecht, Oliver Thiergart, Thomas Robotham, Olli Rummukainen and Emanuel Habets, *International Audio Laboratories, Erlangen, Germany*

TITLE: **Six-Degrees-of-Freedom Binaural Audio Reproduction of First-Order Ambisonics with Distance Information** 203

First-order Ambisonics (FOA) recordings can be processed and reproduced over headphones. They can be rotated to account for the listener's head orientation. However, virtual reality (VR) systems allow the listener to move in six-degrees-of-freedom (6DoF), i.e., three rotational plus three transitional degrees of freedom. Here, the apparent angles and distances of the sound sources depend on the listener's position. We propose a technique to facilitate 6DoF. In particular, a FOA recording is described using a parametric model, which is modified based on the listener's position and information about the distances to the sources. We evaluate our method by a listening test, comparing different binaural renderings of a synthetic sound scene in which the listener can move freely.

5:00 - 5:25 Tuesday, August 21

AUTHORS: Thomas McKenzie, Damian Murphy and Gavin Kearney, *University of York, UK*

TITLE: **Directional Bias Equalisation of First-Order Binaural Ambisonic Rendering** 213

The human auditory system is more accurate at localising sounds in front than at lateral, rear and elevated directions. In virtual reality applications, where Ambisonic audio is presented to the user binaurally (over headphones) in conjunction with a head-mounted display, it is imperative that audio in the frontal direction is as accurate as possible. This paper presents a method for improving frontal high frequency reproduction of binaural Ambisonic rendering through a novel adaptation of the diffuse-field equalisation technique to exploit the non-uniform directional sensitivity of human hearing. The method is evaluated via spectral difference and a height localisation model, and results show improved frontal reproduction at the expense of lateral fidelity.

Workshop 6 & 7

HRTF and Ambisonics

Room: Plato

11:30 - 12:25 Tuesday, August 21

Michel Henein, *VisiSonics*

TITLE: **The Basics of HRTF Personalization** N/A

VR/AR strives to convince the human brain that the virtual world is real. Since sound in the real world is critical to the human experience, virtual sound, based on HRTFs, must be convincing. Due to anthropometric differences between individuals, a generic HRTF set does not provide convincing 3D effect for all humans, which is a major problem for mass VR/AR adoption. Research conducted by the VisiSonics team, first at the University of Maryland, and then at the company for its RealSpace 3D engine has resulted in a few techniques enabling individualized HRTFs to be created quickly, thus helping to provide the most accurate spatial audio experience – as good as being there – for all VR/AR users.

12:35 - 1:30 Tuesday, August 21

Markus Zaunschirm, Daniel Rudrich and Christian Schörkhuber, *IEM, University of Music and Performing Arts, Graz, Austria*

TITLE: **Binaural Ambisonics rendering for headphones** N/A

With the widespread adoption of the Ambisonics format for headphone-based playback of spatial audio the binaural rendering of order-limited Ambisonic signals became of great interest in the fields of virtual reality and 360-degree multimedia productions. Classic binaural renderers adopt decoding strategies from loudspeaker-based rendering (e.g. virtual loudspeaker approach). They either yield strongly direction-dependent coloration (dull in front) when using a dense set of virtual loudspeakers, or they yield strongly sampling-dependent spatial cues and spatial aliasing with a sparse layout which only uses those few virtual loudspeakers required for the Ambisonic order. This workshop compares classic binaural decoding with new cutting-edge binaural Ambisonic renderers. We discuss differences in coloration and localization, and present approaches for improving externalization.

Workshop 8

XR Audio in Diverse Industries

Room: Plato

2:30 - 3:25 Tuesday, August 21

Joel Douek and Benedict Green, *EccoVR, Inc.*

TITLE: **Riding the XR Sound Wave into Diverse Industries** N/A

The nascent technologies of Virtual, Augmented and Mixed reality, although originally driven from the world of entertainment and video games, are extending into an untold number of industries. What this means for anyone working in XR, including on the audio side, is that all these industries are becoming part of our world by extension. It means that career opportunities are multiplying manifold in fascinating new areas where sound never played an obvious role. This expansion of reach has real implications for our professional lives, not only in terms of job opportunities and earning potential, but also the possibility of being at the forefront of new technological and creative developments that will influence the shape of the future.

