



AES 40th INTERNATIONAL CONFERENCE

Spatial Audio: Sense the Sound of Space

NHK Science & Technology Research Laboratories & Tokyo University of the Arts Tokyo, Japan

October 8–10, 2010

Thursday, October 7

18:00–20:00

PRE-CONFERENCE SPECIAL WORKSHOP

Venue: *Tokyo University of the Arts*

Registration Fee: *Free*

Organizers: *AES Japan Student Section, AES 40th Conference Committee, and Tokyo University of the Arts*

Anatomy of 3-D technology

Chair: **Erisa Sato**, Chair of AES Japan Student Section / Tokyo University of the Arts

Panelists: *Sungyoung Kim*, Yamaha Corporation
Seigen Ono, Saidera Paradiso Ltd.

Faculty Adviser: *Toru Kamekawa*, Tokyo University of the Arts

In this workshop we will invite corporate members, manufacturers, and engineers from the areas of professional sound recording and videogame production, allowing students to deepen their understanding in the technology and R&D activities of 3-D audio and visual production in a Q&A setting.

FRIDAY, OCTOBER 8: ALL EVENTS TAKE PLACE AT NHK SCIENCE & TECHNOLOGY RESEARCH LABORATORIES

Friday, October 8

08:55

OPENING REMARKS

Opening Remarks

Friday, October 8

09:00–10:00

KEYNOTE ADDRESS 1

History and Current State of the War between the Two Main Approaches to Surround Sound: Discrete Loudspeaker Feeds versus Hierarchical Matrix

Keynote Speaker: **Angelo Farina**, University of Parma, Parma, Italy

This keynote speech reconstructs the history of the competition, clashing, and hybridization between two opposite concepts, which were proposed for creation of surround sound.

The first approach is the discrete method, which became feasible only when the storage media allowed for complete channel separation of more than 2 channels; in practice, with the advent of the DVD. Conceptually this approach is based on the paradigm “one microphone feeding one loudspeaker,” repeated as many times as the number of loudspeakers employed during playback. Of course, these “microphones” can either be real capsules properly placed and aimed, or the speaker feeds can be created as “virtual microphones,” the most common case is by amplitude-panning a number of dry mono signals obtained by closely-miking musicians and singers. The competitor of this approach was, since the beginning, the usage of a hierarchical system, in which the signals are first matrixed (or encoded), possibly employing a reduced number of channels employed for storage or transmission of the surround

mix, and later dematrixed (or decoded) recreating a number of speaker feeds, which generally can be larger than the number of transmitted channels. This can be very clever, but requires that the signals being matrixed and subsequently dematrixed are perfectly time-aligned (so the microphones should all be coincident).

Between these two extremes, a number of intermediate approaches were developed. The presentation will exploit advantages and disadvantages of the two concepts, providing an historical analysis of the past and an up-to-date information about current development and forthcoming research results.

The presentation will be accompanied by short sound samples prepared in 5.1 format, which will be played during the presentation, if the lecture room is equipped with a surround sound system.

Friday, October 8

10:00–11:30

PAPER SESSION 1: PERCEPTION AND EVALUATION OF SPATIAL AUDIO—PART 1

- 1-1 **Validation of a Simple Spherical Head Model as a Signal Capture Device for Head-Movement-Aware Prediction of Perceived Spatial Impression**—*Chungeun Kim, Russell Mason, Tim Brookes*, Institute of Sound Recording, University of Surrey, Guildford, Surrey, UK

In order to take head movement into account in objective evaluation of perceived spatial impression (including source direction), a suitable binaural capture device is required. A signal capture system was suggested that consisted of a head-sized sphere containing multiple pairs of microphones that, in comparison to a rotating head and torso simulator (HATS), has the potential for improved measurement speed and the capability to measure time varying systems, albeit at the expense of some accuracy. The error introduced by using a relatively simple sphere compared to a more physically accurate HATS was evaluated in terms of three binaural parameters related to perceived spatial impression—interaural time and level differences (ITD and ILD) and interaural cross-correlation coefficient (IACC). It was found that while the error in the IACC measurements was perceptually negligible, the differences in measured ITD and ILD values between the sphere and HATS were not perceptually negligible, although they were reduced slightly when the sphere was mounted on a torso. However, it was found that the sphere-with-torso could give accurate predictions of source location based on ITD and ILD, through the use of a look-up table created from known ITD-ILD-direction mappings. Therefore the validity of the sphere-with-torso as a potential head-movement-aware binaural signal capture device for perceptually relevant measurements of source direction (based on ITD and ILD) and spatial impression (based on IACC) was demonstrated.

- 1-2 **Effects of Individualized Headphone Correction on Front/Back Discrimination of Virtual Sound Sources Displayed Using Individualized Head Related Transfer Functions**—*Abhishek Guru, William Martens, Doheon Lee*, The University of Sydney, Sydney, NSW, Australia

Individualized Head Related Transfer Functions (HRTFs) were used to process brief noise bursts for a 2-interval forced choice (2IFC) front/back discrimination of virtual sound source locations presented via two models of headphones, frequency responses of which could be made nearly flat for each of 21 listeners using individualized headphone correction filters. In order to remove virtual source timbre as a cue for front/back discrimination, spectral centroid of sources processed using rearward HRTFs were manipulated so as to be more or less similar to that of source processed using frontward HRTFs. As this manipulation reduced front/back discrimination to chance levels for 12 out of 21 listeners, performance of 9 listeners showing “good discrimination” was analyzed separately. For these 9 listeners, the virtual sources presented using individualized headphone correction filters supported significantly better front/back discrimination rates than did virtual sources presented without correction to headphone responses.

1-3 Learning to Remediate Sound Localization in the Median Plane Using a Virtual Auditory Display—Kenji Ozawa, Tomohiro Sato, University of Yamanashi, Kofu, Japan

Previous studies have shown the efficacy of sound localization training with non-individualized head-related transfer functions (HRTFs) using a virtual auditory display (VAD) in the horizontal plane. The efficiency of training of sound localization in the upper part of the median plane was experimentally examined in this study. During each training session, noise stimuli for nine elevation angles were prepared, and each of them was presented five times in random order with feedback regarding the correct position. During test sessions, in contrast, music stimuli for ten elevation angles were tested. Just one training session was required to induce a significant learning effect with regard to remediation of sound localization, while a ceiling effect was observed by three sessions. Seven training sessions resulted in the persistence of learning efficacy for over one month. The efficiency of short periods of training will enable us to utilize a VAD with non-individualized HRTFs in various applications.

Friday, October 8

10:00 – 12:00

WORKSHOP 1

Measuring High-Quality Room Impulse Responses for Artistic Applications

Chair: **Wieslaw Woszczyk**, McGill University

Panelists: *Ralph Kessler*, Pinguin
Masahito Ikeda, Yamaha Corporation
Akio Takahashi, Yamaha Corporation

The workshop will focus on techniques for measuring high-quality impulse responses of architectural spaces that find applications in various artistic domains including music production and post-productions, recording, live sound reinforcement, and acoustic support of music performance. The topics will include the selection of sound sources, microphone systems, processing techniques, and techniques for capturing the height information. Examples of productions will illustrate the audible differences between different approaches taken in measuring rooms and reverberant spaces.

Friday, October 8

11:30–13:30

PAPER SESSION 2: SPATIAL RENDERING AND REPRODUCTION—PART 1

2-1 Some Recent Works On Head-Related Transfer Functions and Virtual Auditory Display in China—Bo-Sun Xie,¹ Xiao-Li Zhong,¹ Guang-Zheng Yu,¹ Shan-Qun Guan,² Dan Rao,¹ Zhi-Qiang Liang,¹ Cheng-Yun Zhang¹

¹South China University of Technology, Guangzhou, China

²Beijing University of Posts and Telecommunications, Beijing, China

Head-related transfer function and virtual auditory display are hot issues in research of acoustics, signal processing, and hearing, etc., and have been employed in a variety of applications. In recent years they have received increasing attention in China. This paper reviews the latest developments of head-related transfer function and virtual auditory display in China, especially works accomplished by our group.

