International Conference
Spatial Reproduction
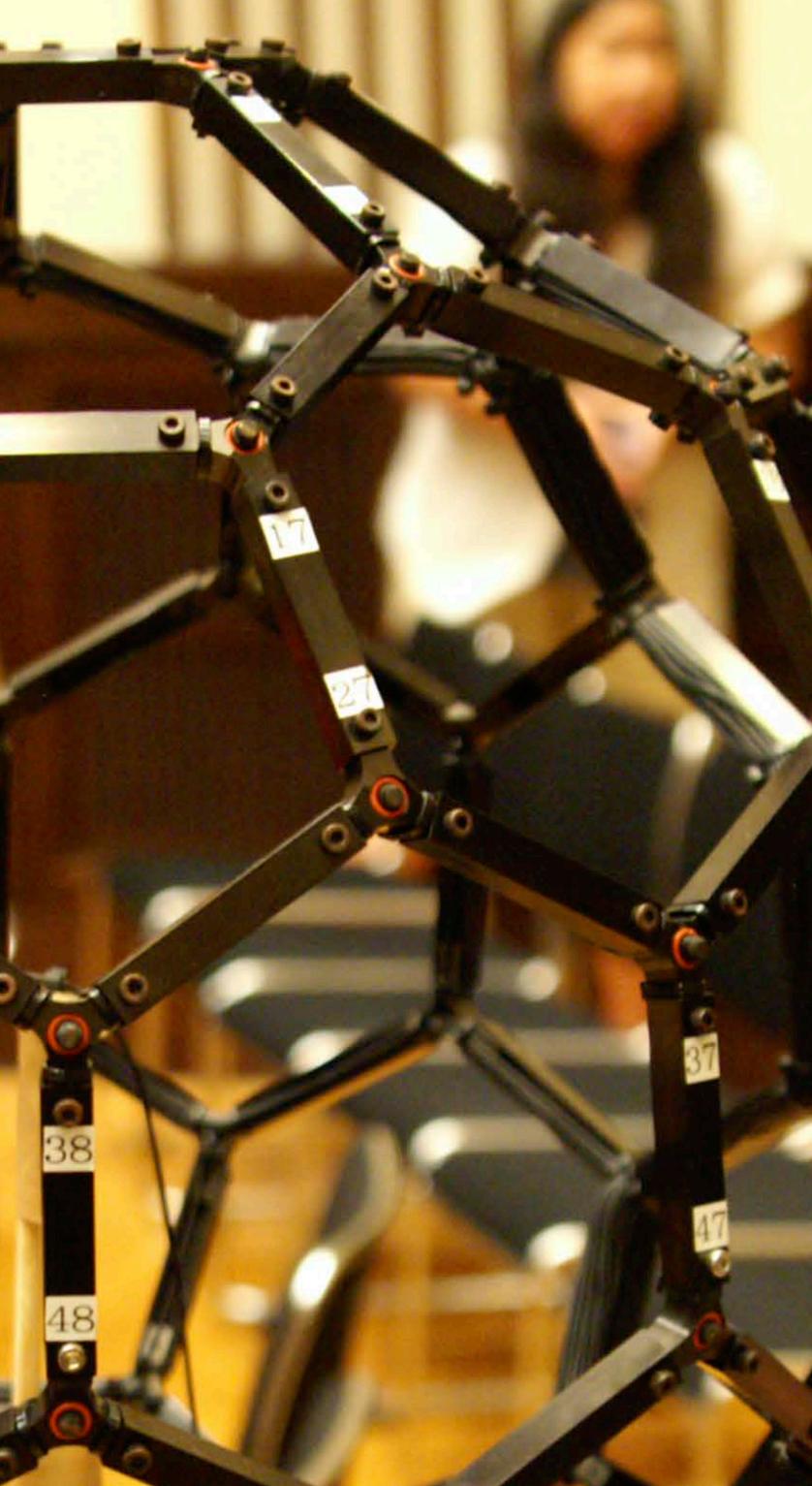## Aesthetics and Science

