

# AES 128<sup>th</sup> Convention Program

May 22 – 25, 2010

Novotel London West, London, UK

The AES has launched a new opportunity to recognize student members who author technical papers. The Student Paper Award Competition is based on the preprint manuscripts accepted for the AES convention.

Twenty-seven student-authored papers were nominated. The excellent quality of the submissions has made the selection process both challenging and exhilarating. This year we are recognizing two papers.

The award-winning student paper will be honored during the Convention, and the student-authored manuscript will be published in a timely manner in the *Journal of the Audio Engineering Society*.

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

(a) The paper was accepted for presentation at the AES 128th Convention.

(b) The first author was a student when the work was conducted and the manuscript prepared.

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

(d) The student will deliver the lecture or poster presentation at the Convention.

\* \* \* \* \*

## The co-winners of the 128th AES Convention Student Paper Award are:

**Perceptual Evaluation of Physical Predictors of the Mixing Time in Binaural Room Impulse Responses**—*Alexander Lindau, Linda Kosanke, Stefan Weinzierl*, Technical University of Berlin, Berlin, Germany  
Convention Paper 8089

*To be presented on Monday, May 24 in Session P17  
—Multichannel and Spatial Audio: Part 1*

**A Subjective Evaluation of the Minimum Audible Channel Separation in Binaural Reproduction Systems through Loudspeakers**—*Yesenia Lacouture Parodi, Per Rubak*, Aalborg University, Aalborg, Denmark  
Convention Paper 8104

*To be presented on Monday, May 24 in Session P19  
—Posters: Psychoacoustics and Listening Tests*

\* \* \* \* \*

## Student Event/Career Development STUDENT SCIENCE SPOT

Saturday, May 22 through Tuesday, May 25  
C4-Foyer

Ongoing presentations of student's non-commercial scientifically challenging works will be given throughout the Convention in an accessible way. They will offer proto-

types or demonstrations, including a student poster session. Within the scope of their presentation, the students will have the opportunity to present themselves and their institutes in an appropriate way.

**Saturday, May 22 09:00 Room Saint Julien**  
**Technical Committee Meeting on Microphones and Applications**

**Session P1 Saturday, May 22**  
**09:30 – 12:30 Room C3**

## HIGH PERFORMANCE AUDIO PROCESSING

Chair: **Neil Harris**, New Transducers Ltd. (NXT), Cambridge, UK

09:30

**P1-1 Model-Driven Development of Audio Processing Applications for Multi-Core Processors**—*Tiziano Leidi*<sup>1</sup>, *Thierry Heeb*<sup>2</sup>, *Marco Colla*<sup>1</sup>, *Jean-Philippe Thiran*<sup>3</sup>  
<sup>1</sup>ICIMSI-SUPSI, Manno, Switzerland  
<sup>2</sup>Digimath, Sainte-Croix, Switzerland  
<sup>3</sup>EPFL, Lausanne, Switzerland

Chip-level multiprocessors are still very young and available forecasts anticipate a strong evolution for the forthcoming decade. To exploit them, efficient and robust applications have to be built with the appropriate algorithms and software architectures. Model-driven development is able to lower some barriers toward applications that process audio in parallel on multi-cores. It allows using abstractions to simplify and mask complex aspects of the development process and helps avoid inefficiencies and subtle bugs. This paper presents some evolutions of Audio n-Genie, an open-source environment for model-driven development of audio processing applications, which has been recently enhanced with support for parallel processing on multi-cores.  
*Convention Paper 7961*

10:00

**P1-2 Real-Time Additive Synthesis with One Million Sinusoids Using a GPU**—*Lauri Savioja*<sup>1,2</sup>, *Vesa Välimäki*<sup>2</sup>, *Julius O. Smith III*<sup>3</sup>  
<sup>1</sup>NVIDIA Research, Helsinki, Finland  
<sup>2</sup>Aalto University School of Science and Technology, Espoo, Finland  
<sup>3</sup>Stanford University, Palo Alto, CA, USA

Additive synthesis is one of the fundamental sound synthesis techniques. It is based on the principle that each sound can be represented as a superposition of sine waves of different frequencies. That task can be done fully parallel and thus it is suitable for GPU (graphics processing unit) implementation. In this paper we show that it is possible to compute over one million unique sine waves in real-time using a current GPU. That performance depends on the applied buffer sizes, but close to the maximum result is reachable already with a buffer of 500 samples.

*Convention Paper 7962*

10:30

**P1-3 A GPGPU Approach to Improved Acoustic Finite Difference Time Domain Calculations—**

*Jamie A. S. Angus, Andrew Caunce, University of Salford, Salford, Greater Manchester, UK*

This paper shows how to improve the efficiency and accuracy of Finite Difference Time Domain acoustic simulation by both calculating the differences using spectral methods and performing these calculations on a Graphics Processing Unit (GPU) rather than a CPU. These changes to the calculation method result in an increase in accuracy as well as a reduction in computational expense. The recent advances in the way that GPU's are programmed (for example using CUDA on Nvidia's GPU) now make them an ideal platform on which to perform scientific computations at very high speeds and very low power consumption.

*Convention Paper 7963*

11:00

**P1-4 Digital Equalization Filter: New Solution to the Frequency Response Near Nyquist and Evaluation by Listening Tests—**

*Thorsten Schmidt,<sup>1</sup> Joerg Bitzer<sup>2</sup>*  
<sup>1</sup>Cube-Tec International, Bremen, Germany  
<sup>2</sup>Jade-University of Applied Sciences, Oldenburg, Germany

Current design methods for digital equalization filter face the problem of a frequency response increasingly deviating from their analog equivalent close to the Nyquist frequency. This paper deals with a new way to design equalization filters, which improve this behavior over the entire frequency range between 0 Hz (DC) and Nyquist. The theoretical approach is shown and examples of low pass, peak-, and shelving-filters are compared to state-of-the-art techniques. Listening tests were made to verify the audible differences and rate the quality of the different design methods.

*Convention Paper 7964*

11:30

**P1-5 Audio Equalization with Fixed-Pole Parallel Filters: An Efficient Alternative to Complex Smoothing—**

*Balázs Bank, Budapest University of Technology and Economics, Budapest, Hungary*

Recently, the fixed-pole design of parallel second-order filters has been proposed to accomplish arbitrary frequency resolution similar to

Kautz filters, at 2/3 of their computational cost. This paper relates the parallel filter to the complex smoothing of transfer functions. Complex smoothing is a well-established method for limiting the frequency resolution of audio transfer functions for analysis, modeling, and equalization purposes. It is shown that the parallel filter response is similar to the one obtained by complex smoothing the target response using a hanning window: a 1/3 octave resolution is achieved by using 3/2 pole pairs per octave in the parallel filter. Accordingly, the parallel filter can be either used as an efficient implementation of smoothed responses, or, it can be designed from the unsmoothed responses directly, eliminating the need of frequency-domain processing. In addition, the theoretical equivalence of parallel filters and Kautz filters is developed, and the formulas for converting between the parameters of the two structures are given. Examples of loudspeaker-room equalization are provided.

*Convention Paper 7965*

12:00

**P1-6 Rapid and Automated Development of Audio Digital Signal Processing Algorithms for Mobile Devices—**

*David Trainor, APTX, Belfast, N. Ireland, UK*

Software applications and programming languages are available to assist audio DSP algorithm developers and mobile device designers, including Matlab/Simulink, C/C++, and assembly languages. These tools provide some assistance for algorithmic experimentation and subsequent refinement to highly-efficient embedded software. However, a typical design flow is still highly iterative, with manual software recoding, translation, and optimization. This paper introduces software libraries and design techniques that integrate existing commercial audio algorithm design tools and permit intuitive algorithmic experimentation and automated translation of audio algorithms to efficient embedded software. These techniques have been incorporated into a new software framework, and the operation of this framework is described using the example of a custom audio coding algorithm targeted to a mobile audio device.

*Convention Paper 7966*

**Session P2**  
**09:30 – 12:30**

**Saturday, May 22**  
**Room C5**

**SOUND REINFORCEMENT AND ROOM ACOUSTICS**

Chair: **Glenn Leembruggen**, Acoustic Directions Pty Ltd., ICE Design, Sydney, Australia, and University of Sydney, Sydney, Australia

09:30

**P2-1 Simultaneous Soundfield Reproduction at Multiple Spatial Regions—**

*Yan Jennifer Wu, Thushara D. Abhayapala, The Australian National University, Canberra, Australia*

Reproduction of different sound fields simultaneously at multiple spatial regions is a complex problem in acoustic signal processing. In this paper we present a framework to recreate 2-D

sound fields simultaneously at multiple spatial regions using a single loudspeaker array. We propose a novel method of finding an equivalent global sound field that consists of each individual desired sound field by the spatial harmonic coefficients translation between the coordinate systems. This method makes full use of the available dimensionality of the sound field. Some fundamental limits are also revealed. Specifically, the number of spatial regions could be reproduced inside the single loudspeaker array as determined by the total dimensionality from each sound field.

*Convention Paper 7967*

10:00

**P2-2 Verification of Geometric Acoustics-Based Auralization Using Room Acoustics Measurement Techniques**—*Aglaia Foteinou,<sup>1</sup> Damian Murphy,<sup>1</sup> Anthony Masinton<sup>2</sup>*

<sup>1</sup>University of York, Heslington, UK

<sup>2</sup>University of York, King's Manor, York, UK

Geometric acoustics techniques such as ray-tracing, image-source, and variants are commonly used in the simulation and auralization of the acoustics of an enclosed space. The shortcomings of such techniques are well documented and yet they are still the methods most commonly used and accepted in architectural acoustic design. This work compares impulse response-based objective acoustic measures obtained from a 3-D model of a medieval English church using a combination of image-source and ray-tracing geometric acoustic methods, with measurements obtained within the actual space. The results are presented in both objective and subjective terms, and include an exploration of optimized boundary materials and source directivity characteristics with problems and limitations clarified.

*Convention Paper 7968*

10:30

**P2-3 Analysis Tool Development for the Investigation of Low Frequency Room Acoustics by Means of Finite Element Method**—*Christos Sevastiadis, George Kalliris, George Papanikolaou, Aristotle University of Thessaloniki, Thessaloniki, Greece*

The sound wave behavior at low frequencies in small sound recording, control, and reproduction rooms, results in resonant sound fields observed in both frequency and spatial domains. Finite Element Method (FEM) software applications provide potential analysis procedures for acoustics problems but there is a lack of tools focusing on room acoustics investigation. The present paper is the result of an attempt to develop a room acoustics modeling tool integrated with a FEM. Simple room construction, common acoustical treatments, and multiple source excitations are involved in modal and steady state analysis procedures in order to simplify the solution of problematic sound fields. The key parameters and features, as long as validation experimental results are presented.

*Convention Paper 7969*

11:00

**P2-4 Subwoofers in Rooms: Experimental Modal Analysis**—*Juha Backman, Nokia Corporation, Espoo, Finland*

The behavior of a loudspeaker in a room depends fully on the coupling of the loudspeaker to the room modes, which are in the subwoofer frequency range individually identifiable. The modes can be determined through computational methods if the surface properties and room geometry are simple enough, but systems with complex properties have to be analyzed experimentally. The paper describes the use of modal analysis techniques, usually applied only to two-dimensional structures, for three-dimensional spaces to determine experimentally the actual modes and their properties.

*Convention Paper 7970*

11:30

**P2-5 Subwoofer Positioning, Orientation, and Calibration for Large-Scale Sound Reinforcement**—*Adam Hill,<sup>1</sup> Malcolm Hawksford,<sup>1</sup> Adam P. Rosenthal,<sup>2</sup> Gary Gand<sup>2</sup>*

<sup>1</sup>University of Essex, Colchester, UK

<sup>2</sup>Gand Concert Sound, Glenview, IL, USA

It is often difficult to achieve even coverage at low-frequencies across a large audience area. To complicate matters, it is desirable to have tight control of the low-frequency levels on the stage. This is generally dealt with by using cardioid subwoofers. While this helps control the stage area, the audience area receives no clear benefit. This paper investigates how careful positioning, orientation, and calibration of a multiple subwoofer system can provide enhanced low-frequency coverage, both in the audience area and on the stage. The effects of placement underneath, on top of, and in front of the stage are investigated as well as the performance of systems consisting of both flown and ground-based subwoofers.

*Convention Paper 7971*

12:00

**P2-6 Live Measurements of Ground-Stacked Subwoofers' Performance**—*Elena Shabalina,<sup>1</sup> Mathias Kaiser,<sup>1</sup> Janko Ramuscak<sup>2</sup>*

<sup>1</sup>RWTH Aachen University, Aachen, Germany

<sup>2</sup>d&b audiotechnik GmbH, Backnang, Germany

Various software-based measurement systems can be used for live sound measurements with music or speech signals in occupied concert halls with sufficient accuracy. However for some special applications as measuring the performance of a ground-stacked bass array component of a multiple component sound reinforcement system in an occupied hall, these systems cannot be used directly, since it is not possible to run a bass array alone during a concert. A simple method to obtain the subwoofers' impulse response using program signal measurements of a full range PA system is proposed. The desired impulse response results from subtraction of full system impulse responses with and without subwoofers running. Required measurement conditions and limitations are discussed.

*Convention Paper 7972*



**Saturday, May 22 10:00 Room Saint Julien**  
**Technical Committee Meeting on Perception**  
**and Subjective Evaluation of Audio Signals**

**Session P3 Saturday, May 22**  
**10:30 – 12:00 C4-Foyer**

**POSTERS: RECORDING, PRODUCTION,**  
**AND REPRODUCTION—MULTICHANNEL**  
**AND SPATIAL AUDIO**

**10:30**

**P3-1 21-Channel Surround System Based on**  
**Physical Reconstruction of a Three**  
**Dimensional Target Sound Field—***Jeongil*

*Seo,<sup>1</sup> Jae-hyoun Yoo,<sup>1</sup> Kyeongok Kang,<sup>1</sup>*  
*Filippo M. Fazl<sup>2</sup>*

<sup>1</sup>ETRI, Daejeon, Korea

<sup>2</sup>University of Southampton, Southampton, UK

This paper presents the 21-channel sound field reconstruction system based on the physical reconstruction of a three dimensional target sound field over the pre-defined control volume. According to the virtual sound source position and intensity, each loudspeaker signal is estimated through convolving with appropriate FIR filter to reconstruct a target sound field. In addition, the gain of FIR filter is only applied to the mid frequency band of a sound source signal to prevent aliasing effects and to save the computational complexity at the high frequency bands. Also the whole filter processing is carried out at the frequency domain to adopt a real-time application. Through the subjective listening tests the proposed system showed better performance on the localization in the horizontal plane comparing with conventional panning method.

*Convention Paper 7973*

**10:30**

**P3-2 Real-Time Implementation of Wave Field**  
**Synthesis on NU-Tech Framework Using**  
**CUDA Technology —***Ariano Lattanzi,<sup>1</sup>*

*Emanuele Ciavattini,<sup>1</sup> Stefania Cecchi,<sup>2</sup>*  
*Laura Romoli,<sup>2</sup> Fabrizio Ferrand<sup>3</sup>*

<sup>1</sup>Leaff Engineering, Ancona, Italy

<sup>2</sup>Università Politecnica delle Marche, Ancona, Italy

<sup>3</sup>Politecnico di Milano, Milan, Italy

In this paper we present a novel implementation of a Wave Field Synthesis application based on emerging NVIDIA Compute Unified Device Architecture (CUDA) technology using NU-Tech Framework. CUDA technology unlocks the processing power of the Graphics Processing Units (GPUs) that are characterized by a highly parallel architecture. A wide range of complex algorithms are being re-written in order to benefit from this new approach. Wave Field Synthesis is a quite new spatial audio rendering technique highly demanding in terms of computational power. We present here results and comparisons between a NU-Tech Plug-In (NUTS) implementing real-time WFS using CUDA libraries and the same algorithm implemented using Intel Integrated Primitives (IPP) Library.

*Convention Paper 7974*

**10:30**

**P3-3 Investigation of 3-D Audio Rendering with**  
**Parametric Array Loudspeakers—***Reuben*

*Johannes, Jia-Wei Beh, Woon-Seng Gan,*

*Ee-Leng Tan, Nanyang Technological*

*University, Singapore*

This paper investigates the applicability of parametric array loudspeakers to render 3-D audio. Unlike conventional loudspeakers, parametric array loudspeakers are able to produce sound in a highly directional manner, therefore reducing inter-aural crosstalk and room reflections. The investigation is carried out by performing objective evaluation and comparison between parametric array loudspeakers, conventional loudspeakers, and headphones. The objective evaluation includes crosstalk measurement and binaural cue analysis using a binaural hearing model. Additionally, the paper also investigates how the positioning of the parametric array loudspeakers affects 3-D audio rendering.

*Convention Paper 7975*

**10:30**

**P3-4 Robust Representation of Spatial Sound in**  
**Stereo-to-Multichannel Upmix—***Se-Woon*

*Jeon,<sup>1</sup> Young-Cheol Park,<sup>2</sup> Seok-Pil Lee,<sup>3</sup>*

*Dae-Hee Youn<sup>1</sup>*

<sup>1</sup>Yonsei University, Seoul, Korea

<sup>2</sup>Yonsei University, Gangwon, Korea

<sup>3</sup>Korea Electronics Technology Institute (KETI), Seoul, Korea

This paper presents a stereo-to-multichannel upmix algorithm based on a source separation method. In the conventional upmix algorithms, panning source and ambient components are decomposed or separated by adaptive algorithm, i.e., least-squares (LS) or least-mean-square (LMS). Separation performance of those algorithms is easily influenced by primary to ambient energy ratio (PAR). Since PAR is time-varying, it causes the energy fluctuation of separated sound sources. To prevent this problem, we propose a robust separation algorithm using a pseudo inverse matrix. And we propose a novel post-scaling algorithm to compensate for the influence of interference with considering desired multichannel format. Performance of the proposed upmix algorithm is confirmed by subjective listening test in ITU 3/2 format.

*Convention Paper 7976*

**10:30**

**P3-5 Ambisonic Decoders; Is Historical Hardware**  
**the Future?—***Andrew J. Horsburgh, D. Fraser*

*Clark, University of the West of Scotland,*

*Paisley, UK*

Ambisonic recordings aim to create full sphere audio fields through using a multi-capsule microphone and algorithms based on a “metatheory” as proposed by Gerzon. Until recently, Ambisonic decoding was solely implemented using hardware. Recent advances in computing power now allow for software decoders to supersede hardware units. It is therefore of interest to determine which of the hardware or software decoders provide the most accurate decoding of Ambisonic

B-format signals. In this paper we present a comparison between hardware and software decoders with respect to their frequency and phase relationships to determine the most accurate reproduction. Results show that software is able to decode the files with little coloration compared to hardware circuits. It is possible to see which implementation of the analog or digital decoders match the behavioral characteristics of an "ideal" decoder.

*Convention Paper 7977*

10:30

**P3-6 Localization Curves in Stereo Microphone Techniques—Comparison of Calculations and Listening Tests Results—Magdalena Plewa, Grzegorz Pyda, AGH University of Science and Technology, Kraków, Poland**

Stereo microphone techniques are a simple and usable solution to reproduce music scenes while maintaining sound direction maintenance. To choose a proper microphone setup one needs to know the "recording angle" that could be easily determined from localization curves. Localization curves in different stereo microphone techniques can be determined with the use of calculations, which take into consideration interchannel level and time differences. Subjective listening tests were carried out to verify the mentioned calculations. Here we present and discuss the comparison between the calculations and the listening tests.

*Convention Paper 7978*

10:30

**P3-7 Investigation of Robust Panning Functions for 3-D Loudspeaker Setups—Johann-Markus Batke, Florian Keiler, Technicolor, Research, and Innovation, Hannover, Germany**

An accurate localization is a key goal for a spatial audio reproduction system. This paper discusses different approaches for audio playback with full spatial information in three dimensions (3-D). Problems of established methods for 3-D audio playback, like the Ambisonics mode matching approach or Vector Base Amplitude Panning (VBAP), are discussed. A new approach is presented with special attention to the treatment of irregular loudspeaker setups, as they are to be expected in real world scenarios like living rooms. This new approach leads to better localization of virtual acoustic sources. Listening tests comparing the new approach with standard mode matching and VBAP will be described in a companion paper.

*Convention Paper 7979*

**Workshop 1**  
10:30 – 12:30

**Saturday, May 22**  
Room C6

**AUDIO NETWORK CONTROL PROTOCOLS**

Chair: **Richard Foss**, Rhodes University, Grahamstown, South Africa

Panelists: *John Grant*, Nine Tiles, Cambridge, UK  
*Robby Gurdan*, Universal Media Access Networks (UMAN), Düsseldorf, Germany

*Stefan Ledergerber*, Harman Pro Audio, Washington, DC, USA

*Philip Nye*, Engineering Arts, Bournemouth, UK

*Andy W. Schmeder*, University of California at Berkeley, Berkeley, CA, USA

Digital audio networks have solved a number of problems related to the distribution of audio within a number of contexts, including recording studios, stadiums, convention centers, theaters, and live concerts. They provide cabling ease, better immunity to interference, and enhanced control over audio routing and signal processing when compared to analog solutions. There exist a number of audio network types and also a number of audio network protocols that define the messaging necessary for connection management and control. The problem with this range of protocol solutions is that a large number of professional audio devices are being manufactured without regard to global interoperability. In this workshop a panel of audio network protocol experts will share the features of audio network protocols that they are familiar with and what features should appear within a common protocol.

**Saturday, May 22 11:00 Room Saint Julien**  
**Technical Committee Meeting on Studio Practices and Production**

**Special Event**  
**AWARDS PRESENTATION AND KEYNOTE ADDRESS**

Saturday, May 22, 12:45 – 13:30 Room C3

**Opening Remarks:**

- Executive Director Roger Furness
- President Diemer de Vries
- Convention Chair Josh Reiss

**Program:**

- AES Awards Presentation
- Introduction of Keynote Speaker by Convention Chair Josh Reiss
- Keynote Address by Masataka Goto

**Awards Presentation**

Please join us as the AES presents Special Awards to those who have made outstanding contributions to the Society in such areas of research, scholarship, and publications, as well as other accomplishments that have contributed to the enhancement of our industry. The awardees are:

**Honorary Member of AES:**

- **George Martin**

**Citation:**

- **Salvador Castaneda Valdes**

**Board of Governors Award**

- **Jan Abildgaard Pedersen**
- **Hiroaki Suzuki**
- **Martin Wöhr**

**Fellowship Award**

- **John Grant**
- **Bozena Kostek**
- **Ville Pulkki**
- **Mark Yonge**

**Silver Medal Award**

- **Ronald Aarts**

**Keynote Speaker**

This year's Keynote Speaker is **Masataka Goto**. Masa- ➔

taka Goto is the leader of the Media Interaction Group at the National Institute of Advanced Industrial Science and Technology (AIST), Japan. In 1992 he was one of the first to start work on automatic music understanding, and has since been at the forefront of research in music technologies and music interfaces based on those technologies. Since 1998 he has also worked on speech recognition interfaces. He has published more than 160 papers in refereed journals and international conferences. Over the past 18 years, he has received 25 awards, including the Young Scientists' Prize, the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology, the Excellence Award in Fundamental Science of the DoCoMo Mobile Science Awards, and the Best Paper Award of the Information Processing Society of Japan (IPSJ). He has served as a committee member of over 60 scientific societies and conferences and was the General Chair of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009) and the Chair of the IPSJ Special Interest Group on Music and Computer (SIGMUS). The title of his speech is "Music Listening in the Future."

