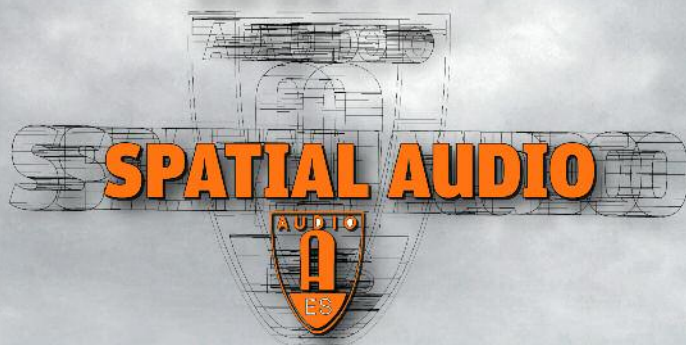


55th International Conference Spatial Audio

AES CONFERENCE ON SPATIAL AUDIO
HELSINKI FINLAND, 27-29 AUG 2014



Chair: Lauri Savioja

27-29 August 2014
Helsinki, Finland

CONFERENCE REPORT

Sponsors



At the end of August 2014, the Audio Engineering Society was welcomed back to Helsinki for the 55th Conference. It was the sixth AES conference to be held in Finland, and the topic was also an AES staple: spatial audio. There have been seven conferences directly concerned with spatial audio and no doubt many more related sessions, workshops, and tutorials at other events. The conference brought together a diverse range of audio scientists from academia and industry (with a 50% split between the two), with interests ranging from the technical reproduction of sound fields, perceptual evaluation, product/service design, musical performance, and more. The wide range of participants (130 attendees from 25 countries) helped to foster a positive atmosphere of discussion, ideation, and collaboration.

The conference took place in the Helsinki Music Centre, which opened in 2011 and is shared by the Sibelius Academy, the Finnish Radio Symphony Orchestra, and the Helsinki Philharmonic Orchestra. The Centre contains a range of performance spaces (the main auditorium was designed by Yasuhisa Toyota and seats 1,700), rehearsal rooms, and recording studios.



Members of the AES 55th Conference organizing committee: from left, Julia Turku, Ville Pulkki, Tapio Lokki, Lauri Savioja, Andrew Goldberg, and Teemu Koski

The conference organizing committee should be commended for hosting such a successful and engaging event, curating an entertaining program of technical and social events. The conference was chaired by Lauri Savioja, Ville Pulkki compiled the technical program, Tapio Lokki acted as papers chair, Teemu Koski was the secretary, Andrew Goldberg organized sponsorship, Julia Turku was treasurer, Kalev Tiits was in charge of the venue, and Florian Camerer helped to organize the panels.

SEMINARS AND DEMONSTRATIONS AT AALTO UNIVERSITY

Proceedings started a day before the conference as 85 participants were treated to an extra afternoon of tutorials and demonstrations hosted by the acoustics group at Aalto University. Ville Pulkki opened the event with a talk about psychoacoustics and technologies of spatial sound. The presentation was intended to offer some conclusions and output at the culmination of a five-year project that had recently finished. The talk focused on the relationship between engineering and psychoacoustics, bridging the gap between spatial sound technology and psychophysics. Pulkki's group was tasked with creating a flexible and loudspeaker-agnostic generic audio format: this was the motivation behind the "DirAC" (directional audio

coding) system, which analyzes and separates directional information and diffuse sound across frequency bands and uses this information in the resynthesis of spatial sound. Pulkki talked about various problems that had been encountered and the resulting solutions that were implemented in order to improve the performance of the system. The engineering work on DirAC was complemented by perceptual investigation of aspects such as the perception of spatially-wide sources and the relationship between audio and video spatial field width.

Following this talk, participants were divided into groups and shown around eight demonstrations in the impressive acoustics facilities at Aalto University. Demos included an immersive audiovisual environment for teleconferencing with 226-degree-horizontal-field-of-view video and 3D audio using DirAC, 12-channel auralizations of concert halls and studio control rooms, and a head-tracked Oculus Rift AV experience, among other examples of technical aspects of the DirAC system.