Tokyo, Japan
6–9 September, 2018

CONFERENCE REPORT

In August of 2018 audio experts, researchers and practitioners gathered in Tokyo to attend the AES International Conference on Spatial Reproduction–Aesthetics and Science. The topic of spatial audio is a mainstay of the AES—this was the seventh conference dedicated to the topic and the second to be held in Tokyo. Since the last conference on spatial audio was held in Helsinki five years ago, there has been an explosion of interest in the topic driven by advances in virtual and augmented reality technology alongside interest from broadcasters in binaural and immersive audio formats. The conference addressed a number of themes including how spatial audio technology can contribute to artistic productions, what technologies and scientific knowledge are required by content creators, and how science and aesthetics can be better connected in this area.

The main event was held at the Senju Campus of Tokyo Denki University. Despite the threat of a typhoon and tropical storms, the conference was a huge success with around 230 delegates attending the conference and an additional 60 delegates registering for the exhibits. The technical paper track included 67 papers covering a range of spatial audio topics including production, game audio, psychoacoustics, ambisonics, and binaural. Running alongside the paper sessions was a tutorial and workshop track that included topics such as microphone techniques for immersive audio capture, object-based audio workflows, and immersive automotive audio. The hall for the workshop sessions was kitted out with an impressive 9.1 system allowing presenters to play back immersive audio content to illustrate their discussions. Lunch and coffee were provided each day giving delegates the opportunity to socialize and discuss the topics presented at the conference. There

Cochairs Yuko Watanabe and Toru Kamekawa open the event.

were demonstrations of the latest 3D audio technology over the three days of the conference, both in dedicated demonstration spaces and in the sponsor booths. Tokyo University of the Arts acted as a satellite venue hosting a welcome concert, demonstrations of 22.2 material, and sessions allowing students to receive critiques on their 3D audio recordings.

A pre-event was held the day before the conference opened. A select group of delegates was given the opportunity to tour the 22.2 multi-channel audio facilities at NHK Broadcasting Center, Shibuya. Following this, delegates were treated to a welcome concert showcasing traditional Japanese music. The concert was held in Studio A at the Tokyo University of the Arts and included a number of pieces for the koto, the national instrument of Japan. In stark contrast to the traditional music, the concert was captured using a variety of state-of-the-art immersive audio capture methods including a spaced array, a soundfield microphone, and an 80-channel array—at the last count, there were a total of 109 microphones in the room.

DAY 1: BINAURAL REPRODUCTION, PSYCHOACOUSTICS, AND THE WOW FACTOR

With the stage set by the concert the previous evening, Hiroshi Yasuda, president of Tokyo Denki University, opened the proceedings with a welcome address charting the evolution and growth of the

Hiroshi Yasuda offers keynote.

Internet of Things—an area of technology that will surely have an impact on how we consume spatial audio in the home in the coming years. With the growth in the number and speed of connected devices, Professor Yasuda stressed the challenges of security and authentication and the important role that speech signal processing could play in these areas.

Following the welcome address, former AES president Jim Anderson gave an entertaining opening keynote about the wow factor in audio reproduction. Using archive recordings dating back to the 1950s, Professor Anderson demonstrated how manufacturers educated the audience about the benefits of

Jim Anderson's keynote explains the wow factor

stereo reproduction over mono. Labels such as Capital Records, Decca, and RCA produced demonstrations of stereo audio designed to wow the listeners at home and convince them of the benefits of the exciting new

Cochair Toru Kamekawa with artists at the opening concert

Delegates gather around one of the many poster presentations.

technology. It was striking how many of these demonstrations were nonmusical scenes—one of the most effective demonstrations was a stereo reproduction found on the Capital Records Stereo Disc of a New Year's Eve crowd in Time Square. Delegates were left to ponder the question of how we as audio scientists and professionals can convince the current audience of the benefits of immersive audio, a theme that would recur in the closing panel discussion.