Music understanding technologies based on signal processing have the potential to enable new ways of listening to music—for everyone. We have developed several Active Music Listening interfaces to demonstrate how end users can benefit from these new technologies. Active Music Listening aims at allowing the user to understand better the music he or she listens to and to actively influence the listening experience. For example, our Active Music Listening interface "SmartMusicKIOSK" has a chorus-search function that enables the user to access directly his or her favorite part of a song (and to skip others) while viewing a visual representation of its musical structure. The software "LyricSynchroniser" can automatically synchronize song lyrics to a recording and highlights the phrase currently sung. A user can easily follow the current playback position and click on a word in the lyrics to listen to it. Given polyphonic sound mixtures taken from available music recordings, our interfaces thus enrich end-users' music listening experiences, and open up new ways of music listening in the future.

**Session P4**  
**14:00 – 18:00**

**Saturday, May 22**  
**Room C3**

### SPATIAL SIGNAL PROCESSING

Chair: **Francis Rumsey**

**14:00**

**P4-1 Classification of Time-Frequency Regions in Stereo Audio**—*Aki Härmä*, Philips Research Europe, Eindhoven, The Netherlands

The paper is about classification of time-frequency (TF) regions in stereo audio data by the type of mixture the region represents. The detection of the type of mixing is necessary, for example, in source separation, upmixing, and audio manipulation applications. We propose a generic signal model and a method to classify the TF regions into six classes that are different combinations of central, panned, and uncorrelated sources. We give an overview of traditional techniques for comparing frequency-domain data and propose a new approach for classification that is based on measures specially trained for the six classes. The performance of the new measures is studied and demonstrated using

synthetic and real audio data.  
*Convention Paper 7980*

**14:30**

**P4-2 A Comparison of Computational Precedence Models for Source Separation in Reverberant Environments**—*Christopher Hummersone, Russell Mason, Tim Brookes*, University of Surrey, Guildford, UK

Reverberation continues to be problematic in many areas of audio and speech processing, including source separation. The precedence effect is an important psychoacoustic tool utilized by humans to assist in localization by suppressing reflections arising from room boundaries. Numerous computational precedence models have been developed over the years and all suggest quite different strategies for handling reverberation. However, relatively little work has been done on incorporating precedence into source separation. This paper details a study comparing several computational precedence models and their impact on the performance of a baseline separation algorithm. The models are tested in a range of reverberant rooms and with a range of other mixture parameters. Large differences in the performance of the models are observed. The results show that a model based on interaural coherence produces the greatest performance gain over the baseline algorithm.  
*Convention Paper 7981*

**15:00**

**P4-3 Converting Stereo Microphone Signals Directly to MPEG-Surround**—*Christophe Tournery*,<sup>1</sup> *Christof Faller*,<sup>1</sup> *Fabian Kuech*,<sup>2</sup> *Jürgen Herre*<sup>2</sup>

<sup>1</sup>Illusonic LLC, Lausanne, Switzerland

<sup>2</sup>Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

We have previously proposed a way to use stereo microphones with spatial audio coding to record and code surround sound. In this paper we are describing further considerations and improvements needed to convert stereo microphone signals directly to MPEG Surround, i.e., a downmix signal plus a bit stream. It is described in detail how to obtain from the microphone channels the information needed for computing MPEG Surround spatial parameters and how to process the microphone signals to transform them to an MPEG Surround compatible downmix.  
*Convention Paper 7982*

**15:30**

**P4-4 Modification of Spatial Information in Coincident-Pair Recordings**—*Jeremy Wells*, University of York, York, UK

A novel method is presented for modifying the spatial information contained in the output from a stereo coincident pair of microphones. The purpose of this method is to provide additional decorrelation of the audio at the left and right replay channels for sound arriving at the sides of a coincident pair but to retain the imaging accuracy for sounds arriving to the front or rear or where the entire sound field is highly correlated.

Details of how this is achieved are given and results for different types of sound field are presented.

*Convention Paper 7983*

16:00

**P4-5 Unitary Matrix Design for Diffuse Jot Reverberators**—*Fritz Menzer, Christof Faller, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*

This paper presents different methods for designing unitary mixing matrices for Jot reverberators with a particular emphasis on cases where no early reflections are to be modeled. Possible applications include diffuse sound reverberators and decorrelators. The trade-off between effective mixing between channels and the number of multiply operations per channel and output sample is investigated as well as the relationship between the sparseness of powers of the mixing matrix and the sparseness of the impulse response.

*Convention Paper 7984*

16:30

**P4-6 Sound Field Indicators for Hearing Activity and Reverberation Time Estimation in Hearing Instruments**—*Andreas P. Streich,<sup>1</sup> Manuela Feilner,<sup>2</sup> Alfred Stirnemann,<sup>2</sup> Joachim M. Buhmann<sup>1</sup>*

<sup>1</sup>ETH Zurich, Zurich, Switzerland

<sup>2</sup>Phonak AG, Stäfa, Switzerland

Sound field indicators (SFI) are proposed as a new feature set to estimate the hearing activity and reverberation time in hearing instruments. SFIs are based on physical measurements of the sound field. A variant thereof, called SFI short-time statistics SFIST2, is obtained by computing mean and standard deviations of SFIs on 10 subframes. To show the utility of these feature sets for the mentioned prediction tasks, experiments are carried out on artificially reverberated recordings of a large variety of sounds encountered in daily life. In a classification scenario where the hearing activity is to be predicted, both SFI and SFIST2 yield clearly superior accuracy even compared to hand-tailored features used in state-of-the-art hearing instruments. For regression on the reverberation time, the SFI-based features yield a lower residual error than standard feature sets and reach the performance of specially designed features. The hearing activity classification is mainly based on the average of the SFIs, while the standard deviation over sub-window is used heavily to predict the reverberation time.

*Convention Paper 7985*

17:00

**P4-7 Stereo-to-Binaural Conversion Using Interaural Coherence Matching**—*Fritz Menzer, Christof Faller, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*

In this paper a method of converting stereo recordings to simulated binaural recordings is presented. The stereo signal is separated into coherent and diffuse sound based on the

assumption that the signal comes from a coincident symmetric microphone setup. The coherent part is reproduced using HRTFs and the diffuse part is reproduced using filters adapting the interaural coherence to the interaural coherence a binaural recording of diffuse sound would have.

*Convention Paper 7986*

17:30

**P4-8 Linear Simulation of Spaced Microphone Arrays Using B-Format Recordings**—*Andreas Walther, Christof Faller, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*

A novel approach for linear post-processing of B-Format recordings is presented. The goal is to simulate spaced microphone arrays by approximating and virtually recording the sound field at the position of each single microphone. The delays occurring in non-coincident recordings are simulated by translating an approximative plane wave representation of the sound field to the positions of the microphones. The directional responses of the spaced microphones are approximated by linear combination of the corresponding translated B-format channels.

*Convention Paper 7987*

**Session P5**  
14:00 – 18:00

**Saturday, May 22**  
**Room C5**

## LOUDSPEAKERS AND HEADPHONES: PART 1

Chair: **Mark Dodd**, GP Acoustics, UK

14:00

**P5-1 Micro Loudspeaker Behavior versus 6½-Inch Driver, Micro Loudspeaker Parameter Drift**—*Bo Rohde Pedersen, Aalborg University Esbjerg, Esbjerg, Denmark*

This study tested micro loudspeaker behavior from the perspective of loudspeaker parameter drift. The main difference between traditional transducers and micro loudspeakers, apart from their size, is their suspension construction. The suspension generally is a loudspeaker's most unstable parameter, and the study investigated temperature drift and signal dependency. We investigated three different micro loudspeakers and compared their behavior to that of a typical bass mid-range loudspeaker unit. There is measured all linear loudspeaker parameters at different temperatures.

*Convention Paper 7988*

14:30

**P5-2 Modeling a Loudspeaker as a Flexible Spherical Cap on a Rigid Sphere**—*Ronald Aarts,<sup>1,2</sup> Augustus J. E. M. Janssen<sup>2</sup>*

<sup>1</sup>Philips Research Europe, Eindhoven, The Netherlands

<sup>2</sup>Technical University Eindhoven, Eindhoven, The Netherlands

It has been argued that the sound radiation of a loudspeaker is modeled realistically by assuming



the loudspeaker cabinet to be a rigid sphere with a moving rigid spherical cap. Series expansions, valid in the whole space on and outside the sphere, for the pressure due to a harmonically excited, flexible cap with an axially symmetric velocity distribution are presented. The velocity profile is expanded in functions orthogonal on the cap rather than on the whole sphere. This has the advantage that only a few expansion coefficients are sufficient to accurately describe the velocity profile. An adaptation of the standard solution of the Helmholtz equation to this particular parametrization is required. This is achieved by using recent results on argument scaling in orthogonal Zernike polynomials. The efficacy of the approach is exemplified by calculating various acoustical quantities with particular attention to certain velocity profiles that vanish at the rim of the cap to a desired degree. These quantities are: the sound pressure, polar response, baffle-step response, sound power, directivity, and acoustic center of the radiator. The associated inverse problem, in which the velocity profile is estimated from pressure measurements around the sphere, is feasible as well since the number of expansion coefficients to be estimated is limited. This is demonstrated with a simulation.

*Convention Paper 7989*

15:00

**P5-3 Closed and Vented Loudspeaker Alignments that Compensate for Nearby Room**

**Boundaries**—*Jamie A. S. Angus*, University of Salford, Salford, Greater Manchester, UK

The purpose of this paper is to present a method of designing loudspeakers in which the presence of the nearest three boundaries are taken into account in the design of the low frequency speaker. The paper will first review the effects of the three boundaries. It will then discuss how these effects might be compensated for. The paper will then examine the low frequency behavior of loudspeaker drive units and make suggestions for new alignments, which take account of the boundaries. It will conclude with some simulation examples.

*Convention Paper 7990*

15:30

**P5-4 Point-Source Loudspeaker Reversely-Attached Acoustic Horn: Its Architecture, Acoustic Characteristics and Application to HRTF Measurements**

—*Takahiro Miura, Teruo Muraoka, Tohru Ifukube*, The University of Tokyo, Tokyo, Japan

It is ideal to measure acoustic characteristics by point-source sound. In the case when simultaneous recording of single source by multiple microphones located at under 1 m from the source, it is difficult to regard the loudspeaker as a point-source. In this paper we propose a point-source loudspeaker whose radiation diameter is smaller than 2 cm. The loudspeaker is designed to attach the mouse of hyperbolic horn to the diaphragm of a loudspeaker unit. Directional patterns of the proposed was measured at a distance of 50 cm from the radiation point in

anechoic chamber. As a result, the difference of directional intensity at the frequency range of 20 – 700 Hz were within 3 dB at any combination of azimuth and elevation. At the frequency range over 700 Hz, difference of azimuthal directional intensity were within 10 dB while that of the elevational ones were within 20 dB.

*Convention Paper 7991*

16:00

**P5-5 The Low-Frequency Acoustic Center: Measurement, Theory, and Application—**

*John Vanderkooy*, University of Waterloo, Waterloo, Ontario, Canada

At low frequencies the acoustic effect of a loudspeaker becomes simpler as the wavelength of the sound becomes large relative to the cabinet dimensions. One point acoustically acts as the center of the speaker at the lower frequencies. Measurements and acoustic boundary-element simulations verify the concept. Source radiation can be expressed as a multipole expansion, consisting of a spherical monopolar portion and a significant dipolar part, which becomes zero when the acoustic centre is chosen as the origin. Theory shows a strong connection between diffraction and the position of the acoustic center. General criteria are presented to give the position of the acoustic center for different geometrical cabinet shapes. Polar plots benefit when the pivot point is chosen to be the acoustic center. For the first of several applications we consider a subwoofer, whose radiation into a room is strongly influenced by the position of the acoustic center. A second application that we consider is the accurate reciprocity calibration of microphones, for which it is necessary to know the position of the acoustic center. A final application is the effective position of the ears on the head at lower frequencies. Calculations and measurements show that the acoustic centers of the ears are well away from the head.

*Convention Paper 7992*

16:30

**P5-6 Prediction of Harmonic Distortion Generated by Electro-Dynamic Loudspeakers Using Cascade of Hammerstein Models—**

*Marc Rebillat*,<sup>1,2</sup> *Romain Hennequin*,<sup>3</sup> *Etienne Corteel*,<sup>4</sup> *Brian F. G. Katz*<sup>1</sup>

<sup>1</sup>LIMSI-CRNS, Orsay, France

<sup>2</sup>LMS (CNRS, École Polytechnique), Palaiseau, France

<sup>3</sup>Institut TELECOM, TELECOM ParisTech, Paris, France

<sup>4</sup>sonic emotion, Oberglatt (Zurich), Switzerland

Audio rendering systems are always slightly nonlinear. Their nonlinearities must be modeled and measured for quality evaluation and control purposes. Cascade of Hammerstein models describes a large class of nonlinearities. To identify the elements of such a model, a method based on a phase property of exponential sine sweeps is proposed. A complete model of nonlinearities is identified from a single measurement. Cascade of Hammerstein models corresponding to an electro-dynamic loudspeaker are identified this way. Harmonic distortion is afterward pre-



dicted using the identified models. Comparisons with classical measurement techniques show that harmonic distortion is accurately predicted by the identified models over the entire audio frequency range for any desired input amplitude.  
*Convention Paper 7993*

17:00

**P5-7 Characterizing Studio Monitor Loudspeakers for Auralization**—*Ben Supper*, Focusrite Audio Engineering Ltd., High Wycombe, UK

A method is presented for obtaining the frequency and phase response, directivity pattern, and some of the nonlinear distortion characteristics of studio monitor loudspeakers. Using a specially-designed test signal, the impulse response and directivity pattern are measured in a small recording room. A near-field measurement is also taken. An algorithm is presented for combining the near- and far-field responses in order to compute out the early reflections of the room. Doppler distortion can be calculated using recorded and measured properties of the loudspeaker. The result is a set of loudspeaker impulse and directional responses that are detailed enough for convincing auralization.  
*Convention Paper 7994*

17:30

**P5-8 Automated Design of Loudspeaker Diaphragm Profile by Optimizing the Simulated Radiated Sound Field with Experimental Validation**—

*Patrick Macey*,<sup>1</sup> *Kelvin Griffiths*<sup>2</sup>

<sup>1</sup>PACSYS Limited, Nottingham, UK

<sup>2</sup>Harman International Automotive Division, Bridgend, UK

Loudspeaker designers often aim to reduce acoustic effects of diaphragm resonance from the frequency and power response. However, even in the domain of simulation models, this often requires considerable trial and error to produce a satisfactory outcome. A modeling platform to simulate loudspeakers is presented, which can automatically cycle through vast permutations guided by a carefully considered objective function until convergence is achieved, and importantly, the entire iterative effort is accommodated by the computer. Experience and knowledge must make an initial evaluation of which factors are important in achieving a desired result a priori. Once prepared, the basis of a rapid loudspeaker development tool is formed that can return a tangible resource profit when compared to successive manual efforts to optimize loudspeaker components.  
*Convention Paper 7995*

**Session P6**  
14:00 – 15:30

**Saturday, May 22**  
C4-Foyer

**POSTERS: AUDIO EQUIPMENT  
AND EMERGING TECHNOLOGIES**

14:00

**P6-1 Study and Evaluation of MOSFET  $R_{ds(ON)}$ –Impedance Efficiency Losses in High Power**

**Multilevel DCI-NPC Amplifiers**—*Vicent Sala*, *German Ruiz*, *Luis Romeral*, UPC-Universitat Politecnica de Catalunya, Terrassa, Spain

This paper justifies the usefulness of multilevel power amplifiers with DCI-NPC (Diode Clamped Inverter– Neutral Point Converter) topology in applications where size and weight needs were optimized. These amplifiers can work at high frequencies thereby reducing the size and weight of the filter elements. However, it is necessary to study, analyze, and evaluate the efficiency losses because this amplifier has double the number of switching elements. This paper models the behavior of the MOSFET  $R_{ds(ON)}$  in a DCI-NPC topology for different conditions.  
*Convention Paper 7996*  
*Paper presented by German Ruiz.*

14:00

**P6-2 Modeling Distortion Effects in Class-D Amplifier Filter Inductors**—*Arnold Knott*,<sup>1</sup> *Tore Stegenborg-Andersen*,<sup>1</sup> *Ole C. Thomsen*,<sup>1</sup> *Dominik Bortis*,<sup>2</sup> *Johann W. Kolar*,<sup>2</sup> *Gerhard Pfaffinger*,<sup>3</sup> *Michael A. E. Andersen*<sup>1</sup>

<sup>1</sup>Technical University of Denmark, Lyngby, Denmark

<sup>2</sup>Swiss Federal Institute of Technology in Zurich, Zurich, Switzerland

<sup>3</sup>Harman/Becker Automotive Systems GmbH, Straubing, Germany

Distortion is generally accepted as a quantifier to judge the quality of audio power amplifiers. In switch-mode power amplifiers various mechanisms influence this performance measure. After giving an overview of those, this paper focuses on the particular effect of the nonlinearity of the output filter components on the audio performance. While the physical reasons for both, the capacitor and the inductor induced distortion are given, the practical in-depth demonstration is done for the inductor only. This includes measuring the inductors performance, modeling through fitting and resulting into simulation models. The fitted models achieve distortion values between 0.03 % and 0.20 % as a basis to enable the design of a 200 W amplifier.  
*Convention Paper 7997*

14:00

**P6-3 Multilevel DCI-NPC Power Amplifier High-Frequency Distortion Analysis through Parasitic Inductance Dynamic Model**—*Vicent Sala*, *German Ruiz*, *E. López*, *Luis Romeral*, UPC-Universitat Politecnica de Catalunya, Terrassa, Spain

The high frequency distortion sources in DCI-NPC (Diode Clamped Inverter-Neutral Point Converter) amplifiers topology are studied and analyzed. It has justified the need for designing a model that contains the different parasitic inductive circuits that presents dynamically this kind of amplifier, as a function of the combination of its active transistors. By means of a proposed pattern layout we present a dynamic model of the parasitic inductances of the amplifier Full-Bridge DCI-NPC, and this is used to propose some simple rules for the optimal designing of layouts for these types of amplifiers.

Simulation and experimental results are presented to justify the proposed model, and the affirmations and recommendations are given in this paper.

*Convention Paper 7998*

*Paper presented by German Ruiz.*

14:00

- P6-4 How Much Gain Should a Professional Microphone Preamp Have?—Douglas McKinnie**, Middle Tennessee State University, Murfreesboro, TN, USA

Many tradeoffs are required in the design of microphone preamplifier circuits. Characteristics such as noise figure, stability, bandwidth, and complexity may be dependent upon the gain of the design. Three factors determine the gain required from a microphone preamp: sound-pressure level of the sound source, distance of the microphone from that sound source (within the critical distance), and sensitivity of the microphone. This paper is an effort to find a probability distribution of the gain settings used with professional microphones. This is done by finding the distribution of max SPL in real use and by finding the sensitivity of the most commonly used current and classic microphones.

*Convention Paper 7999*

14:00

- P6-5 Equalizing Force Contributions in Transducers with a Partitioned Electrode—Libor Husník**, Czech Technical University in Prague, Prague, Czech Republic

A partitioned electrode in an electrostatic transducer can present among others a possibility for making the transducer with the direct D/A conversion. Nevertheless, partitioned electrodes, the sizes of which are proportional to powers of 2 or terms of other convenient series, do not have the corresponding force action on the membrane. The reason is the membrane does not vibrate in a piston-like mode and electrode parts close to the membrane periphery do not excite membrane vibrations in the same way as the elements near the center. The aim of this paper is to suggest equalization of force contributions from different partitioned electrodes by varying their sizes. Principles presented here can also be used for other membrane-electrode arrangements.

*Convention Paper 8000*

14:00

- P6-6 Low-End Device to Convert EEG Waves to MIDI—Adrian Attard Trevisan,<sup>1</sup> Lewis Jones<sup>2</sup>**  
<sup>1</sup>St. Martins Institute of Information Technology, Hamrun, Malta  
<sup>2</sup>London Metropolitan University, London, UK

This research provides a simple and portable system that is able to generate MIDI output based on the inputted data collected through an EEG collecting device. The context is beneficial in many ways, where the therapeutic effects of listening to the music created by the brain waves documents many cases of treating health problems. The approach is influenced by the interface described

in the article “Brain-Computer Music Interface for Composition and Performance” by Eduardo Reck Miranda, where different frequency bands trigger corresponding piano notes through, and the complexity of, the signal represents the tempo of the sound. The correspondence of the sound and the notes have been established through experimental work, where data of participants of a test group were gathered and analyzed, putting intervals for brain frequencies for different notes. The study is an active contribution to the field of the neurofeedback, by providing criteria tools for assessment.

*Convention Paper 8001*

*Paper will be presented by Lewis Jones*

14:00

- P6-7 Implementation and Development of Interfaces for Music Performance through Analysis of Improvised Dance Movements—Richard Hoadley**, Anglia Ruskin University, Cambridge, UK

Electronic music, even when designed to be interactive, can lack performance interest and is frequently musically unsophisticated. This is unfortunate because there are many aspects of electronic music that can be interesting, elegant, demonstrative, and musically informative. The use of dancers to interact with prototypical interfaces comprising clusters of sensors generating music algorithmically provides a method of investigating human actions in this environment. This is achieved through collaborative work involving software and hardware designers, composers, sculptors, and choreographers who examine aesthetically and practically the interstices of these disciplines. This paper investigates these interstices.

*Convention Paper 8002*

14:00

- P6-8 Violence Prediction through Emotional Speech—José Higuera-Soler, Roberto Gil-Pita, Enrique Alexandre, Manuel Rosa-Zurera**, Universidad de Alcalá, Alcalá de Henares, Madrid, Spain

Preventing violence takes an absolute necessity in our society. Whether in homes with a particular risk of domestic violence, as in prisons or schools, there is a need for systems capable of detecting risk situations, for preventive purposes. One of the most important factors that precede a violent situation is an emotional state of anger. In this paper we discuss the features that are required to provide decision makers dedicated to the detection of emotional states of anger from speech signals. For this purpose, we present a set of experiments and results with the aim of studying the combination of features extracted from the literature and their effects over the detection performance (relationship between probability of detection of anger and probability of false alarm) of a neural network and a least-square linear detector.