2-2 Sound Field Reproduction by Wavefront Synthesis Using Directly Aligned Multi Point Control—Noriyoshi Kamado,¹ Haruhide Hokari,² Shoji Shimada,² Hiroshi Saruwatari,¹ Kiyohiro Shikano¹

¹Nara Institute of Science and Technology (NAIST), Nara, Japan

²Nagaoka University of Technology, Nagaoka, Niigata, Japan

In this paper we present a comparative study on directly aligned multi point controlled wavefront synthesis (DMCWS) and wave field synthesis (WFS) for the realization of a high-accuracy sound reproduction system, and the amplitude and phase characteristics of the wavefronts generated by DMCWS and WFS are assessed on the computer simulations and measurement in actual environments. First, the results of computer simulations revealed that the wavefront in DMCWS has wide applicability in both the spatial and frequency domains with small amplitude and phase errors, particularly above the spatial aliasing frequency in WFS. Next, we developed wavefront measurement system and measured a DMCWS wavefront with this system and proposed algorithm. The results of measurements clarified the effect of a reflected wave and the frequency characteristics of a loudspeaker. Also, DMCWS has wide applicability in frequency domains in actual environments. From these findings, we concluded the advantageousness of DMCWS compared with WFS.

2-3 Local Sound Field Synthesis by Virtual Secondary Sources—Sascha Spors, Jens Ahrens, Deutsche Telekom Laboratories, Technical University of Berlin, Berlin, Germany

Sound field synthesis techniques like Wave Field Synthesis and Higher-Order Ambisonics aim at the physical synthesis of a desired sound field over an extended listening area. However, for practical setups the accuracy up to which the desired sound field can be synthesized over an extended area is limited. For certain applications it is desirable to limit the spatial extent of the listening area in order to increase the accuracy within this limited region for a given loudspeaker arrangement. Local sound field synthesis aims at a higher accuracy within a local listening area. An approach to local sound field synthesis is presented that is based on the concept of using virtual loudspeakers that are placed more densely around the local listening area than the existing loudspeakers. The approach is illustrated using Wave Field Synthesis as an example.

2-4 Comparison of Higher Order Ambisonics and Wave Field Synthesis with Respect to Spatial Discretization Artifacts in Time Domain—Jens Ahrens, Hagen Wierstorf, Sascha Spors, Deutsche Telekom Laboratories, Technische Universität Berlin, Berlin, Germany

We present a time domain analysis and comparison of spatial discretization artifacts in near-field compensated higher order Ambisonics and wave field synthesis. Simulations of both methods on the same circular loudspeaker array are investigated and the results are interpreted in terms of fundamental psychoacoustical properties of the human auditory system, most notably the precedence effect. It can be shown that both methods exhibit fundamentally different properties regarding the synthesized first arriving wave fronts as well as additional correlated wave fronts (echoes). The properties of both types of wave fronts are a consequence of the combination of the spatial bandwidth of the loudspeaker driving function and the fact that a finite number of spatially discrete loudspeakers are employed.

Friday, October 8

12:00 – 13:30

WORKSHOP 2

**Space Builder: A Comprehensive Production Tool
for 22.2 Channel Sound Design**

Co-chairs: **Kimio Hamasaki**, NHK
Wieslaw Woszczyk, McGill University

Panelists: *Richard King*, McGill University
Doyuen Ko, McGill University
Brett Leonard, McGill University
Kentaro Matsui, NHK

The researchers from McGill University's Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) and the NHK Science and Technology Laboratories will present a convolution-based Space Builder developed for ambient sound design and flexible spatial processing applied in 22.2 channel audio productions. They will show hardware and software implementations, detailed system architecture with a dedicated GUI, and present mixes mastered using the system. The goal is to deliver a production tool that easily lends a high quality spatial treatment in three dimensions to recordings originally made in mono, stereo, or 5-channel surround, without height. The attendees will have a chance to try the Space Builder user interface and to hear the results.

Friday, October 8

14:30–17:30

**PAPER SESSION 3: SPATIAL RENDERING
AND REPRODUCTION—PART 2**

3-1 A Design Tool to Produce Optimized Ambisonic Decoders
—*David Moore*,¹ *Jonathan Wakefield*²

¹Glasgow Caledonian University, Glasgow, UK
²University of Huddersfield, Huddersfield, UK

This paper describes a tool for designing Ambisonic surround sound decoders. The tool is highly flexible and provides a decoder designer with powerful features to enable the design of a decoder to their specific requirements. The tool employs computer search to find decoder parameters that best meet design criteria specified in a multi-objective fitness function. Features include: objective range-removal and importance, even performance by angle, performance that correlates with human spatial resolution, and frequency dependent and independent decoders of different orders. Performance can be optimized for a single listener or multiple off-center listeners. The current tool works for 5.0 surround sound; however it can be extended to other horizontal-only and 3-D configurations. Results are shown that demonstrate the tool's capability and flexibility for various scenarios.

3-2 Sound Image Localization Tests of a 21-Channel Surround System—*Jae-Hyoun Yoo, Jeongil Seo, Kyeongok Kang*, ETRI, Yuseong-gu, Daejeon, Korea

A 21-channel sound field reconstruction system has been developed with physical reconstruction of a three dimensional target sound field over the pre-defined control volume. We performed subjective listening tests of distance and vertical localization compared with VBAP (Vector Based Amplitude Panning), and the results showed that LSM (Least Squared Method) had better performance on distance localization than VBAP, and similar performance in vertical localization with VBAP.

3-3 A Novel Design for a Gel-Type DML Transducer Incorporating a Solid Panel Projecting Multiple Independent Sound Sources—*Minsung Cho*,^{1,2} *Elena Prokofieva*,¹ *Mike Barker*,¹ *Jordi Munoz*²

¹Edinburgh Napier University, Edinburgh, UK
²SFX Technologies Ltd., Edinburgh, UK

The gel-type DML transducer (referred to as the gel transducer in this paper) excites a panel to radiate sound waves through

the gel surround. The panel breaks into different characteristic modes of vibration as frequency increases. At high frequencies above 1,600 Hz, the extent of the moving area of the panel begins to reduce and becomes localized on the position of the gel transducer. This results in both the sound radiating area and volume velocity being reduced so that the directivity of the sound field narrows. As a result, a single panel with the two gel transducers attached can radiate independent sound sources with minimal acoustic cancellations at high frequencies. The current paper reports on this effect using theoretical and practical approaches.

3-4 Sound Generators Using Electroactive Elastomer for Multichannel Audio—*Takehiro Sugimoto*,¹ *Kazuho Ono*,¹ *Yuichi Morita*,² *Kosuke Hosoda*,² *Daisaku Ishii*,² *Akio Ando*¹

¹NHK Science and Technology Research Laboratories, Setagaya-ku, Tokyo, Japan

²Foster Electric Co., Ltd., Akishima, Tokyo, Japan

To develop sound generators suitable for multichannel audio we studied the transformation efficiency of electroactive elastomer (EAE), which is soft material that can be transformed by applying voltage. From the results of our analysis, we proposed two types of sound generators, both of which are lightweight because they do not use conventional driving parts. The first type is a cylindrical sound generator that radiates sound omnidirectionally in the horizontal plane using the EAE's flexibility. The second type is a push-pull sound generator with a comparatively improved frequency response by more effective use of the transformation of EAE. The obtained acoustic characteristics and future applications are discussed in this paper.

3-5 A New Method for B-Format to Binaural Transcoding—*Svein Berge, Natasha Barrett*, Berges Allmenndigitalte Rådgivningstjeneste, Oslo, Norway

A frequency-domain parametric method for transcoding first-order B-format signals to a binaural format is introduced. The method provides better spatial sharpness than linear methods allow. A high angular resolution planewave decomposition of the B-format establishes two independent direction estimates per time/frequency bin. This alleviates the requirement that the sound sources in a mix are W-disjoint orthogonal, implicit in previous nonlinear methods. The characteristics and causes of audible artifacts are discussed. Methods are introduced that suppress the different types of artifacts. A listening test is presented that ranks the sound quality of the method between third-order and fifth-order linear ambisonics systems.

3-6 A Virtual Acoustic Film Dubbing Stage—*Stephen Smyth, Michael Smyth, Steve Cheung, Lorr Kramer*, Smyth Research LLC, Camarillo, CA, USA

By capturing personalized binaural room responses within a film dubbing stage a highly accurate three dimensional virtualization of the stage acoustics is possible using standard stereo headphones. For the first time a virtual dubbing stage can be captured and brought to the desktop audio workstation. This paper focuses on one particular virtualization technology, SVS, that discusses some of the virtualization issues relating to dubbing stages, describes how the technology addresses these issues, and highlights some of the remaining problems of the virtualization technique. Finally, the actual measurement of a large film theater is described.