Workshop 9

XR Audio and the Music World

Room: Plato

4:00 - 4:55 Tuesday, August 21

Joel Douek, *EccoVR, Inc.*

TITLE: **Can XR Rekindle the Music World?** N/A

Against the often bleak picture of music industry and the diminishing value of music, XR offers a ray of hope. With the personal presence, proximity and free agency the VR listener has, a sense of intimacy, wonderment and privilege can be restored to the musical experience. That magical connection can be regained as we stand (albeit virtually) in the presence of our musical heroes or be present at the birth of a song. AR, MR & Social VR can further enrich musical experiences with opportunities for interaction and deeper information. This talk will showcase some recent music-driven XR experiences and explore the future potential of these new platforms and technologies to breath life into the creation and appreciation of music.

Paper Session 7

Virtual Acoustics & Environment Modeling

Room: Michelangelo

Session Chair: Ramani Duraiswami

11:30 - 11:55 Wednesday, August 22

AUTHORS: **Sara R. Martin and U. Peter Svensson**, *Acoustics Research Centre, Dep. of Electronic Systems, Norwegian University of Science and Technology, Trondheim, Norway*

TITLE: **Modeling sound sources with non-convex shapes using an edge diffraction approach** 222

This paper explores the modeling of sound radiation from vibrating structures, representing the acoustic environment with Green's functions. A fictive convex hull is created that encloses the vibrating structure, and different subdomains will be created at the structure's indents. Each boundary between those subdomains and the convex exterior is then discretized, employing "virtual pistons". The radiation impedance of those virtual pistons can be computed efficiently for the external convex domain with the edge diffraction model in [J. Acoust. Soc. Am. 133, pp. 3681-3691, 2013]. The impedance contributions of the indenting subdomains can be computed theoretically, by means of modal analysis or by any common numerical method. An open shoebox-shaped object is presented and analyzed as an example.

12:00 - 12:25 Wednesday, August 22

AUTHORS: **Anne Heimes, Muhammad Imran and Michael Vorländer**, *Institute of Technical Acoustics RWTH University, Germany*

TITLE: **Real-Time Building Acoustics Auralization of Virtual Environments** N/A

In this study we propose a framework for auralization of sound transmission in virtual buildings and develop the building acoustics filters. This paper describes the calculations for airborne sound insulation metrics based on ISO-12354 and comprehends the auralization process of sound transmission through the flanking paths of building elements. The insulation filters describe the sound transmission between dwellings by partitions and by flanking structures to estimate the transfer functions between the sources and receivers. In Unity 3D, a virtual scenario is created that consists of a multi-storey resident flat with different types of building elements. The sound insulation filters are implemented into virtual scenario to introduce more realism in auralization of buildings acoustics in virtual spaces.

12:30 - 12:55 Wednesday, August 22

AUTHORS: **Keith Godin, Ryan Rohrer, John Snyder and Nikunj Raghuvanshi**, *Microsoft*

TITLE: **Wave Acoustics in a Mixed Reality Shell** 232

We demonstrate the first integration of wave acoustics in a virtual reality operating system. The Windows mixed reality shell hosts third-party applications inside a 3D virtual home, propagating sound from these applications throughout the environment to provide a natural user interface. Rather than applying manually-designed reverberation volumes or ray-traced geometric acoustics, we use wave acoustics which robustly captures cues like diffracted occlusion and reverberation propagating through portals while reducing the design and maintenance burden. We describe our rendering implementation, materials-based design techniques, reverberation tuning, dynamic range management, and temporal smoothing that ensure a natural listening experience across unpredictable audio content and user motion.

Paper Session 8

Applications in VR/AR

Room: Michelangelo

Session Chair: Jean-Marc Jot

2:00 - 2:25 Wednesday, August 22

AUTHORS: **Jonathan Mathews and Jonas Braasch**, *Rensselaer Polytechnic Institute*

TITLE: **Real-Time Source-Tracking Spherical Microphone Arrays for Immersive Environments** 240

Spherical microphone arrays have attained considerable interest in recent years for their ability to decompose three-dimensional soundfields. This paper details real-time capabilities of a source-tracking system composed of a beamforming array and multiple lavalier microphones. Using the lavalier microphones for source identification, a particle filter can be implemented to allow independent tracking of the orientation of multiple sources simultaneously. This source identification and tracking mechanism is utilized in an immersive lab space. In conjunction with networked audiovisual equipment, the system can generate a real-time virtual representation of sound sources for a more dynamic telematic experience.