*Convention Paper 8003*

14:00

- P6-9 FoleySonic: Placing Sounds on a Timeline through Gestures—David Black,<sup>1</sup> Kristian**

Gohlke,<sup>1</sup> Jörn Loviscach<sup>2</sup>

<sup>1</sup>University of Applied Sciences, Bremen,  
Bremen, Germany

<sup>2</sup>University of Applied Sciences, Bielefeld,  
Bielefeld, Germany

The task of sound placement on video timelines is usually a time-consuming process that requires the sound designer or foley artist to carefully calibrate the position and length of each sound sample. For novice and home video producers, friendlier and more entertaining input methods are needed. We demonstrate a novel approach that harnesses the motion-sensing capabilities of readily available input devices, such as the Nintendo Wii Remote or modern smart phones, to provide intuitive and fluid arrangement of samples on a timeline. Users can watch a video while simultaneously adding sound effects, providing a near real-time workflow. The system leverages the user's motor skills for enhanced expressiveness and provides a satisfying experience while accelerating the process.  
*Convention Paper 8004*

14:00

**P6-10 A Computer-Aided Audio Effect Setup Procedure for Untrained Users—***Sebastian*

*Heise,<sup>1</sup> Michael Hlatky,<sup>1</sup> Jörn Loviscach<sup>2</sup>*

<sup>1</sup>University of Applied Sciences, Bremen,  
Bremen, Germany

<sup>2</sup>University of Applied Sciences, Bielefeld,  
Bielefeld, Germany

The number of parameters of modern audio effects easily ranges in the dozens. Expert knowledge is required to understand which parameter change results in a desired effect. Yet, such sound processors are also making their way into consumer products, where they overburden most users. Hence, we propose a procedure to achieve a desired effect without technical expertise based on a black-box genetic optimization strategy: Users are only confronted with a series of comparisons of two processed examples. Learning from the users' choices, our software optimizes the parameter settings. We conducted a study on hearing-impaired persons without expert knowledge, who used the system to adjust a third-octave equalizer and a multi-band compressor to improve the intelligibility of a TV set.

*Convention Paper 8005*

**Workshop 2**  
14:00 – 16:00

**Saturday, May 22**  
Room C2

**AES42 AND DIGITAL MICROPHONES**

Chair: **Helmut Wittek**, SCHOEPS Mikrofone GmbH,  
Karlsruhe, Germany

Panelists: *Claudio Becker-Foss*, DirectOut, Mittweida,  
Germany  
*Stephan Flock*, DirectOut, Mittweida,  
Germany  
*Christian Langen*, Schoeps Mikrofone,  
Karlsruhe, Germany  
*Stephan Peus*, Georg Neumann GmbH,

Berlin, Germany

*Gregor Zielinski*, Sennheiser, Wedemark,  
Germany

The AES42 interface for digital microphones is not yet widely used. This can be due to the relatively new appearance of digital microphone technology, but also a lack of knowledge and practice with digital microphones and the corresponding interface. The advantages and disadvantages have to be communicated in an open and neutral way, regardless of commercial interests, on the basis of the actual need of the engineers.

Along with an available "White paper" about AES42 and digital microphones, which is aimed a neutral in-depth information and was compiled from different authors, the workshop intends to bring to light facts and prejudices on this topic.

**Tutorial 1**  
14:00 – 15:30

**Saturday, May 22**  
Room C6

**DO-IT-YOURSELF SEMANTIC AUDIO**

Presenter: **Jörn Loviscach**, Fachhochschule Bielefeld  
(University of Applied Sciences), Bielefeld,  
Germany

Content-based music information retrieval (MIR) and similar applications require advanced algorithms that often overburden non-expert developers. However, many building blocks are available—mostly for free—to significantly ease software development, for instance of similarity search methods, or to serve as components for ad-hoc solutions, for instance in forensics or linguistics. This tutorial looks into software libraries/frameworks (e.g., MARSYAS and CLAM), toolboxes (e.g., MIRtoolbox), Web-based services (e.g., Echo Nest), and stand-alone software (e.g., The Sonic Visualizer) that help with the extraction of audio features and/or execute basic machine learning algorithms. Focusing on solutions that require little to no programming in the classical sense, the tutorial's major part consists in live demos of hand-picked routes to roll one's own semantic audio application.

**Saturday, May 22** 14:00 **Room Saint Julien**  
**Technical Committee Meeting on Coding of Audio Signals**

**Saturday, May 22** 14:00 **Room Mouton Cadet**  
**Standards Committee Meeting SC-02-02 Digital Input/Output Interfacing**

**Saturday, May 22** 15:00 **Room Saint Julien**  
**Technical Committee Meeting on Archiving, Restoration, and Digital Libraries**

**Tutorial 2**  
15:30 – 17:00

**Saturday, May 22**  
Room C1

**MASTERING FOR BROADCAST**

Presenter: **Darcy Proper**, Galaxy Studios, Mol, Belgium

Mastering has often had an aura of mystery around it. Those "in the know" have always regarded it as a vital and necessary last step in the process of producing a record. Those who have never experienced it have often had only a vague idea of what good mastering could achieve. However, the loudness race in recent years has put the mastering community under pressure; on one side from the producers or labels who want their product

louder than the competition and on the other side from the artists or mixers who don't want their work smashed into a lifeless "brick" and maxed out by excessive use of limiter plug-ins.

Darcy Proper is a multi-Grammy-winning mastering engineer whose golden ears (and hands) have put the finishing touches on a vast array of high profile records, among many others those from Steely Dan. She will talk about her approach to her work and will also demo examples with various degrees of compression with a legacy broadcast processor as the final piece of gear in the signal chain, simulating a radio broadcast. The audience is then able to experience the effects and artifacts that different compression levels will cause at the consumer's end.

**Workshop 3**  
**16:00 – 18:00**

**Saturday, May 22**  
**Room C6**

### **CALIBRATION OF ANALOG REPRODUCING EQUIPMENT FOR DIGITIZATION PROJECTS**

Chair: **George Brock-Nannestad**, Patent Tactics, Gentofte, Denmark

Panelists: *Sean W. Davies*, S.W. Davies Ltd., Aylesbury, UK  
*Andrew Pearson*, British Library Sound Archive, London, UK  
*Peter Posthumus*, Post Sound Ltd., Newton Abbot, UK

Analog reproduction equipment sees its last use in the transfer of audio content to digital files. In order to obtain the optimal signal from the carrier (tape and disc) the equipment has to be adjusted. In order to obtain traceability the equipment has to be calibrated. Generations of technicians that did this on a regular basis are vanishing quickly, and making available calibration materials, such as the AES-S001-064 Coarse-groove Calibration disc, is not enough when the know-how for its use is lacking. This workshop aims to redress this situation, in that it provides a firm theoretical underpinning of practical demonstrations, in which the basic principles of calibration and alignment will be discussed along with logging the activity. Keywords are: calibration and alignment of analog tape reproducers, including equipment for tapes that have been encoded, such as Dolby-A and Dolby SR. Furthermore calibration and alignment of gramophone record reproducing equipment will be discussed. Active discussion is encouraged and may spill over into optical sound as found on films.

**Saturday, May 22**      **16:00**      **Room Saint Julien**  
**Technical Committee Meeting on Hearing and Hearing Loss Prevention**

**Session P7**      **Saturday, May 22**  
**16:30 – 18:00**      **C4-Foyer**

### **POSTERS: LOUDSPEAKERS AND HEADPHONES**

**16:30**

**P7-1**      **Theoretical and Experimental Comparison of Amplitude Modulation Techniques for Parametric Loudspeakers—Wei Ji, Woon-Seng Gan, Peifeng Ji**, Nanyang Technological University, Singapore

Due to the self-demodulation property of finite amplitude ultrasonic waves propagating in air, a highly focused sound beam can be reproduced. This phenomenon is known as the parametric array in air. However, because of the nonlinearity effect of ultrasonic waves, the reproduced sound wave suffers from high distortion. Several amplitude modulation techniques have been previously proposed to mitigate this problem. Currently, no comprehensive study has been carried out to evaluate the performance of these amplitude modulation techniques. This paper attempts to provide theoretical and experimental studies of the effectiveness in mitigating the distortion from different amplitude modulation techniques. Objective measurements, such as sound pressure level and total harmonic distortion are used to compare the performance of these techniques.  
*Convention Paper 8006*

**16:30**

**P7-2**      **Improvement of Sound Quality by Means of Ultra-Soft Elastomer for the Gel Type Inertia Driven DML Type Transducer—Minsung Cho,<sup>1,2</sup> Elena Prokofieva,<sup>1</sup> Jordi Munoz,<sup>2</sup> Mike Barker<sup>1</sup>**  
<sup>1</sup>Edinburgh Napier University, Edinburgh, UK  
<sup>2</sup>SFX Technologies Ltd., Edinburgh, UK

Unlike standard DML transducers, the gel-type inertia driven transducer (referred to as the gel transducer in this paper) designed as a mini woofer DML-type transducer, transfers its piston movement to a panel through the gel surround, thereby generating a transverse wave. This mechanism induces maximum movement of a magnet assembly, which boosts the force of the moving voice coil so that sound pressure level increases as the acceleration of the panel increases. This effect is proportional to sound pressure level at low and medium frequencies in the range 50 Hz to 1000 Hz. Furthermore it is found that the gel surround prevents reflected transverse waves from the panel into interfering with the piston waves generated by the transducer. Results of the stiffness testing of the gel surround are presented together with data for the displacement and acceleration of the panel with the gel transducer attached. These data are compared to acoustical outputs.  
*Convention Paper 8007*

**16:30**

**P7-3**      **Electrical Circuit Model for a Loudspeaker with an Additional Fixed Coil in the Gap—Daniele Ponteggia,<sup>1</sup> Marco Carlisi,<sup>2</sup> Andrea Manzin<sup>3</sup>**  
<sup>1</sup>Studio Ing. Ponteggia, Terni, Italy  
<sup>2</sup>Ing. Carlisi Marco, Como, Italy  
<sup>3</sup>18 Sound - Division of A.E.B. S.r.l., Cavriago, Italy

A previous paper by some of the authors investigated a solution to minimize the loudspeaker inductance based on an additional fixed coil positioned in the gap with two additional terminals. This device is referred as A.I.C. (Active Impedance Control). In this paper a suitable electrical circuit model for such loudspeaker with four terminals and two coils is proposed.  
*Convention Paper 8008*



16:30

- P7-4 Measurement of the Nonlinear Distortions in Loudspeakers with a Broadband Noise—**  
*Rafal Siczek, Andrzej B. Dobrucki, Wrocław University of Technology, Wrocław, Poland*

The paper presents application of digital filters for measurement of nonlinear distortion in loudspeakers using Wolf's method and broadband noise as the exciting signal. The results of simulation of the measurement process for loudspeaker with various nonlinearities of BI factor and suspension stiffness are presented. The influence of various exciting signals, e.g., white and pink noise as well as the parameters of digital filters have been tested. The results of measurement of an actual loudspeaker are also presented.

*Convention Paper 8009*

16:30

- P7-5 The Fundamentals of Loudspeaker Radiation and Acoustic Quality—**  
*Grzegorz P. Matusiak, New Smart System (AC Systems), Poznan, Poland and Smørum, Denmark*

The loudspeaker designer faces many challenges when designing a new loudspeaker system. The proposed solution is to replace the mathematically complicated series circuit of the radiation impedance by a new one, that can easily be introduced into a loudspeaker circuit in the whole frequency range. It facilitates a designing process and enables to learn more about the loudspeaker. The fundamental resonance in air is lower than in vacuum and depends on the radiating surface, as ACOUSTIC QUALITY, which also depends on frequency.  $1/QT=1/QE+1/QM+1/QA$ . The parameters affect the SPL and efficiency response. Investigated structures: pulsating sphere, baffled circular, square and rectangular piston, infinite strip, unflanged pipe-piston, finally horn (0-infinity Hz).

*Convention Paper 8010*

16:30

- P7-6 Analysis of Low Frequency Audio Reproduction via Multiple Low-BI Loudspeakers—**  
*Theodore Altanis, John Mourjopoulos, University of Patras, Patras, Greece*

Small loudspeakers with low force factors (BI) behave as narrow bandwidth filters with high quality factors and can be tuned at low resonance frequencies. It is possible to optimize low frequency reproduction (in the sub-woofer range) using a number of such drivers. This paper employs a set of such low BI loudspeakers with different eigenfrequencies to achieve efficient reproduction of low frequency signals and examines the aspects of such implementation comparing the results to an ideal sub-woofer system and also via perceptual evaluation of these systems.

*Convention Paper 8011*

16:30

- P7-7 Circular Loudspeaker Array with Controllable Directivity—**  
*Martin Møller,<sup>1</sup> Martin Olsen,<sup>1</sup>*

*Finn T. Agerkvist,<sup>1</sup> Jakob Dyreby,<sup>2</sup> Gert Kudahl Munch<sup>2</sup>*

<sup>1</sup>Technical University of Denmark, Lyngby, Denmark

<sup>2</sup>Bang & Olufsen A/S, Struer, Denmark

Specific directivity patterns for circular arrays of loudspeakers can be achieved by utilizing the concept of phase-modes, which expands the directivity pattern into a series of circular harmonics. This paper investigates the applicability of this concept applied on a loudspeaker array on a cylindrical baffle, with a desired directivity pattern, which is specified in the frequency interval of 400–5000 Hz. From the specified frequency independent directivity pattern filter transfer functions for each loudspeaker are determined. The sensitivity of various parameters, related to practical implementation, is also investigated by introducing filter errors for each array element. Measurements on a small scale model using 2 inch drivers are compared with simulations, showing good agreement between experimental and predicted results.

*Convention Paper 8012*

### **Student Event/Career Development OPENING AND STUDENT DELEGATE ASSEMBLY MEETING – PART 1**

Saturday, May 22, 17:00 – 18:00  
Room C2

Chair: **Miroslav Jakovljevic**

Vice Chair: **Veit Winkler**

The first Student Delegate Assembly (SDA) meeting is the official opening of the Convention's student program and a great opportunity to meet with fellow students from all corners of the world. This opening meeting of the Student Delegate Assembly will introduce new events and election proceedings, announce candidates for the coming year's election for the Europe/International Regions, announce the finalists in the recording competition categories, hand out the judges' sheets to the nonfinalists, and announce any upcoming events of the Convention. Students and student sections will be given the opportunity to introduce themselves and their activities, in order to stimulate international contacts.

All students and educators are invited to participate in this meeting. Election results and Recording Competition and Poster Awards will be given at the Student Delegate Assembly Meeting—Part 2 on Tuesday, May 25, at 14:00.

### **Saturday, May 22 14:00 Room Mouton Cadet Standards Committee Meeting SC-02-08 Audio-File Transfer and Exchange**

### **Special Event MIXER PARTY**

Saturday, May 22, 18:00 – 19:30  
C4-Foyer

A Mixer Party will be held on Saturday evening to enable Convention attendees to meet in a social atmosphere to catch up with friends and colleagues from the world of audio. There will be a cash bar.

### **Special Event ORGAN RECITAL BY GRAHAM BLYTH**

Saturday, May 22, 20:00 – 21:30  
Temple Church, Temple, London

Graham Blyth's traditional organ recital will be given on the organ of the famous Temple Church, a late 12th century church in London located between Fleet Street and the River Thames, built for and by the Knights Templar as their English headquarters. In modern times, two Inns of Court (Inner Temple and Middle Temple) both use the church. It is famous for its effigy tombs and for being a round church. It was heavily damaged during the Second World War but has been largely restored. The church contains two organs: a chamber organ built by Robin Jennings in 2001 (a three-stop continuo organ), and a four manual Harrison & Harrison organ. The Harrison organ will be completely renovated from July 2011 until Easter 2013.

The featured works are by Hubert Parry, George Thalben-Ball, Charles Villiers Stanford, Frank Bridge, Percy Whitlock, Louis Vierne, and Leon Beëllmann.

Graham Blyth was born in 1948, began playing the piano aged 4 and received his early musical training as a Junior Exhibitioner at Trinity College of Music in London, England. Subsequently, at Bristol University, he took up conducting, performing Bach's St. Matthew Passion before he was 21. He holds diplomas in Organ Performance from the Royal College of Organists, The Royal College of Music and Trinity College of Music. In the late 1980s he renewed his studies with Sulemita Aronowsky for piano and Robert Munns for organ. He gives numerous concerts each year, principally as organist and pianist, but also as a conductor and harpsichord player. He made his international debut with an organ recital at St. Thomas Church, New York in 1993 and since then has played in San Francisco (Grace Cathedral), Los Angeles (Cathedral of Our Lady of Los Angeles), Amsterdam, Copenhagen, Munich (Liebfrauentempel), Paris (Madeleine and St. Etienne du Mont) and Berlin. He has lived in Wantage, Oxfordshire, since 1984 where he is currently Artistic Director of Wantage Chamber Concerts and Director of the Wantage Festival of Arts.

He divides his time between being a designer of professional audio equipment (he is a co-founder and Technical Director of Soundcraft) and organ related activities. In 2006 he was elected a Fellow of the Royal Society of Arts in recognition of his work in product design relating to the performing arts.

## Student Event/Career Development

### STUDENT PARTY

Saturday, May 7, 20:00 –22:00

TBD

Information will be given at the Student Delegate Assembly —Part 1.

**Session P8**  
**09:00 – 13:00**

**Sunday, May 23**  
**Room C3**

## MUSIC ANALYSIS AND PROCESSING

Chair: **David Malham**, University of York, York, UK

**09:00**

- P8-1 Automatic Detection of Audio Effects in Guitar and Bass Recordings—Michael Stein,<sup>1</sup> Jakob Abeßer,<sup>1</sup> Christian Dittmar,<sup>1</sup> Gerald Schuller<sup>2</sup>**  
<sup>1</sup>Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany  
<sup>2</sup>Ilmenau University of Technology, Ilmenau, Germany

This paper presents a novel method to detect and distinguish 10 frequently used audio effects in recordings of electric guitar and bass. It is based on spectral analysis of audio segments located in the sustain part of previously detected guitar tones. Overall, 541 spectral, cepstral and harmonic features are extracted from short time spectra of the audio segments. Support Vector Machines are used in combination with feature selection and transform techniques for automatic classification based on the extracted feature vectors. With correct classification rates up to 100 percent for the detection of single effects and 98 percent for the simultaneous distinction of 10 different effects, the method has successfully proven its capability—performing on isolated sounds as well as on multitimbral, stereophonic musical recordings.

*Convention Paper 8013*

**09:30**

### **P8-2 Time Domain Emulation of the Clavinet—**

*Stefan Bilbao,<sup>1</sup> Matthias Rath<sup>2</sup>*

<sup>1</sup>University of Edinburgh, Edinburgh, UK

<sup>2</sup>Technische Universität Berlin, Berlin, Germany

The simulation of classic electromechanical musical instruments and audio effects has seen a great deal of activity in recent years, due in part to great recent increases in computing power. It is now possible to perform full emulations of relatively complex musical instruments in real time, or near real time. In this paper time domain finite difference schemes are applied to the emulation of the Hohner Clavinet, an electromechanical stringed instrument exhibiting special features such as sustained hammer/string contact, pinning of the string to a metal stop, and a distributed damping mechanism. Various issues, including numerical stability, implementation details, and computational cost will be discussed. Simulation results and sound examples will be presented.

*Convention Paper 8014*

**10:00**

### **P8-3 Polyphony Number Estimator for Piano Recordings Using Different Spectral Patterns**

*—Ana M. Barbancho, Isabel Barbancho, Javier Fernandez, Lorenzo J. Tardón, Universidad de Málaga, Málaga, Spain*

One of the main tasks of a polyphonic transcription system is the estimation of the number of voices, i.e., the polyphony number. The correct estimation of this parameter is very important for polyphonic transcription systems, this task has not been discussed in depth in the known transcription systems. The aim of this paper is to propose a novel estimation method of the polyphony number for piano recordings. This new method is based on the use of two different types of spectral patterns: single-note patterns and composed-note patterns. The usage of composed-note patterns in the estimation of the polyphony number and in the polyphonic detection process has not been previously reported in the literature.

*Convention Paper 8015*

10:30

**P8-4 String Ensemble Vibrato: A Spectroscopic Study**—*Stijn Mattheij*, AVANS University, Breda, The Netherlands

A systematic observation of the presence of ensemble vibrato on early twentieth century recordings of orchestral works has been carried out by studying spectral line shapes of individual musical notes. Broadening of line shapes was detected in recordings of Beethoven's Fifth Symphony and Brahms's Hungarian Dance no. 5; this effect was attributed to ensemble vibrato. From these observations it may be concluded that string ensemble vibrato was common practice in orchestras from the continent throughout the twentieth century. British orchestras did not use much vibrato before 1940.

*Convention Paper 8016*

11:00

**P8-5 Influence of Psychoacoustic Roughness on Musical Intonation Preference**—*Julián Villegas*,<sup>1</sup> *Michael Cohen*,<sup>1</sup> *Ian Wilson*,<sup>1</sup> *William Martens*<sup>2</sup>

<sup>1</sup>University of Aizu, Aizu, Japan

<sup>2</sup>University of Sydney, Sydney, NSW, Australia

An experiment to compare the acceptability of three different music fragments rendered with three different intonations is presented. These preference results were contrasted with those of isolated chords also rendered with the same three intonations. The least rough renditions were found to be those using Twelve-Tone Equal-Temperament (12-TET). Just Intonation (JI) renditions were the roughest. A negative correlation between preference and psychoacoustic roughness was also found.

*Convention Paper 8017*

11:30

**P8-6 Music Emotion and Genre Recognition Toward New Affective Music Taxonomy**—*Jonghwa Kim*, *Lars Larsen*, University Augsburg, Augsburg, Germany

Exponentially increasing electronic music distribution creates a natural pressure for fine-grained musical metadata. On the basis of the fact that a primary motive for listening to music is its emotional effect, diversion, and the memories it awakens, we propose a novel affective music taxonomy that combines the global music genre taxonomy, e.g., classical, jazz, rock/pop, and rap, with emotion categories such as joy, sadness, anger, and pleasure, in a complementary way. In this paper we deal with all essential stages of automatic genre/emotion recognition system, i.e., from reasonable music data collection up to performance evaluation of various machine learning algorithms. Particularly, a novel classification scheme, called consecutive dichotomous decomposition tree (CDDT) is presented, which is specifically parameterized for multi-class classification problems with extremely high number of class, e.g., sixteen music categories in our case. The average recognition accuracy of 75% for the 16 music categories shows a realistic possibility of the affective music

taxonomy we proposed.  
*Convention Paper 8018*

12:00

**P8-7 Perceptually-Motivated Audio Morphing: Warmth**—*Duncan Williams*, *Tim Brookes*, University of Surrey, Guildford, UK

A system for morphing the warmth of a sound independently from its other timbral attributes was coded, building on previous work morphing brightness only, and morphing brightness and softness. The new warmth-softness-brightness morpher was perceptually validated using a series of listening tests. A multidimensional scaling analysis of listener responses to paired-comparisons showed perceptually orthogonal movement in two dimensions within a warmth-morphed and everything-else-morphed stimulus set. A verbal elicitation experiment showed that listeners' descriptive labeling of these dimensions was as intended. A further "quality control" experiment provided evidence that no "hidden" timbral attributes were altered in parallel with the intended ones. A complete timbre morpher can now be considered for further work and evaluated using the tri-stage procedure documented here.