Friday, October 8

14:30–17:30

WORKSHOP 3

Periphony—More than Just Over Your Head

Chair: **Jeff Levison**, SurroundExpert.com

Panelists: *Stuart Bowling*, Dolby Laboratories
Kimio Hamasaki, NHK
Ulrike Kristina Schwarz, Bayerischer Rundfunk
Wilfried Van Baelen, Galaxy Studios
Helmut Wittek, Schoeps Mikrofone GmbH
Wieslaw Woszczyk, McGill University

A growing number of playback formats are being made available—either in experimental systems or real life possibilities—for playback of audio that not only is in surround sound but also incorporates additional channels to create the sensation of height thus maximizing what some have referred to as the sensations of envelopment and engulfment. Systems in use include UHDTV, IMAX, Dolby ProLogic IIz, Chesky 2+2+2, DTS-HD, Ambisonics—plus other experimental methods. Great interest has been rekindled in these systems with the success of 3-D motion pictures and now the introduction of 3-D video to the consumer at home.

Presentations will include discussions regarding scientific research, classical recording, pop music production, radio-style drama, expanded cinema, and theatrical stage sound design. A brief history will be offered for the uninitiated to understand the basic premises at work in periphony. The panelists will offer playback of samples of their work in formats ranging from 5.1+2 to 22.2.

Friday, October 8 17:30

SPECIAL EVENT

Super Hi-Vision and 22.2 Multichannel Sound Demonstrations

Friday, October 8 19:30

SPECIAL EVENT

Welcome Concert: Japanese Traditional Music and Multichannel Sound Music

SATURDAY, OCTOBER 9: ALL EVENTS TAKE PLACE AT TOKYO UNIVERSITY OF THE ARTS

Saturday, October 9 08:55

AES PRESIDENT ADDRESS

Saturday, October 9 09:00–10:00

KEYNOTE ADDRESS 2

Audio Displays and Microphone Arrays for Active Listening

Keynote Speaker: **Yōiti Suzuki**, Tohoku University

To realize future communications interactively with a high sense-of-presence, it is important to recall that we humans are active creatures, moving through the environment to acquire accurate spatial information. For instance, in terms of spatial hearing, humans usually make slight head and body movements unconsciously, even when trying to keep still while listening. Actually, such movement is known to be effective for improving the precision of auditory spatial recognition. We designate this style of listening as active listening. Therefore, it is particularly important that sound systems to synthesize sound fields, which we call auditory displays, be responsive to a listener's movement, at least to a listener's head rotation. Auditory displays matching the motions of active listening are therefore eagerly sought for use in future communications. In this presentation, we first show that a sound field that is synthesized to a listener's movement in a responsive manner significantly enhances the listener's perceived sense-of-presence. Our auditory display, which is responsive to a listener's movement based on binaural reproduction architecture, was used for this experiment. Then we introduce our high-definition small spherical microphone array based on Symmetrical object with ENchased Zillion microphones (SENZI) architecture and its implementation with 252 channels of microphones. SENZI can sense spatial sound information

comprehensively and precisely so that the sensed (recorded) sounds can be suitably reproduced by auditory displays that are responsive to a listener's movement, beyond place and time. Finally, we introduce a high-definition auditory display based on high-order Ambisonics (HOA) architecture with the fifth order: the highest order realized to date. This system is implemented with a surrounding loudspeaker array of 157 loudspeakers. These systems are expected to be useful to realize new and advanced communications systems providing high sense-of-presence. Moreover, such systems are expected to be useful as experimental systems to accumulate new knowledge related to human perceptions, which is crucially important for the advancement of communications.

Acknowledgments: Parts of this research are supported by Tohoku University GCOE program CERIAES, Grants-in-Aid for Specially Promoted Research (no. 19001004) to Suzuki from JSPS, and by SCOPE (no. 082102005) to Sakamoto from MIC Japan.

Saturday, October 9

10:00–11:00

PAPER SESSION 4: APPLICATIONS OF SPATIAL AUDIO

4-1 Applying Spatial Audio to Human Interfaces: 25 Years of NASA Experience—*Durand R. Begault,¹ Elizabeth M. Wenzel,¹ Martine Godfroy,^{1,2} Joel D. Miller,^{1,3} Mark R. Anderson^{1,3}*

¹Human Systems Integration Division, NASA Ames
Research Center, Moffett Field, CA, USA

²San José State University Foundation, San José, CA, USA

³Dell Services-Perot Systems, Plano, TX, USA

From the perspective of human factors engineering, the inclusion of spatial audio within a human-machine interface is advantageous from several perspectives. Demonstrated benefits include the ability to monitor multiple streams of speech and non-speech warning tones using a “cocktail party” advantage and for aurally-guided visual search. Other potential benefits include the spatial coordination and interaction of multimodal events and evaluation of new communication technologies and alerting systems using virtual simulation. Many of these technologies were developed at NASA Ames Research Center, beginning in 1985. This paper reviews examples and describes the advantages of spatial sound in NASA-related technologies, including space operations, aeronautics, and search and rescue. The work has involved hardware and software development as well as basic and applied research.

4-2 “GABRIEL”: Geo-Aware BRroadcasting for In-Vehicle Entertainment and Localizability—*Julian Villegas, Michael Cohen*, University of Aizu, Aizu-Wakamatsu, Japan

We have retrofitted a vehicle with location-aware advisories /announcements, delivered via wireless headphones for passengers and “nearphones” or bone-conduction headphones for the driver. Our prototype differs from other projects exploring spatialization of the aural information: besides the commonly used landmarks to trigger audio stream delivery, our prototype uses geo-located virtual sources to synthesize spatial soundscapes. Intended as a “proof of concept” and tested for future research, our development features multilingual tourist information, navigation instructions, and traffic advisories rendered simultaneously.

Saturday, October 9

10:00 – 12:00

WORKSHOP 4

3-D Sound

Co-Chairs: **Kazuho Ono, Thomas Sporer**

Panelists: *Diemer de Vries*, Delft University of Technology
Kimio Hamasaki, NHK Science and Technology
Research Laboratories
Shiro Ise, Kyoto University

Toshiyuki Kimura, National Institute of Information
and Communications Technology
Frank Melchior, IOSONO GmbH
Sascha Spors, Deutsche Telekom Laboratories/TU Berlin

Current state of 3-D sound includes various ideas and technologies. This workshop overviews the current state of these technologies. WFS, Ambisonics, 22.2ch audio, together with other innovative ideas such as focusing sound and boundary control method, etc., will be presented and discussed.

Saturday, October 9

11:00

POSTER SESSION

P-1 Developing Common Attributes to Evaluate Spatial Impression of Surround Sound Recording—Toru

Kamekawa, Atsushi Marui, Tokyo University of the Arts,
Tokyo, Japan

For the evaluation of spatial impression, several attributes are used. However, in investigating more critical evaluation of spatial impression such as surround microphone setting, it is very difficult to share common meanings for each perceptual attribute. The authors attempted to elicit common attributes from surround sound recordings by triadic elicitation procedure. Three attributes, “brightness,” “temporal separability,” and “spatial homogeneity” were elicited. Pair-wise comparison was implemented to evaluate five different microphone placements for surround recordings using these attributes. From the results of ANOVA, significant differences between microphone placements and effect of subjects’ individual differences were observed at all attributes. After removing the subjects who had circular triads and by using cluster analysis procedure, 60 to 70 percent (depending on the attributes) of the professional subjects remained. This is more stable compared to the student subjects for all three attributes. It is suggested that training is necessary for naïve listeners to share the same meanings of these attributes. Focusing one of the elicited attributes related to spatial impression, “spatial homogeneity,” which has significant differences between a pair of recording excerpts, authors studied correspondence with physical factors. D/R (the direct to reverb ratio) was calculated from running IACC, which was measured from a binaural recording of music and filtered into third-octave bands. It is hypothesized that differences of mean D/R at certain frequency band (around 3 kHz), what were measured between different listening positions, correspond to “spatial homogeneity.”

P-2 Standardization of PEAQ-MC: Extension of ITU-R BS.1387-1 to Multichannel Audio—Jusith Liebetrau, Thomas Sporer,

Sven Kämpf, Sebastian Schneider, Fraunhofer Institute for
Digital Media Technology IDMT, Ilmenau, Germany

PEAQ (Perceptual Evaluation of Audio Quality) is a standardized algorithm for the objective measurement of perceived audio quality. It predicts the perceived audio quality of mono and stereo audio files as listeners would do in a subjective listening test according to ITU-R BS.1116-1. Unfortunately this prediction is not intended for multichannel material, such as 5.1 or beyond. Additionally, the quality estimation of the standard does not consider spatial artifacts. Members of ITU-R Working Party 6C are developing an extension toward multichannel compatibility, as well as an integration of the modeling of spatial artifacts assessments. In this paper the current status of the standardization progress is illustrated. The concept of one of the three proposals currently under consideration is explained and the results of a first verification test conducted by ITU-R are presented.

