



Virtual, Synthetic, and Entertainment Audio

AES 22ND INTERNATIONAL CONFERENCE

Espoo, Finland
June 15-17, 2002



Jyri Huopaniemi (left) and Nick Zacharov, conference cochairs



The south of Finland in the middle of summer is quite different from north-of-the-Arctic-Circle Rovaniemi, the site of the AES 16th International Conference, in April 1999. During the AES 22nd International Conference in Espoo in June there were warm sea breezes off the Gulf of Finland instead of snowdrifts; instead of huskies and snowmobiles there were sailboats and catamarans nearby. The conference organizing team, however, undeterred by their earlier experience of running an AES conference, were familiar faces drawn chiefly from the AES Finnish Section.

The HUT (Helsinki University of Technology) acoustics laboratory of Professor Matti Karjalainen, technical chair of the conference, continues to spawn enthusiastic and high-quality researchers, not least among these the two energetic conference cochairs, Jyri Huopaniemi and Nick Zacharov. They created a highly topical theme for this event, aiming to bring together scientists working in the related fields of virtual, synthetic, and entertainment audio for a peer-reviewed conference of exceptionally good quality. Unlike most AES conferences, they decided to require authors to propose full papers instead of just abstracts. Papers Chair Vesa Välimäki arranged for anonymous reviewers to read and select the most suitable, offer comments, and suggest revisions before

publication. This, they believed, would be a key factor in increasing the quality of the conference, and by all accounts this bold move was rewarded. The proceedings contain 47 papers selected from 57 submissions, including four invited papers from key experts in the field—Jens Blauert, Jürgen Herre, Xavier Rodet, and Peter Svensson.

Held in the futuristic-looking computer science building at HUT in Espoo, a major city just across the water from Helsinki, the conference ran June 15–17 and consisted of papers sessions, posters, demonstrations, and social events. Additional support was provided by sponsors Bang & Olufsen, Genelec, HUT, Nokia, and Yamaha (F-Music). The conference provided ample opportunities for delegates from 23 countries to benefit from each other's knowledge and expertise. The theme of the conference emphasized vital new areas of research concerned with audio creation and processing in the artificial rather than the natural sound domain. Here one could see the future of the audio industry leaping ahead before one's eyes into virtual worlds, augmented reality, synthetic 3-D environments, and new conjunctions of art, cognitive science, and engineering.

OPENING DINNER

On the Friday evening before the start of the conference, delegates were treated to an excellent buffet dinner at ➔



INVITED AUTHORS



Xavier Rodet



Jürgen Herre



Peter Svensson



Jens Blauert

Nokia Research Center in Helsinki. The dinner followed an introduction to the conference during which a novel musical instrument, the electric kantele, was described and played. This traditional Finnish stringed instrument has been electrified as part of a national project to promote and preserve it. The inventor of the electric kantele and his young son performed on different instruments to show how the kantele can be adapted to rock music as well as traditional music.

VIRTUAL AND AUGMENTED REALITY

Virtual worlds require virtual acoustics, and the ways in which computers can be used to model virtual acoustic spaces were admirably explained in an invited paper by Peter Svensson and Ulf Kristiansen of the Norwegian University of Science and Technology. Svensson explained how different methods can be used to simulate reflections from room surfaces of different types, outlining the relative merits of these as well as approaches to auralization—the creation of a rendered audible example of sound in the modeled space.

Also in this session Olivier Delerue from IRCAM in France described work undertaken in a European project called Listen, concerned with the artificial enhancement of natural auditory environments. He explained how virtual sound scenes can be used in museum spaces to modify the experience of the visitors by tracking their positions and rendering binaural signals, such as audio enhancements of paintings, to their earphones. A listener can approach a painting and hear related auditory stimuli that appear to be coming from the painting.

Another example of augmented reality, in which a form of hybrid sound reproduction integrates sound reproduced over loudspeakers with that reproduced through earphones, was described by Christian Müller-Tomfelde of Fraunhofer. This produces a public signal heard by everyone in the environment and a private signal heard only by the individual wearing headphones. Natural localization, for example, could be augmented by artificial room impression. Miniature hearing-aid-style earphones were discussed as a future practical tool that might be used to reproduce the required signals, coupled wirelessly by something like Bluetooth to a mobile computing device as a means of avoiding the need for obtrusive wired headphones. The concept of a science-fiction-like auditory implant and receiver can be envisaged as a step further from this, by which natural listening could be coupled with artificial reproduction of remotely transmitted signals.