The afternoon was rounded off with a talk by Tapio Lokki who detailed his work on the evaluation of concert hall acoustics, considering the mapping between physical factors, sensory evaluation, and preference. The talk provided an interesting overview of sensory evaluation methods, which are often borrowed from food and wine science and can include individual assessors developing their own descriptive vocabulary with the use of advanced statistical methodologies to assess the relationship between physical parameters, perceptual ratings, and the assessors. Lokki also gave an interesting description of his tour of European concert halls in order to capture room impulse responses: "we stole the acoustics, packed it, and put it in a bus."

KEYNOTE LECTURES

Each of the three days of the conference started with an invited presentation from a distinguished researcher in the field, carefully selected by the organizing committee to cover the engineering, perceptual, and musical aspects of spatial sound technologies.

On day one, Sascha Spors gave a talk entitled "The Adventure of Spatial Sound Reproduction," which was the perfect way to kick off the conference.

The talk included an outline of the history of spatial audio reproduction, starting with the invention of the telephone in 1861 all the way through to modern multichannel systems and panning algorithms. Sascha touched upon a number of points that would be revisited in presentations throughout the conference: the fact that it is often necessary to create the desired impression rather than to exactly duplicate a real listening experience; the fact that it is important not to overlook timbral factors in a spatial audio system; and the relationships between signal processing, perception, and acoustics, which when integrated give a complete view of spatial sound reproduction. During the questions after the talk participants were introduced to the "Catchbox," a throwable microphone



Sascha Spors identifies some key areas of spatial audio during his keynote lecture.



Sakari Tervo presents a demonstration of auralizations of concert halls and studio control rooms in the largest of three anechoic chambers at Aalto University.



MorrowSound ran a 3D audio-visual demo over Oculus Rift during the conference.

the diversity of participants and the outlook of the AES. As an “acousmatic” composer (i.e., music that is made in a studio, performed over loudspeakers, exists only as a recording, and has an unlimited sound palette), Harrison described his *modus operandi* as being flexible, listening based, and not based around prior planning, and suggested that audio engineering can facilitate this approach (for example, such composition is only possible due to the development of sound recording and storage). Harrison introduced the loudspeaker layouts used in the Birmingham BEAST (Birmingham ElectroAcoustic Sound Studio) project, which has over one hundred loudspeakers at its disposal for composition and performance. It was interesting to note that the motivation for loudspeaker positioning was empirical and listening-based, and that in this particular use of spatial sound technology, it was important to create the sensation of space and movement rather than extreme accuracy of spatial audio field reproduction.

The final invited presentation was given by Piotr Majdak, who spoke about “sound localization beyond the horizontal plane.” Majdak spoke about localization mechanisms in humans, starting from the early work by Lord Rayleigh to more modern theories and methods. Binaural methods and head-related transfer functions (HRTFs) were the subject of much discussion during the conference, particularly the need for individualized HRTF measurements; Majdak discussed the search for a metric for evaluation of HRTFs and detailed work that had considered the “spectral difference” between a pair of measurements.

that made it relatively easy for the microphone to be moved around to auditorium, as well as providing a source of entertainment as the soft box was flung across the room.

Jonty Harrison’s keynote talk on the second day was a good example of

The three invited presentations served as a good background for the detailed investigations of the substantial paper and poster program, which featured 24 presentations and 16 posters.

SPATIAL SOUND TECHNIQUES

Day one of the conference featured two paper sessions on spatial sound techniques,”chaired by Sakari Tervo (Aalto University).

Philippe-Aubert Gauthier from the University of Sherbrooke spoke about the benefits of accurate simulation of spatial sound fields when investigating safety in industrial environments and the need for nondestructive methods for capturing such environments. He introduced a combination of spot microphones and microphone arrays that were used to separate foreground and background sounds. Simulation results were promising and the work was due to be continued with real recordings reproduced over wave field synthesis (WFS). The following paper, presented by Symeon Delikaris-Manias of Aalto University, also considered a method for separating acoustic sources, introducing modifications to the cross pattern coherence algorithm (CroPaC), which had been demonstrated at Aalto the previous day. The next two papers concentrated on methods for reproducing sound fields. Philip Coleman (University of Surrey) introduced planarity panning as a way of producing a spatial sound field. This method focuses sound energy on a target zone while also constraining the angle of arrival and was shown to accurately position five virtual loudspeakers. Naomichi Yanagidate (Universities of Southampton and Chuo) compared methods for generating personal sound zones in a car cabin, proposing a method based on acoustic contrast control that also constrained the variance in sound pressure throughout the target zone and presenting simulation results.