P-3 Spatial Audio Analysis Based on Perceptually Empirical Mode Decomposition—ChingShun Lin, YungCheng Chao,

National Taiwan University of Science and Technology, Taipei,
Taiwan

The goal of an immersion system is to create the illusion of proximity for people in different areas. To achieve this, it is essential to pick-up and regenerate all the crucially visual and aural information that is perceptible by human senses. One of the important decisions in any audio processing system is to choose appropriate features. The criteria may count on how exactly the signals are represented and how easily the following process can be performed. In addition, it is more challenging if we would like to simulate the acoustical characteristics around a specific instrument based on the universal recording. The totally distinct spectral characteristics generated by the target instrument should dominate the recording, whereas the subtly surrounding audio clues must be preserved as well. In this paper we analyze such spatial audio and demonstrate how it can be systematically computed by the perception-based empirical mode decomposition. As an example of our method, analysis for reference microphone signal and its counterpart that simulates what a spot microphone would pick up near a percussive instrument is presented.

P-4 Improved ITD Estimation in Reverberant Environments —

Gavin Kearney, Damien Kelly, Frank Boland, Trinity College,
Dublin, Ireland

In this paper we present an improved model for Interaural Time Difference (ITD) estimation in reverberant environments. The Phase Transform (PHAT) weighting function, used in generalized cross correlation is investigated here for application to Interaural Cross Correlation (IACC) measurements from binaural microphones. A binaural model is developed to accommodate the method, and demonstrations of its implementation in a real reverberant room show improved ITD estimation over non-weighted IACC, when compared to subjective listening results.

P-5 Wave Front Reconstruction Using the Spatial Covariance Matrices Method—Hiroki Hagiwara,¹ Yoshinori Takahashi,^{1,2}

Kazunori Miyoshi¹

¹Kogakuin University, Tokyo, Japan

²Tokyo Metropolitan College of Industrial Technology,
Tokyo, Japan

This paper describes a sound-field reproduction method called a Co-Variance Method (CVM) that reproduces a spatial covariance matrix of the original sound field without information of sound-source locations. Multichannels of reproduced sound in a listening area were made from the recorded signals in the original field according to the adoptively optimized process where the quasi-Newton method is utilized in order to get the solution. Numerical simulation results confirmed that the method works well in principle, and consequently reconstructed wavefronts are similar to the original ones. Relationship between the number of microphones or loudspeakers and reproduction errors are investigated. It is shown that the phase error diminishes as the number of pairs for microphones and loudspeakers increases more than that of positions set for the covariance analysis.

P-6 Inverse Wave Propagation in Wave Field Synthesis—

Shoichi Koyama, Yusuke Hiwasaki, Ken'ichi Furuya, Yoichi Haneda, NTT Cyber Space Laboratories, Musashino-Shi,
Tokyo, Japan

A combined method of wave field synthesis and inverse wave-field propagation that enables the recreation of sound images in front of a loudspeaker array is presented. By use of inverse wave propagation, the sound pressure at the virtual receiving plane, which is in the backward direction of wave propagation from the actual receiving plane, can be estimated. This estimation is an analytical method based on the wave equation typically used in holography. Shifting the virtual receiving plane backward from the actual receiving plane corresponds to shift-

ing the reconstruction plane forward from the secondary sources. As a result, this method has the advantage of being able to reproduce virtual primary sources in front of secondary sources even when the position or directivity of the primary sources is unknown; this was not possible by the conventional focused-source technique. Numerical simulation results are also presented to show the efficacy of the proposed method.

P-7 Sound Field Equalization by Active Acoustic Impedance

Control—Jyunji Hagio,¹ Akihiro Kakiuchi,^{1,2} Akira Omoto¹

¹Kyushu University, Minami-ku, Fukuoka, Japan

²Currently with NHK, Tokyo, Japan

This study examined sound field equalization using active control. The specific acoustic impedance, the ratio of sound pressure, and particle velocities were adopted as a source of control. Providing the impedance in the aimed direction was controlled to be the characteristic impedance of the medium, the propagating plane wave in any particular direction was expected. In addition, the directional characteristic of the sound propagation was also controlled by adjusting the weight of ratios of pressure and velocity in three orthogonal directions. The results of numerical simulations indicated the potential efficiency of the proposed method. Furthermore, results of subjective experiments showed the possibility of the proposed control changing the perceived direction of the incoming sound.

P-8 Interactive Enhancement of Stereo Recordings Using Time-Frequency Selective Panning—Maximo Cobos, Jose J. Lopez, Universidad Politecnica de Valencia, Valencia, Spain

Localization of sounds in physical space plays a very important role in multiple audio related disciplines, such as music, sound art, or sound editing for audiovisual productions. The most well known technique for providing such spatial impression is stereo panning, which creates a virtual location of a sound source by distributing its energy between two independent channels during the mixing process. However, once all the sound events have been mixed, re-distributing source locations to widen the stereo image is not certainly an easy task. Motivated by this problem, we propose a source spatialization technique based on the time-frequency processing of the stereo mixture. The energy distribution over the stereo panorama is modified according to a nonlinear warping function, providing a widening effect that enhances the stereo experience without degrading the sound quality and preserving the original conception of the mixing engineer.

P-9 Sound Field Reproduction Applied to Flight Vehicles Sound Environments—Cédric Camier, Philippe-Aubert Gauthier, Yann Pasco, Alain Berry, Université de Sherbrooke, Sherbrooke, Quebec, Canada and McGill University, Montreal, Quebec, Canada

This paper proposes a preliminary theoretical study for sound field and sound environment reproduction in flight vehicles. A fully-coupled cavity, cylindrical shell, and exterior radiation model approximates an aircraft cabin mock-up. Material and geometry characteristics are inspired by measurements performed on a cabin mock-up. The sound field reproduction is based on reproduction error minimization at a microphone array positioned in the cavity. Two reproduction systems, based on actuators or loudspeakers are simulated in order to compare their feasibility and performance. The model linking excitator strength with the pressure on the spatially extended array region is developed in a matricial form. The promising results obtained in terms of reproduced pressure in the array region in both cases presume the reliability of such dedicated systems.

P-10 Microphone Configurations for Teleconference Application of Directional Audio Coding and Subjective Evaluation—Jukka Ahonen, Aalto University School of Science and Technology, Aalto, Finland

Directional Audio Coding (DirAC) is a spatial-sound processing technique where the arrival direction and diffuseness of sound are analyzed in frequency-bands from microphone signals, transmitted with one or multiple audio channels, and utilized for various purposes in synthesis. Among other applications, DirAC has been used in low bit-rate teleconferencing to provide spatial separation of remote talkers corresponding to reality. Here, the use of different microphone configurations, consisting of omnidirectional or directional microphones, is discussed in DirAC teleconferencing. The audio quality with different microphone techniques for DirAC teleconferencing is evaluated by the subjective listening tests, and the results are presented in this paper.

P-11 Real-Time Tracking of Speech Sources Using Binaural Audio and Orientation Tracking—Marko Takanen, Matti Karjalainen, Aalto University School of Science and Technology, Espoo, Aalto, Finland

This paper presents a method for real-time estimation of the directions of speech sources from captured binaural audio. Accurate direction estimates are required in order to embed the sound sources correctly to the auditory environment of the far-end user in telecommunication between two augmented reality audio (ARA) users. The dependency of the estimation accuracy on the orientation of the near-end user is avoided in this method by combining the information from an orientation tracker to the direction estimates. The results from the anechoic experiments illustrate that the presented method can estimate the direction(s) of non-simultaneous speech source(s) in real-time, and that head movement improves the estimation accuracy of sources on the sides of the user.