Paper presentations began with two sessions on binaural reproduction. The sessions opened with a talk from Ryo Matsuda from Kyoto University on the robustness of dynamic crosstalk cancellation for different arrangements of loudspeakers. Using acoustic measurements in a spherical loudspeaker array, loudspeakers located on opposite lateral sides of the listener were found to be the optimal positions for two-speaker crosstalk cancellation. Helene Bahu from Dysonics presented a paper on the prediction of ITD parameters using geometric head models with offset ears. It was shown that shifting the ears from antipodal locations in a simple spherical head model resulted in a reduction of the global ITD error. Bosun Xie from South China University presented a study on the effect of dynamic and spectral cues on vertical localization showing that both types of cues have a contribution. Annika Neidhardt from TU Ilmenau presented a study investigating the detectability of nearby surfaces via echolocation in a virtual environment. It was found that the detection rate increased with decreasing distance to the surface.

After a short break, the second binaural session opened with a paper by César Salvador from Tohoku University presenting a dataset of near-distance HRTFs calculated using the boundary element method. Insight into individual distance-related features was provided through spherical and circular Fourier analysis of the dataset.

After the paper session on binaural, the conference participants had the opportunity to explore the first poster session of the conference. A short plenary session allowed the presenters to give a brief overview of their posters. The topics covered in the session included real-time FDTD simulation of room acoustics for VR, the effects of individualized HRTFs on distance perception, the use of decorrelation filters to control apparent source size in ambisonics, real-time sound field rendering, source separation using image signal processing, and the effect of the number of virtual loudspeakers on binaural rendering.

The paper track wrapped up for the day with two sessions on psychoacoustics. Tom Wühle and Sebastian Merchel from TU Dresden opened the first session with papers on different psychoacoustic aspects of sound projectors. The first of these papers presented an experiment to quantify the level attenuation of the direct sound from a beam steering array necessary to ensure the projected sound remained in its intended position. The second of these papers investigated the mid-frequency attenuation necessary for the position of the projected sound to remain in a stable position. Following this, James Woodcock from the University of Salford presented a perceptual evaluation of media device orchestration—the idea of using an ad hoc arrays of personal devices to augment a spatial audio experience. The evaluation showed that this method led to a better overall quality of listening experience than standard five-channel reproduction outside of the sweet spot listening position.

The second psychoacoustics session began with a presentation by Pablo Delgado from International Audio Laboratories Erlangen on an energy- and frequency-dependent model for interaural level difference distortions. Misaki Hasuo from WOWOW Inc. presented work on the effect of sound source characteristics on spatial impression in multichannel audio reproduction. The work suggested that the ability to distinguish between different numbers of reproduction channels is influenced by the type of sound source. The final paper of the session was presented by Sungyoung Kim on behalf of William Martens and was on the effect of head-rolling on the discrimination of spatial diffuseness in with- and without-height reproduction. The results of a 2-AFC experiment suggested listeners could detect a decrease in spatial diffuseness when they were allowed to roll their heads in the without-height condition.

Running alongside the paper track were a series of informative workshops and tutorials. In the first tutorial session, Kimio Hamasaki led a discussion with Jim Anderson and Ulrike Schwartz from Anderson Audio on their recent experiences recording using various 3D recording formats. The second tutorial session was led by Hyunkook Lee from the University of Huddersfield and focused on the psychoacoustics of 3D sound recording and reproduction. The first workshop, facilitated by Mick Sawaguchi from UNAMAS-Label and Hideo Irimajiri from WOWOW Inc., described immersive sound production techniques used in Japan. Finally, Will Howie from McGill University gave a workshop presenting new techniques for music recording and mixing for 22.2 reproduction.

Hyunkook Lee leads a tutorial on the psychoacoustics of 3D sound.

Kimio Hamasaki (center) in conversation with Jim Anderson and Ulrike Schwartz.