Philip Jackson demonstrates his catching skills to ask a question using the Catchbox microphone.



The conference attracted 130 participants from 25 countries.



The audience listens intently during a presentation.

The session continued in the afternoon with two papers about HRTFs. Bosun Xie (South China University of Technology) presented a method for simplifying the personalization of HRTFs, analyzing similarity using a Wavelet decomposition, and Dinesh Manocha (University of North Carolina) presented a fast method for computing personalized HRTFs taking only 20 minutes on a desktop computer.

The spatial sound techniques session continued on day two, with the first paper presented by Andreas Franck (Fraunhofer), who spoke about an efficient frequency-domain crossfading method for interpolation of HRTFs, which is required for binaural synthesis of dynamic acoustic scenes. The following papers focused on loudspeaker-based reproduction methods, with Davide Scaini (Barcelona Media) and Tobias Weller (Macquarie University) talking about ambisonics followed by another presentation from the University of Sherbrooke by Anthony Bolduc who used WFS for sound field reproduction of vibroacoustic models. The last paper of the spatial sound techniques sessions was presented by Mikko-Ville Laitinen (Aalto University), who proposed a modified amplitude-panning law with a frequency-dependent gain normalization factor for ensuring the constant perceived loudness of a panned source, generating many questions from the interested audience.

SPATIAL SOUND ENGINEERING

The spatial sound engineering session contained four papers with a range of technical and operational outlooks. Panagiotis Charalampous (Cyprus University of Technology) compared various tree-traversal algorithms, which are an integral part of geometrical acoustic methods of sound propagation simulation. Results suggested that best-first methods are good for real-time applications, while depth-first methods are good for offline. Akio Ando (University of Toyama) gave an in-depth account of Makita's 1962 theory of sound localization and built upon this to derive a model of sound localization in multiloudspeaker reproduction. Jose J. Lopez (University of Valencia) presented a more application-based paper, introducing a system for teleconferencing over a mobile phone that gave the user various spatialization options in order to improve clarity and immersion in a conference call. Finally, Francis Stevens (York University) presented a method for capturing spatial impulse responses, comparing between swept-sine and starter-pistol excitation methods and suggesting the benefits of each.

SPATIAL SOUND PSYCHOPHYSICS

The final two paper sessions were titled Spatial Sound Psychophysics. Presentations delved into a range of issues regarding perception of spatial sound. The first paper, presented by Alexander Adami (International Audio Laboratories, Erlangen), considered down-mixing. A MUSHRA test was used to validate the performance of a new method in which coherent signal parts are identified and suppressed in one channel. The method was found to be preferred when results were aggregated, although there were some pronounced program dependencies. Olli Santala (Aalto University) presented some detailed

results of listening tests performed with pulse sequences distributed over an array of loudspeakers, suggesting that spatio-temporal hearing resolution is surprisingly low. Hagen Wierstorf (TU Berlin) investigated coloration in various WFS arrangements, again stressing the importance of timbral quality in a spatial audio system. Michael Scoeffler (International Audio Laboratories, Erlangen) looked into the influence of replay systems and preference for the audio material on the overall listening experience, and grouped listeners into two categories: "song likers" and "multichannel likers." The last paper before lunch saw a return to HRTFs with Kyla McMullen (Clemson University) presenting results about the consistency of selection of HRTFs over time; listeners were found to be fairly consistent in their selections, leading to a need to study similarities between selected HRTFs.