P-12 On the Influence Coding Method on Japanese Speech Intelligibility in Virtual 3-D Audio Space—Yosuke Kobayashi,¹ Kazuhiro Kondo,¹ Kiyoshi Nakagawa,¹ Yukio Iwaya²

¹Yamagata University, Yamagata, Japan

²Tohoku University, Miyagi, Japan

In this paper we investigated the influence of stereo coding on the 3-D audio for Japanese. We encoded localized test samples using joint stereo and parametric stereo of the HE-AAC encoder at identical coding rates. The Japanese word intelligibility test employed was the Japanese Diagnostic Rhyme Tests (JDRT). First, we localized the speaker in front of the listener at an arbitrary distance a (1.00a). Next, we compared the effect of noise located at a distance of 0.25a from the listener at one of the angles 15 degrees apart on the horizontal plane. The result showed that the target speech cannot be separated from the noise for any stereo coding when the noise was in front of speaker between azimuths of +30 deg. to -30 deg. However, at other azimuths, the intelligibility scores were far better. Stereo coding shows degraded intelligibility compared to the reference at any noise azimuths. However, joint stereo was shown to be constantly better compared to parametric coding, suggesting that the former is the stereo coding of choice for transmission of localized 3-D audio.

P-13 Acoustic Measurement System for 3-D Loudspeaker Set-Ups—Andreas Silzle, Matthias Lang, Jose Angel Pineda Pardo, Oliver Thiergart, Giovanni Del Galdo, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

For any reproducible listening set-up it is crucial to verify whether the reproduction system is operating properly. This is a tedious and cumbersome task, as state-of-the-art listening rooms are equipped with intricate signal processing chains and a higher number of loudspeakers arranged at very specific positions. Building an automatic test-system with an adequate accuracy and reliability represents a rather challenging engineering problem. This paper presents a multi-loudspeaker test-

system, which accomplishes this task, realized by carefully combining existing measurement techniques. The thorough validation of the developed test-system indicates an accuracy in determining the position of ± 3 degree, the distance of ± 4 cm, and the magnitude response of ± 1 dB, for each loudspeaker in the listening room with a total measurement time of less than 10 s per loudspeaker.

P-14 Spatial Audio and Reverberation in an Augmented Reality Game Sound Design—*Natasa Paterson, Fionnuala Conway*, University of Dublin, Trinity College, Dublin, Ireland

In this paper we describe the sound design of a location based mobile phone game and investigate the role of spatialized audio and reverberation. Particular attention was given to the effect of audio on immersion and emotional engagement of participants. We present the sound design, implementation and evaluation of our working prototype, *Viking Ghost Hunt* and the results and interpretations obtained. Evaluation of the game was undertaken over a three-day period with the participation of 19 subjects in which the prototype was evaluated for game immersion and engagement. Further tests of a subset of participants was undertaken in order to test the specific audio parameters of spatialized audio and reverberation and their effect on game immersion and engagement. The results indicated that audio and specifically reverberation, play an important role in immersing a player within the game space.

Saturday, October 9

13:00–16:00

PAPER SESSION 5: SURROUND SOUND WITH HEIGHT

5-1 Perception of Sound Image Elevation in Various Acoustic Environments—*Kentaro Matsui, Akio Ando*, NHK Science & Technology Research Laboratories, Setagaya-ku, Tokyo, Japan

To investigate the discrimination threshold of sound image elevation for a three-dimensional sound system, we conducted subjective evaluation experiments using a paired comparison method. Experiments on the discrimination of sound image elevation with loudspeakers were done in three chambers, each of which had a different reverberation time. Experiments with a headphone system were also done, in which stimuli were recorded binaurally. The experiments showed that (1) when the loudspeakers were set laterally to the subject, the elevation of the perceived sound image increased linearly according to the elevation of the sound source; (2) when the loudspeakers were set in front of or behind the subject, the perceptual resolution deteriorated as the sound source ascended. The experiments also indicated that there is no relation between the room reverberation time and perceptual resolution, and that the deterioration of perceptual resolution is conspicuous in headphone listening.

5-2 Investigating Listeners' Localization of Virtually Elevated Sound Sources—*Sungyoung Kim, Masahiro Ikeda, Yusuke Ono, Akio Takahashi*, Yamaha Corporation, Iwata, Shizuoka, Japan

The object of this study was to experimentally observe and compare the perceived directions of elevated sound sources in two conditions: one reproduced from a real loudspeaker and another from a virtually manipulated loudspeaker using a newly proposed transaural crosstalk cancellation. A total of twelve listeners evaluated perceived directions of various sound sources through a direct estimation of azimuth and elevation angle. The results showed that virtually elevated sound sources were generally perceived as lower compared to physically elevated ones, possibly due to the discrepancy between the Head Related Transfer Function (HRTF) used and the listeners' HRTFs. Subsequent analysis showed that localization of both conditions was influenced by the type and the bandwidth of the stim-

uli, yet not by the condition whether or not a listener had a reference position.

5-3 Influence of the Listening Position in the Perception of Elevated Sources in Wave-Field Synthesis—*Jose J. Lopez,¹ Maximo Cobos,¹ Basilio Pueo²*

¹Universidad Politecnica de Valencia, Valencia, Spain

²Universidad de Alicante, Alicante, Spain

This paper describes some effects related to the perception of elevated sources in Wave-Field Synthesis using HRTF elevation cues. In this recently proposed hybrid system, the conventional WFS approach is used to achieve localization in the horizontal plane, whereas elevation effects are simulated by means of spectral elevation cues. Since the simulated HRTF cues are the same within the listening area, the height of the virtual source changes depending on the listening position. Thus, different listeners perceive different source heights, having a perception that changes when they move around the listening area. Experiments aimed at investigating this effect are presented.

5-4 The Effect of Processing Two Overhead Channels on the Spatial Impression of Virtual Rooms Rendered in Eight-Channel Surround Sound—*Wieslaw Woszczyk, Doyuen Ko, David Benson, Brett Leonard*, McGill University, Montreal, Quebec, Canada

In eight-channel surround sound reproduction, two of the channels are elevated above the listener and their signals are transformed to provide three listening conditions for the evaluation of spatial impression. The elevation channels are either unprocessed, convolved with a short impulse response of a dummy head, or convolved with a dummy head response and also equalized. The six horizontal channels remain the same for each of the overhead conditions and the loudness of the overhead channels is calibrated to be equal for all three conditions. Two anechoic monophonic sound sources are convolved with eight-channel impulse responses previously measured in two large rooms using eight microphones with the overhead information captured by a pair of bidirectional spaced-apart microphones angled 90° between them and pointing diagonally upward. Listening tests conducted with ten expert subjects show a dependence of spatial impression (height, immersion, preference) on the nature of overhead signals, on the program (sound source), and the choice of room.

5-5 Evaluating Candidate Sets of Head-Related Transfer Functions for Control of Virtual Source Elevation—*Hyun Jo,¹ Youngjin Park,¹ William Martens²*

¹KAIST, Daejeon, Korea

²The University of Sydney, Sydney, NSW, Australia

The performance of 3 candidate sets of generic Head-Related Transfer Functions (HRTFs) was evaluated in terms of differences in their ability to create distinct elevation angles for virtual sources located on a listener's median plane. In comparison to directional errors observed when 6 listeners were presented with sources processed using their own individually measured Head-Related Impulse Responses (HRIRs), the errors associated with the 3 alternative candidate sets were tested for 7 virtual source target elevation angles on the median plane, spanning the upper hemifield with 30-degree resolution from front to back. One candidate set was that of a selected subject's HRIRs that had been modified through a customization procedure designed to provide optimal elevation performance for that subject (which was the selected 'representative' subject denoted 'CH' in Hwang, et al. [12]). Another candidate set was that formed by taking the mean over the optimized HRIRs of 9 subjects that resulted from that same customization procedure. The third set was generated by taking the mean over the HRIRs of 43 subjects found in the CIPIC database. Performance under identical conditions was also observed for

sources convolved with the appropriate individually customized HRIR for each of the 6 listeners, and also using a familiar KEMAR dataset to serve as a reference. Localization performance was quantified using the 3 error metrics of absolute polar angle error, vertical polar angle error, and front-back confusion rate. Only in terms of front-back confusion rate was performance shown to be inferior for virtual sources processed using the mean of the HRIRs from the CIPIC dataset when compared to either those processed using the representative subject's customized HRIRs, or using mean of customized HRIRs of 9 subjects. Although the study investigated virtual sources varying in elevation on the median-plane only, the results may generalize to the whole upper hemifield.

5-6 Live Production of 22.2 Multichannel Sound for Sports Programs—Tsuyoshi Hinata, Yuichi Otakeyama, Hiromi Sueishi, Japan Broadcasting Corporation (NHK), Tokyo, Japan

The NHK Outside Broadcast Division has produced many live 5.1 surround sound broadcasts of sports events using high-definition television (HDTV) broadcasting. With future live broadcasting in mind, this paper covers methods of realizing surround sound that takes advantage of the characteristics of the 22.2 multichannel sound system, the methods of realizing a superior special presentation, and the problems involved in the live mixing of 22.2 multichannel sound for sports programs produced with the Super Hi-Vision system developed by the NHK Science and Technology Research Laboratories.