DAY 2: GAME AUDIO, AMBISONICS, AND MORE PSYCHOACOUSTICS

Akio Ando's keynote discussed rendering methods.

The second day kicked off with a keynote from Akio Ando on rendering methods for multi-channel audio. The keynote discussed recent advances in multichannel audio, and framed the 22.2 format as primarily a production format—a high-resolution reference that conceptually decouples the production and reproduction environments. As the production system is now abstracted from the reproduction system, rendering methods are required for reproduction of immersive content over different systems. Akio advocated using different rendering approaches for the direct and diffuse components of a soundfield. Expanding on this, he discussed the upmix method used by NHK that consists of decomposing the soundfield into direct and reverberant components and rendering these components separately.

The first paper session of the day was on the topic of game audio. Will Bailey from the University of Salford opened the session with a paper investigating the correlation between personality traits and externalization of binaural audio in virtual environments. The study suggested that cognitive factors should be considered in assessments of immersive audio in virtual environments. This was followed by two talks by authors from CAPCOM. Takahiro Kitagawa spoke about a method for applying acoustic features to sounds in virtual environments that is computationally more efficient than conventional ray-casting methods. The proposed method consisted of using layered quad tree grids to quickly determine the position of sounds in the environments and the use of busses to apply acoustic features representing the connection between different virtual spaces. Kentaro Nakashima presented a method for rendering close sources in a real-time binaural game engine. Their approach was supported by listening experiments showing that the proposed method resulted in closer-sounding sources than conventional processing methods.

Brian Katz from Sorbonne Université presented a study assessing the impact of individualized HRTFs on player performance in a VR shooter game. Participants played the game with their best-match and worst-match HRTFs. The results indicated that performance in the game improved when participants used their best-match HRTF. Matthew Boerum from McGill University presented a study investigating the influence of virtual environments on the performance of participants in a panning task. No significant differences in performance were found between the task conditions suggesting that participants could perform the task equally well in real and virtual environments.

The first of two e-Brief poster sessions followed the game audio paper session. The topics of the posters included acoustic characterization of headphones, recording experiments with tetrahedral microphones, gestural controllers for 3D panning, recording and mixing techniques for ambisonics, localization in the median plane using specular reflection shape features, automatic movie scene classification based on audio signal features, near-field source rendering in VR, sound installations using sound field synthesis, loudspeaker positions for near-field binaural reproduction without crosstalk cancellation, discrimination of high-resolution audio, a large-scale 3D loudspeaker array driven by a 1-bit signal, and soundfield reproduction using prism-type arrangements of loudspeakers.

Matthieu Parmentier presented workshops on object-based audio.

The third psychoacoustics paper session of the conference opened with a talk by Sungyoung Kim from Rochester Institute of Technology on a study investigating the effect of room effects on the perception of immersive audio. An attribute-rating experiment was conducted using binaural recordings of 22.2 and stereo reproduction in four different rooms. The experiment revealed that differences between the rooms could be described by two perceptual factors relating to spatial dimension and spectral/spatial integration. Following this, Thomas Lund from Genelec gave a fascinating overview of physiological, clinical, and psychological research relating to auditory perceptual bandwidth.

The final paper sessions of the day were on ambisonics. Pierre Lecomte from Centre Lyonnais d'Acoustique presented a method for directional filtering of ambisonic scenes. Tanner Upthegrove from Virginia Tech presented work integrating the Reaper DAW and a Max patch that implemented VBAP, HOA, and 3D reverb rendering. The system was shown to work in a diverse range of spatial audio venues. Pierre Grandjean from Université de Sherbrooke presented work deriving a rule for the size of the reproduction area in spherical harmonic sound field synthesis in terms of directional metrics such as particle velocity and acoustic intensity. Jorge Treviño from Tohoku University showed how a technique from perturbation theory called series acceleration could be used to increase the listening area of HOA reproduction. Finally, Frederico Pereira from The University of Sydney presented an evaluation of a system for augmented audio reality using an over-ear speaker system. The results of the objective and subjective evaluation showed that the proposed system achieved a good degree of precision in the perceived direction of rendered sources.

