AES 30TH INTERNATIONAL CONFERENCE

Intelligent Audio Environments

Saariselkä, Finland
March 15–17, 2007

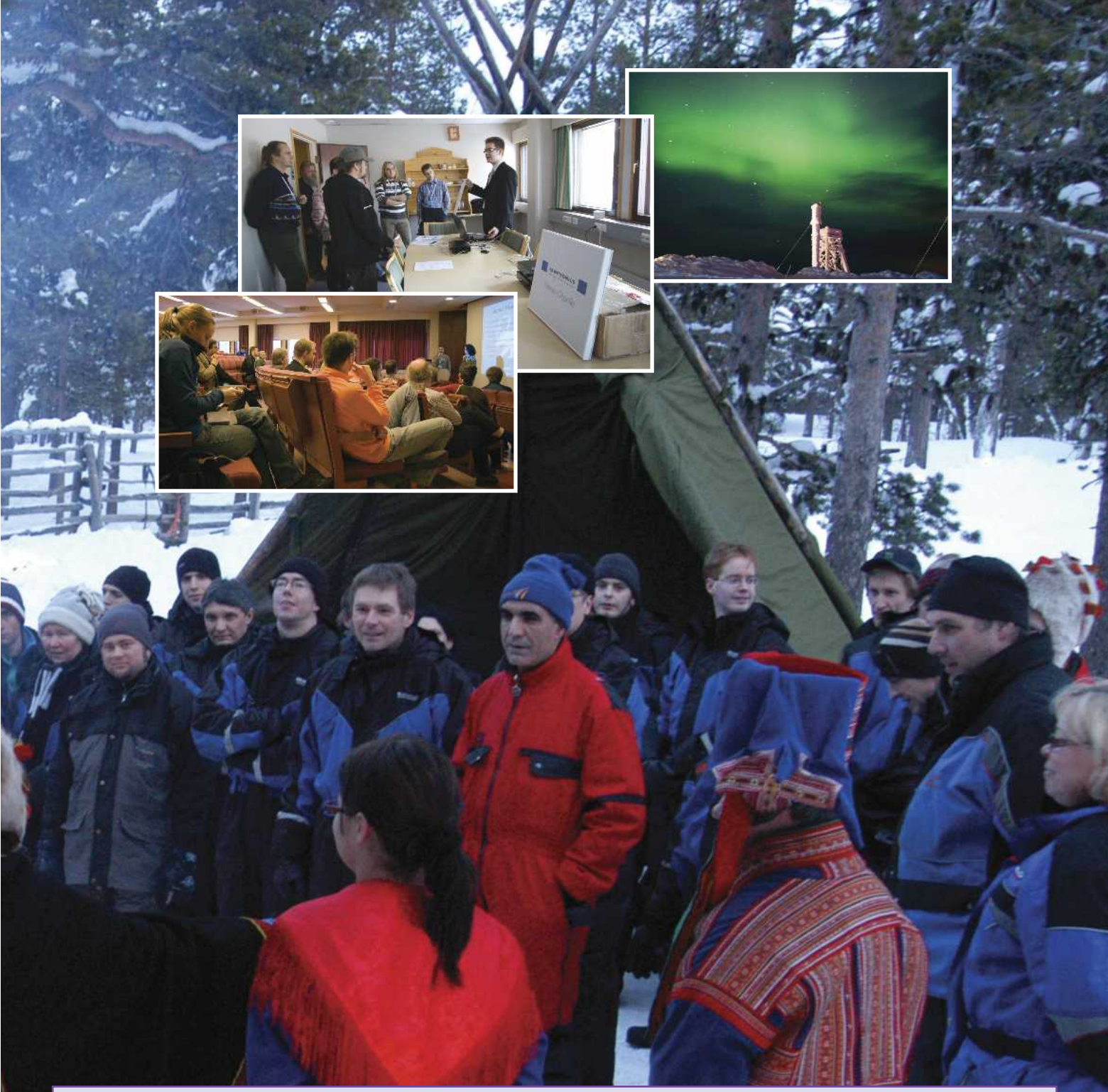
Juha Merimaa
conference cochair

Tapio Lokki
conference cochair

The AES 30th International Conference, *Intelligent Audio Environments*, was held in Saariselkä, Finland, 160 miles above the Arctic Circle. It was billed as the most northern AES conference yet, and this is a record it is likely to keep unless a future conference is arranged on an iceberg in the Arctic Ocean. Despite the snow and ice outside, the Finnish organizing committee, chaired by Juha Merimaa and Tapio Lokki, arranged a very warm welcome for the conference attendees, which included a "Finnish Experience" evening as seen in the photo on this page (more on this event later).

Ville Pulkki, papers chair, organized the papers according to a range of related aspects, such as automated separation of scene components, spatial rendering techniques, synthesis of virtual acoustical environments, and mobile audio applications.

The conference opened with a keynote speech by Jyri Huopaniemi of Nokia. From a background of mobile and com-



munication audio, he set out his views on the main developments in audio over the last ten years and used this to try to predict what will be important in the future. Among the things he mentioned were the increase of user-created content available via the Internet, the development and application of audio metadata (both manually and automatically created), various ways of interacting with audio devices, and immersive environments that stimulate multiple senses.

Huopaniemi observed that there are a number of challenges that may hinder the adoption of new technologies in mobile audio, not least the fact that additional processing will more rapidly drain the already limited battery power that is available. He finished by summarizing his key tips for future development: we should innovate rather than simply adapt current methods, and we should balance the push of technological advances with the pull of the wants and needs of consumers.

SOUND QUALITY

Attention then turned to an important aspect for some consumers, the perceived quality of sound. In Yong Choi presented a paper that considered methods to improve models of perceived sound quality, such as PEAQ. Currently, these models only consider a single channel of audio. He presented research that introduces binaural cues, such as variations in ➥

Authors

John Mourjopoulos