SOUND SYNTHESIS

Xavier Rodet of IRCAM is a name well known to experts and novices alike in the field of sound synthesis. Among the highlights of this session, his invited paper “Present State and Future Challenges of Synthesis and Processing of the Singing Voice” provided an illuminating review of the field with entertaining examples of different approaches used over the years. Showing examples of both solo voice and choral singing synthesis, Rodet explained the need for a complex interaction between science, musical creativity, and engineering, a theme that was to recur many times throughout the conference both in informal discussions and formal presentations. Naturalness of the synthetic creation was held to be an important goal by many, although some challenged the need for artificial creations to be natural, proposing that perhaps some sounds are only perceived to be unnatural because we are unfamiliar with them.

“Modeling Bill’s Gait: Analysis and Parametric Synthesis of Walking Sounds” was the tongue-in-cheek title of the presentation by another of the well-known figures in music synthesis, Perry Cook of Princeton University. In a rapid-fire and captivating talk Cook showed how a parametric synthesis approach called PhOLIE (Physically Oriented Library of Interactive Effects) can be used to model walking sounds like footsteps on gravel, for use in movie sound effects. He developed a novel, hand-held control box interfaced to his laptop computer that enabled various parameters of the synthesizer to be adjusted so as to alter the nature of the crunching sounds resulting from the statistical modeling of virtual particles rattling around in a virtual com- ➔

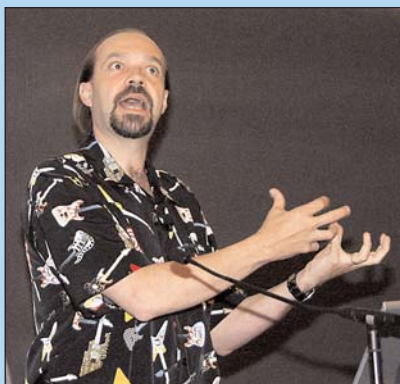


partment or against each other. Sounds such as that of maracas can also be created by shaking the device because it contains an accelerometer to detect motion. He explained the value of the approach in efficient sound creation for games, where PCM samples might require too much memory or processing time.

In one of the demo rooms running during the conference Mikael Laurson of the Sibelius Academy together with Vesa Välimäki and Cumhur Erkut of HUT showed a novel approach to the production of virtual acoustic guitar music. This used a physical synthesis model to generate the guitar sound and an original score-based interface for controlling the compositional input.

3-D AUDIO TECHNOLOGIES

A key factor in the creation and analysis of artificial environments is spatial-representation techniques. HRTF (head-related transfer function) estimation from head models and interpolation in moving sound were among the themes developed in papers during this session. There was also an interesting paper from Ole Kirkeby of Nokia on a simple stereo-widening network for allowing reproduction of loudspeaker stereo material over headphones. Using basic crossfeed-with-delay processing to simulate the interaural crosstalk that takes place in loudspeaker listening, the aim was to get the stereo image out of the head. In fact during the conference a live subjective experiment was conducted, comparing different stereo-enhancement algorithms for headphone listening, all designed in various ways to improve the reproduction of loudspeaker stereo over headphones. Unfortunately, as Gaetan Lorho of Nokia explained later in the conference in his paper on a round robin test using such systems, none of a range of so-called enhancement algorithms succeeded in being rated more highly on average than an unprocessed stereo signal, perhaps because of listener familiarity. Some discussion ensued between the delegates about the reasons why similar circuits had not proved more popular over the years, and it was agreed that they had possibly arrived before their time during an era when headphone use was not so widespread.



Author Perry Cook

Stadium-style seating in lecture hall provided excellent visibility.



Carlos Avendano's paper on frequency domain techniques for 2-to-5 channel upmixing provided a fascinating insight into what can be done to separate ambience information from discretely panned sources in stereo signals. He demonstrated fine control over an algorithm that could separate ambience to the rear loudspeakers of a 5-channel mix as well as repanning individual amplitude-panned sources in the front image. The ambience extraction process even enabled reverberation levels to be increased or reduced overall in the mix.

AUDIO CODING TECHNIQUES

Jürgen Herre of Fraunhofer, a leading figure in the field of low bit-rate audio coding, presented the invited paper "Audio Coding—An All-Round Entertainment Technology." ➔



Delegates check in at registration desk.