2:30 - 2:55 Wednesday, August 22

AUTHORS: **Rishabh Gupta, Rishabh Ranjan, Woon-Seng Gan**, *Nanyang Technological University, Singapore*, **Jianjun He**, *Maxim Integrated Product Inc*

TITLE: **On the use of closed back headphones for active hear-through equalization in augmented reality applications** 246

Augmented Reality (AR) audio refers to techniques where virtual sounds are superimposed with real sounds to produce immersive digital content. Headphones are widely used in consumer devices for playback of virtual sounds. However, for AR audio, an important step is to make sure that headphones allow external sounds to pass through naturally using Hear-Through (HT) processing. In this paper, an investigation of HT design for real sounds using closed-back circumaural headphones equipped with two pairs of microphones was conducted. An adaptive filtering algorithm was used to derive the equalization filter. Experimental result shows close match of the equalized signal to reference open ear listening. Subjective study was carried out to compare the spatial and timbre sound quality of HT mode.

3:00 - 3:25 Wednesday, August 22

AUTHORS: **Karlheinz Brandenburg**, *Fraunhofer IDMT & Technical University Ilmenau, Germany*, **Estefanía Cano, Hanna Lukashevich, Thomas Köllmer**, *Fraunhofer IDMT, Ilmenau, Germany*, **Annika Neidhardt, Florian Klein, Ulrike Sloma and Stephan Werner**, *Technical University Ilmenau, Germany*

TITLE: **Plausible Augmentation of Auditory Scenes Using Dynamic Binaural Synthesis for Personalized Auditory Realities** 258

Personalized Auditory Realities (PARTy) have been introduced at a recent conference (Karlheinz Brandenburg et al.: Personalized Auditory Realities, DAGA 2018, Garching, Germany). In this proposed system, a real auditory environment is analyzed to find separate audio objects. These may be isolated using source separation techniques. Such objects can then be manipulated, virtual sound objects added and the whole scene can be rendered to fit into the real auditory environment. This paper presents the basic idea of PARTy and then focuses on the dynamic binaural synthesis system which is used to reproduce ear signals fitting into the real environment. The viability of the system is evaluated for the features localization, externalization, and overall quality in terms of a spatial auditory perception.

Paper Session 8 (continued)

Applications in VR/AR

Room: Michelangelo

Session Chair: Jean-Marc Jot

3:30 - 3:55 **Wednesday, August 22**

AUTHORS: **Andrea Genovese, Gabriel Zalles, Gregory Reardon and Agnieszka Roginska, *New York University***

TITLE: **Acoustic perturbations in HRTFs measured on Mixed Reality Headsets** 268

Materials that obstruct the path of acoustic waveforms in free-field to the human ear, may introduce distortions that can modify the natural Head-Related Transfer Functions. In this paper, the effect of wearing commercially available Head-Mounted Displays for Mixed and Augmented Reality has been measured via a dummy head mannequin. Such spectral distortions may be relevant for mixed reality environments where real and virtual sounds mix together in the same auditory scene. The analysis revealed that the measured HMDs affected the fine structure of the HRTF ($> 3-6$ kHz) and also introduced non-negligible distortions in the interaural level difference range mostly at the contralateral ear. Distortion patterns in HRTFs and cue modifications are reported and discussed.

4:00 - 4:25 **Wednesday, August 22**

AUTHORS: **Zeynep Özcan, *Istanbul Technical University*, Anıl Çamcı, *University of Michigan***

TITLE: **An Augmented Reality Music Composition Based on the Sonification of Animal Behavior** 284

In this paper, we discuss the immersive sonification of an artificial ecosystem in the form of an interactive sound art piece, named Proprius. We first offer an overview of existing work that utilizes ecological models in compositional and sonification contexts. We then describe the behavioral and ethological models utilized in Proprius. We evaluate the musical characteristics of animal behaviors, and discuss our approach to sonifying them in the context of an interactive augmented reality composition. We provide details of our system in terms of how it implements ecological simulation, immersive audio, and embodied interaction.