*Convention Paper 8019*

12:30

**P8-8 A Novel Envelope-Based Generic Dynamic Range Compression Model**—*Adam Weisser*, Oticon A/S, Smørum, Denmark

A mathematical model is presented, which reproduces typical dynamic range compression, when given the nominal input envelope of the signal and the compression constants. The model is derived geometrically in a qualitative approach and the governing differential equation for an arbitrary input and an arbitrary compressor is found. Step responses compare well to commercial compressors tested. The compression effect on speech using the general equation in its discrete version is also demonstrated. This model applicability is especially appealing to hearing aids, where the input-output curve and time constants of the nonlinear instrument are frequently consulted and the qualitative theoretical effect of compression may be crucial for speech perception.

*Convention Paper 8020*

**Session P9**  
**09:00 – 13:00**

**Sunday, May 23**  
**Room C5**

**ROOM AND ARCHITECTURAL ACOUSTICS**

Chair: **Ben Kok**, Consultant, The Netherlands

09:00

**P9-1 On the Air Absorption Effects in a Finite Difference Implementation of the Acoustic Diffusion Equation Model**—*Juan M. Navarro*,<sup>1</sup> *José Escolano*,<sup>2</sup> *José J. Lopez*<sup>3</sup>

<sup>1</sup>San Antonio's Catholic University of Murcia,



Guadalupe, Spain

<sup>2</sup>University of Jaén, Linares, Spain

<sup>3</sup>Universidad Politécnica de Valencia, Valencia, Spain

In room-acoustics modeling, the sound atmospheric attenuation is a critical phenomenon that has to be taken into account to obtain correct predictions, especially when high frequencies and large enclosures are simulated. A finite difference scheme implementation of the diffusion equation model with a mixed boundary condition is evaluated, including the air absorption within the room. This paper focuses on investigating the performance of this implementation for room-acoustics simulation. In particular, the stability condition is developed to compare the features of the solution both with and without the air absorption term. Moreover, the correct behavior of the numerical implementation has been studied comparing predicted results using different surfaces and air absorption coefficients.

*Convention Paper 8021*

09:30

**P9-2 A Comparison of Two Diffuse Boundary Models Based on Finite Differences Schemes**

—José Escolano,<sup>1</sup> Juan M. Navarro,<sup>2</sup> Damian T. Murphy,<sup>3</sup> Jeremy J. Wells,<sup>3</sup> José J. López<sup>4</sup>

<sup>1</sup>University of Jaén, Linares, Spain

<sup>2</sup>San Antonio's Catholic University of Murcia, Guadalupe, Spain

<sup>3</sup>University of York, York, UK

<sup>4</sup>Universidad Politécnica de Valencia, Valencia, Spain

In room acoustics, the reflection and scattering of sound waves at the boundaries strongly determines the behavior within the enclosed space itself. Therefore, the accurate simulation of the acoustic characteristics of a boundary is an important part of any room acoustics prediction model, and in particular diffuse reflections at a boundary is one of the most important properties to model correctly for an accurate and perceptually natural result. This paper presents a comparison of two simulation models for the phenomenon of diffuse reflection at a boundary: one based on introducing physical variations at the boundary itself, and another based on the use of a diffusion equation model.

*Convention Paper 8022*

10:00

**P9-3 Considerations for the Optimal Location and Boundary Effects for Loudspeakers in an Automotive Interior**—Roger Shively,<sup>1</sup> Jeff Bailey,<sup>1</sup> Jérôme Halley,<sup>2</sup> Lars Kurandt,<sup>2</sup> François Malbos,<sup>3</sup> Gabriel Ruiz,<sup>4</sup> Alfred Svobodnik<sup>5</sup>

<sup>1</sup>Harman International, Novi, MI, USA

<sup>2</sup>Harman International, Karlsbad, Germany

<sup>3</sup>Harman International, Chateau du Loir, France

<sup>4</sup>Harman International, Bridgend, Wales, UK

<sup>5</sup>Harman International, Vienna, Austria

Referencing earlier work by the authors on the boundary effects in an automotive vehicle interior (AES Convention Paper 4245, May 1996) on the mid-to-high frequency timbral changes in the sound field due to the proximity to loudspeakers

of reflective, semi-rigid surfaces, modeling of midsize loudspeakers in the interior of an automobile is reported on, as well as modeled results for a specific case are given.

*Convention Paper 8023*

10:30

**P9-4 Acoustics of the Restored Petruzzelli Theater**

—Marco Facondini,<sup>1</sup> Daniele Ponteggia<sup>2</sup>

<sup>1</sup>TanAcoustics Studio, Pesaro, Italy

<sup>2</sup>Studio Ing. Ponteggia, Terni, Italy

Petruzzelli theater in Bari has been recently restored after the disastrous fire of 1991 had seriously damaged the building. The restoration has focused on the aesthetics and functionality of the room, in particular giving great attention to improving the acoustics of the theater. This paper reviews the acoustical design process that has been carried out using a computer model of the hall and measurements during the restoration process. Objective indexes from measurements of the renewed theater are compared with literature suggested values and with similar halls.

*Convention Paper 8024*

11:00

**P9-5 Identification of a Room Impulse Response Using a Close-Microphone Reference Signal**

—Elias Kokkinis, John Mourjopoulos, University of Patras, Patras, Greece

The identification of room impulse responses is very important in many audio signal processing applications. Typical system identification methods require access to the original source signal that is seldom available in practice, while blind system identification methods require prior knowledge of the statistical properties of the source signal that, for audio applications, are not available. In this paper a new system identification method is proposed using a close-microphone reference signal, and it is shown that it accurately identifies real room impulse responses.

*Convention Paper 8025*

11:30

**P9-6 Acoustic Impulse Response Measurement Using Speech and Music Signals**—John Usher, Barcelona Media, Barcelona, Spain

Continuous measurement of room impulse responses (RIRs) in the presence of an audience has many applications for room acoustics: in-situ loudspeaker/room equalization; teleconferencing; and for architectural acoustic diagnostics. A continuous analysis of the RIR is often preferable to a single measurement, especially with non-stationary room characteristics such as from changing atmospheric or audience conditions. This paper discusses the use of adaptive filters updated according to the NLMS algorithm for fast, continuous in-situ RIR acquisition; particularly when the input signal is music or speech. We show that the dual-channel FFT (DCFFT) method has slower convergence and is less robust to colored signals such as music and speech. Data is presented comparing the NLMS and the DCFFT methods and we show that the



adaptive filter approach provides RIRs with high accuracy and high robustness to background noise using music or speech signals.  
*Convention Paper 8026*

12:00

**P9-7 Evaluation of Late Reverberant Fields in Loudspeaker Rendered Virtual Rooms—**  
*Wieslaw Woszczyk, Brett Leonard, Doyuen Ko, McGill University, Montreal, Quebec, Canada*

Late reverberant decay was isolated from high-resolution impulse response measured in a church that has excellent acoustics suitable for music recording. The measured multichannel impulse response (IR) was compared to two synthetic decays (SIR) built from Gaussian noise and modeled on the original IR. The synthetic decays were created to be similar to the original response in the rate of decay and spectral weighting in order to make listening comparisons between them. Four different transition points (100 ms, 200 ms, 300 ms, 400 ms) were chosen to crossfade between the original early response and the synthetic decays. Listeners auditioned the synthetic and measured rooms convolved with three anechoic monophonic sources. The results of tests conducted within immersive surround sound environment with height help to verify whether shaped random noise is a suitable substitute for late reverberation in high-resolution simulations of room acoustics.  
*Convention Paper 8027*  
*This paper is presented by Brett Leonard.*

12:30

**P9-8 Finite Difference Room Acoustic Modeling on a General Purpose Graphics Processing Unit—**  
*Alexander Southern,<sup>1,2</sup> Damian Murphy,<sup>1</sup> Guilherme Campos,<sup>3</sup> Paulo Dias<sup>3</sup>*

<sup>1</sup>University of York, York, UK  
<sup>2</sup>Aalto University School of Science and Technology, Aalto, Finland  
<sup>3</sup>University of Aveiro, Aveiro, Portugal

Detailed and convincing walkthrough auralizations of virtual rooms requires much processing capability. One method of reducing this requirement is to pre-calculate a data-set of room impulse responses (RIR) at locations throughout the space. Processing resources may then focus on RIR interpolation and convolution using the dataset as the virtual listening position changes in real-time. Recent work identified the suitability of wave-based models over traditional ray-based approaches for walkthrough auralization. Despite the computational saving of wave-based methods to generate the RIR dataset, processing times are still long. This paper presents a wave-based implementation for execution on a general purpose graphics processing unit. Results validate the approach and show that parallelization provides a notable acceleration.  
*Convention Paper 8028*

**Session P10**  
**09:00 – 10:30**

**Sunday, May 23**  
**C4-Foyer**

**POSTERS: AUDIO PROCESSING—ANALYSIS AND SYNTHESIS OF SOUND**

09:00

**P10-1 Cellular Automata Sound Synthesis with an Extended Version of the Multitype Voter Model—**  
*Jaime Serquera, Eduardo R. Miranda, University of Plymouth, Plymouth, UK*

In this paper we report on the synthesis of sounds with cellular automata (CA), specifically with an extended version of the multitype voter model (MVM). Our mapping process is based on DSP analysis of automata evolutions and consists in mapping histograms onto sound spectrograms. This mapping allows a flexible sound design process, but due to the non-deterministic nature of the MVM such process acquires its maximum potential after the CA run is finished. Our extended version model presents a high degree of predictability and controllability making the system suitable for an in-advance sound design process with all the advantages that this entails, such as real-time possibilities and performance applications. This research focuses on the synthesis of damped sounds.  
*Convention Paper 8029*

09:00

**P10-2 Stereophonic Rendering of Source Distance Using DWM-FDN Artificial Reverberators—**  
*Saul Maté-Cid, Hüseyin Hacıhabiboglu, Zoran Cvetkovic, King's College London, London, UK*

Artificial reverberators are used in audio recording and production to enhance the perception of spaciousness. It is well known that reverberation is a key factor in the perception of the distance of a sound source. The ratio of direct and reverberant energies is one of the most important distance cues. A stereophonic artificial reverberator is proposed that allows panning the perceived distance of a sound source. The proposed reverberator is based on feedback delay network (FDN) reverberators and uses a perceptual model of direct-to-reverberant (D/R) energy ratio to pan the source distance. The equivalence of FDNs and digital waveguide mesh (DWM) scattering matrices is exploited in order to devise a reverberator relevant in the room acoustics context.  
*Convention Paper 8030*

09:00

**P10-3 Separation of Music+Effects Sound Track from Several International Versions of the Same Movie —**  
*Antoine Liutkus,<sup>1</sup> Pierre Leveau<sup>2</sup>*  
<sup>1</sup>Télécom ParisTech, Paris, France  
<sup>2</sup>Audionamix, Paris, France

This paper concerns the separation of the music+effects (ME) track from a movie soundtrack, given the observation of several international versions of the same movie. The approach chosen is strongly inspired from existing stereo audio source separation and especially from spatial filtering algorithms such as DUET that can extract a constant panned source from a mixture very efficiently. The problem is indeed similar for we aim here at separating the ME track, which is the common background of all international versions of the movie soundtrack. The algorithm has been adapted to a number of channels greater than 2. Preprocessing tech-

niques have also been proposed to adapt the algorithm to realistic cases. The performances of the algorithm have been evaluated on realistic and synthetic cases.

*Convention Paper 8031*

09:00

**P10-4 A Differential Approach for the Implementation of Superdirective Loudspeaker Array**—*Jung-Woo Choi, Youngtae Kim, Sangchul Ko, Jungho Kim*, SAIT, Samsung Electronics Co. Ltd., Gyeonggi-do, Korea

A loudspeaker arrangement and corresponding analysis method to obtain a robust superdirective beam are proposed. The superdirectivity technique requires precise matching of the sound sources modeled to calculate excitation patterns and those used for the loudspeaker array. To resolve the robustness issue arising from the modeling mismatch error, we show that the overall sensitivity to the model-mismatch error can be reduced by rearranging loudspeaker positions. Specifically, a beam pattern obtained by a conventional optimization technique is represented as a product of robust delay-and-sum patterns and error-sensitive differential patterns. The excitation pattern driving the loudspeaker array is then reformulated such that the error-sensitive pattern is only applied to the outermost loudspeaker elements, and the array design that fits to the new excitation pattern is discussed.

*Convention Paper 8032*

09:00

**P10-5 Improving the Performance of Pitch Estimators**—*Stephen J. Welburn, Mark D. Plumbley*, Queen Mary University of London, London, UK

We are looking to use pitch estimators to provide an accurate high-resolution pitch track for resynthesis of musical audio. We found that current evaluation measures such as gross error rate (GER) are not suitable for algorithm selection. In this paper we examine the issues relating to evaluating pitch estimators and use these insights to improve performance of existing algorithms such as the well-known YIN pitch estimation algorithm.

*Convention Paper 8033*

09:00

**P10-6 Reverberation Analysis via Response and Signal Statistics**—*Eleftheria Georganti, Thomas Zarouchas, John Mourjopoulos*, University of Patras, Patras, Greece

This paper examines statistical quantities (i.e., kurtosis, skewness) of room transfer functions and audio signals (anechoic, reverberant, speech, music). Measurements are taken under various reverberation conditions in different real enclosures ranging from small office to a large auditorium and for varying source-receiver positions. Here, the statistical properties of the room responses and signals are examined in the frequency domain. From these properties, the rela-

tionship between the spectral statistics of the room transfer function and the corresponding reverberant signal are derived.

*Convention Paper 8034*

09:00

**P10-7 An Investigation of Low-Level Signal Descriptors Characterizing the Noise-Like Nature of an Audio Signal**—*Christian Uhle*, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

This paper presents an overview and an evaluation of low-level features characterizing the noise-like or tone-like nature of an audio signal. Such features are widely used for content classification, segmentation, identification, coding of audio signals, blind source separation, speech enhancement, and voice activity detection. Besides the very prominent Spectral Flatness Measure various alternative descriptors exist. These features are reviewed and the requirements for these features are discussed. The features in scope are evaluated using synthetic signals and exemplarily real-world application related to audio content classification, namely voiced-unvoiced discrimination for speech signals and speech detection.

*Convention Paper 8035*

09:00

**P10-8 Algorithms for Digital Subharmonic Distortion**—*Zlatko Baracskai*,<sup>1,2</sup> *Ryan Stables*<sup>1</sup>  
<sup>1</sup>Birmingham City University, Birmingham, UK  
<sup>2</sup>University of Birmingham, Birmingham, UK

This paper presents a comparison between existing digital subharmonic generators and a new algorithm developed with the intention of having a more pronounced subharmonic frequency and reduced harmonic, intermodulation and aliasing distortions. The paper demonstrates that by introducing inversions of a waveform at the minima and maxima instead of the zero crossings, the discontinuities are mitigated and various types of distortion are significantly attenuated.

*Convention Paper 8036*

**Workshop 4**  
09:00 – 11:00

**Sunday, May 23**  
**Room C1**

**BLU-RAY AS A HIGH RESOLUTION AUDIO FORMAT FOR STEREO AND SURROUND**

Chair: **Stefan Bock**, EBB-msm-studios GmbH, Munich, Germany

Panelists: *Simon Heyworth*, SAM, Chagford, UK  
*Morten Lindberg*, 2L, Oslo, Norway  
*Johannes Müller*, msm-studios, Munich, Germany  
*Crispin Murray*, Metropolis, London, UK  
*Ronald Prent*, Galaxy Studios, Mol, Belgium

The decision for the Blu-ray disc as the only HD packaged media format also offering up to 8 channels of uncompressed high resolution audio has at least eliminated one of the obstacles of getting high-res surround sound music to the market. The concept of utilizing Blu-

ray as a pure audio format will be explained, and Blu-ray will be positioned as successor to both SACD and DVD-A. The operational functionality and a double concept of making it usable both with and without screen will be demonstrated by showing a few products that are already on the market.

**Tutorial 3** **Sunday, May 23**  
**09:00 – 11:00** **Room C2**

### HEARING AND HEARING LOSS PREVENTION

Presenter: **Benj Kanters**, Columbia College, Chicago, IL, USA

The Hearing Conservation Seminar and HearTomorrow.Org are dedicated to promoting awareness of hearing loss and conservation. This program is specifically targeted to students and professionals in the audio and music industries. Experience has shown that engineers and musicians easily understand the concepts of hearing physiology, as many of the principles and theories are the same as those governing audio and acoustics. Moreover, these people are quick to understand the importance of developing their own safe listening habits, as well as being concerned for the hearing health of their clients and the music-listening public. The tutorial is a 2-hour presentation in three sections: first, an introduction to hearing physiology; the second, noise-induced loss; and third, practicing effective and sensible hearing conservation.

**Sunday, May 23** **09:00** **Room Saint Julien**  
**Technical Committee Meeting on Spatial Audio**

**Sunday, May 23** **10:00** **Room Saint Julien**  
**Technical Committee Meeting on Transmission and Broadcasting**

**Workshop 5** **Sunday, May 23**  
**11:00 – 12:45** **Room C6**

### APPLICATIONS OF TIME-FREQUENCY PROCESSING IN SPATIAL AUDIO

Chair: **Ville Pulkki**, Aalto University School of Science and Technology, Aalto, Finland

Panelists: *Christof Faller*, Illusonic LLC, Lausanne, Switzerland  
*Jean-Marc Jot*, DTS Inc., Scotts Valley, CA, USA  
*Christian Uhle*, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

The time-frequency resolution of human hearing has been taken into account for a long time in perceptual audio codecs. Recently, the spatial resolution of humans has also been exploited in time-frequency processing. This has already led into some commercial applications. 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. The application areas discussed in the workshop include, at least, audio coding, microphone techniques, upmixing, directional microphones, and studio effects.

**Special Event**  
**AES/APRS—LIFE IN THE OLD DOGS YET—**  
**PART ONE: KEEPING STUDIOS ALIVE**  
Sunday, May 23, 11:00 – 13:00  
Room C2

Moderator: **Andrew Leyshon**, Nottingham University

Panelists: *Mark Anders*, Bug Music, UK  
*Malcolm Atkin*, formerly AIR and Sphere Studios  
*Ian Brenchley*, MD, Metropolis Studios  
*Paul Brown*, Paul Brown Management  
*Jonathan Smith*, GM, Abbey Road Studios

This Special Event examines the trends in the professional recording business in recent years and invites a new generation of recording entrepreneurs to share ideas often driven by the expanding needs of contemporary clients. Professor Andrew Leyshon from Nottingham University presents his thought-provoking analysis of the changing worldwide recording climate. This is followed by a distinguished panel of recording stakeholders discussing current and possible future recording business models.

**Student Event/Career Development**  
**CAREER/JOB FAIR**

Sunday, May 23, 11:00 – 13:00  
C4-Foyer

The Career Fair will feature several companies from the exhibit floor. All attendees of the Convention, students and professionals alike, are welcome to come talk with representatives from the companies and find out more about job and internship opportunities in the audio industry. Bring your resume!

**Sunday, May 23** **11:00** **Room Saint Julien**  
**Technical Committee Meeting on Human Factors in Audio Systems**

**Sunday, May 23** **11:00** **Room Mouton Cadet**  
**Standards Committee Meeting SC-05-02 Audio Connectors**

**Tutorial 4** **Sunday, May 23**  
**11:30 – 13:00** **Room C1**

**CANCELED**

**Sunday, May 23** **12:00** **Room Saint Julien**  
**Technical Committee Meeting on Signal Processing**

**Sunday, May 23** **12:30** **Room Mouton Cadet**  
**Standards Committee Meeting SC-05-05 Grounding and EMC Practices**

**Sunday, May 23** **13:00** **Room Saint Julien**  
**Technical Committee Meeting on Audio Recording and Mastering Systems**

**Student Event/Career Development**  
**BENEFIT OF ACCREDITATION—JAMES**

Sunday, May 23, 13:15 – 13:45  
Room C5

Moderator: **Phil Harding**  
**Dave Ward**

*Will a Qualification Get You a Job in Music Production?* Not so long ago, to get a career in music production meant you starting at the bottom in a recording studio pushing a broom and making the tea, slowly absorbing the skills until you eventually graduated from tape op to assistant engineer, finally reaching the rank of producer. Now there are hundreds of degree courses and thousands of graduates looking for jobs in far fewer commercial studios than existed before —has anything changed? Is it worth the investment of putting yourself through college if you then have to start at the bottom again. And are there any jobs anyway?

JAMES is dedicated to making links between Education and the Media Industries. JAMES is the education arm of the Association of Professional Recording Services (APRS), the Music Producers Guild (MPG) and the UK Screen Association. JAMES is made up of dedicated industry professionals who wish to ensure that many years of professional experience are not lost to future generations.

**Session P11**  
**13:30 – 18:30**

**Sunday, May 23**  
**Room C3**

### NETWORK, INTERNET, AND BROADCAST AUDIO

Chair: **Bob Walker**, Consultant, UK

**13:30**

**P11-1 A New Technology for the Assisted Mixing of Sport Events: Application to Live Football Broadcasting**—*Giulio Cengarle, Toni Mateos, Natanael Olaiz, Pau Arumí*, Barcelona Media Innovation Centre, Barcelona, Spain

This paper presents a novel application for capturing the sound of the action during a football match by automatically mixing the signals of several microphones placed around the pitch and selecting only those microphones that are close to, or aiming at, the action. The sound engineer is presented with a user interface where he or she can define and move dynamically the point of interest on a screen representing the pitch, while the application controls the faders of the broadcast console. The technology has been applied in the context of a three-dimensional surround sound playback of a Spanish first-division match.  
*Convention Paper 8037*

**14:00**

**P11-2 Recovery Time of Redundant Ethernet-Based Networked Audio Systems**—*Maciej Janiszewski, Piotr Z. Kozłowski*, Wrocław University of Technology, Wrocław, Poland

Ethernet-based networked audio systems has become more popular among audio system designers. One of the most important issues that is available from networked audio system is the redundancy. The system can recover after different types of failures—cable or even device failure. Redundancy protocols implemented by audio developers are different than protocols known from computer networks, but both may be used in an Ethernet-based audio system. The most important attribute of redundancy in the audio system is a recovery time. This paper is a summary of research that was done at Wrocław

University of Technology. It shows the recovery time after different types of failures, with different network protocols implemented, for all possible network topologies in CobraNet and EtherSound systems.