Invited Speaker

Christof Faller
Invited Speaker

Matti Karjalainen

Markus Vaalgamaa
Invited Speaker

Stephanie Bertet

Huseyin Hacihabiboglu

Jyri Huopaniemi
Invited Speaker

Katja Hoffmann
de Linares

Nicolas Tsingos

Jean-Marc Jot

Lukasz Litwic

interaural differences, to attempt to predict spatial degradations of compressed audio. This was followed by a paper presented by Ulrich Reiter that considers how tasks can distract the listener from listening critically. He conducted a subjective experiment where listeners were asked to make judgments of the reverb time of stimuli while remembering spoken numbers. He showed that as the difficulty of the memory task increased, the accuracy of the judgments of reverb time decreased. He suggested that this effect could be exploited by including interactive tasks to enable simplifications to be made to the synthesis of virtual acoustical environments. In doing so this should prevent the user from continuously monitoring the audio quality.

PERCEPTION AND CONTENT ANALYSIS

The next two sessions covered the perception and analysis of audio scenes, mostly with the aim of extracting meaningful features from audio signals. These features can then be used either to divide the scene into its component parts or to categorize and label the audio for later searching and rendering.

Toni Hirvonen presented a paper that investigates physiologically-inspired methods of detecting the interaural level difference. He tested this model by measuring a number of signals that had been used in previous subjective experiments into binaural masking level differences. He found that the new model matched the subjective results well. The perception and detection of the end of musical notes was considered by Russell Mason. Through a combination of subjective listening tests and objective analyses, he found that the addition of reverberation to anechoic music signals effectively extended the perceived duration of the notes and that the perceived offset time matched the point at which the signal dropped below 30 dB with respect to the peak level of each signal. The next paper, presented by Lukasz Litwic, attempts to separate the components of an auditory scene through the use of neural networks to provide a nonlinear combination of lower-level time–frequency cues. This method also allows the lower-level cues to be weighted by their relative importance.

The paper presented by Peter ➥

Coffee breaks provided time for networking and follow-up questions to authors.