Saturday, October 9 13:00 – 15:00

WORKSHOP 5

Surround Recording for Music

Presenter: **Akira Fukada, NHK**

What is the goal of surround recording? This is the eternal question for all the engineers who are trying this area. This workshop is expected to be a hands-on session demonstrating the recording process between musicians and recording engineers through the actual recording of a jazz trio performance. Participants can experience what will happen during the recording process in the studio or control room or both.

Performers: **Aaron Choulai** (piano), **Yasushi Fukumori** (drums), **Akiyoshi Shimizu** (bass)

Saturday, October 9 15:00 – 16:00

WORKSHOP 6

New Theoretical Model of Sound Field Diffusion

Presenter: **Toshiki Hanyu, Nihon University**

A new theoretical model for quantitatively characterizing sound field diffusion based on scattering coefficient and absorption coefficient of walls was developed. The concepts of equivalent scattering area, equivalent scatter reflection area, average scattering coefficient and average scatter reflection coefficient are introduced in order to express all walls' capability of scatter in a room. Using these concepts and the mean free path, scatter-to-absorption ratio, mean scatter time and diffusion time are defined in order to evaluate degree of diffusion of a space. Furthermore the effect of spatial scattering objects to sound field diffusion is formulated. In addition the time variation of specular and scattered components in a room impulse response is formulated. The verification of these characterization methods was performed with computer simulations based on the sound ray tracing method. The results supported that the ideas presented are basically valid. Because the method for measuring the scattering coefficient has already been defined by an ISO standard, it is possible to prepare a database of the coefficients gradually. Therefore one can design the degree of sound field diffusion by applying the equations presented.

Friday, October 8

16:00–17:30

PAPER SESSION 6: SIGNAL PROCESSING AND CODING

6-1 Multichannel Audio Coding Based on Minimum Audible Angles—Adrien Daniel,^{1,2} Rozenn Nicol,¹

Stephen McAdams²

¹Orange Labs, Lannion, France

²McGill University, Montreal, Quebec, Canada

The method described in this paper provides a scheme for encoding multichannel audio signals representing a spatial auditory scene based on human sound perception in space. It relies on a psychoacoustic model based on measures of minimum audible angles (MAA) in the presence of distracting sound sources. A compression gain is obtained by truncating the order of representation of the auditory scene in the Higher-Order Ambisonics (HOA) domain according to the psychoacoustic model. Numeric simulations were conducted in order to link the error of representation of the field with an angular distortion of the apparent direction of the sound sources.

6-2 Signal Models and Upmixing Techniques for Generating Multichannel Audio—Mark S. Vinton, Mark F. Davis, Charles Q. Robinson, Dolby Laboratories, San Francisco, CA, USA

Most systems for upmixing stereo content have traditionally used sums and differences of source signals, an arrangement referred to as matrixing. Matrix-based upmixers have evolved from passive operation to sophisticated active matrix designs and have achieved widespread commercial use. This paper introduces a new algorithm for upmixing from two to five channels using a hybrid of both scale factor and variable matrix techniques. The algorithm applies equally well to both Lt/Rt-encoded and conventional stereo programs. It improves on traditional methods, providing better reproduction of the front sound stage, while anchoring center images for off center listeners and creating more compelling ambience and envelopment. Data from subjective listening tests are provided to support these conclusions.

6-3 Three-Dimensional Sound Field Analysis with Directional Audio Coding Based on Signal Adaptive Parameter Estimators—Oliver Thiergart, Giovanni Del Galdo, Magdalena Prus, Fabian Kuech, Fraunhofer Institute for Integrated Circuits, IIS, Erlangen, Germany

Directional audio coding (DirAC) provides an efficient description of spatial sound in terms of an audio downmix signal and parametric side information, namely the direction of arrival and diffuseness of sound. The sound scene can be reproduced based on this information with any audio reproduction system such as multichannel playback or binaural rendering. Input to the DirAC analysis are acoustic signals, e.g., captured by a microphone array. The accuracy of the DirAC parameter estimation can suffer from a low signal-to-noise ratio (SNR) and a high temporal variance of the input signals. To handle these problems, this contribution proposes signal adaptive parameter estimators that increase the estimation accuracy by considering the SNR and the stationarity interval of the input. Simulations show that the DirAC analysis is significantly improved.

Saturday, October 9

16:00 – 17:30

WORKSHOP 7

Mixing and Production of Spatial Audio

Mixing with Perspective

Presenter: **Florian Camerer, ORF, Austria**

Concerts of classical music, as well as operas, have been a part of broadcast programming since the beginning of television. The aesthetic relationship between sound and picture plays an important part in the satisfactory experience of the consumer. The question how far the audio

perspective (if at all!) should follow the video angle (or vice versa) has always been a subject of discussion among sound engineers and producers. In the course of a diploma work this aspect has been investigated systematically. One excerpt of the famous New Year's Concert (from 2009) has been remixed into four distinctly different surround sound versions. Close to 80 laymen who expressed an interest in classical music had the task of judging these versions to the same picture if they found the audio perspective appropriate to the video or not.

A Disruptive Production Tool in the Workflow of Spatial Audio Masters

Presenter: **Wilfried Van Baelen**, Galaxy Studios, Belgium

This presentation will highlight the many practical, and technical (sometimes hard to remedy) issues that have to be solved to enable spatial audio formats in the market. Content owners, audio engineers, and consumers each have their own specific requirements, and have become very demanding when it comes to compatibility combined with high quality audio. Based on presenter's pioneering work since many years in the audio and film industry, he will share with you his vision, strategy and vast experience that led to the development of a unique production tool allowing for an easy workflow to create spatial, discrete 3-D Sound Masters that are backwards compatible with existing main standards and formats without any audible concession.

Saturday, October 9

18:00

SPECIAL EVENT: BANQUET

**Dinner Cruise—Japanese Cuisine
in a Japanese Houseboat**

**SUNDAY, OCTOBER 10: ALL EVENTS TAKE PLACE
AT TOKYO UNIVERSITY OF THE ARTS**

Sunday, October 10

09:00–10:00

KEYNOTE ADDRESS 3

Space Concept in the Contemporary Music

Keynote Speaker: **Mikako Mizuno**, Nagoya City University

In the presentation the space idea in the contemporary music will be discussed. The term "Contemporary Music" has a limited meaning, that is, the music of avant-garde style, especially created by the composers of the second half of the twentieth century.

The relationship between space and music can be discussed only when the composers have some technical method to realize their spatial ideas. The main method here to be presented is that of "Tekkokan," which was produced by the up and coming composer at that time, Toru Takemitsu in the World's Fair 1970. His special idea was realized by the architectural design and the huge loudspeaker system controlled by in original way. Another example that realizes the musical idea as architecture will be also discussed, including the musical pieces by Iannis Xenakis and Luigi Nono.

Sunday, October 10

10:00–12:00

PAPER SESSION 7: PERCEPTION AND EVALUATION OF SPATIAL AUDIO—PART 2

7-1 Perceptual Localization of a Phantom Sound Image

for Ultrahigh-Definition TV—Young Woo Lee, Sunmin Kim,
Samsung Electronics Co., Ltd., Suwon, Gyeonggi-do, Korea

This paper presents a localization perception of a phantom sound image for ultrahigh-definition TV with respect to various loudspeaker configurations: two-horizontal, two-vertical, and triplet loudspeakers. Vector base amplitude panning algorithm with modification for non-equidistant loudspeaker setup is applied to create the phantom sound image. In order to practically study the localization performance in real situations, the listening tests were conducted at the on-axis and off-axis positions in normal listening rooms. A method of adjustment that can reduce the ambiguity of a perceived angle is exploited to

evaluate the angles of octave-band signals. The subjects changed the panning angle until the real sound source and virtually panned source were coincident. A spatial blurring can be measured by examining the differences of the panning angles perceived with respect to each band. The listening tests show that the triplet panning method has better performance than vertical panning in view of perceptual localization and spatial blurring at both on-axis and off-axis positions.