*Convention Paper 8038*

**14:30**

**P11-3 Upping the Auntie: A Broadcaster's Take on Ambisonics**—*Chris Baume, Anthony Churnside*, British Broadcasting Corporation Research & Development, London, UK

This paper considers Ambisonics from a broadcaster's point of view: to identify barriers preventing its adoption within the broadcast industry and explore the potential advantages were it to be adopted. This paper considers Ambisonics as a potential production and broadcast technology and attempts to assess the impact that the adoption of Ambisonics might have on both production workflows and the audience experience. This is done using two case studies: a large-scale music production of "The Last Night of the Proms" and a smaller scale radio drama production of "The Wonderful Wizard of Oz." These examples are then used for two subjective listening tests: the first to assess the benefit of representing height allowed by Ambisonics and the second to compare the audience's enjoyment of first order Ambisonics to stereo and 5.0 mixes.  
*Convention Paper 8039*

**15:00**

**P11-4 Audio-Video Synchronization for Post-Production over Managed Wide-Area Networks**—*Nathan Brock,<sup>1</sup> Michelle Daniels,<sup>1</sup> Steve Morris,<sup>2</sup> Peter Otto<sup>1</sup>*

<sup>1</sup>University of California San Diego, La Jolla, CA, USA

<sup>2</sup>Skywalker Sound, Marin County, CA, USA

A persistent challenge with enabling remote collaboration for cinema post-production is synchronizing audio and video assets. This paper details efforts to guarantee that the sound quality and audio-video synchronization over networked collaborative systems will be measurably the same as that experienced in a traditional facility. This includes establishing a common word-clock source for all digital audio devices on the network, extending transport control and time code to all audio and video assets, adjusting latencies to ensure sample-accurate mixing between remote audio sources, and locking audio and video playback to within quarter-frame accuracy. We will detail our instantiation of these techniques at a demonstration given in December 2009 involving collaboration between a film editor in San Diego and a sound designer in Marin County, California.  
*Convention Paper 8040*

**15:30**

**P11-5 A Proxy Approach for Interoperability and Common Control of Networked Digital Audio Devices**—*Osedum P. Igumbor, Richard J. Foss*, Rhodes University, Grahamstown, South Africa

This paper highlights the challenge that results



from the availability of a large number of control protocols within the context of digital audio networks. Devices that conform to different protocols are unable to communicate with one another, even though they might be utilizing the same networking technology (Ethernet, IEEE 1394 serial bus, USB). This paper describes the use of a proxy that allows for high-level device interaction (by sending protocol messages) between networked devices. Furthermore, the proxy allows for a common controller to control the disparate networked devices.  
*Convention Paper 8041*

16:00

**P11-6 Network Neutral Control over Quality of Service Networks—Philip Foulkes,**

*Richard Foss, Rhodes University, Grahamstown, South Africa*

IEEE 1394 (FireWire) and Ethernet Audio/Video Bridging are two networking technologies that allow for the transportation of synchronized, low-latency, real-time audio and video data. Each networking technology has its own methods and techniques for establishing stream connections between the devices that reside on the networks. This paper discusses the interoperability of these two networking technologies via an audio gateway and the use of a common control protocol, AES-X170, to allow for the control of the parameters of these disparate networks. This control is provided by a software patchbay application.  
*Convention Paper 8042*

16:30

**P11-7 Relative Importance of Speech and Non-Speech Components in Program Loudness Assessment—Ian Dash,<sup>1</sup> Mark Bassett,<sup>2</sup>**

*Densil Cabrera<sup>2</sup>*

<sup>1</sup>Australian Broadcasting Corporation, Sydney, NSW, Australia;

<sup>2</sup>The University of Sydney, Sydney, NSW, Australia

It is commonly assumed in broadcasting and film production that audiences determine soundtrack loudness mainly from the speech component. While intelligibility considerations support this idea indirectly, the literature is very short on direct supporting evidence. A listening test was therefore conducted to test this hypothesis. Results suggest that listeners judge loudness from overall levels rather than speech levels. A secondary trend is that listeners tend to compare like with like. Thus, listeners will compare speech loudness with other speech content rather than with non-speech content and will compare loudness of non-speech content with other non-speech content more than with speech content. A recommendation is made on applying this result for informed program loudness control.  
*Convention Paper 8043*

17:00

**P11-8 Loudness Normalization in the Age of Portable Media Players—Martin Wolters,<sup>1</sup> Harald Mundt,<sup>1</sup> Jeffrey Riedmiller<sup>2</sup>**

<sup>1</sup>Dolby Germany GmbH, Nuremberg, Germany

<sup>2</sup>Dolby Laboratories Inc., San Francisco, CA, USA

In recent years, the increasing popularity of portable media devices among consumers has created new and unique audio challenges for content creators, distributors as well as device manufacturers. Many of the latest devices are capable of supporting a broad range of content types and media formats including those often associated with high quality (wider dynamic-range) experiences such as HDTV, Blu-ray or DVD. However, portable media devices are generally challenged in terms of maintaining consistent loudness and intelligibility across varying media and content types on either their internal speaker(s) and/or headphone outputs. This paper proposes a nondestructive method to control playback loudness and dynamic range on portable devices based on a worldwide standard for loudness measurement as defined by the ITU. In addition the proposed method is compatible to existing playback software and audio content following the Replay Gain ([www.replaygain.org](http://www.replaygain.org)) proposal. In the course of the paper the current landscape of loudness levels across varying media and content types is described and new and nondestructive concepts targeted at addressing consistent loudness and intelligibility for portable media players are introduced.  
*Convention Paper 8044*

17:30

**P11-9 Determining an Optimal Gated Loudness Measurement for TV Sound Normalization—**

*Eelco Grimm,<sup>1</sup> Esben Skovenborg,<sup>2</sup> Gerhard Spikofski<sup>3</sup>*

<sup>1</sup>Grimm-Audio, Eindhoven, The Netherlands

<sup>2</sup>tc-electronic, Risskov, Denmark

<sup>3</sup>Institute of Broadcast Technology, Berlin, Germany

Undesirable loudness jumps are a notorious problem in television broadcast. The solution consists in switching to loudness-based metering and program normalization. In Europe this development has been led by the EBU P/LOUD group, working toward a single target level for loudness normalization applying to all genres of programs. P/LOUD found that loudness normalization as specified by ITU-R BS.1770-1 works fairly well for the majority of broadcast programs. However, it was realized that wide loudness-range programs were not well-aligned with other programs when using ITU-R BS.1770-1 directly, but that adding a measurement-gate provided a simple yet effective solution. P/LOUD therefore conducted a formal listening experiment to perform a subjective evaluation of different gate parameters. This paper specifies the method of the subjective evaluation and presents the results in term of preferred gating parameters.  
*Convention Paper 8154*

18:00

**P11-10 Analog or Digital? A Case-Study to Examine Pedagogical Approaches to Recording Studio Practice—Andrew King, University of Hull, Scarborough, North Yorkshire, UK**

This paper explores the use of digital and analog ➡

mixing consoles in the recording studio over a single drum kit recording session. Previous research has examined contingent learning, problem-solving, and collaborative learning within this environment. However, while there have been empirical investigations into the use of computer-based software and interaction around and within computers, this has not taken into consideration the use of complex recording apparatus. A qualitative case study approach was used in this investigation. Thirty hours of video data was captured and transcribed. A preliminary analysis of the data shows that there are differences between the types of problems encountered by learners when using either an analog or digital mixing console.  
*Convention Paper 8045*

**Session P12**  
**14:00 – 17:30**

**Sunday, May 23**  
**Room C5**

## NOISE REDUCTION AND SPEECH INTELLIGIBILITY

Chair: **Rhonda Wilson**, Meridian Audio, Huntingdon, UK

**14:00**

**P12-1 Monolateral and Bilateral Fitting with Different Hearing Aids Directional Configurations**—*Lorenzo Picinali*,<sup>1</sup> *Silvano Prosser*<sup>2</sup>  
<sup>1</sup>De Montfort University, Leicester, UK  
<sup>2</sup>Università di Ferrara, Ferrara, Italy

Hearing aid bilateral fitting in hearing impaired subjects raises some problems concerning the interaction of the perceptual binaural properties with the directional characteristics of the device. This experiment aims to establish whether and to which extent in a sample of 20 normally hearing subjects the binaural changes in the speech-to-noise level ratio (s/n), caused by different symmetrical and asymmetrical microphone configurations, and different positions of the speech signal (frontal or lateral), could alter the performances of the speech recognition in noise. Speech Reception Thresholds (SRT) in noise (simulated through an Ambisonic virtual sound field) have been measured monolaterally and bilaterally in order to properly investigate the role of the binaural interaction in the perception of reproduced signals.  
*Convention Paper 8046*

**14:30**

**P12-2 Acoustic Echo Cancellation for Wideband Audio and Beyond**—*Shreyas Paranjpe*, *Scott Pennock*, *Phil Hetherington*, QNX Software Systems Inc. (Wavemakers), Vancouver, BC, Canada

Speech processing is finally starting the transition to wider bandwidths. The benefits include increased intelligibility and comprehension and a more pleasing communication experience. High quality full-duplex Acoustic Echo Cancellation (AEC) is an integral component of a hands-free speakerphone communication system because it allows participants to converse in a natural manner (and without a headset!) as they would in

person. Some high-end Telepresence systems already achieve life-like communication but are computationally demanding and prohibitively expensive. The challenge is to develop a robust Acoustic Echo Canceler (AEC) that processes full-band audio signals with low computational complexity and reasonable memory consumption for an affordable Telepresence experience.  
*Convention Paper 8047*

**15:00**

**P12-3 Adaptive Noise Reduction for Real-Time Applications**—*Constantin Wiesener*,<sup>1</sup> *Tim Flohrer*,<sup>2</sup> *Alexander Lerch*,<sup>2</sup> *Stefan Weinzierl*<sup>1</sup>  
<sup>1</sup>TU Berlin, Berlin, Germany  
<sup>2</sup>zplane.development, Berlin, Germany

We present a new algorithm for real-time noise reduction of audio signals. In order to derive the noise reduction function, the proposed method adaptively estimates the instantaneous noise spectrum from an autoregressive signal model as opposed to the widely-used approach of using a constant noise spectrum fingerprint. In conjunction with the Ephraim and Malah suppression rule a significant reduction of both stationary and non-stationary noise can be obtained. The adaptive algorithm is able to work without user interaction and is capable of real-time processing. Furthermore, quality improvements are easily possible by integration of additional processing blocks such as transient preservation.  
*Convention Paper 8048*

**15:30**

**P12-4 Active Noise Reduction in Personal Audio Delivery Systems; Assessment Using Loudness Balance Methods**—*Paul Darlington*, *Pierre Guuu*, Phitek Systems (Europe) Sarl, Lausanne, Switzerland

Subjective methods developed for rating passive hearing protectors are inappropriate to measure active noise reduction of earphones or headphones. Additionally, assessment using objective means may produce misleading results, which do not correlate well with wearer experience and do not encompass human variability. The present paper describes application of "loudness balance" methods to the estimation of active attenuation of consumer headphones and earphones. Early results suggest that objective measures of the active reduction of pressure report greater attenuations than those implied by estimates of perceived loudness, particularly at low frequency.  
*Convention Paper 8049*

**16:00**

**P12-5 Mapping Speech Intelligibility in Noisy Rooms**—*John F. Culling*,<sup>1</sup> *Sam Jelfs*,<sup>1</sup> *Mathieu Lavandier*<sup>2</sup>  
<sup>1</sup>Cardiff University, Cardiff, UK  
<sup>2</sup>Université de Lyon, Vaux-en-Velin Cedex, France

We have developed an algorithm for accurately predicting the intelligibility of speech in noise in a reverberant environment. The algorithm is based

on a development of the equalization-cancellation theory of binaural unmasking, combined with established prediction methods for monaural speech perception in noise. It has been validated against a wide range of empirical data. Acoustic measurements of rooms, known as binaural room impulse responses (BRIRs) are analyzed to predict intelligibility of a nearby voice masked by any number of steady-state noise maskers in any spatial configuration within the room. This computationally efficient method can be used to generate intelligibility maps of rooms based on the design of the room.  
*Convention Paper 8050*

16:30

**P12-6 Further Investigations into Improving STI's Recognition of the Effects of Poor Frequency Response on Subjective Intelligibility—***Glenn Leembruggen,<sup>1</sup> Marco Hippler,<sup>2</sup> Peter Mapp<sup>3</sup>*

<sup>1</sup>Acoustic Directions Pty Ltd., ICE Design, Sydney, Australia, and University of Sydney, Sydney, Australia

<sup>2</sup>University of Applied Sciences Cologne, Cologne, Germany

<sup>3</sup>Peter Mapp and Associates, Colchester, UK

Previous work has highlighted deficiencies in the ability of the STI metric to satisfactorily recognize the subjective loss of intelligibility that occurs with sound systems having poor frequency responses, particularly in the presence of reverberation. In a recent paper we explored the changes to STI values resulting from a range of dynamic speech spectra taken over differing time lengths with different filter responses. That work included determining the effects on STI values of three alternative spreading functions simulating the ear's upward masking mechanism. This paper extends that work and explores the effects on STI values of two masking methods used in MPEG-1 audio coding.  
*Convention Paper 8051*

17:00

**P12-7 Real-Time Speech-Rate Modification Experiments —***Adam Kupryjanow, Andrzej Czyzewski,* Gdansk University of Technology, Gdansk, Poland

An algorithm designed for real-time speech time scale modification (stretching) is proposed, providing a combination of typical synchronous overlap and add based time scale modification algorithm and signal redundancy detection algorithms that allow to remove parts of the speech signal and replace them with the stretched speech signal fragments. Effectiveness of signal processing algorithms are examined experimentally together with the resulting sound quality.  
*Convention Paper 8052*

**Workshop 6**  
14:00 – 15:45

**Sunday, May 23**  
Room C2

**HOW DO WE EVALUATE HIGH RESOLUTION FORMATS FOR DIGITAL AUDIO?**

Chair: **Hans van Maanen,** Temporal Coherence, The Netherlands

Panelists: *Peter Craven,* Algol Applications Ltd., Steyning, UK  
*Milind Kunchur,* University of South Carolina, SC, USA  
*Thomas Sporer,* Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany  
*Menno van der Veen,* Ir. Bureau Vanderveen, Zwolle, The Netherlands  
*Wieslaw Woszczyk,* McGill University, Montreal, Quebec, Canada

Since the introduction of the High Resolution Formats for Digital Audio (e.g. SACD, 192 kHz / 24 bit), there has been discussion about the audibility of these formats, compared to the CD format (44.1 kHz / 16 bit). What difference does high sample rate and bit depth make in our perception? Can we hear tones above 20 kHz? Can we perceive quantization errors in 16-bit audio? Does high sample rate make a difference in our phase resolution? Are we even asking the right questions? Controlled, scientific listening tests have mostly given ambiguous or inconclusive results, yet a large number of consumers, using "high-end" audio equipment, prefer the sound from the "high resolution" formats over the CD. The workshop will start with introductory notes from the panel members, discussing the differences between "analog" and first-generation digital formats, address some of the paradoxes of the CD format, present results on "circumstantial" evidence and subjective testing, show results on the audibility of the human hearing, which cannot be explained by the commonly accepted 20 kHz upper limit and discuss the problems and pitfalls of "scientific" listening tests, where possible illustrated with demonstrations.

These introductory notes should provoke a discussion with the audience about the audibility of the improvements of the "high resolution" formats. We attempt to reach consensus, where possible, regarding what is known and what is not with respect to our ability to perceive the differences between standard and high resolution audio. We further discuss the paradigms of testing for evaluating the quality and perception of high resolution audio, how to structure the tests, how to configure the testing environment, and how to analyze the results.

The outcome of the workshop should also be to find the way forward by identifying the bottlenecks which—at this moment—hamper the further implementation of the "high resolution" formats for "high-end" audio as these formats create an opportunity for the audio industry as a whole: better sources stimulate the development of better reproduction systems.

**Special Event**  
**AES/APRS/MPG PLATINUM PRODUCER**  
**BEHIND THE GLASS**

Sunday, May 23, 14:00 – 15:45  
Room C1

Moderator: **Howard Massey**

Panelists: Jon Cohen  
*Jon Kelly*  
*Stephen Lipson*  
*Russ Titelman*

An informal conversation with legendary British and American producers about the journey to the intersection where art and commerce meet. Topics will include tips and techniques for eliciting the best performance from an



artist; realizing the artist's vision without busting the budget; and the changing role of the producer in today's increasingly homegrown environment.

*Howard Massey* is a leading audio industry consultant, technical writer, and author of *Behind the Glass* and *Behind the Glass Volume II*, two collections of in-depth interviews with top record producers and audio engineers widely used in recording schools the world over. He also co-authored legendary Beatles engineer Geoff Emerick's acclaimed 2006 memoir, *Here, There, and Everywhere: My Life Recording the Music of the Beatles*.

*This event will be prefaced by a special presentation to Sir George Martin OBE.*

### Student Event/Career Development EDUCATION FAIR

Sunday, May 23, 14:00 – 16:00  
C4-Foyer

Institutions offering studies in audio (from short courses to graduate degrees) will be represented in a "table top" session. Information on each school's respective programs will be made available through displays and academic guidance. There is no charge for schools to participate. Admission is free and open to all Convention attendees.

**Sunday, May 23                      14:00      Room Saint Julien**  
**Technical Committee Meeting on Electro Magnetic Compatibility**

**Sunday, May 23                      15:00      Room Saint Julien**  
**Technical Committee Meeting on Semantic Audio Analysis**

**Workshop 7    Sunday, May 23**  
**16:00 – 19:00    Room C6**

### THE WORK OF FORENSIC SPEECH AND AUDIO ANALYSTS

Chair: **Peter French**, JP French Associates, York, UK, University of York, York, UK

Panelists: *Philip Harrison*, JP French Associates, York, UK, University of York, York, UK

This workshop provides an illustrated introduction to the various categories of work carried out by forensic speech and audio analysts. In attempting to convey the variety, scope, and flavor of the work, methodological issues, principles, and developments are discussed in respect of key areas of forensic investigation. These include speaker profiling, the analysis of recordings of criminal speech samples in order to assemble a profile of regional, social, ethnic and other information concerning speakers; speaker comparison testing, the comparison of recorded voice samples using auditory-phonetic and acoustic methods as well as automatic systems to assist with identification; content analysis, the use of phonetic and acoustic techniques to decipher contentious areas of evidential recordings; procedures for evaluating ear-witness testimony; sound propagation testing, the reconstructions of crime scenes and the measurement of sound from given points across distances to assess what might have been heard by witness in various positions; recording authentication, the examination of evidential recordings for signs of tampering or falsification. Some of the points developed are exemplified by recordings and

analyses arising from important and high profile criminal cases. The presentation will be followed by a discussion session where questions and points are invited from those attending.

**Tutorial 5**  
**16:00 – 18:00**

**Sunday, May 23**  
**Room C2**

### SPATIAL AUDIO REPRODUCTION: FROM THEORY TO PRODUCTION

Presenters: **Frank Melchior**, IOSONO GmbH, Erfurt, Germany  
**Sascha Spors**, Deutsche Telekom AG Laboratories, Berlin, Germany

Advanced high-resolution spatial sound reproduction systems like Wave Field Synthesis (WFS) and Higher-Order Ambisonics (HOA) are being used increasingly. Consequently more and more material is being produced for such systems. Established channel-based production processes from stereophony can only be applied to a certain extent. In the future, a paradigm shift toward object-based audio production will have to take place in order to cope for the needs of systems like WFS. This tutorial spans the bridge from the physical foundations of such systems, over their practical implementation toward efficient production processes. The focus lies on WFS, however the findings will also be applicable to other systems. The tutorial is accompanied by practical examples of object-based productions for WFS.

**Student Event/Career Development**  
**RECORDING COMPETITION—PART 1**  
**(WORLD/FOLK, JAZZ/BLUES, POP/ROCK)**  
Sunday, May 23, 16:00 – 19:00  
Room C1

The Student Recording Competition is a highlight at each convention. A distinguished panel of judges participates in critiquing finalists of each category in an interactive presentation and discussion. This event presents stereo recordings in these categories:

- Jazz/Blues 16:00 to 17:00
- World/Folk 17:00 to 18:00
- Pop/Rock 18:00 to 19:00

The top three finalists in each category present a short summary of their techniques and intentions, and play their projects for all who attend. Meritorious awards will be presented at the closing Student Delegate Assembly Meeting (SDA-2) on Tuesday afternoon.

It's a great chance to hear the work of your colleagues from other educational institutions. Everyone learns from the judges' comments even if your project isn't one of the finalists, and it's a great chance to meet other students and faculty.

Judges include: *Stereo – Jazz/Blues*: Jim Anderson, Dave McLaughlin, Jim Kaiser. *Stereo – World/Folk*: Darcy Proper, Andres Mayo, Mark Drews. *Stereo – Pop/Rock*: Gary Gottlieb, Hugh Robjohns, Barry Marshall.

**Sunday, May 23                      16:00      Room Saint Julien**  
**Technical Committee Meeting on Fiber Optics for Audio**

**Sunday, May 23                      16:00      Room Mouton Cadet**  
**Standards Committee Meeting SC-04-04 Microphone Measurement and Characterization**



**POSTERS: ROOM ACOUSTICS, SOUND REINFORCEMENT, AND INSTRUMENTATION**

16:30

**P13-1 Adaptive Equalizer for Acoustic Feedback Control**—*Alexis Favrot, Christof Faller, Illusonic LLC, Lausanne, Switzerland*

Acoustic feedback is a recurrent problem in audio applications involving amplified closed loop systems. The goal of the proposed acoustic feedback control algorithm is to adapt an equalizer, applied to the microphone signal, to prevent feedback before the effect of it is noticeable. This is achieved by automatically decreasing the gain of an equalizer at frequencies where feedback likely will occur. The equalization curve is determined using information contained in an adaptively estimated feedback path. A computationally efficient implementation of the proposed algorithm, using short-time Fourier transform, is described.

*Convention Paper 8053*

16:30

**P13-2 The Meshotron: A Network of Specialized Hardware Units for 3-D Digital Waveguide Mesh Acoustic Model Parallelization**—*Guilherme Campos, Sara Barros, University of Aveiro, Aveiro, Portugal*

This paper presents the project of a computing network—the Meshotron—specifically designed for large-scale parallelization of three-dimensional Digital Waveguide Mesh (3-D DWM) room acoustic models. It discusses the motivation of the project, its advantages, and the architecture envisaged for the application-specific hardware (ASH) units to form the network. The initial stages involve the development of a software prototype based on the rectangular mesh topology, using appropriate hardware simulation tools, and the design and test of FPGA-based scattering units for air and boundary nodes.