Lennox took a philosophical and perceptually-oriented approach to creating auditory scenes. He started with an overview of the philosophy of our perception of acoustical environments and came to the conclusion that we need more perceptually relevant controls of spatial attributes, such as direct manipulation of trajectories and spaces, rather than relying on the traditional and somewhat unintuitive combination of pairwise panners and reverberation units.

Taking a musical point of view, Pedro Ponce de Leon presented research that attempts to identify saxophonists from recordings of their performances. Different musicians have different, recognizable playing styles. This research tried to mimic this categorization. Low-level internote and intranote features were calculated for the performances of three tunes by three saxophonists. By examining factors such as attack level, sustain duration, legato, spectral centroid, and other parameters, it was possible to discriminate between the three musicians with a high level of accuracy.

The final paper on content analysis was presented by Sampo Vesa. He described methods of clustering environmental sounds based on common features. He recorded the background noise of an office and used different algorithms to cluster the resulting sounds. Using a genetic algorithm to undertake the clustering resulted in good separation of the different sounds, mainly based on the differing temporal characteristics.

POSTERS

The posters session allowed for in-depth discussion of a range of subjects related to the conference topic. A number of posters considered the perception of acoustical environments, including consideration of the precedence effect of sounds arriving from the side of a listener, and acoustical measurement techniques aimed at minimal perceptual annoyance. For signal capture, one poster examined the use of a single rotating microphone for analyzing the acoustical environment, and at the other end of the signal chain another modeled the attenuation caused by in-ear headphones used for virtual reality rendering.

HIGH-RESOLUTION SPATIAL REPRODUCTION

The first paper on the second day was presented by Stephanie Bertet and concerned the use of higher order Ambisonics to position sounds around 12 loudspeakers on the horizontal plane. She described how listeners were asked to adjust the position of an acoustic marker to match a sound in a specific direction. The time taken to achieve this was used to evaluate the accuracy of the reproduction. It was found that it took less time to achieve the task with a fourth-order system compared to a first-order system and that binaural measurements showed that the range of interaural time differences created by the system was higher for the higher order systems.

The two following papers covered issues related to wavefield synthesis. Sascha Spors investigated the problems caused by the listener being too close to the loudspeaker array. Through simulation he showed that wavefield synthesis systems create side-effects in the nearfield. Khoa-Van Nguyen presented a paper that investigates the use of conventional loudspeakers and distributed mode (flat panel) loudspeakers for wavefield synthesis. One characteristic of flat-panel loudspeakers is that the off-axis response is uncorrelated compared to the on-axis response. A diffuse filter was created to simulate this effect on conventional loudspeakers. Then a subjective experiment was conducted to evaluate the effect that this has on the stimuli. He found that the main subjective effect was a change in the perceived distance, with the flat-panel loudspeakers having the furthest perceived distance and the unfiltered conventional loudspeakers having the closest distance.

FLEXIBLE REPRODUCTION

The next session covered reproduction systems that can adapt based on the characteristics of the source signals or the listening environment. Sunmin Kim presented a virtual surround system that adapts based on the position of the listener. He experimented with a number of methods to detect the position of the listener. He concluded that the best results can be obtained by including an ultrasonic transmitter in the remote control. This, when combined with the infrared transmitter, can be used to derive the position in terms of distance and angle.

The rendering of virtual auditory scenes, such as those used in computer games, usually makes use of a large number of individual recordings, positioning each individually. Nicolas Tsingos presented an alternative method, which involves recording natural soundscapes, separating these into point sources and background noise, and rendering these indi- ➥

AES Student Paper Competition

The AES has launched a new opportunity to recognize student members who author technical papers. The competition is based on the preprint manuscripts accepted for presentation at the technical programs of AES conventions.

Arijit Biswas (right), the first winner of the AES Student Paper Award, and Rob Maher, AES 121st Convention papers cochair

Nominees for the Student Paper Award are required to meet the following qualifications:

(a) The paper is accepted for presentation by the papers chairs at the convention.

(b) The first author is an AES student member when the work is conducted and the manuscript prepared.

(c) The student author's affiliation listed in the manuscript is an accredited educational institution.

(d) The student will present the lecture or poster presentation at the convention.