Paper Session 9

HRTF Modeling

Room: Michelangelo

Session Chair: Ravish Mehra

5:00 - 5:25 Wednesday, August 22

AUTHORS: **Christopher Buchanan**, *Signum Audio, UK*, **Michael Newton**, *Acoustics & Audio Group, University of Edinburgh, UK*

TITLE: **Dynamic Balanced Model Truncation of the Spherical Transfer Function For Use in Structural HRTF Models** 293

The Spherical Transfer Function (STF) has previously been used in structural HRTF modelling as an analytical approximation to the human head. Versions based on both spherical and spheroidal solid bodies have been incorporated into a range of systems, such as the well known Brown & Duda [Brown 1998] structural model. STF-based models provide a way to simulate frequency dependent head shadowing (ILD) and time delay (ITD) effects, which can form the foundation for structural HRTF representation. We derive and implement a customizable approximation of the STF based on Balanced Model Truncation, and utilize its inherent modular characteristics to synthesize binaural signals from monaural input with relatively low cost implications.

5:30 - 5:55 Wednesday, August 22

AUTHORS: **David Romblom and Helene Bahu**, *Dysonics, San Francisco*

TITLE: **A Revision and Objective Evaluation of the 1-Pole 1-Zero Spherical Head Shadowing Filter** 303

Structural models of Head Related Transfer Functions attempt to decompose complex acoustic phenomena into constituent signal processing models. The work of Brown and Duda modeled the head with two elements: a pure delay estimated from a ray-tracing formula by Woodworth and a 1-pole 1-zero shadowing filter. The ray-tracing formula is valid for frequencies above $\sim 2\text{kHz}$ while interaural time delay (ITD) is perceptually significant below $\sim 1.5\text{kHz}$. The frequency-dependence of the phase-derived ITD_p has been shown by many authors but is not accounted for by ray-tracing assumptions. As such, the shadowing filter must account for the low frequency variation in ITD_p. Using a numerical approximation of Rayleigh's solution as a reference, this paper evaluates and revises the work of Brown and Duda.

6:00 - 6:25 Wednesday, August 22

AUTHORS: **Fabian Brinkmann and Stefan Weinzierl**, *Technical University of Berlin, Germany*

TITLE: **Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition** 311

Head-related transfer functions (HRTFs) are the basis for virtual auditory reality, and a larger number of HRTFs is required for a perceptually transparent representation. Consequently, HRTFs are commonly reconstructed from a reduced data set by means of interpolation. The spherical harmonics transform (SHT), which decomposes HRTF sets into weighted orthogonal basis functions, is an particularly promising approach for this, because it yields a spatially continuous HRTF representation. Because the SHT has to be order truncated in practice, it is of interest to find HRTF representations that concentrate the HRTF energy at low SH orders. This study compares previous approaches to a decomposition based on time aligned HRIRs for their suitability to reduce the required SH order.

Paper Session 9 (continued)

HRTF Modeling

Room: Michelangelo

Session Chair: Ravish Mehra

6:30 - 6:55 Wednesday, August 22

AUTHORS: **Faiyadh Shahid, Nikhil Javeri, Kapil Jain and Shruti Badhwar**, *EmbodimentVR*

TITLE: **AI DevOps for large-scale HRTF prediction and evaluation: an end to end pipeline** 321

Bringing truly immersive 3D audio experiences to the end user requires a fast and a user friendly method of predicting HRTFs. While machine learning based approaches for HRTF prediction hold potential, it can be challenging to determine the best workflow for deployment given the iterative nature of data preprocessing, feature extraction, prediction and performance evaluation. Here, we describe an automated, end to end pipeline for HRTF prediction and evaluation that simultaneously tracks the data, code and model, allowing for a comparison of existing and new techniques against a single benchmark.