*Convention Paper 8054*

16:30

**P13-3 A Hybrid Approach for Real-Time Room Acoustic Response Simulation**—*Andrea Primavera,<sup>1</sup> Lorenzo Palestini,<sup>1</sup> Stefania Cecchi,<sup>1</sup> Francesco Piazza,<sup>2</sup> Marco Moschetti<sup>2</sup>*

<sup>1</sup>Università Politecnica delle Marche, Ancona, Italy

<sup>2</sup>Korg Italy, Osimo, Italy

Reverberation is a well known effect particularly important for music listening especially for recorded and live music. Generally, there are two approaches for artificial reverberation: the desired signal can be obtained by convolving the input signal with a measured impulse response (IR) or a synthetic one. Taking into account the advantages of both approaches, a hybrid artificial reverberation algorithm is presented. The early reflections are derived from a real IR, truncated considering the calculated mixing time, and the

reverberation tail is obtained considering the Moorer's structure. The parameters defining this structure are derived from the analyzed IR, using a minimization criteria based on Simultaneous Perturbation Stochastic Approximation (SPSA). The obtained results showed a high-quality reverberator with a low computational load.

*Convention Paper 8055*

16:30

**P13-4 Sensor Networks for Measurement of Acoustic Parameters of Classrooms**—*Suthikshn Kumar, PESIT, Bangalore, India*

Measurement of acoustic parameters such as signal-to-noise ratio, reverberation time, and background noise is important in order to optimize classroom configuration and architecture. The classroom learning environment can thus be improved. The lecturers and students find comfort and assurance in knowing that the acoustic parameters of the classroom have been measured and indicate good acoustics. We propose sensor networks for acoustic measurement applications. The sensor network consists of an array of inexpensive sensor nodes which communicate and aggregate the acoustic parameters. The paper presents the sensor node architecture and illustrates the typical usage of the sensor networks for the measurement of acoustic parameters in classrooms.

*Convention Paper 8056*

[This paper will not be presented but is available for purchase.]

16:30

**P13-5 In Situ Directivity Measurement of Flush-Mounted Loudspeakers in a Non-Environment Listening Room**—*Daniel Fernández-Comesaña,<sup>1</sup> Paúl Rodríguez-García,<sup>1</sup> Soledad Torres-Guijarro,<sup>2</sup> Antonio Pena<sup>3</sup>*

<sup>1</sup>Institute of Sound and Vibration Research (ISVR), Southampton, UK

<sup>2</sup>Laboratorio Oficial de Metroloxía de Galicia (LOMG), Tecnópole Ourense, Spain

<sup>3</sup>ETSE Telecomunicación, Universidade de Vigo, Vigo, Spain

Directivity is one important parameter to define the behavior of a loudspeaker. There are many techniques and standards about directivity measurements in anechoic chambers, but in situ measurements of flush-mounted loudspeakers show some specific problems. This paper develops a procedure to measure directivity under the special conditions of a non-environment listening room, introducing the techniques utilized, the problems found with the proposed solutions, and discussing the limitations of the process. The existence of reflections, baffling effects due to adjacent walls and a comparison to theoretical models of the radiation of a piston are discussed.

*Convention Paper 8057*

16:30

**P13-6 Measurement of Loudspeakers without Using an Anechoic Chamber Utilizing Pulse-Train Measurement Method**—*Teruo Muraoka,<sup>1</sup>*

Takahiro Miura,<sup>1</sup> Haruhito Shimura,<sup>2</sup>  
Hiroshi Akino,<sup>2</sup> Tohru Ifukube<sup>1</sup>

<sup>1</sup>The University of Tokyo, Tokyo, Japan

<sup>2</sup>Audio-Technica Corporation, Machida, Japan

Generally, loudspeakers are measured in an anechoic chamber. However, those chambers are very expensive, which makes the sound engineer's easy acoustical measurements almost impossible. Thus the authors devised a new measurement without using any anechoic chamber. Traditionally, loudspeakers are measured in an anechoic chamber by installing in a proper speaker enclosure and by driving with swept sinusoidal wave. Loudspeaker's radiated sound is detected with an omnidirectional reference microphone, and the output of the microphone is displayed graphically. If a shotgun microphone is employed instead of omnidirectional microphone, an anechoic chamber will become unnecessary. Furthermore, if pulse-train is used as a test signal, noise in surrounding circumstances will be reduced by applying a synchronous-averaging algorithm. Based upon this idea, the authors measured loudspeaker in any acoustical circumstances and obtained very similar data with that by the traditional method.

*Convention Paper 8058*

*Paper presented by Takahiro Miura.*

16:30

**P13-7 Acoustic Space Sampling and the Grand Piano in a Non-Anechoic Environment: A Recordist-Centric Approach to the Musical Acoustic Study**—Grzegorz Sikora,<sup>1,2</sup>

Brett Leonard,<sup>1,2</sup> Martha de Francisco,<sup>1,2</sup>

Douglas Eck<sup>2,3,4</sup>

<sup>1</sup>McGill University, Montreal, Quebec, Canada

<sup>2</sup>Centre for Interdisciplinary Research in Music, Media, and Technology, Montreal, Quebec, Canada

<sup>3</sup>Université de Montréal, Montreal, Quebec, Canada

<sup>4</sup>International Laboratory for Brain, Music, and Sound Research, Montreal, Quebec, Canada

A novel approach to instrument acoustics research is presented in which the instrument is coupled with a room and both are measured as a single acoustical system, in juxtaposition to many anechoic and computer simulated studies of musical acoustics. The technique is applied to the ubiquitous concert grand piano, where spectral information is gathered through the process of "acoustic space sampling" (AcSS), using more than 1,330 microphones. The physical data is then combined with psychoacoustic predictors to generate a map of timbre. This map is compared to the preference of expert listeners, thereby correlating the physical measures obtained through acoustic space sampling to the application of the recording engineer.

*Convention Paper 8059*

**Sunday, May 23 17:30 Room Mouton Cadet  
Standards Committee Meeting SC-04-01 Acoustics  
and Sound Source Modeling**

**Sunday, May 23 17:30 Room Mouton Cadet  
Standards Committee Meeting SC-03-12 Forensic  
Audio**

**Special Event  
OPEN HOUSE OF THE TECHNICAL COUNCIL  
AND THE RICHARD C. HEYSER MEMORIAL  
LECTURE**

Sunday, May 23, 19:00 – 20:30

Room C2

Lecturer: **Brian C. J. Moore**

The Heyser Series is an endowment for lectures by eminent individuals with outstanding reputations in audio engineering and its related fields. The series is featured twice annually at both the United States and European AES Conventions. Established in May 1999, The Richard C. Heyser Memorial Lecture honors the memory of Richard Heyser, a scientist at the Jet Propulsion Laboratory, who was awarded nine patents in audio and communication techniques and was widely known for his ability to clearly present new and complex technical ideas. Heyser was also an AES governor and AES Silver Medal recipient.

The Richard C. Heyser distinguished lecturer for the 128th AES Convention is Brian C. J. Moore, Professor of Auditory Perception, Department of Experimental Psychology, University of Cambridge. The topic of his lecture is "Hearing Loss and Hearing Aids."

Hearing loss affects more than 10 percent of the adult population in most countries and is especially prevalent among the elderly. The most common form of hearing loss arises from dysfunction of the cochlea in the inner ear. In most cases, the only form of treatment is via hearing aids or (for profound losses) cochlear implants. In this lecture Moore will review some of the perceptual consequences of hearing loss, which involve much more than just loss of sensitivity to weak sounds. He will then describe the signal processing that is performed in hearing aids and will consider the extent to which hearing aids "compensate" for hearing loss. Possible avenues for the future will be discussed.

Brian Moore is Professor of Auditory Perception in the University of Cambridge. He is a Fellow of the Royal Society, the Academy of Medical Sciences, and the Acoustical Society of America, and an Honorary Fellow of the Belgian Society of Audiology and of the British Society of Hearing Aid Audiologists. He is President of the Association of Independent Hearing Healthcare Professionals (UK). He is an Associate Editor of the *Journal of the Acoustical Society of America* and is a member of the Editorial Boards of *Hearing Research*, and *Audiology and Neurotology*. He has written or edited 14 books and over 500 scientific papers and book chapters. In 2003 he was awarded the Acoustical Society of America Silver Medal in physiological and psychological acoustics. In 2004 he was awarded the first "International Award in Hearing" from the American Academy of Audiology. He has twice been awarded the Littler Prize of the British Society of Audiology. In 2008 he received the Award of Merit from the Association for Research in Otolaryngology and the Hugh Knowles Prize for Distinguished Achievement from Northwestern University. He is Wine Steward of Wolfson College Cambridge.

Moore's presentation will be followed by a reception hosted by the AES Technical Council.

**Session P14  
09:00 – 12:00**

**Monday, May 24  
Room C3**

**SPATIAL AUDIO PERCEPTION**

Chair: **Russell Mason**, University of Surrey, Guildford, Surrey, UK

**09:00**

**P14-1 Perceptual Assessment of Delay Accuracy**

**and Loudspeaker Misplacement in Wave Field Synthesis**—*Jens Ahrens, Matthias Geier, Sascha Spors*, Technische Universität Berlin, Berlin, Germany

The implementation of simple virtual source models like plane and spherical waves in wave field synthesis (WFS) employs delays that are applied to the input signals. We present a formal experiment evaluating the perceptual consequences of different accuracies of these delays. Closely related to the question of delay accuracy is the accuracy of the loudspeaker positioning. The second part of the presented experiment investigates the perceptual consequences of improperly placed loudspeakers. Dynamic binaural room impulse response-based simulations of a real loudspeaker array are employed and a static audio scene setup is considered.  
*Convention Paper 8068*

09:30

**P14-2 Perceptual Evaluation of Focused Sources in Wave Field Synthesis**—*Matthias Geier, Hagen Wierstorf, Jens Ahrens, Ina Wechsung, Alexander Raake, Sascha Spors*, Technische Universität Berlin, Berlin, Germany

Wave Field Synthesis provides the possibility to reproduce virtual sound sources located between the loudspeaker array and the listener. Such sources are known as focused sources. A previously published study including an informal listening test has shown that the reproduction of focused sources is subject to audible artifacts, especially for large loudspeaker arrays. The combination of the time-reversal nature of focused sources and spatial sampling leads to pre-echoes. The perception of these artifacts is quite different depending on the relative listener position. This paper describes a formal test that was conducted to verify the perceptual relevance of the physical properties found in previous papers.  
*Convention Paper 8069*

10:00

**P14-3 Experiments on the Perception of Elevated Sources in Wave-Field Synthesis Using HRTF Cues**—*Jose J. Lopez,<sup>1</sup> Maximo Cobos,<sup>1</sup> Basilio Pueo<sup>2</sup>*

<sup>1</sup>Technical University of Valencia, Valencia, Spain  
<sup>2</sup>University of Alicante, Alicante, Spain

Wave-Field Synthesis (WFS) is a spatial sound reproduction technique that has attracted the interest of many researchers in the last decades. Unfortunately, although WFS has been shown to provide excellent localization accuracy, this property is restricted to sources located on the horizontal plane. Recently, the authors proposed a hybrid system that combines HRTF-based spectral filtering with WFS. This system makes use of the conventional WFS approach to achieve localization in the horizontal plane, whereas elevation effects are simulated by means of spectral elevation cues. This paper provides a review of the proposed method together with a compilation of the last experiments carried out to evaluate the perception of elevated sources in this novel system.  
*Convention Paper 8070*

10:30

**P14-4 Comparison of the Width of Sound Sources in 2-Channel and 3-Channel Sound Reproduction**—*Munhum Park, Aki Härmä, Steven van de Par, Georgia Tryfou*, Philips Research Laboratories, Eindhoven, The Netherlands

In this paper we present the result of listening tests where the width of the sound stage was compared between conventional 2-channel stereophony and 3-channel reproduction with an additional center loudspeaker. When listeners were seated at the axis of symmetry, there was no significant difference between the two cases. In off axis positions there was a clear trend that the perceived image width of the noise is influenced by the way an additional interfering stimulus is reproduced, which was found to depend on the correlation of the noise. The results suggest that the spatial attribute of one sound image may be affected by another, for which possible explanations based on the principles of binaural hearing will be discussed.  
*Convention Paper 8071*

11:00

**P14-5 Study of the Effect of Source Directivity on the Perception of Sound in a Virtual Free-Field**—*Sorrel Hoare, Alex Southern, Damian Murphy*, University of York, York, UK

In the context of soundscape evaluation, an area that has yet to be explored satisfactorily concerns faithful modeling and auralization of outdoor spaces. A common limitation of popular acoustic modeling techniques relates to source characterization, an aspect of which considered here is source directivity. In terms of auralization, source directivity may be linked to perceived spatial sound quality. In a preliminary listening test a selection of audio samples are auralized in a virtual free-field where the directivity pattern of the source is varied. The results confirm the significance of this characteristic on the acoustic perception, thus validating the extension of this research to more complex models.  
*Convention Paper 8072*

11:30

**P14-6 Audio Spatialization for The Morning Line**—*David G. Malham, Tony Myatt, Oliver Larkin, Peter Worth, Matthew Paradis*, University of York, York, UK

The Morning Line is a large-scale outdoor sculpture with a multi-dimensional sound system consisting of 41 small weatherproof main loudspeakers and 12 subs, configured in 6 surround arrays, or "rooms," distributed throughout the sculpture. The irregular nature of these arrays necessitated the use of VBAP panning on the sculpture but when preparing works to be played on it, Ambisonics is more often used. This paper describes the hardware and software developed for the sculpture, theories of spatial audio perception that prompted the design approach, as well as the production facilities provided for the composers whose works are performed on The Morning Line.  
*Convention Paper 8073*



Session P15  
09:00 – 12:00

Monday, May 24  
Room C5

10:00

**LOUDSPEAKERS AND HEADPHONES: PART 2**

Chair: **John Vanderkooy**, University of Waterloo,  
Waterloo, Ontario, Canada

09:00

**P15-1 Chameleon Subwoofer Arrays—Generalized Theory of Vected Sources in a Closed Acoustic Space—Adam J. Hill, Malcolm O. J. Hawksford**, University of Essex, Colchester, UK

An equalization model is presented that seeks optimal solutions to wide area low-frequency sound reproduction in closed acoustic spaces. The methodology improves upon conventional wisdom by incorporating a generalized subwoofer array where individual frequency dependent loudspeaker polar responses are described by complex spherical harmonic frequency dependent functions. Multi-point system identification is performed using three-dimensional finite-difference time-domain simulation with optimization applied to seek global equalization represented by a set of orthogonal transfer functions applied to each spherical harmonic of each subwoofer within the array. The system is evaluated within a three-dimensional virtual acoustic space using both time and frequency domain metrics.

*Convention Paper 8074*

09:30

**P15-2 Dynamical Measurement of the Effective Radiation Area SD—Wolfgang Klippel**,<sup>1</sup>  
*Joachim Schlechter*<sup>2</sup>

<sup>1</sup>University of Dresden, Dresden, Germany  
<sup>2</sup>Klippel GmbH, Dresden, Germany

The effective radiation area SD is one of the most important loudspeaker parameters because it determines the acoustical output (SPL, sound power) and efficiency of the transducer. This parameter is usually derived from the geometrical size of the radiator considering the diameter of half the surround area. This conventional technique fails for microspeakers and headphone transducers where the surround geometry is more complicated and the excursion varies not linearly versus radius. The paper discusses new methods for measuring the SD more precisely. The first method uses a laser sensor and microphone to measure the voice coil displacement and the sound pressure generated by the transducer while mounted in a sealed enclosure. The second method uses only mechanical vibration and geometry of the radiator measured by using a laser triangulation scanner. The paper checks the reliability and reproducibility of conventional and the new methods and discusses the propagation of the measurement error on the T/S parameters using text box perturbation technique and other derived parameters (sensitivity).

*Convention Paper 8075*

**P15-3 Modeling Acoustic Horns with FEA—**

*David J. Murphy*,<sup>1</sup> *Rick Morgans*<sup>2</sup>

<sup>1</sup>Krix Loudspeakers, Hackham, SA, Australia

<sup>2</sup>Cyclopic Energy, Adelaide, SA, Australia

Simulations for acoustic horns have been developed using Finite Element Analysis (FEA) in both ANSYS and COMSOL modeling software. The FEA method solves the ideal, linear wave equation, so the effects of nonlinearity and viscosity are not inherently simulated. Quarter models were used in each case as the acoustic horns were not axi-symmetric. The size and shape of the horns have been informed by long-standing design parameters. The simulation results have been compared with acoustic measurements, and discrepancies investigated. While it was found that beam angle (dispersion) characteristics were relatively robust, it was useful in both cases to improve the accuracy of the beam angle, SPL, and impedance simulations by the application of frequency dependent damping to models of the compression driver.

*Convention Paper 8076*

10:30

**P15-4 Electroacoustic Measurements for High Noise Environment Intercom Headsets—**

*Stelios Potirakis*,<sup>1</sup> *Nicolas-Alexander Tatlas*,<sup>2</sup>  
*Maria Rangoussi*<sup>1</sup>

<sup>1</sup>Technological Education Institute of Piraeus,  
Aigaleo-Athens, Greece

<sup>2</sup>University of Patras, Patras, Greece

Intercom headsets are mandatory communication apparatus in high noise environment (HNE) conditions. Although military intercom headsets are typically used under extreme environmental conditions, a standard performance evaluation method exists only for the earphone elements. A systematic methodology for the measurement and performance evaluation of HNE headsets has recently been proposed based on the use of Head and Torso Simulator (HATS), addressing both signal reproduction and noise reduction issues. In this paper an improvement of the specific method is proposed concerning the headset electroacoustic reproduction measurements. The proposed enhancement refers to the use of impulse response measurements for the extraction of the specific characteristics as a function of frequency.

*Convention Paper 8077*

11:00

**P15-5 Phase, Polarity, and Delay or Why a Loudspeaker Crossover Is Not a Time Machine—***Ian Dash*,<sup>1</sup> *Fergus Fricke*<sup>2</sup>

<sup>1</sup>Australian Broadcasting Corporation, Sydney,  
NSW, Australia

<sup>2</sup>University of Sydney, Sydney, NSW, Australia

Phase delay and group delay functions are well known in lumped system theory but they are not well understood, particularly in the context of non-minimum phase lumped systems. This paper presents four paradoxes that arise from an overly-literal interpretation of the phase delay function. The paradoxes are very simply illustrat-



ed using the example of a first order loudspeaker crossover. A new system response model is presented to resolve these paradoxes. The model is extended to higher order systems and to non-minimum phase systems. Applications and implications for system analysis are discussed.  
*Convention Paper 8078*

11:30

**P15-6 Time and Level Localization Curves for a Regularly -Spaced Octagon Loudspeaker Array**—*Laurent S. R. Simon, Russell Mason, University of Surrey, Guildford, Surrey, UK*

Multichannel microphone array designs often use the localization curves that have been derived for 2-0 stereophony. Previous studies showed that side and rear perception of phantom image locations require somewhat different curves. This paper describes an experiment conducted to determine localization curves using an octagonal loudspeaker setup. Various signals with a range of interchannel time and level differences were produced between pairs of adjacent loudspeakers, and subjects were asked to evaluate the perceived sound event's direction and its locatedness. The results showed that the curves for the side pairs of adjacent loudspeakers are significantly different to the front and rear pairs. The resulting curves can be used to derive suitable microphone techniques for this loudspeaker setup.

*Convention Paper 8079*

**Workshop 8**  
09:00 – 11:00

**Monday, May 24**  
Room C2

**INTERACTING WITH SEMANTIC AUDIO—BRIDGING THE GAP BETWEEN HUMANS AND ALGORITHMS**

Chair: **Michael Hlatky**, University of Applied Sciences, Bremen, Bremen, Germany

Panelists: *Masataka Goto*, Media Interaction Group, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan  
*Anssi Klapuri*, Queen Mary University of London, London, UK  
*Jörn Loviscach*, Fachhochschule Bielefeld, University of Applied Sciences, Bielefeld, Germany  
*Yves Raimond*, BBC Audio & Music Interactive, London, UK

Technologies under the heading Semantic Audio have undergone a fascinating development in the past few years. Hundreds of algorithms have been developed; first applications have made their way from research into possible mainstream application. However, the current level of awareness among prospective users and the amount of actual practical use do not seem to live up to the potential of semantic audio technologies. We argue that this is more an issue concerning interface and interaction than a problem concerning the robustness of the applied algorithms or a lack of need in audio production. The panelists of this workshop offer ways to improve the usability of semantic audio techniques. They look into current applications in off-the-shelf products, discuss the use in a variety of specialized applications such as cus-

tom-tailored archival solutions, demonstrate and showcase their own developments in interfaces for semantic audio, and propose future directions in interface and interaction development for semantic audio technologies ranging from audio file retrieval to intelligent audio effects.

The second half of this workshop includes hands-on interactive experiences provided by the panel.

**Workshop 9**  
09:00 – 10:30

**Monday, May 24**  
Room C6

**REDUNDANT NETWORKS FOR AUDIO**

Chair: **Umberto Zanghieri**, ZP Engineering srl, Rome, Italy

Panelists: *Marc Brunke*, Optocore, Munich, Germany  
*David Myers*, Audinate, Ultimo NSW, Australia  
*Michel Quaix*, Digigram, Montbonnot, France  
*Al Walker*, Midas Klark Teknik, Kidderminster, UK

Redundancy within a digital network for audio transport has specific requirements when compared to redundancy for usual data networks, and for this reason several practical implementations offer dual network ports. Ideas and solutions from current formats will be presented, detailing requirements specific to usage cases, such as digital audio transport for live events or for install. The discussion will include also non-IP audio networks. Different topologies and switchover issues will be presented, and practical real-world examples will be shown.

**Tutorial 6**  
09:00 – 10:15

**Monday, May 24**  
Room C1

**CLASSICAL MUSIC WITH PERSPECTIVE**

Presenter: **Sabine Maier**, Tonmeister, Vienna, 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 versions (in stereo and surround sound). 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.

In this tutorial the experimental procedure as well as the results will be discussed. Examples of the different mixes will be played.

**Monday, May 24** 09:00 **Room Saint Julien**  
**Technical Committee Meeting on High Resolution Audio**

**Monday, May 24** 10:00 **Room Saint Julien**  
**Technical Committee Meeting on Audio Forensics**

**Session P16**  
**10:30 – 12:00**

**Monday, May 24**  
**C4-Foyer**

**POSTERS: AUDIO PROCESSING—AUDIO CODING AND MACHINE INTERFACE**

**10:30**

**P16-1 Trajectory Sampling for Computationally Efficient Reproduction of Moving Sound Sources**

**Sources**—*Nara Hahn, Keunwoo Choi, Hyunjoo Chung, Koeng-Mo Sung*, Seoul National University, Seoul, Korea

Reproducing moving virtual sound sources has been addressed in a number of spatial audio studies. The trajectories of moving sources are relatively oversampled if they are sampled in temporal domain, as this leads to inefficiency in computing the time delay and amplitude attenuation due to sound propagation. In this paper methods for trajectory sampling are proposed that use the spatial property of the movement and the spectral property of the received signal. These methods reduce the number of sampled positions and reduce the computational complexity. Listening tests were performed to determine the appropriate downsampling rate that depends not only on the trajectory but on the frequency content of the source signal.