The nominated student papers will be judged by selected members of the AES *Journal* Review Board who are experts in the fields related to the topics of the individual papers.

The quality of the winning paper must meet the standard requirements of the *Journal* review process. No award will be given at a convention if none of the student papers are considered to meet this requirement. For instance, while the quality of the papers presented by students at the recent 122nd Convention in Vienna was very good, none met the requirements for *Journal* publication. Arijit Biswas was the first winner of the AES Student Paper Award (see photo). His (and coauthor's Albertus C. den Brinker) paper, "Perceptually Biased Linear Prediction," was published in the December 2006 (Vol. 54, No. 12) *Journal*.

If a student paper is selected to receive the award it will be announced during the convention, and the student-authored manuscript will be published in a timely manner in the ***Journal of the Audio Engineering Society***.

If you wish to enter the competition, be sure to check the Student Paper Award box when you load your proposal on the AES convention proposal-submission sites.




Demos included those by, clockwise from top left, Panphonics, Nicolas Tsingos, Peter Eastty of Oxford Digital, and Ville Pulkki's directional audio coding (DirAC).

vidually. He used a spaced array of microphones in which the positions were logged and entered into the analysis software by taking multiple photographs. By calculating the correlation between pairs of microphones, the positions of sound sources can be determined. The background noise was separated from this based on the assumption that this was stationary and spatially diffuse. The point sources were reproduced by extracting the related time–frequency segments recorded by the closest microphone and rendering this using an appropriate 3-D spatialization technique; the background noise was reproduced using first-order Ambisonics.

Christof Faller presented an invited paper in which he reviewed a range of matrix surround sound systems. He went through the history of spatial audio up to the early matrix surround systems, and he set out the requirements of a matrix surround sound system. He showed the limitations of passive matrix systems, especially in terms of accurately rendering signals to the rear loudspeakers, and showed a number of methods for actively extracting the ambient sound from the complete signal. His presentation was supported by a demonstration where he showed the limitations of passive systems and the performance enhancements possible through more complex analysis and rendering techniques.

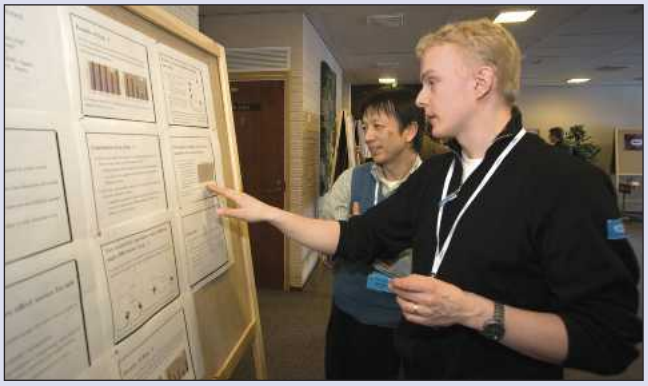
DEMONSTRATIONS

There were a number of other demonstrations at the conference that gave examples of cutting-edge audio technology. Ville Pulkki showed a teleconferencing system based on directional audio coding (DirAC, also see paper on this topic on page 503 and article on spatial audio on page 537). Using an array of four budget microphones, the system encodes the spatial characteristics of the soundfield in terms of the diffuseness of the signal and the direction of energy flow in each time–frequency region. This information can be sent together with a mono audio signal, and the spatial information can be used to render the soundfield at the receiving end. By processing the spatial rendering, the system can make use of spatial separation to avoid problems with annoying echoes and feedback.

Panphonics demonstrated their electrostatic loudspeaker

Tapio Lokki and Julia Turku greet attendees at the registration desk.

systems, which make use of signal processing to enhance the perceived sound. Two main components in the processing are methods to increase the perceived bass response and to widen the spatial image through the use of crosstalk cancellation methods. The latter enables the production of a wide stereo image from a relatively small single unit.



Poster sessions allowed authors to provide in-depth information on their research: clockwise from top left, Banu Gunel, Adrian Vasilache, and Xihong Wu.