Friday evening welcome reception at Nokia Research Center

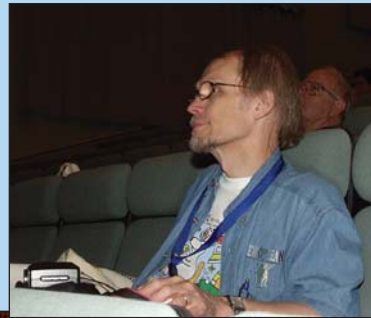


Vesa Välimäki,
papers chair



Tapio Lokki, venue officer

Matti Karjalainen,
technical chair



Concentrating primarily on MPEG, he showed how recent extensions to the standard enabled one to recover the HF content of lower sampling rate material using limited helper information transmitted alongside the main audio. The improvement in quality was quite noticeable in 48-kbit/s stereo material compared with basic MP3 coding. He pointed out that there was little prospect of further noticeable savings in the bit rate for transparent-quality coding—128 kbit/s for stereo using MPEG 2 AAC was still the best that could be achieved. Nonetheless, good quality coding still has further potential for bit-rate savings. With regard to audio transmission over wireless connections, he pointed out that the robustness of the channel (in fact, the lack of it) is the main problem, so audio coding schemes need to build in extra robustness of their own to accommodate this, independent of the transport stream. This is something that can be accommodated within MPEG-4.

Michael Goodwin of Creative Advanced Technology Center proceeded to explain ways in which coded audio material can be postprocessed in computing devices, since it is likely that coded audio will need to coexist with an increasing range of processes such as spatialization, mixing, and effects. He showed that some of this processing can be achieved in the so-called compressed domain by unpacking the coded audio data and processing it in the frequency domain before transforming the resulting mixed audio signals back to the time domain. In this way fewer transforms might be needed than the number of source signals originally received. Parametric coding and resynthesis techniques that are only just beginning to be exploited provide further opportunities for savings in computational complexity at the decoder.

A remarkable example of a 3-D virtual environment called EVE (Experimental Virtual Environment) was demonstrated by Lauri Savioja and his team from HUT. A Silicon Graphics Onyx computer the size of a small refrigerator with 2 Gbytes of RAM drove a number of projectors to create a 3-D rendering of virtual spaces within a small cube, provided the viewer wore suitable shutter glasses. The viewer could navigate around a virtual lecture theater or rove around the inside of a chemical molecule after getting over the initial sense of spatial disorientation or vertigo. A VBAP (vector-base amplitude-panning) spatialization system with



Consulting at registration desk: from left, Martti Rahkila, webmaster and IT officer; Roger Furness, AES executive director; Juha Merimaa, secretary; and David Isherwood, publications officer.

14 loudspeakers had also been installed for audio rendering, enabling viewers to direct the position of a virtual ball in the space, with audio that followed it.

POSTER SESSION

A lively poster session was conducted on Sunday afternoon. Authors presented their posters to describe their research and provide small demonstrations, made possible in many cases through the convenience of laptop computers and headphones. A number of interesting crossover projects could be observed involving artistic and technical concepts, including Andrew Paterson's "Stratigraphical Sound in 4-D Space," which provides a novel archaeological approach to the use of layered strata in the authoring of sounds for different phases of an interactive audio-visual experience.

SUBJECTIVE AND OBJECTIVE EVALUATION

The human receiver is the ultimate judge of the quality of any audio system, and the papers in this session dealt with experiments that involved subjective evaluation of sound quality or spatialization. The paper by Matti Gröhn and Tapio Lokki of HUT on static and dynamic sound source localization in a virtual room dealt with the ability of subjects to track a moving sound source that traced various shapes in the virtual 3-D environment; this concept was discussed earlier in the description of the EVE demonstration. Using the VBAP algorithm they found that there was



Carlos Avendano gives demo on frequency domain techniques.



Numerous, high-quality papers on display at Sunday afternoon's poster session offer delegates more food for thought.

a tendency for the auditory judgments of listeners to be elevated compared with the visual stimulus; also the trajectories of sources tended to be pulled toward the loudspeaker positions. They found that localization blur was greater with panned sources than when originating from a single loudspeaker, which is not unexpected.

COMPUTATIONAL AUDITORY SCENE ANALYSIS

Arguably the most distinguished participant of the 22nd Conference was Professor Jens Blauert of the Institute for Communication Acoustics. His work has spawned numerous Ph.D. theses as well as his own masterwork, *Spatial Hearing*. Blauert's invited paper, "Instrumental Analysis and Synthesis of Auditory Scenes: Communication Acoustics," was an inspiring evaluation of the current state of acoustics as it relates to information technology. At the end of the 1800s Lord Rayleigh's exposition of the physical laws of acoustics was considered complete, but then the vacuum triode came along and audio could be amplified, bringing with it a whole new field of electroacoustics. Then computers came along and changed things again. Blauert was the first professor in his laboratory to have a computer, and many of his colleagues could not understand why he would want one, but his foresight paid off.