*Convention Paper 8080*

**10:30**

**P16-2 Audio Latency Measurement for Desktop Operating Systems with Onboard Soundcards**

**Sources**—*Yonghao Wang*,<sup>1,2</sup> *Ryan Stables*,<sup>2</sup> *Joshua Reiss*<sup>1</sup>

<sup>1</sup>Queen Mary University of London, London, UK

<sup>2</sup>Birmingham City University, Birmingham, UK

Using commodity computers in conjunction with live music digital audio workstations (DAW) has become increasingly more popular in recent years. The latency of these DAW audio processing chains for some application such as live audio monitoring has always been perceived as a problem when DSP audio effects are needed. With “High Definition Audio” being standardized as the onboard soundcard’s hardware architecture for personal computers, and with advances in audio APIs, the low latency and multichannel capability has made its way into home studios. This paper will discuss the results of latency measurements of current popular operating systems and hosts applications with different audio APIs and audio processing loads.

*Convention Paper 8081*

**10:30**

**P16-3 Error Control Techniques within Multidimensional-Adaptive Audio Coding Algorithms**

**Sources**—*Neil Smyth*, APTX, Belfast, N. Ireland, UK

Multidimensional-adaptive audio coding algorithms can adapt multiple performance measures to the demands of different audio applications in real-time. Depending on the transmission or storage environment, audio processing applications require forms of error control to maintain acceptable audio quality. By definition, multidimensional-

adaptive audio coding utilizes numerous error detection, correction, and concealment techniques. However, such techniques also have implications for other relevant performance measurements, such as coded bit-rate and computational complexity. This paper discusses the signal-processing tools used by a multidimensional-adaptive audio coding algorithm to achieve varying levels of error control while the fundamental structure of the algorithm is also varying. The effects and trade-offs on other coding performance measures will also be discussed.

*Convention Paper 8082*

**10:30**

**P16-4 Object-Based Audio Coding Using Non-Negative Matrix Factorization for the Spectrogram Representation**

**Sources**—*Joonas Nikunen, Tuomas Virtanen*, Tampere University of Technology, Tampere, Finland

This paper proposes a new object-based audio coding algorithm, which uses non-negative matrix factorization (NMF) for the magnitude spectrogram representation and the phase information is coded separately. The magnitude model is obtained using a perceptually weighted NMF algorithm, which minimizes the noise-to-mask ratio (NMR) of the decomposition, and is able to utilize long term redundancy by an object-based representation. Methods for the quantization and entropy coding of the NMF representation parameters are proposed and the quality loss is evaluated using the NMR measure. The quantization of the phase information is also studied. Additionally we propose a sparseness criteria for the NMF algorithm, which is set to favor the gain values having the highest probability and thus the shortest entropy coding word length, resulting to a reduced bit rate.

*Convention Paper 8083*

**10:30**

**P16-5 Cross-Layer Rate-Distortion Optimization for Scalable Advanced Audio Coding**

**Sources**—*Emmanuel Ravelli, Vinay Melkote, Tejaswi Nanjundaswamy, Kenneth Rose*, University of California at Santa Barbara, Santa Barbara, CA, USA

Current scalable audio codecs optimize each layer of the bit-stream successively and independently with a straightforward application of the same rate-distortion optimization techniques employed in the non-scalable case. The main drawback of this approach is that the performance of the enhancement layers is significantly worse than that of the non-scalable codec at the same cumulative bit-rate. We propose in this paper a novel optimization technique in the Advanced Audio Coding (AAC) framework wherein a cross-layer iterative optimization is performed to select the encoding parameters for each layer with a conscious accounting of rate and distortion costs in all layers, which allows for a trade-off between performance at different layers. Subjective and objective results demonstrate the effectiveness of the proposed approach and provide insights for bridging the gap with the non-scalable codec.

*Convention Paper 8084*

10:30

**P16-6 Issues and Solutions Related to Real-Time TD-PSOLA Implementation**—*Sylvain Le Beux*,<sup>1</sup> *Boris Doval*,<sup>2</sup> *Christophe d'Alessandro*<sup>1</sup>  
<sup>1</sup>LIMSI-CNRS, Université Paris-Sud XI, Orsay, France  
<sup>2</sup>LAM-IJLRA, Université Paris, Paris, France

This paper presents a procedure adaptation for the calculation of TD-PSOLA algorithm when the processing of pitch-shifting and time-stretching coefficients needs to be achieved in real-time at every new synthesis pitch mark. In the scope of standard TD-PSOLA algorithm, modification coefficients are defined from the analysis time axis whereas for real-time applications, pitch and duration control parameters need to be sampled at synthesis time. This paper will establish the theoretical correspondence between both approaches. Another issue related to real-time context concerns the trade-off between the latency required for processing and the type of analysis window used.  
*Convention Paper 8085*  
*Paper presented by Boris Doval.*

10:30

**P16-7 Integrating Musicological Knowledge into a Probabilistic Framework for Chord and Key Extraction**—*Johan Pauwels*, *Jean-Pierre Martens*, Ghent University, Ghent, Belgium

In this paper a formerly developed probabilistic framework for the simultaneous detection of chords and keys in polyphonic audio is further extended and validated. The system behavior is controlled by a small set of carefully defined free parameters. This has permitted us to conduct an experimental study that sheds a new light on the relative importance of musicological knowledge in the context of chord extraction. Some of the obtained results are at least surprising and, to our knowledge, never reported as such before.  
*Convention Paper 8086*

10:30

**P16-8 A Doubly Sparse Greedy Adaptive Dictionary Learning Algorithm for Music and Large-Scale Data**—*Maria G. Jafari*, *Mark D. Plumbley*, Queen Mary University of London, London, UK

We consider the extension of the greedy adaptive dictionary learning algorithm that we introduced previously to applications other than speech signals. The algorithm learns a dictionary of sparse atoms, while yielding a sparse representation for the speech signals. We investigate its behavior in the analysis of music signals and propose a different dictionary learning approach that can be applied to large data sets. This facilitates the application of the algorithm to problems that generate large amounts of data, such as multimedia and multichannel application areas.  
*Convention Paper 8087*

**Student Event/Career Development  
RECORDING COMPETITION—PART 2  
(STEREO CLASSICAL, SURROUND CLASSICAL  
AND NON-CLASSICAL)**  
Monday, May 24, 10:30 – 13:30  
Room C1

The Student Recording Competition is a highlight at each convention. A distinguished panel of judges participates in critiquing finalists of each category in an interactive presentation and discussion. This event presents stereo and surround recordings in these categories:

- Stereo Classical 10:30 to 11:30
- Surround Classical 11:30 to 12:30
- Surround Non-Classical 12:30 to 13:30

The top three finalists in each category present a short summary of their techniques and intentions, and play their projects for all who attend. Meritorious awards will be presented at the closing Student Delegate Assembly Meeting (SDA-2) on Tuesday afternoon.

It's a great chance to hear the work of your colleagues from other educational institutions. Everyone learns from the judges' comments even if your project isn't one of the finalists, and it's a great chance to meet other students and faculty.

Judges include: *Stereo – Classical*: Tony Falkner, Ken Blair, David Bowles. *Surround – Classical*: Philip Hobbs, Tin Jonker, Bastian Kuijtit. *Surround – Non-Classical*: Ron Prent, Thor Legvold, Bill Crabtree.

**Workshop 10  
11:00 – 13:00**

**Monday, May 24  
Room C6**

## A CURRICULUM FOR GAME AUDIO

Chair: **Richard Stevens**, Leeds Metropolitan University, UK

Panelists: *Dan Bardino*, Creative Services Manager, Sony Computer Entertainment Europe Limited, London, UK  
*Andy Farnell*, Author of *Designing Sound*  
*David Mollerstedt*, DICE, Stockholm, Sweden  
*Dave Raybould*, Leeds Metropolitan University, Leeds, UK  
*Nia Wearn*, Staffordshire University, Staffordshire, UK

How do I get work in the games industry? Anyone involved in the discussions that follow this question in forums, conferences, and workshops worldwide will realize that many students in Higher Education who are aiming to enter the sector are not equipped with the knowledge and skills that the industry requires. In this workshop a range of speakers will discuss, and attempt to define, the various roles and related skillsets for audio within the games industry and will outline their personal route into this field. The panel will also examine the related work of the IASIG Game Audio Education Working Group in light of the recent publication of its Game Audio Curriculum Guidelines draft. This will be a fully interactive workshop inviting debate from the floor alongside discussion from panel members in order to share a range of views on this important topic.

**Special Event  
AES/APRS—LIFE IN THE OLD DOGS YET—  
PART TWO: QUALITY AND QUANTITY—  
A COSTLY DILEMMA**  
Monday, May 24, 11:00 – 13:00  
Room C2

Moderator: **George Massenburg**, GML LLC

Panelists: *Barry Fox*  
*Rob Kelly*, Air/Strongroom Studios

Legendary audio polymath George Massenburg presents ➡



his take on the value of aspiring to the heights of quality. This is followed by a panel discussing whether a way can be found to square the circle between the desire to achieve long-term, multi-format, re-exploitable catalog and short-term production cost-cutting. Clearly, quality is not simply a matter of applying technology. The panel will ponder on how the need to sustain realistic investment in recording budgets may be a challenge to many studio clients and how much do recording costs impact upon quality and success?

**Monday, May 24 11:00 Room Saint Julien**  
**Technical Committee Meeting on Acoustics and Sound Reinforcement**

**Monday, May 24 12:00 Room Saint Julien**  
**Technical Committee Meeting on Network Audio Systems**

**Monday, May 24 12:30 Room Mouton Cadet**  
**Standards Committee Meeting SC-02-01 Digital Audio Measurement Techniques**

**Monday, May 24 13:00 Room Saint Julien**  
**Technical Committee Meeting on Automotive Audio**

**Session P17 Monday, May 24**  
**14:00 – 17:30 Room C3**

### MULTICHANNEL AND SPATIAL AUDIO: PART 1

Chair: **Wieslaw Woszczyk**, McGill University,  
Montreal, Quebec, Canada

**14:00**

**P17-1 Individualization of Dynamic Binaural Synthesis by Real Time Manipulation of ITD**—*Alexander Lindau, Jorgos Estrella, Stefan Weinzierl*, Technical University of Berlin, Berlin, Germany

The dynamic binaural synthesis of acoustic environments is usually constrained to the use non-individual impulse response datasets, measured with dummy heads or head and torso simulators. Thus, fundamental cues for localization such as interaural level differences (ILD) and interaural time differences (ITD) are necessarily corrupted to a certain degree. For ILDs, this is a minor problem as listeners may swiftly adapt to spectral coloration at least as long as an external reference is not provided. In contrast, ITD errors can be expected to lead to a constant degradation of localization. Hence, a method for the individual customization of dynamic binaural reproduction by means of real time manipulation of the ITD is proposed. As a prerequisite, subjectively artifact free techniques for the decomposition of binaural impulse responses into ILD and ITD cues are discussed. Finally, based on listening test results, an anthropometry-based prediction model for individual ITD correction factors is presented. The proposed approach entails further improvements of auditory quality of real time binaural synthesis.  
*Convention Paper 8088*

**14:30**

**P17-2 Perceptual Evaluation of Physical Predictors of the Mixing Time in Binaural Room Impulse Responses**—*Alexander Lindau, Linda Kosanke, Stefan Weinzierl*, Technical University of Berlin, Berlin, Germany

The mixing time of room impulse responses denotes the moment when the diffuse reverberation tail begins. A diffuse sound field can physically be defined by (1) equi-distribution of acoustical energy and (2) a uniform acoustical energy flux over the complete solid angle. Accordingly, the perceptual mixing time is the moment when the diffuse tail cannot be distinguished from that of any other position in the room. This provides an opportunity for reducing the length of binaural impulse responses that are dynamically exchanged in virtual acoustic environments (VAEs). Numerous model parameters and empirical features for the prediction of perceptual mixing time in rooms have been proposed. This paper aims at a perceptual evaluation of all potential estimators. Therefore, binaural impulse response data sets were collected with an adjustable head and torso simulator for a representative sample of rectangular-shaped rooms. Prediction performance was evaluated by linear regression using results of a listening test where mixing times could be adaptively altered in real time to determine a just audible transition time into a homogeneous diffuse tail. Regression formulae for the perceptual mixing time are presented, conveniently predicting perceptive mixing times to be used in the context of VAEs.  
*Convention Paper 8089*

**15:00**

**P17-3 HRTF Measurements with a Continuously Moving Loudspeaker and Swept Sines**—*Ville Pulkki, Mikko-Ville Laitinen, Ville Pekka Sivonen*, Aalto University School of Science and Technology, Aalto, Finland

An apparatus is described, which is designed to measure head-related transfer functions (HRTFs) for audio applications. A broadband, two-driver loudspeaker is rotated around the subject with continuous movement, and responses are measured with a swept-sine technique. Potential error sources are discussed and quantified, and it is shown that the responses are almost identical to responses measured with a static, small single-driver loudspeaker. It is also shown that the method can be used to measure a large number of HRTFs in a relatively short time period.  
*Convention Paper 8090*

**15:30**

**P17-4 In Situ Microphone Array Calibration for Parameter Estimation in Directional Audio Coding**—*Oliver Thiergart, Giovanni Del Galdo, Maja Taseska, Jose Angel Pineda Pardo, Fabian Kuech*, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

Directional audio coding (DirAC) provides an efficient representation of spatial sound using a downmix audio signal and parametric information, namely direction-of-arrival (DOA) and dif-



fuseness of sound. Input to the DirAC analysis are B-format signals, usually obtained via microphone arrays. The DirAC parameter estimation is impaired when phase mismatch between the array sensors occurs. We present an approach for the in situ microphone array calibration solely based on the DirAC parameters. The algorithm aims at providing consistent parameter estimates rather than matching the sensors explicitly. It does neither require to remove the sensors from the array, nor depend on a priori knowledge such as the array size. We further propose a suitable excitation signal to assure robust calibration in reverberant environments.

*Convention Paper 8093*

16:00

**P17-5 Sound Field Recording by Measuring Gradients** —*Mihailo Kolundzija*,<sup>1</sup> *Christof Faller*,<sup>1</sup> *Martin Vetterli*<sup>1,2</sup>

<sup>1</sup>Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

<sup>2</sup>University of California at Berkeley, Berkeley, CA, USA

Gradient-based microphone arrays, the horizontal sound field's plane wave decomposition, and the corresponding circular harmonics decomposition are reviewed. Further, a general relation between directivity patterns of the horizontal sound field gradients and the circular harmonics of any order is derived. A number of example differential microphone arrays are analyzed, including arrays capable of approximating the sound pressure gradients necessary for obtaining the circular harmonics up to order three.

*Convention Paper 8092*

16:30

**P17-6 Evaluation of a Binaural Reproduction System Using Multiple Stereo-Dipoles**—*Yesenia Lacouture Parodi*, *Per Rubak*, Aalborg University, Aalborg, Denmark

The sweet spot size of different loudspeaker configurations was investigated in a previous study carried out by the authors. Closely spaced loudspeakers showed a wider control area than the standard stereo setup. The sweet spot with respect to head rotations showed to be especially large when the loudspeakers are placed at elevated positions. In this paper we describe a system that uses the characteristics of the loudspeakers placed above the listener. The proposed system is comprised of three pairs of closely spaced loudspeakers: one pair placed in front, one placed behind, and one placed above the listener. The system is based on the idea of dividing the sound reproduction into regions to reduce front-back confusions and enhance the virtual experience without the aid of a head tracker. A set of subjective experiments with the intention of evaluating and comparing the performance of the proposed system are also discussed.

*Convention Paper 8091*

17:00

**P17-7 Conditioning of the Problem of a Source Array Design with Inverse Approach**—

*Jeong-Guon Ih*,<sup>1</sup> *Wan-Ho Cho*<sup>1,2</sup>

<sup>1</sup>KAIST, Daejeon, Korea

<sup>2</sup>Chuo University, Bunkyo, Tokyo, Japan

An inverse approach based on the acoustical holography concept can be effectively applied to the acoustic field rendered for achieving a target sound field given as a relative response distribution of sound pressure. To implement this method, the source configuration should be determined a priori, and a meaningful inverse solution of an ill-conditioned transfer matrix should be obtained. To choose efficient source positions that are almost mutually independent, the redundancy detection algorithm like the effective independence method was employed to decide the proper positions for a given number of sources. In this way, an efficient and stable filter set for a source array in controlling the sound field can be obtained. An interior domain with irregularly shaped boundaries was adopted as the target field to control for testing the suggested inverse method.

*Convention Paper 8094*

**Session P18**

**14:00 – 18:00**

**Monday, May 24**

**Room C5**

**AUDIO CODING AND COMPRESSION**

Chair: **Jamie Angus**, University of Salford, Salford, UK

14:00

**P18-1 Identification of Perceptual Attributes for the Assessment of Low Bit Rate Multichannel Audio Codecs**—*Paulo Marins*, Capes, Brasilia, Brazil

This paper describes an experiment that aimed to identify independent perceptual attributes detailing the distortions introduced by low bit rate multichannel audio codecs. This study consisted of a verbal elicitation experiment in which the Repertory Grid Technique (RGT) method was employed. Four attributes were identified and labeled. The independent perceptual attributes uncovered were: "coding and high frequency noise," "spatial image clarity," "scene width," and "tone color." The overall importance of each of the independent perceptual attributes identified was quantified. It was found that "coding and high frequency noise" was the attribute with the highest overall importance, followed by "spatial image clarity," "scene width," and "tone color."

*Convention Paper 8095*

*Paper will not be presented but is available for purchase.*

14:30

**P18-2 High-Level Sound Coding with Parametric Blocks** —*Daniel Möhlmann*, *Otthein Herzog*, Universität Bremen, Bremen, Germany

This paper proposes a new parametric encoding model for sound blocks that is specifically designed for manipulation, block-based comparison, and morphing operations. Unlike other spectral models, only the temporal evolution of the dominant tone and its time-varying spectral

envelope are encoded, thus greatly reducing perceptual redundancy. All sounds are synthesized from the same set of model parameters, regardless of their length. Therefore, new instances can be created with greater variability than through simple interpolation. A method for creating the parametric blocks from an audio stream through partitioning is also presented. An example of sound morphing is shown and applications of the model are discussed.

*Convention Paper 8096*

15:00

**P18-3 Exploiting High-Level Music Structure for Lossless Audio Compression**—*Florin Ghido*, Tampere University of Technology, Tampere, Finland

We present a novel concept of “noncontiguous” audio segmentation by exploiting the high-level music structure. The existing lossless audio compressors working in asymmetrical mode divide the audio into quasi-stationary segments of variable length by recursive splitting (MPEG-4 ALS) or by dynamic programming (asymmetrical OptimFROG) before computing a set of linear prediction coefficients for each segment. Instead, we combine several variable length segments into a group and use a single set of linear prediction coefficients for each group. The optimal algorithm for combining has exponential complexity and we propose a quadratic time approximation algorithm. Integrated into asymmetrical OptimFROG, the proposed algorithm obtains up to 1.20% (on average 0.23%) compression improvements with no increase in decoder complexity.

*Convention Paper 8097*

15:30

**P18-4 Interactive Teleconferencing Combining Spatial Audio Object Coding and DirAC Technology**—*Jürgen Herre, Cornelia Falch, Dirk Mahne, Giovanni del Galdo, Markus Kallinger, Oliver Thiergart*, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

The importance of telecommunication continues to grow in our everyday lives. An ambitious goal for developers is to provide the most natural way of audio communication by giving users the impression of being located next to each other. MPEG Spatial Audio Object Coding (SAOC) is a technology for coding, transmitting, and interactively reproducing spatial sound scenes on any conventional multi-loudspeaker setup (e.g., ITU 5.1). This paper describes how Directional Audio Coding (DirAC) can be used as recording front-end for SAOC-based teleconference systems to capture acoustic scenes and to extract the individual objects (talkers). By introducing a novel DirAC to SAOC parameter transcoder, a highly efficient way of combining both technologies is presented that enables interactive, object-based spatial teleconferencing.

*Convention Paper 8098*

16:00

**P18-5 A New Parametric Stereo- and Multichannel Extension for MPEG-4 Enhanced Low Delay**

**AAC (AAC-ELD)**—*María Luis Valero,<sup>1</sup> Andreas Hölzer,<sup>2</sup> Markus Schnell,<sup>1</sup> Johannes Hilpert,<sup>1</sup> Manfred Lutzky,<sup>1</sup> Jonas Engdegård,<sup>3</sup> Heiko Purnhagen,<sup>3</sup> Per Ekstrand,<sup>3</sup> Kristofer Kjörling<sup>3</sup>*

<sup>1</sup>Fraunhofer Institute for Integrated Circuits IIS,

Erlangen, Germany

<sup>2</sup>DSP Solutions GmbH & Co., Regensburg, Germany

<sup>3</sup>Dolby Sweden AB, Stockholm, Sweden

ISO/MPEG standardizes two communication codecs with low delay: AAC-LD is a well established low delay codec for high quality communication applications such as video conferencing, tele-presence, and Voice over IP. Its successor AAC-ELD offers enhanced bit rate efficiency being an ideal solution for broadcast audio gateway codecs. Many existing and upcoming communication applications benefit from the transmission of stereo or multichannel signals at low bitrates. With low delay MPEG Surround, ISO has recently standardized a low delay parametric extension for AAC-LD and AAC-ELD. It is based on MPEG Surround technology with specific adaption for low delay operation. This extension comes along with a significant improved coding efficiency for transmission of stereo and multichannel signals.

*Convention Paper 8099*

16:30

**P18-6 Efficient Combination of Acoustic Echo Control and Parametric Spatial Audio Coding**—*Fabian Kuech, Markus Schmidt, Meray Zourub*, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

High-quality teleconferencing systems utilize surround sound to provide natural communication experience. Directional Audio Coding (DirAC) is an efficient parametric approach to capture and reproduce spatial sound. It uses a monophonic audio signal together with parametric spatial cue information. For reproduction, multiple loudspeaker signals are determined based on the DirAC stream. To allow for hands-free operation, multichannel acoustic echo control (AEC) has to be employed. Standard approaches apply multichannel adaptive filtering to address this problem. However, computational complexity constraints and convergence issues inhibit practical applications. This paper proposes an efficient combination of AEC and DirAC by explicitly exploiting its parametric sound field representation. The approach suppresses the echo components in the microphone signals solely based on the single channel audio signal used for the DirAC synthesis of the loudspeaker signals.

*Convention Paper 8100*

17:00

**P18-7 Sampling Rate Discrimination: 44.1 kHz vs. 88.2 kHz**—*Amandine Pras, Catherine Guastavino*, McGill University, Montreal, Quebec, Canada

It is currently common practice for sound engineers to record digital music using high-resolution formats, and then down sample the files to 44.1 kHz for commercial release. This study

aims at investigating whether listeners can perceive differences between musical files recorded at 44.1 kHz and 88.2 kHz with the same analog chain and type of AD-converter. Sixteen expert listeners were asked to compare 3 versions (44.1 kHz, 88.2 kHz, and the 88.2 kHz version down-sampled to 44.1 kHz) of 5 musical excerpts in a blind ABX task. Overall, participants were able to discriminate between files recorded at 88.2 kHz and their 44.1 kHz down-sampled version. Furthermore, for the orchestral excerpt, they were able to discriminate between files recorded at 88.2 kHz and files recorded at 44.1 kHz.