A system was demonstrated by Carneal and colleagues that uses multiple microphone arrays to identify and track sound sources. Each array has its own processing unit that locates signals, which are then combined across a number of arrays to give a robust estimation of the positions and directions of motion of various sound sources.

A demonstration was given by Nicolas Tsingos that showed real-time audio rendering within a computer game environment. He showed how the system can automatically select and prioritize audio sources in a complex scene and how spatial rendering can be simplified by grouping sources based on the complexity of the scene as well as the relative positions of the sources and observer.

Other demonstrations included a room adaptation system by Genelec and a generalized audio up-mix system to create surround sound from 2-channel audio by Jean-Marc Jot.

PANEL DISCUSSION

Nick Zacharov organized a panel discussion that covered a wide range of topics, including the meaning of the conference title, *Intelligent Audio Environments*, the creation of metadata, and the differing needs of mobile technologies compared to conventional hi-fi consumer audio. A common thread was the need for scalable solutions to audio—to be able to create media that can be rendered on anything from a low-bandwidth headphone-based mobile device up to a high-quality multichannel reproduction system. Peter Eastty observed that the audio industry is being subsumed by the computer industry, with the view that audio is just another form of data and that we need to educate the computer industry to do audio processing properly.

Each panelist was asked to give his view of the future of audio. Christof Faller opined that closed, standardized systems block innovation and that we need to develop open and flexible systems that can adapt to the latest technology. Jean-Marc Jot expanded on this by asking for systems that can transport the correct settings from the author to the user, so that the listener does not have to do anything to set up the reproduction system. This prompted more discussion about the relative merits of open systems that can be updated versus standardized systems that are potentially more reliable and consistent. The discussion was summarized by Brian Katz, stating that there needs to be a better match between audio development and consumer needs.

Panel discussion: from left, moderator Nick Zacharov, Jean-Marc Jot, John Morjopoulos, Erlendur Karlsson, Brian Katz, Christof Faller, and Peter Eastty.

ROOM ACOUSTICS AND MODELING

The following session considered methods for simulating and modeling reverberant environments. The first two papers examined methods for synthesizing reverberation—one using a modeling technique and one using a high-level approximation. Dirk Schröder presented a system that calculates virtual

Saturday night's banquet included an evocative performance by internationally renowned accordionist Maria Kalaniemi, who let Ville Pulkki (left) jump in with an impromptu vocal performance.

acoustical environments in real time over a head-tracked binaural reproduction system. This uses a ray-tracing algorithm to give a more precise estimate of the characteristics of the late reverberation. Matti Karjalainen presented a paper that described the use of "velvet noise" (noise with a smoother amplitude envelope) to generate artificial reverberation, using listening tests to show that this gives better results than random impulses for a given impulse density.

Huseyin Hacihabiboglu discussed the use of digital waveguide meshes to simulate the acoustics of rooms. These are usually excited at a single point in the mesh, which acts effectively as an ideal omnidirectional point source. The presented research extended this by considering the directivity of the sound source and developing methods to take this into account in the model. The final paper, presented by Christian Uhle, considered methods of upmixing monophonic recordings to surround sound. He used non-negative matrix factorization to separate the direct sound from the ambience, and used these separated signals to create a 5.1 rendition.

VOICE OVER IP

Markus Vaalgamaa gave an invited presentation on voice over IP (VoIP) systems, starting with the history and discussing the benefits, challenges, and solutions for such a system. He laid out the problems caused by internet protocol (IP) data transfer, including variable latency and lost data packets. There are a number of methods that can be used to get around these problems, including time stretching data packets to attempt to conceal gaps in the audio. He also summarized research into the range of overall latency that can be tolerated without annoyance to the talkers.

SOCIAL EVENTS

The Finnish conferences are becoming renowned for their interesting and unusual social events, and this conference was no exception. On Thursday the committee arranged a "Finnish Experience" evening, where the delegates could experience

The Finnish Experience: On Friday evening attendees got to sample the culture of the indigenous Sami people and Lapland winter sports: clockwise from top left, Sami songstress, snowmobiling, a dinner of reindeer prepared and served by chefs in traditional costumes, huge split-log fire burned long and late into the night.