7-2

**Speech Intelligibility in Teleconference Application
of Directional Audio Coding—Jukka Ahonen, Ville Pulkki,**
Aalto University School of Science and Technology, Aalto,
Finland

Directional Audio Coding (DirAC), which is a method to parametrize the directional sound field, can be applied to a low bit-rate teleconferencing. The direction and diffuseness of the sound field are analyzed from microphone signals within frequency-bands in one end, transmitted to the other end as a metadata in a single channel with an audio signal, and used to reproduce spatial sound. In this paper one- and two-dimensional arrays of omnidirectional microphones to provide input signals for DirAC teleconferencing are reviewed. A listening test to measure speech intelligibility in DirAC teleconferencing was conducted, when both one- and two-dimensional microphone arrays were utilized. The results for the test are presented in this paper.

7-3

**Evaluation of a Speech-Based and Binaural Speech
Transmission Index—Anton Schlesinger,¹ Juan-Pablo
Ramirez,² Marinus M. Boone¹**

¹Delft University of Technology, Delft, The Netherlands

²Deutsche Telekom Laboratories, Berlin Institute of
Technology, Berlin, Germany

A speech-based and binaural Speech Transmission Index is presented and evaluated in a variety of acoustical degradations and spatial conditions. The proposed method facilitates the assessment of speech intelligibility in classical room acoustics and electroacoustics by simply comparing a binaural speech recording in adverse conditions with its clean original. Both the binaural processing stage and the speech-based Speech Transmission Index method are effective and computationally fast realizations. The central part of the binaural processor forms a cross-correlation stage that is designed to replicate psychoacoustic data of binaural interaction. Supplemented with the head shadow effect, which is generated in a "better-ear" fashion, a fair amount of the binaural advantage in speech intelligibility is modeled. An evaluation of the method was performed in a battery of listening tests. These tests incorporate different degradations, as e.g., stationary noise and fluctuation noise, a set of nonlinear signal alterations, including a speech enhancement processor, and a multitude of spatial configurations with different room acoustics and with up to four interferers. As a result, the objective method offers a stable prediction of the subjective results in binaural speech intelligibility throughout most of the linear degradations. In spite of that, the full account of the binaural advantage is not achieved by the current implementation of the method, which suggests further research.

7-4

**Listening and Conversational Quality of Spatial Audio
Conferencing—Alexander Raake, Claudia Schlegel, Matthias
Geier, Jens Ahrens, Deutsche Telekom Laboratories, TU
Berlin, Berlin, Germany**

We present the results of a listening and a conversation tests on the quality of spatial and non-spatial audio conferences. To this aim, we have developed conversation test scenarios for audio conferences with three remote participants in order to carry out quality evaluation tests for audio-conferences that are comparable with similar scenarios for traditional one-to-one telephone conversation assessment. We have applied the test scenarios during a conversation test, to (i) validate the test

scenarios, (ii) in a realistic usage context measure the advantages of spatial versus non-spatial audio conferencing, and in relation with the quality impact due to the transmitted speech bandwidth, and (iii) provide recordings of conferences for later use in listening tests. In the conversation test, we have compared different bandwidths (narrowband/NB, 300-3400 Hz; wideband/WB, 50-7000 Hz; full-band/FB, 20-22000 Hz), spatial versus non-spatial headphone-based rendering, and channels with and without talker echo. In a subsequent listening test using recorded conferences, we have attempted to assess the quality of spatial and non-spatial audio-conferencing in a more detailed fashion, including aspects such as speaker identification and memory.

Sunday, October 10

10:00 – 12:00

WORKSHOP 8

The Art and Practice of Multichannel Field Recording

Chair: **Charles Fox**, University of Regina

Panelists: *Florian Camerer*, ORF

Yasuo Hijikata, Field recordist

Mick Sawaguchi, Mick Sound Lab.

Carrying their audio kit in a backpack, sound recordists have been trekking into the wild, meeting the unique demands of multichannel field recording in the natural environment with creativity and skills that continue to push the boundaries of sound recording. The extreme weather and remote locations have not prevented these adventurous audio practitioners from developing and experimenting with recording methods, achieving high quality results that provide unique, engaging sonic experiences for audiences. The “The Art and Practice of Multichannel Field Recording” panel will share their knowledge in multichannel, immersive location recording research and methods that are an invaluable part of creating the three-dimensional audio experience.

Sunday, October 10

12:00

SPECIAL EVENT

The Surround-Scape Lunchtime Concert

Presenters: **Shiro Murakami**, Tokyo University of the Arts
Mick Sawaguchi, Mick Sound Lab.

What Is “Surround-Scape”?

You may have heard of the term “sound scape,” and surround-scape is related to that. The concept is a combination of Nature surround sound and Music. Mick Sawaguchi has been recording surround sound out in nature since 2000, concentrating on specific themes such as wind, waves, forest sounds, or the sounds of the seasons so far. A careful selection of these sounds forms the basis for musicians who play live hearing those sounds, combining composed elements, and improvisation—all inspired by the natural surround sound ambiances. The Surround-scape will let you feel the “Power of the Earth” and give you a “Hug by Nature”!

The lunch concert will consist of a few pieces of Sawaguchi’s surround clips with piano music played by Shiro Murakami who is a fourth year student of Tokyo University of the Arts. The venue will have a surround sound loudspeaker setup around the seating area, the piano will be placed in the middle of the studio. Enjoy a peaceful atmosphere and feel a hug by Nature Surround Sound after lunch!

Sunday, October 10

13:00–14:00

PAPER SESSION 8: MICROPHONE AND MIXING TECHNIQUES

8-1 Sound Field Extrapolation: Inverse Problems, Virtual

Microphone Arrays, and Spatial Filters—*Philippe-Aubert Gauthier, Cédric Camier, Yann Pasco, Éric Chabatte, Alain Berry*, Université de Sherbrooke, Sherbrooke, Quebec, Canada and McGill University, Montreal, Quebec, Canada

Sound field extrapolation is useful for measurement, description, and characterization of sound environments and sound fields that must be reproduced using spatial sound systems such as Wave Field Synthesis, Ambisonics, etc. In this paper two methods are compared: inverse problems and virtual microphone arrays with filtering in the cylindrical harmonics domain. The goal was to define and identify methods that could accommodate to various non-uniform sensor arrays (i.e., non array-specific methods) and that are less sensitive to measurement noise. According to the results presented in this paper, it seems that the method based on the inverse problem with Tikhonov regularization is less sensitive to measurement noise.

8-2 A Spherical Microphone Array for Synthesizing Virtual Directive Microphones in Live Broadcasting and in Postproduction

—*Angelo Farina,¹ Andrea Capra,¹ Lorenzo Chiesi,¹ Leonardo Scopece²*

¹University of Parma, Parma, Italy

²RAI CRIT Research and Technology Innovation Center, Turin, Italy

The paper describes the theory and the first operational results of a new multichannel recording system based on a 32-capsule spherical microphone array. Up to 7 virtual microphones can be synthesized in real-time, choosing dynamically the directivity pattern (from cardioid to 6th-order ultradirective) and the aiming. A graphical user’s interface allows for moving the virtual microphones over a 360-degree video image. The system employs a novel mathematical theory for computing the matrix of massive FIR filters, which are convolved in real time and with small latency thanks to a partitioned convolution processor.

Sunday, October 10

13:00 – 14:30

WORKSHOP 9

New Spatial Audio Coding Methods Based on Time-Frequency Processing

Chair: **Ville Pulkki**, Aalto University

Panelists: *Jean-Marc Jot*, DTS, Inc.

Juha Vilkamo, Fraunhofer IIS

The time-frequency resolution of human hearing has been taken into account for long time in perceptual audio codecs. Recently, the spatial resolution of humans has also been exploited in time-frequency processing. This has already lead into new methods to represent parametrically multichannel audio files and B-format recordings in time-frequency domain. This leads to drastic decrease in bit rate, and in reproduction of B-format recordings, also to increased audio quality compared to traditional methods. This workshop covers the capabilities and incapacities of human spatial hearing and the audio techniques that exploit these features. Typically the techniques are based on estimation of directional information for each auditory frequency channel, which information is then used in further processing.

Sunday, October 10

14:00–16:30

PAPER SESSION 9: SPATIALIZATION AND REVERBERATION

9-1 Binaural Reverberation Using Two Parallel Feedback

Delay Networks—*Fritz Menzer*, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Binaural reverberators often use a tapped delay line to model early reflections, to which a delayed late reverberation tail is added. This paper introduces a reverberator structure where

the early reflections are modeled using a feedback delay network (FDN) and the late reverberation is modeled by a second FDN in parallel. The impulse responses of both FDNs are highly overlapping, simulating the presence of diffuse reverberation from the beginning of the impulse response. A particular feature of this reverberator is the reproduction of first and second order reflections using only head related transfer functions (HRTFs) for the directions of the first order reflections, allowing to reduce the computational complexity.