Frank Melchior (BBC) kicked off the final session, carefully outlining the statistical analysis undertaken on the results of an experiment looking at the plausibility of binaural synthesis with non-individualized HRTFs. The listeners found it difficult to determine the headphone simulation from the loudspeaker reproduction, although a small statistical difference was observed. Riitta Väänänen (Nokia) introduced a voice-guided navigation system for pedestrians, using a mobile phone and an



Frank Melchior of the BBC presents a paper in the spatial sound psychophysics session.

augmented reality audio headset. In user experience testing, the device was found to help participants reach their destination more easily as well as interact with the environment, device, or other people. Marko Takanen (Aalto University) used a binaural auditory model to predict localization error and coloration artifacts in simulated WFS and ambisonics systems, finding a good correlation with previously collected perceptual results. The last paper of the conference was presented by Stephan Werner (Technical University of Ilmenau), who investigated the effect of contextual factors on the perceived externalization of binaural synthesis, suggesting that externalization is improved by visibility of a congruent listening environment.

POSTERS

The sixteen posters were presented over a buffet dinner on the first evening, allowing ample time and space for discussion. Before the start of the session, each author was asked to give a one minute overview of the content of their poster. There must have been some budding soap opera writers among the authors as cliffhangers were used to draw in the crowds.

A number of the posters focussed on binaural technologies: Jesper Udesen (GN-Resound) echoed Stephan Werner's finding that the presence of visual cues effects externalization in binaural synthesis; Catarina Mendonça showed similar results for sound localization with nonindividualized HRTFs; Chengyun Zhang introduced a system for dynamic binaural rendering of 5.1 surround sound; and Tomi Huttunen and colleagues from various institutions presented a comparison of methods for rapidly generating individualized HRTFs.

Crossmodel interactions were also explored by Michele Geronazzo, who looked at audio-haptic exploration of virtual objects, and Samuel Moulin, who investigated audio-visual congruence in virtual environments created with WFS.

Neofytos Kaplanis (Bang & Olufsen) presented a detailed literature review of attributes for evaluation of reverberation in small rooms, along with a synthesized overview of the fundamental attributes and their subcategories.

The remaining posters covered a wide range of different topics including microphone array capture and beamforming, different sound field rendering methods, spatial room impulse responses, and surround sound encoding.

WORKSHOPS

The technical program also featured two workshops in the form of chaired panel discussions with time for questions from the audience.

The first workshop was entitled "Spatial Audio in Broadcasting and Music: Where Are We Going?" Florian Camerer (Austrian Broadcasting) chaired the discussion, with Bosse Ternström (Swedish Radio), Samuli Liikanen (Finnish Broadcasting Company), and Wilfried Van Baelen (Auro Technologies) making up the panel. Florian opened proceedings with some facts about Finland before posing the questions "where have we come from?" and "where are we going?" with spatial audio in broadcasting, asking the panelists about their first experience with surround sound. There was generally a great deal of agreement between the panelists and a number of general points stood out. Audio professionals are keen to work in new spatial formats as

technology becomes available, but it is normally challenging as there is no increase in budget to make multiple mixes or produce for a higher number of loudspeakers. The panelists also agreed that 5.1 surround is currently the most common format; while Ternström commented that 5.1 content is always preferable to stereo, Van Baelen noted that 5.1 is considerably less good than true 3D sound. An interesting question-and-answer session followed the culmination of the workshop, with many contributions from members of the audience. Tapio Lokki started proceedings by asking how many members of the audience had a properly set up 5.1 system at home; this was in the region of 10%, although there was some disagreement that 5.1 did not have a wide user base. Further questions considered binaural audio with head-tracking,

broadcasting for car audio, and distribution over the internet, as well as a comment from Angelo Farina on the current direction of spatial audio broadcasting in Italy.