30th Conference Committee: from left, Lauri Savioja, Juha Merimaa, Julia Turku, Tapio Lokki, Ville Pulkki, Nick Zacharov, and Mikko Peltola.

reindeer rides, sledding, snowshoe walking, Sami music, and traditional Finnish food and drink.

The banquet was held on Friday evening in a mountain-top restaurant overlooking Saariselkä. The conference delegates dined on traditional Finnish dishes such as elk, snow grouse, and cloudberries while watching the sun set over the snowy landscape. Entertainment was provided by Maria Kalaniemi (see www.mariakalaniemi.com) a Finnish vocalist and virtuoso accordionist, with occasional vocal input from members of the conference committee. The committee even conspired with the gods of Lapland to organize a fabulous display of the Aurora Borealis after dinner, with the mountain providing a perfect viewpoint for the phenomena (see photo on page 519).

DIGITAL AUDIO DELIVERY

The first session of the final day of the conference covered methods of delivering the audio to the end user in a range of applications. Arnault Nagle presented a method to reduce the data in a VoIP teleconference system. By routing all the signals through a central bridge rather than individually connecting each user, a masking prediction can be used to only pass on the perceivable audio data from the sum of the signals, thus reducing the total amount of bandwidth.

Jukka Ahonen presented a paper that describes the teleconference system based on DirAC that was demonstrated by Ville Pulkki as previously mentioned. He described the system used in the demonstration, as well as steps taken in the practical realization of the system, such as comparative microphone measurements. Jean-Marc Jot presented a paper that discusses the problems of converting surround sound between different loudspeaker layouts and upmixing from 2-channel stereo to nonstandard loudspeaker layouts. He outlined an approach that calculates the energy vector in each time–frequency division and then reproduces this appropriately for the given loudspeaker layout. He also gave demonstrations of the performance of the system, comparing the interchannel separation of this system with conventional passive systems.

APPLICATIONS

With the large number of established audio formats for film, there is no room left for new formats, so Katja Hoffman de Linares examined methods to synchronize additional audio formats. The approach attempts this by tracking the analog optical audio track on the film, which gives reasonable success except during periods of silence. Antoine Tarault presented a paper that looks at the effect of 3-D audio on interactive location tasks. Listeners were required to find a light in a darkened room using a head-tracked remote camera and microphone system. The 3-D audio rendering system was varied to observe how this affects the time taken to complete the task. Tarault found that a binaural rendering was most effective, though there were still problems with elevation perception.

Tapio Lokki described a system that was created to add artificial reverberation to a lecture room that has a low natural reverberation time. He altered the characteristics of the reverberation, including changing the directional characteristics of the artificial reverberation. Listeners were asked to draw the perceived dimensions of the acoustical space, which showed that the perceived shape of the artificial acoustics can be varied effectively.

The final paper of the session was presented by Alexander Carot and Alain Renaud. They set out the criteria for different types of network music performance and then went into more detail about live bilateral interaction (two or more performers in different locations playing together over a network). They discussed the difficulties in this work, such as coping with latency issues and jitter, and mentioned methods that can be used to get around these problems.

The conference was concluded with an invited presentation by John Mourjopoulos. He talked about the limitations of digital wireless audio systems, starting with an overview of current wireless local area networks (WLANs). He followed this through to the possibility of a complete digital system (including digital transducers) that can adapt to the reproduction environment. His long-term view is that it will be possible to develop a system that can adapt to its environment, allowing the consumer to place loudspeakers wherever it is convenient, with the system configuring and optimizing itself for the positions and the acoustical environment. (The CD-ROM of conference papers can be purchased online at www.aes.org/publications/conf.cfm.)

Thanks were given to all the delegates, presenters, sponsors, the organizing committee, and the AES staff, all of whom contributed to a very educational and successful conference. It lived up to its advance billing as the “coolest” AES conference, not only because of its location 160 miles above the Arctic Circle in snow-covered Lapland, but also because of the truly unique social atmosphere created so artfully by the AES Finnish section.