In the tutorial and workshop track, Hyunkook Lee from Huddersfield University led a panel discussion on microphone techniques for immersive audio recording. Toru Kamekawa chaired a workshop on upmix and downmix techniques that provided delegates with an insight into the current opportunities and challenges in bringing immersive sound to auto-

Panel discussion on microphone techniques for immersive recording

Above left: Delegates enjoy an enticing array of Japanese food and wine.
Above middle: Hiroaki "Bike" Suzuki addresses the crowd.
Above right: Toasts raised during the banquet.
Right: An animated discussion over drinks between, from left, Thomas Lund, Matthieu Parmentier and Florian Camerer.

CONFERENCE BANQUET

motive audio. Finally, Matthieu Parmentier from France TV presented two workshops on object-based audio showcasing some of the production tools and methods currently used in object-based workflows.

The conference banquet was held in the evening and went ahead as planned despite the threat of a typhoon disrupting proceedings. Following speeches by the conference organizers, delegates were treated to a remarkable array of Japanese food along with wine and beers. The conference provided a great opportunity for delegates to socialize and discuss what they had heard at the conference in the preceding days.

DAY 3: SIGNAL PROCESSING, PRODUCTION, AND WRAPPING UP

The final day of the conference began with a paper session on signal and spatial processing. The session was accompanied by Tom Ammermann's Kraftwerk and Booka Shade demonstrations, which could even be heard from the tutorial session on the floor below. The first paper was by Philippe-Aubert Gauthier from Université de Sherbrooke and investigated the use of the latent group lasso as a cost function for soundfield reproduction. Following this, Andreas Franck from the University of Southampton introduced the Versatile Interactive Scene Renderer—an open source software framework for audio processing. An application example was presented demonstrating how the framework could be used to develop a panning algorithm, from development and offline testing to subjective evaluation. The final paper of the session, presented by Atsushi Marui from Tokyo University of the Arts, was on an experiment to evaluate a real-time reverberation enhancement system. The effect of microphone distance on various perceptual

Philippe-Aubert Gauthier

attributes was investigated through a paired comparison experiment.

After a short break, the paper session on production commenced with a paper by Will Howie from McGill University evaluating microphone techniques for 9.1 reproduction. The study evaluated spaced, coincident, and near-coincident microphone techniques and found that the two spaced techniques were generally rated higher than the other techniques on the perceptual attributes investigated in the experiment. Christoph Sladeczek from Fraunhofer spoke about a new method for mastering object-based audio. Their system utilized mastering objects that applied processes to multiple objects in an object-based scene. Aki Mäkivirta from Genelec presented a study investigating the use of spatial averaging in the equalization of immersive audio systems. For the low RT control room used in the study, no significant differences were found between an equalization method using spatial averaging and a single point measurement.

The final poster session of the conference included e-Briefs on individualized binaural reproduction using the virtual artificial head technique, production of immersive content using intensity-based techniques, an immersive audio-visual installation, filter selection for boundary surface control, a method for simulating movement in multichannel reverberation, a compensation method for bone conduction headphones, validation of a 3D mesh of the KEMAR dummy head for BEM simulation, the spatial impression of source widening for binaural reproduction, a reciprocity method for binaural measurement, and a 96-channel immersive reproduction system based on boundary surface control.

The final paper session of the conference was on production and audio systems. The session began with a talk by Gerard Erruz from Eurecat on sound-space choreography. A software tool was presented that controlled sound sources through sets of three-dimensional movements. Etienne Hendrickx from the University of Brest presented the results of an experiment on audiovisual coherence during live performances. In a perceptual evaluation of two different mixes of the same performance, it was found that listeners preferred the mix with audiovisual coherence when a video was presented but a spatially uncon-

Panelists after discussing keys for connecting science and aesthetics in 3D audio reproduction: from left, Akira Omoto, Jim Anderson, Wilfried van Baelen, Yuko Watanabe, Hyunkook Lee, Thorsten Weigelt, and Toru Kamekawa.