The second workshop focussed on 3D audio quality evaluation. Andreas Silzle (Fraunhofer IIS) acted as chair, while Poppy Crum (Dolby) and Nick Zacharov (DELTA) made up the panel. Silzle started by introducing the aim of 3D audio evaluation: to collect describing parameters of the audio signal processing chain in order to optimize algorithms. He went on to look at the link between quality elements (the physical domain) and quality features or attributes (the perceptual domain). Silzle summarized test methodologies designed to establish links between the two domains, and talked briefly about efforts toward standardization. Nick Zacharov went into greater detail about the various levels of evaluation between objective and subjective using an updated version of the "filter model." He further defined the concept of attributes, suggesting that a word becomes an "attribute" when its function can be statistically demonstrated, and spoke about the statistical tools available for evaluating both attributes and

assessors. Zacharov also spoke about another key part of the sensory evaluation process: the test subjects. He suggested that an expert assessor is one who can demonstrate repeatability, discrimination, and agreement. The presentation finished with the assertion that many studies have been performed for the purposes of attribute elicitation and called for a movement toward agreement in the community about the scales that are used. Poppy Crum detailed some test methodologies in use at Dolby, emphasizing the importance of standardized tests but also the need to devise new methodologies where appropriate and suggesting that it is important to correctly report experimental procedures rather than just falling back on the standards where they have



Delegates discuss Angelo Farina's (center) poster, "Spatial Sound Recording with Dense Microphone Arrays."



Florian Camerer (right) introduces the panelists in the spatial audio broadcasting workshop, from left, Wilfried Van Baelen, Samuli Liikanen, and Bosse Ternström.

not been followed accurately. It was clear from the experience at Dolby that expectation moderates listener judgments—this is another theme that was seen in papers throughout the conference. Again, a lively discussion followed the workshop with questions about agreement in attributes where listeners don't necessarily agree on the listening experience, the concept of naturalness versus plausibility, and listener expectation.

The two workshops left delegates with a lot to talk about during coffee breaks throughout the conference and think about in terms of their research and operational practice.

SOCIAL EVENTS

As well as an interesting and diverse technical program, the conference organizers treated participants to two enjoyable social events and an informative tour of the venue.

The afternoon of the first day featured a break for outdoor activities. Conference attendees were split into teams and led off to various activity stations, where friendly competition provided a lighthearted way for participants to get to know each other better. Activities included sheet volleyball, throwing frisbees at trees, horseshoe throwing, knot tying, and blow-dart shooting (the paper session chairs surely took interest in this as a potential method of stopping presen-

ters from running over their allotted time). As the events drew to a close, the rain that had been threatening finally appeared and forced a retreat into a very small tent, where the overall winners were presented with prizes and coffee and pastries were served. A break for exercise and fresh air was great for keeping everyone energized before the afternoon continued with more papers and a workshop.

The conference banquet was held on the second night at Restaurant Botta, a short walk from the conference venue. Following a tour of the Helsinki Music Centre, the banquet kicked off with a performance from Pink Twins, a Helsinki-based duo comprising brothers Juha and Vesa Vehviläinen, who perform audio-visual “improvised digital soundscapes.” The performance gave something for delegates to think about and discuss after Jonty Harrison's invited lecture earlier in the day; while the music may not have been to everyone's taste, it was certainly an immersive and entertaining spatial audio experience. Pink Twins returned for a second performance after a delicious three-course meal, and the socializing continued into the evening.

SUMMARY

Overall, the conference was very well organized, smoothly run, and of immense value to all participants. The conference committee must be commended on putting together a diverse and interesting program, as well as fostering a productive environment in a desirable setting with engaging entertainment throughout. It is clear that the area of spatial audio is flourishing in research and in industry, and that the AES is at the forefront of technical and operational advances in this area. Some key points that came up throughout the conference included the improvement of timbral quality within spatial audio systems, the proliferation of binaural audio and particularly individualized HRTFs, and the need for careful and scientific perceptual evaluation and modeling. Attendees were treated to numerous interesting demonstrations and given ample opportunity for networking and discussion with new and old contacts. The conference sponsors should be thanked for their support: Genelec, Nokia, Neumann, Auro, and Morrow Sound (who ran a demonstration during many of the coffee breaks), as well as Aalto University and the Sibelius Academy. The AES will look forward to another visit to Finland in the years to come.



Delegates in the concert hall during the Helsinki Music Centre tour



The conference banquet at Restaurant Botta



Delegates get competitive with a game of sheet volleyball at the outdoor social event.

Editor's note: a USB drive or downloadable PDF of the conference papers can be purchased online at www.aes.org/publications/conf.cfm