Computational auditory scene analysis (CASA) is said to be the analysis equivalent of virtual reality (VR) which is concerned with synthesis of scenes. Both deal with a parametric version of the world. CASA requires cognitive models of human perceptual processes, and ➡



Delegates enjoy refreshments at coffee break.



Excellent lunches were enjoyed throughout conference at Helsinki University of Technology.



At banquet: clockwise from above, AES President Garry Margolis (inset) and Roger Furness compliment conference committee; band called Underground Hammers provide musical entertainment; Nick Zacharov and Jyri Huopaniemi listen as Martti Rahkila tells delegates that conference planning racked up tens of thousands of email messages.

audio signals, such as the spatial qualities of early and late reverberation and the beat of music.

SOCIAL EVENTS

The Finns know how to enjoy themselves, and the conference organizers did their best to ensure that delegates did too. The Finnish Experience, anticipated by some with trepidation, proved to be a thoroughly enjoyable evening for all. It started with a concert given by students from the Sibelius Academy. They played a variety of Finnish music of different styles and with different combinations of players who were joined by a singer for some Sibelius songs. The environment for this was the relatively new recital hall of the Helsinki Conservatory,



whose acoustic designer, Henryk Møller, was on hand to describe how he had created hard diffusing surfaces molded out of gypsum that also had an interesting visual appearance. The relatively high reverberation time and lateral efficiency of the hall gave it a pleasing spaciousness that enhanced the delightful playing of the enthusiastic and talented musicians.

The clubhouse of the Finnish Sauna Society on the shore of the Gulf of Finland was the locale for the remainder of the evening. Delegates signed up for one of two sessions: the first group using international rules wore

the audio world needs to embrace cognitive engineering if it is to move forward. Referring to the problem of integrating psychology with engineering, Blauert pointed out that in the liberal arts you are the king of the hill if you find a new problem, whereas in engineering the goal is solving the problem. We somehow need to bridge the divide between these paradigms.

The remainder of the session on CASA was concerned primarily with issues of blind identification of features in

swimming suits, while later in the evening the second group of sauna veterans went “au naturelle” according to Finnish rules. Consisting of three smoke saunas, two wood saunas, and a number of more modern systems, the clubhouse provided an ideal venue for the assembled company to experience one of the great pleasures of Finnish life. After allowing the penetrating heat of the sauna to soothe and relax aching muscles, delegates could take a bracing swim in the frigid Gulf of Finland. The food and drink of an outdoor



Delegates enjoy toast on boat ride to Suomenlinna Island.



Finnish Experience: after exhilarating saunas and refreshing swim in Gulf of Finland, delegates enjoy barbecue at Finnish Sauna Society.

barbeque provided abundant nourishment for repeated sauna-swims.

On the following evening, the team arranged a boat trip to the nearby island of Suomenlinna. A representative of the city of Espoo thanked the AES for its choice of the city for the conference. The restaurant in the fort on the island proved an excellent venue for the conference banquet and the subsequent entertainment, which was provided by an interesting band called Underground Hammers. After a sumptuous dinner, Zacharov and Huopaniemi praised the numerous crew members and the other members of the conference committee: Juha Merimaa, secretary; David Isherwood, publications officer; Martti Rahkila, webmaster and IT officer; Tapio Lokki, venue officer; Juha Backman, sponsors coordinator; and Kari Jaksola, treasurer. They thanked the delegates, the authors, invited speakers, and AES officers for their support.



Musical highlight of conference was concert at Helsinki Conservatory by students from Sibelius Academy.

As Jens Blauert implied at the end of his invited paper, the topics raised at the conference and the range of problems yet to be addressed in the integration of cognitive science, information technology, art, audio engineering, and acoustics prove that there is a long and interesting future ahead of us.

A CD-ROM and a printed version of The Proceedings of the AES 22nd International Conference are available for order on www.aes.org, or from any AES office.

Entire team was recognized for producing outstanding conference: from left, Tapio Lokki, David Isherwood, Jyri Huopaniemi, Mikka Tikander, Matti Karjalainen, Vesa Välimäki, Juha Merimaa, Nick Zacharov, Henri Penttinen, Martti Rahkila, Juha Backman, Marko Takanen, Matti Airas, Ville Pulkki, Paulo Esquef, Kalle Koivuniemi, and Hanna Järveläinen. (Kari Jaksola, Miikka Huttunen, Petri Mäntysalo, Sampo Vesa, and Raimo Lauren missed photo.)