- 9-2 Space Builder: An Impulse Response-Based Tool for Immersive 22.2 Channel Ambiance Design**—*Wiesław Woszczyk, Brett Leonard, Doyuen Ko, McGill University, Montreal, Quebec, Canada*

The convolution-based Space Builder employs segments of impulse responses to construct flexible spatial designs using an intuitive graphic interface. The system uses multiple low-latency convolution engines loading data from a library of multichannel impulse responses, a 128-channel MADI router and mixer operating at 24/96 resolution, and a MIDI controller. The controller reveals different levels of complexity depending on the needs of the user. The design, architecture, and functionality of the modules and the system are described. The system provides spatial up-conversion from 1 to 24 channels and has capacity to expand the number of channels beyond 22.2. Future applications for the system, including live sound applications, will be presented.

- 9-3 Sparse Frequency-Domain Reverberator**—*Juha Vilkkamo, Bernhard Neugebauer, Jan Plogsties, Fraunhofer IIS, Erlangen, Germany*

Numerous applications require realistic and computationally efficient late reverberation. In this paper the perceptually relevant properties of reverberation are identified, and a novel frequency transform domain reverberator that fulfills these properties is proposed. Listening test results confirm that the proposed reverberator has a perceptual quality equivalent to the ideal solution of decaying Gaussian noise in frequency bands.

- 9-4 Design and Implementation of an Interactive Room Simulation for Wave Field Synthesis**—*Frank Melchior,¹ Christoph Sladeczek,² Andreas Partzsch,¹ Sandra Brix²*
¹IOSONO GmbH, Erfurt, Germany
²Fraunhofer IDMT, Ilmenau, Germany

This paper describes a novel concept and implementation of reproduction system independent room simulation. The system is based on circular array measurements that deliver high spatial resolution impulse responses. Using this data a reproduction system independent interaction method has been developed. A detailed description of the system and interaction methods will be given. Furthermore the first implementation of the system for Wave Field Synthesis will be described in detail.

- 9-5 Creation of Binaural Impulse Responses for High-Presence Reproduction**—*Yakashi Nishi,¹ Shintarou Hiji²*
¹University of Kitakyushu, Fukuoka, Japan
²Kyushu Regional Police Bureau Info-Communications Division, Fukuoka, Japan

Geometrical acoustic simulation is time consuming and can only achieve fixed impulse responses. A new system to create binaural impulse responses, whose acoustical characteristics can be arbitrarily specified, is proposed. Psychologically important parameters such as reverberation time, clarity index, and inter-aural cross-correlation can be independently and significantly controlled. The fundamental procedures to realize this system are described. Simulation experiments were conducted to verify the availability and effectiveness of the algorithm.

Sunday, October 10

14:30 – 15:30

WORKSHOP 10

The Importance of the Direct to Reverberant Ratio in the Perception of Distance, Localization, Clarity, and Envelopment

Presenter: **David Griesinger**

The Direct to Reverberant ratio (D/R)—the ratio of the energy in the first wave front to the reflected sound energy—is absent from most discussions of room acoustics. Yet only the direct sound (DS) provides information about the localization and distance of a sound source. This paper discusses the perception of DS in a reverberant field depends on the D/R and the time delay between the DS and the reverberant energy. Threshold data for DS perception will be presented, and the implications for listening rooms, hall design, and electronic enhancement will be discussed. We find that both clarity and envelopment depend on DS detection. In listening rooms the direct sound must be at least equal to the total reflected energy for accurate imaging. As the room becomes larger (and the time delay increases) the threshold goes down. Some conclusions: typical listening rooms benefit from directional loudspeakers, small concert halls should not have a shoe-box shape, early reflections need not be lateral, and electro acoustic enhancement of late reverberation may be vital in small halls.

Sunday, October 10

16:30–18:00

PAPER SESSION 10: MEASUREMENT AND ANALYSIS OF 3-D SOUND

- 10-1 Theoretical Study and Numerical Analysis of 3-D Sound Field Reproduction System Based on Directional Microphones and Boundary Surface Control**—*Toshiyuki Kimura, National Institute of Information and Communications Technology, Koganei, Tokyo, Japan*

Three-dimensional sound field reproduction using directional microphones and wave field synthesis can be used to synthesize wave fronts in a listening area using directional microphones and loudspeakers placed at the boundary of the area; the position of the loudspeakers is the same as that of microphones in this technique. Thus, it is very difficult to construct an audio-visual virtual reality system using this technique because the screen or display of the visual system cannot be placed at the position of loudspeakers. In order to reproduce the 3-D sound field of the listening area even when the loudspeakers are not placed at the boundary of the area, this paper proposes a 3-D sound field reproduction system using directional microphones and boundary surface control. Results of a computer simulation show that the proposed system can reproduce the 3-D sound field in the listening area more accurately than the conventional system.

- 10-2 Characteristics of Near-Field Head-Related Transfer Function for KEMAR**—*Guang-Zheng Yu, Bo-Sun Xie, Dan Rao, South China University of Technology, Guangzhou, China*

A spherical dodecahedron sound source was designed to approximate a point source. The resulted sound source is approximately non-directivity below 10 kHz; the multiple scattering caused by sound source are negligible, and low-frequency characteristics of sound source are improved. By using the sound source, the near-field head-related transfer function (HRTF) database of a Knowles Electronic Manikin for Acoustic Research (KEMAR) has been established. Based on the database, characteristics of the near-field HRTFs were analyzed and compared with those of the far-field HRTFs in frequency domain and time domain respectively. Finally, the variations of binaural localization cues including interaural level difference (ILD) and interaural time difference (ITD) with source distance were analyzed, so as to evaluate the distance

localization cues contained in the near-field HRTFs.

10-3 Application of Synchronous Averaging Method in Frequency Domain for Asynchronous System to Sound Distance Measurement—Hiroshi Koide, Qiusheng Xie, Kouichi Tsuchiya, Tomohiko Endo, Akira Ebisawa, Shokichiro Hino, Etani Electronics Co., Ltd., Tokyo, Japan

In this paper, first we introduce the principle of synchronous averaging in frequency domain that is applicable to asynchronous systems where the transmitting side and receiving side are not synchronized. Then we discuss the possibility of applying synchronous averaging for measuring the distance of real or virtual sound sources. If the position of real or virtual sound sources in asynchronous systems can be identified, then there will be no need to consider the wiring of transmitting side and receiving side. This works well in applications of measurement at two or more points over a large space. Here, we identify the virtual sound source in asynchronous systems by using an adjacent four points method.

Sunday, October 10

15:30 – 18:00

WORKSHOP 11

In Pursuit of Spatialized Sound in Games

Co-chairs: **Steven P. Martz**, THX Ltd.
Kazutaka Someya, Art Spirits Inc.

Panelists: *Kanako Kakino*, Namco Bandai Games Inc.
Tetsukazu Nakanishi, Namco Bandai Games Inc.
Masayuki Sato, Square Enix Co. Ltd.
Astushi Suganuma, Square Enix Co., Ltd.
Kazuya Takimoto, Capcom Co., Ltd.

Recently, it's been possible to create highly interactive surround environments in games due to the rapid technological advancements in the latest game consoles. Now we can enjoy high quality surround sound with a very realistic experience.

We usually create spatialized sound by mixing individual sound elements and use sound effects to get them to fit each scene of the game in real-time as it progresses. In order to accomplish that, we've been trying to combine various raw sounds and make full use of current technology. We, game audio producers, would like to create a sense of reality in spatialized sound. We think is a very important role for us.

Therefore, in this workshop, we'd like to focus on a "pursuit of spatialized sound in games" and take a good look at various examples of spatialized sounds that are based on current technological approaches and techniques. Also, we'll examine a range of issues and perspectives in game audio. All of our presenters are well-known game-audio creators with various backgrounds and work in Japan, USA, and Europe. Each panelist will supply an original presentation, which will be related to each culture and background. We plan to have a discussion as well as a Q&A session with the audience at the end of our workshop, hoping to enrich their view of game audio.

Sunday, October 10

18:00

CLOSING REMARKS

CONFERENCE SPONSORS



Ballad



FOSTER

FUJITSU TEN



OTARITEC

Pioneer

Panasonic
ideas for life

ROCK ON PRO



SINCE 1975
sona



藝大フレンズ