Workshops 10 & 11

Dialogue for VR/AR and MPEG-H

Room: Plato

11:30 - 12:25 Wednesday, August 22

Jelle van Mourik, Simon Gumbleton and Nick Ward-Foxton, *Sony Interactive Entertainment Europe*

TITLE: Recording and implementing dialogue for VR applications N/A

Rendering voice recordings in a 3-dimensional virtual environment presents many technical and creative challenges, many different from traditional dialogue recording. Based on the learnings from the production of several VR game titles, this talk will discuss an end-to-end workflow of getting dialogue into a VR application. It discusses best-practice techniques with respect to the recording stage. Additionally, all relevant signal paths, such as HRTF processing of the direct sound and reflections, and spatial processing of reverberation are discussed. Particular attention is given to first-person dialogue, which provides an extra challenge, as the lack of proprioception of speech complicates the player's suspension of disbelief that the voice is emanating from themselves. Finally, future improvements and topics for investigation are suggested.

12:35 - 1:30 Wednesday, August 22

Patrick Flangan, *THX Ltd.*

TITLE: MPEG-H for Gaming AR/VR/MR. Using HOA and Objects to deliver immersive audio. N/A

A new way to deliver immersive audio for gaming, VR, AR, and MR. Currently the content available on the market is 3 order Ambisonics at best. Here we present MPEG-H that is capable of delivering up 6th order HOA. A fundamental shift to traditional channel-based audio playback where the device say that you must render to 7.1. In the meantime, all the content is consumed on headphones.

Workshops 12 & 13

Environments and Spatial Computing

Room: Plato

2:30 - 3:25 Wednesday, August 22

Anne Heimes, Muhammad Imran and Michael Vorländer, *Institute of Technical Acoustics RWTH Aachen University*

TITLE: Real-Time Implementation of Building Acoustic Auralization for Virtual Reality Environments N/A

We implement a framework for real-time auralization of sound transmission for virtual environments. In Unity 3D, a virtual building is created consisting of a resident flat with different types of building elements. The demo shows real-time airborne sound insulation auralization that comprehends the sound transmission filter for flanking paths of building elements. The insulation filters describe the sound transmission between dwellings by partitions and by flanking structures. The real-time performance of implementation is evaluated in terms of computational costs. Therefore, real-time auralization of virtual buildings is possible for evaluation of acoustics of building structures in immersive manners. This framework allows the users to perform tasks of daily life of work under conditions of usual behavior and movement in virtual reality.

3:35 - 4:30 Wednesday, August 22

Anastasia Devana, *Magic Leap*

TITLE: The 4 Domains of Audio in Spatial Computing N/A

Audio techniques for visual media have developed and matured alongside mediums like film, television and video games. Audio professionals in these mediums enjoy established paradigms and sophisticated tools. But the advancement of technologies like Spatial Computing has introduced a brand new medium, with a whole new set of audio challenges, considerations, and possibilities. Creating compelling audio in this medium requires audio designers to change the paradigm or how they think about sound, and to seek out new vocabulary, tools, and processes. In this talk I will introduce the concept of 4 domains of audio, as a prism through which sound designers can address the challenges and take advantage of the unique opportunities of this medium.

Workshops 14 & 15

Immersive Audio for VR/AR

Room: Plato

5:00 - 5:55 Wednesday, August 22

Tim Gedemer and Francois LaFleur, *Source Sound, Inc.*

TITLE: Keeping Up With Visual Advances: Creating Innovative Audio for VR/AR - The Struggle is Real N/A

While the world continues to generate copious amounts of traditional audio / video content, the battle to capture future media consumers is on. With higher quality VR/AR/MR experiences being produced every day, audio must keep up it's end of the bargain by providing sound to match stunning visual advances occurring in these new mediums. This must be realized through the continued advancement of audio technology by researchers and product developers, in concert with the application of that technology by the creative minds working in the field. In this series of three case studies: Pixar's *Coco VR*, Dreamscape's *Alien Zoo*, and CNN's VR news portal, we share information on technical and creative challenges encountered during each project's spatial audio production.

6:05 - 7:00 Wednesday, August 22

Varun Nair, Jon Ojeda, Andrew Boyd, *Facebook*

TITLE: Immersive Audio At Facebook N/A

This panel discussion with the Facebook Sound Design and Audio engineering teams will focus on the process of designing audio for immersive media at scale. Four short presentations - sound design process, audio for AR, audio for VR, and audio for 360 media - will be followed by a discussion. The presentations will cover Facebook's unique approach to sound design at scale, lessons in designing immersive audio across a variety of formats and experiences, and the challenges in designing audio at the cutting edge.