Delegates experience one of the many immersive demonstrations presented during the conference.

strained mix when no video was presented. James Woodcock from the University of Salford presented a framework for intelligent rendering in object-based audio. The framework allowed the integration of perceptual models and rules based on expert knowledge to render object-based audio to different speaker layouts. Richard Hughes, also from the University of Salford, presented an overview of dual frequency amplitude panning along with a perceptual evaluation of moving sources for different vector base panning methods and ambisonics. Sebastian Schlecht from International Audio Laboratories Erlangen presented a method for computing the feedback matrix for feedback delay network spatial reverberation. Ephraim Hahn from McGill University presented a study investigating the impact of 3D audio on the listener's emotional response. The study found that 3D audio increases the listener's emotional arousal when listening to classical music excepts.

The day's tutorial and workshop track began with a session from Tom Ammermann who discussed immersive 3D audio production philosophies and workflows though demonstrations of his work on recent Kraftwerk and Booka Shade releases. Following this, Charles Nichols presented a workshop on the development of the Cube multimedia research lab—a facility at Virginia Tech Institute for Creativity, Arts, and Technology that includes a 134.6-speaker system along with direction speakers, 360° video projection, motion capture, and a VR system. The session included the performance of his electric violin playing. Atsushi Ohta from Bandai Namco Studios Inc. presented

a workshop on the current and future challenges of spatial audio expression in VR games. After the lunch break, Sungyoung Kim chaired a tutorial session on how lessons learned from concert hall design on spatial hearing can inform acoustic enhancement systems. The tutorial included an on-site demonstration of the acoustic enhancement system in Yamaha Ginza Hall. The final workshop of the day was chaired by Tomoya Kishi from CAPCOM. Specialists from game, acoustics, VR audio, and postproduction compared the creative approach of 3D audio and emphasized the importance of hybridizing aesthetics and science in game audio.

Proceedings concluded with a panel discussion that posed the question "What are the keys for connecting science and aesthetics in 3D audio reproduction?" A major theme that emerged from the discussion was the need for more communication between audio researchers, practitioners, and consumers. Coming back full circle to Jim Anderson's opening keynote, the panelists stressed the need to convince consumers and practitioners of the benefits of immersive audio, just as was done for stereo in the 1950s and 1960s. The discussion ended on a high note—there was general agreement that the field of spatial audio reproduction had progressed at an impressive rate since the last conference five years ago, and there was enthusiasm for another conference on the topic in the future.

Overall, the conference was a great success. The organizing committee should be congratulated for putting on such a well-organized and interesting event. The conference sponsors should also be thanked for supporting the event as well as putting on demonstrations of their technology: Acoustic Field Inc., Audio Brains, Beetech Inc., Eclipse By Denso Ten, Evixar Inc., Fraunhofer IIS, Genelec Japan Inc., Media Integration, Inc., MQA Ltd., Optis Japan K.K., Otaritec Corporation, Sennheiser Japan K.K., Sona Corporation, Sony Corporation, Tacsystem Inc., Trigence Semiconductor, and Yamaha Music Japan Co. Ltd.

Conference committee members. Front row (L–R): Akira Nishimura, Toru Kamekawa, Yuko Watanabe, Takashi Mikami, Yusuke Ikeda. Back row (L–R): Tomoya Kishi, Yoichiro Mikami, Kazuma Hoshi, Masayuki Mimura, Masataka Nakahara, Hyunkook Lee, Akira Omoto, Hiroyuki Takahashi

Editor's note: AES Members can access the conference papers free of charge via the AES E-library at http://www.aes.org/publications/conferences/