*Convention Paper 8101*

17:30

**P18-8 Comparison of Multichannel Audio Decoders for Use in Mobile and Handheld Devices—**

*Manish Nema, Ashish Malot, Nokia India Pvt. Ltd., Bangalore, Karnataka, India*

Multichannel audio provides immersive experience to listeners. Consumer demand coupled with technological improvements will drive consumption of high-definition content in mobile and handheld devices. There are several multichannel audio coding algorithms, both, proprietary ones like Dolby Digital, Dolby Digital Plus, Windows Media Audio Professional (WMA Pro), Digital Theater Surround High Definition (DTS-HD), and standard ones like Advanced Audio Coding (AAC), MPEG Surround, available in the market. This paper presents salient features/coding techniques of important multichannel audio decoders and a comparison of these decoders on key parameters like processor complexity, memory requirements, complexity/features for stereo playback, and quality/coding efficiency. The paper also presents a ranking of these multichannel audio decoders on the key parameters in a single table for easy comparison.

*Convention Paper 8102*

*Paper will be presented by Marina Villanueva-Barreiro.*

**Session P19**  
14:00 – 15:30

**Monday, May 24**  
C4-Foyer

**POSTERS: PSYCHOACOUSTICS  
AND LISTENING TESTS**

14:00

**P19-1 Auditory Perception of Dynamic Range in the Nonlinear System—***Andrew J. R. Simpson, Simpson Microphones, West Midlands, UK*

This paper is concerned with the perception of dynamic range in the nonlinear system. The work is differentiated from the generic investigation of “sound quality,” which is usually associated with studies of nonlinear distortion. The proposed hypothesis suggests that distortion products generated within the compressive type nonlinear system are able to act as loudness compensator for the associated amplitude compression in the perceived loudness function. The hypothesis is tested using the Time-Varying

Loudness model of Glasberg and Moore (2002) and further tested using AB scale hidden reference type listening test methods. The Loudness Overflow Effect (LOE) is introduced and bandwidth is shown to be a significant limiting factor. Results and immediate implications are briefly discussed.

*Convention Paper 8103*

14:00

**P19-2 A Subjective Evaluation of the Minimum Audible Channel Separation in Binaural Reproduction Systems through**

**Loudspeakers—***Yesenia Lacouture Parodi, Per Rubak, Aalborg University, Aalborg, Denmark*

To evaluate the performance of crosstalk cancellation systems the channel separation is usually used as a parameter. However, no systematic evaluation of the minimum audible channel separation has been found in the literature known by the authors. This paper describes a set of subjective experiments carried out to evaluate the minimum channel separation needed such the binaural signals with crosstalk are perceived to be equal to the binaural signals reproduced without crosstalk. A three alternative-forced-choice discrimination experiment, with a simple adaptive algorithm with weighed up-down method was used. The minimum audible channel separation was evaluated for the listeners placed at symmetric and asymmetric positions with respect to the loudspeakers. Eight different stimuli placed at two different locations were evaluated. Span angles of 12 and 60 degrees were also simulated. Results indicate that in order to avoid lateralization the channel separation should be below  $-15$  dB for most of the stimuli and around  $-20$  dB for broad-band noise.

*Convention Paper 8104*

14:00

**P19-3 Evaluation of Speech Intelligibility in Digital Hearing Aids—***Lorena Álvarez, Leticia Vaqueró,*

*Enrique Alexandre, Lucas Cuadra, Roberto Gil-Pita, University of Alcalá, Alcalá de Henares, Madrid, Spain*

This paper explores the feasibility of using the Speech Intelligibility Index (SII) to evaluate the performance of a digital hearing aid. This standardized measure returns a number, between zero and unity, which can be interpreted as the proportion of the total speech information available to the listener, and correlates with the intelligibility of the speech signal. The paper will focus on the use of the SII as a metric from which to compare the performance of two different hearing aids, in terms of speech intelligibility. From the purpose of this work, experiments employing data from four real subjects with mild-to-profound hearing losses, when using these different hearing aids, will be done. Results will show how the use of the SII can lead to a better selection of a hearing aid in the detriment of others, while avoiding the need for making extensive subjective listening tests.

*Convention Paper 8105*



14:00

**P19-4 Method to Improve Speech Intelligibility in Different Noise Conditions**—*Rogerio G. Alves, Kuan-Chieh Yen, Michael C. Vartanian, Sameer A. Gadre*, CSR - Cambridge Silicon Radio, Auburn Hills, MI, USA

Mobile communication applications have to address various environmental noise situations. In order to improve the quality of voice communication, not only an effective noise reduction algorithm for the far-end user is wanted, but also an algorithm that helps improve intelligibility for the near-end user in different environmental noise situations is desired. Due to it, the goal of this paper is to improve the overall voice communication experience of mobile device users by introducing a method to improve intelligibility by increasing perceptual loudness of the received speech signal in accordance with the environmental noise. Please note that, due to the characteristics of the mobile devices application, a method capable of working in real time with low computational complexity is highly desired.  
*Convention Paper 8106*

14:00

**P19-5 Contemporary Theories of Tinnitus Generation, Diagnosis and Management Practices**—*Stamatia Staikoudi*, Queen Margaret University, Edinburgh, Scotland, UK

Tinnitus can be defined as the perception of a sound in the head or the ears, in the absence of external acoustic stimulation. It is a symptom experienced by more than seven million people across the UK and many more worldwide, including children. Its pitch may vary from individual to individual, and it can be described as ringing, whistling, humming, or buzzing amongst other. We will be looking at contemporary theories for its generation, current methods of diagnosis, and management practices.  
*Convention Paper 8107*

14:00

**P19-6 Analytical and Perceptual Evaluation of Nonlinear Devices for Virtual Bass System**—*Nay Oo, Woon-Seng Gan*, Nanyang Technological University, Singapore

Nonlinear devices (NLDs) are generally used in virtual bass systems (VBS). The prime objective is to extend the low frequency bandwidth psychoacoustically by generating a series of harmonics in the upper-bass and/or mid-frequency range where loudspeakers can reproduce well. However, these artificially added harmonics introduce intermodulation distortion and may change the timbre of the sound tracks. In this paper nine memoryless NLDs are studied based on objective analysis and subjective listening tests. The objectives of this paper are (1) to quantify the spectral contents of NLDs when fed by single-tone; (2) to find out which type of NLDs is best for psychoacoustics bass enhancement through subjective listening tests and objective GedLee nonlinear distortion metric; and (3) to investigate whether there is any correlation between subjective listening tests results and

objective performance scores.  
*Convention Paper 8108*

**Workshop 11**  
14:00 – 15:45

**Monday, May 24**  
**Room C1**

**SURROUND FOR SPORTS**

Presenters: **Martin Black**, Senior Sound Supervisor & Technical Consultant, BSkyB, UK  
**Peter Davey**, Audio Quality Supervisor at Beijing 2008 Olympics and Vancouver 2010 Olympics  
**Ian Rosam**, 5.1 Audio Quality Supervisor for FIFA World Cup, Euro 2008, Beijing 2008 and Vancouver 2010 Olympics

Surround for Sports is an increasingly important area of multichannel audio. It is a de-facto standard for large-scale productions like the Olympics or the football world and European championships.

The presenters, both experienced audio supervisors for such events, will touch on a variety of subjects regarding surround sound design for sports and many practical issues like: crowd/audience; field of play FX; game sounds, e.g., ball kicks; competitors, e.g., curling; referees/umpires, e.g., rugby/tennis; board sounds, e.g., darts/basketball; scoring/timing, e.g., fencing buzzers, boxing time bell; commentators/reporters out of vision; commentators/presenters/reporters in vision. Where those elements should be in a 5.1 mix will be discussed as well as: use of center channel, use of the LFE, HDTV stereo fold-down in a set top box, Dolby E, metadata, bass management, and last but not least localization of sounds with 5.1 and human hearing.

**Tutorial 7**  
14:00 – 15:45

**Monday, May 24**  
**Room C6**

**SCREEN CURRENT INDUCED NOISE (SCIN), SHIELDING, AND GROUNDING—MYTHS VS. REALITY**

Presenters: **Bruce C. Olson**, Olson Sound Design, Brooklyn Park, USA  
**John Woodgate**, J M Woodgate and Associates, Rayleigh, UK

Since the landmark series of AES *Journal* articles on Screen Current Induced Noise (SCIN), Shielding, and the Pin 1 problem, were published in June of 1995 there has been much discussion of the results and their implications for audio systems. This tutorial will demonstrate SCIN and show a Spice model that helps to explain what is happening. These effects are often confused with grounding and shielding schemes so we will also explain which interactions are real and which ones are mythical.

**Tutorial 8**  
14:00 – 15:45

**Monday, May 24**  
**Room C2**

**LES PAUL: WE USE HIS INNOVATIONS EVERY DAY**

Presenter: **Barry Marshall**, New England Institute of Art, Brookline, MA, USA

The many contributions of the late Les Paul to the art and technology of recording were often overlooked in media coverage of his passing last year at the age of 94. While his importance as a musician and as an innovator



of the solid body electric guitar certainly deserve wide praise and respect, his developments in recording continue to be used by producers, engineers and musicians at virtually every recording session. Producer and educator Barry Marshall looks at the career of Les Paul and plays some of his ground-breaking recordings as a sideman, as a solo artist, and as a part of the Les Paul and Mary Ford duo act. Special emphasis in the presentation will be placed on the way that Les's musicianship and musical instincts drove his technical breakthroughs.

### Student Event/Career Development MENTORING SESSIONS

Monday, May 24, 14:00 – 15:45  
C4-Foyer

Students are invited to sign-up for an individual meeting with a distinguished mentor from the audio industry. The opportunity to sign up will be given at the end of the opening SDA meeting. Any remaining open spots will be posted in the student area. All students are encouraged to participate in this exciting and rewarding opportunity for individual discussion with industry mentors.

**Monday, May 24 14:00 Room Saint Julien**  
**Technical Committee Meeting on Loudspeakers and Headphones**

**Monday, May 24 14:30 Room Mouton Cadet**  
**Standards Committee Meeting SC-03-06 Digital Library and Archive Systems**

**Monday, May 24 15:00 Room Saint Julien**  
**Technical Committee Meeting on Audio for Telecommunications**

**Workshop 12 Monday, May 24**  
**16:00 – 18:00 Room C6**

### AUDIO HISTORY, ARCHIVING, AND RESTORATION

Chair: **Sean W. Davies**, S.W. Davies Ltd., Aylesbury, UK

Panelists: *Ted Kendall*, Sound Restorer, Kington, UK  
*John Liffen*, The Science Museum, London, UK  
*Will Prentice*, The National Sound Archive, British Library, London, UK  
*Nadia Walaskowitz*, Phonogram Archive, Vienna, Austria

This workshop covers: (a) the preservation of historic audio equipment, both as exhibits and as working apparatus required for playback of historic formats; (b) the arrangement and management of archives/collections; (c) the preservation of the audio content for distribution.

**Workshop 13 Monday, May 24**  
**16:00 – 18:00 Room C2**

### LOUDNESS IN BROADCASTING— THE NEW EBU RECOMMENDATION R128

Chair: **Andrew Mason**, BBC R&D, London, UK

Panelists: *Jean-Paul Moerman*, Salzbrenner Stageteck Media Group, Buttenheim, Germany  
*Richard van Everdingen*, Dutch Broadcasting Loudness Committee

The EBU group P/LOUD is approaching the final stage of its work that will result in recommendations that will have a profound effect on any audio production in broadcasting. The gradual switch from peak to loudness normalization combined with a new maximum true peak level and the usage of the descriptor "loudness range" allows for the first time to fully characterize the audio part of a program. More importantly it has the potential to solve the most frequent complaint of the listeners, that of severe level inconsistencies. This is the first time that the new EBU loudness recommendation R128 is presented in detail, alongside a detailed introduction to the subject as well as practical case studies.

### Student Event/Career Development EDUCATION FORUM PANEL

Monday, May 24, 16:00 – 18:00  
Room C6

Chair: **Gary Gottlieb**

Panelists: *Chuck Ainlay*  
*Akira Fukada*  
*George Massenburg*  
*Ulrike Schwarz*

### How Does It Sound Now? The Evolution of Audio

One day Chet Atkins was playing guitar when a woman approached him. She said, "That guitar sounds beautiful." Chet immediately quit playing. Staring her in the eyes he asked, "How does it sound now?" The quality of the sound in Chet's case clearly rested with the player, not the instrument, and the quality of our product ultimately lies with us as engineers and producers, not with the gear we use. The dual significance of this question, "How does it sound now," informs our discussion, since it addresses both the engineer as the driver and the changes we have seen and heard as our business and methodology have evolved through the decades. Let's start by exploring the methodology employed by the most successful among us when confronted with new and evolving technology. How do we retain quality and continue to create a product that conforms to our own high standards? Is this even possible for students in today's market, with the option of true apprenticeships all but gone? How can students develop chops, aesthetics, and technical knowledge? How can they become prepared to thrive and produce a quality product? Can we help give educators the tools they need to address the issue of continuity of quality? This may lead to other conversations about the musicians we work with, the consumers we serve, and the differences and similarities between their standards and our own. How high should your standards be? How should it sound now? How should it sound tomorrow?

**Monday, May 24 16:00 Room Saint Julien**  
**Technical Committee Meeting on Audio for Games**

**Monday, May 24 16:00 Room Mouton Cadet**  
**Standards Committee Meeting SC-03-07 Audio Metadata**

Session P20  
16:30 – 18:00

Monday, May 24  
C4-Foyer

**POSTERS: AUDIO CONTENT MANAGEMENT—  
AUDIO INFORMATION RETRIEVAL**

16:30

**P20-1 Complexity Scalable Perceptual Tempo Estimation From HE-AAC Encoded Music—**

*Danilo Hollosi,<sup>1</sup> Arijit Biswas<sup>2</sup>*

<sup>1</sup>Ilmenau University of Technology, Ilmenau, Germany

<sup>2</sup>Dolby Germany GmbH, Nürberg, Germany

A modulation frequency-based method for perceptual tempo estimation from HE-AAC encoded music is proposed. The method is designed to work on fully-decoded PCM-domain; the intermediate HE-AAC transform-domain after partial decoding; and directly on HE-AAC compressed-domain using Spectral Band Replication (SBR) payload. This offers complexity scalable solutions. We demonstrate that SBR payload is an ideal proxy for tempo estimation directly from HE-AAC bit-streams without even decoding them. A perceptual tempo correction stage is proposed based on rhythmic features to correct for octave errors in every domain. Experimental results show that the proposed method significantly outperforms two commercially available systems, both in terms of accuracy and computational speed.

*Convention Paper 8109*

16:30

**P20-2 On the Effect of Reverberation on Musical Instrument Automatic Recognition—**

*Mathieu Barthelet, Mark Sandler, Queen Mary University of London, London, UK*

This paper investigates the effect of reverberation on the accuracy of a musical instrument recognition model based on Line Spectral Frequencies and K-means clustering. One-hundred-eighty experiments were conducted by varying the type of music databases (isolated notes, solo performances), the stage in which the reverberation is added (learning, and/or testing), and the type of reverberation (3 different reverberation times, 10 different dry-wet levels). The performances of the model systematically decreased when reverberation was added at the testing stage (by up to 40%). Conversely, when reverberation was added at the training stage, a 3% increase of performance was observed for the solo performances database. The results suggest that pre-processing the signals with a dereverberation algorithm before classification may be a means to improve musical instrument recognition systems.

*Convention Paper 8110*

16:30

**P20-3 Harmonic Components Extraction in Recorded Piano Tones—**

*Carmine Emanuele Cella, Università di Bologna, Bologna, Italy*

It is sometimes desirable, in the purpose of analyzing recorded piano tones, to remove from the

original signal the noisy components generated by the hammer strike and by other elements involved in the piano action. In this paper we propose an efficient method to achieve such result, based on adaptive filtering and automatic estimation of fundamental frequency and inharmonicity; the final method, applied on a recorded piano tone, produces two separate signals containing, respectively, the hammer knock and the harmonic components. Some sound examples to listen for evaluation are available on the web as specified in the paper.

*Convention Paper 8111*

16:30

**P20-4 Browsing Sound and Music Libraries by Similarity—**

*Stéphane Dupont,<sup>1</sup> Christian Frisson,<sup>2</sup> Xavier Siebert,<sup>1</sup> Damien Tardieu<sup>1</sup>*

<sup>1</sup>Université de Mons, Mons, Belgium

<sup>2</sup>Université Catholique de Louvain, Louvain-la-Neuve, Belgium

This paper presents a prototype tool for browsing through multimedia libraries using content-based multimedia information retrieval techniques. It is composed of several groups of components for multimedia analysis, data mining, interactive visualization, as well as connection with external hardware controllers. The musical application of this tool uses descriptors of timbre, harmony, as well as rhythm and two different approaches for exploring/browsing content. First, a dynamic data mining allows the user to group sounds into clusters according to those different criteria, whose importance can be weighted interactively. In a second mode, sounds that are similar to a query are returned to the user, and can be used to further proceed with the search. This approach also borrows from multi-criteria optimization concept to return a relevant list of similar sounds.

*Convention Paper 8112*

16:30

**P20-5 On the Development and Use of Sound Maps for Environmental Monitoring—**

*Maria Rangoussi,<sup>1</sup> Stelios M. Potirakis,<sup>1</sup> Ioannis Paraskevas,<sup>1</sup> Nicolas-Alexander Tatlas<sup>2</sup>*

<sup>1</sup>Technological Education Institute of Piraeus,

Aigaleo-Athens, Greece

<sup>2</sup>University of Patras, Patras, Greece

The development, update, and use of sound maps for the monitoring of environmental interest areas is addressed in this paper. Sound maps constitute a valuable tool for environmental monitoring. They rely on networks of microphones distributed over the area of interest to record and process signals, extract and characterize sound events and finally form the map; time constraints are imposed by the need for timely information representation. A stepwise methodology is proposed and a series of practical considerations are discussed to the end of obtaining a multi-layer sound map that is periodically updated and visualizes the sound content of a “scene.” Alternative time-frequency-based features are investigated as to their efficiency within the framework of a hierarchical classification structure.

16:30

**P20-6 The Effects of Reverberation on Onset Detection Tasks**—*Thomas Wilmering, György Fazekas, Mark Sandler*, Queen Mary University of London, London, UK

The task of onset detection is relevant in various contexts such as music information retrieval and music production, while reverberation has always been an important part of the production process. The effect may be the product of the recording space or it may be artificially added, and, in our context, destructive. In this paper we evaluate the effect of reverberation on onset detection tasks. We compare state-of-the-art techniques and show that the algorithms have varying degrees of robustness in the presence of reverberation depending on the content of the analyzed audio material.

*Convention Paper 8114*

16:30

**P20-7 Segmentation and Discovery of Podcast Content**—*Steven Hargreaves, Chris Landone, Mark Sandler, Panos Kudumakis*, Queen Mary University of London, London, UK

With ever increasing amounts of radio broadcast material being made available as podcasts, sophisticated methods of enabling the listener to quickly locate material matching their own personal tastes become essential. Given the ability to segment a podcast that may be in the order of one or two hours duration into individual song previews, the time the listener spends searching for material of interest is minimized. This paper investigates the effectiveness of applying multiple feature extraction techniques to podcast segmentation and describes how such techniques could be exploited by a vast number of digital media delivery platforms in a commercial cloud-based radio recommendation and summarization service.

*Convention Paper 8115*

**Monday, May 24 16:00 Room Mouton Cadet Standards Committee Meeting SC-03-04 Storage and Handling of Media**

**Special Event  
BANQUET**

Monday, May 24, 20:00 – 22:30  
Kew Bridge Steam Museum  
Green Dragon Lane  
Brentford, Middlesex

This year the Banquet will take place at one of London's most spectacular and dramatic venues—the Kew Bridge Steam Museum. The museum is set within a magnificent and atmospheric Victorian waterworks, where industry sits side by side with elegance. Guests enter via the Grand Junction Engine House and can enjoy a venue with a difference. Dinner is served inside the Steam Hall while the majestic steam engines work around you.

The ticket price includes all food and drinks and the bus to the museum and back. Buses will leave Novotel at 19:00 and 19:30.

**Session P21  
09:00 – 13:00**

**Tuesday, May 25  
Room C3**

**MULTICHANNEL AND SPATIAL AUDIO: PART 2**

Chair: **Ronald Aarts**, Philips Research Laboratories, Eindhoven, The Netherlands

09:00

**P21-1 Center-Channel Processing in Virtual 3-D Audio Reproduction over Headphones or Loudspeakers**—*Jean-Marc Jot, Martin Walsh*, DTS Inc., Scotts Valley, CA, USA

Virtual 3-D audio processing systems for the spatial enhancement of recordings reproduced over headphones or frontal loudspeakers generally provide a less compelling effect on center-panned sound components. This paper examines this deficiency and presents virtual 3-D audio processing algorithm modifications that provide a compelling spatial enhancement effect over headphones or loudspeakers even for sound components localized in the center of the stereo image, ensure the preservation of the timbre and balance in the original recording, and produce a more stable “phantom center” image over loudspeakers. The proposed improvements are applicable, in particular, in laptop and TV set audio systems, mobile internet devices, and home theater “soundbar” loudspeakers.

*Convention Paper 8116*

09:30

**P21-2 Parametric Representation of Complex Sources in Reflective Environments**—*Dylan Menzies*, De Montfort University, Leicester, UK

Aspects of source directivity in reflective environments are considered, including the audible effects of directivity and how these can be reproduced. Different methods of encoding and production are presented, leading to a new approach to extend parametric encoding of reverberation, as described in the DIRAC and MPEG formats, to include the response to source directivity.

*Convention Paper 8118*

10:00

**P21-3 Analysis and Improvement of Pre-Equalization in 2.5-Dimensional Wave Field Synthesis**—*Sascha Spors, Jens Ahrens*, Technische Universität Berlin, Berlin, Germany

Wave field synthesis (WFS) is a well established high-resolution spatial sound reproduction technique. Typical WFS systems aim at the reproduction in a plane using loudspeakers enclosing the plane. This constitutes a so-called 2.5-dimensional reproduction scenario. It has been shown that a spectral correction of the reproduced wave field is required in this context. For WFS this correction is known as pre-equalization filter. The derivation of WFS is based on a series



of approximations of the physical foundations. This paper investigates on the consequences of these approximations on the reproduced sound field and in particular on the pre-equalization filter. An exact solution is provided by the recently presented spectral division method and is employed in order to derive an improved WFS driving function. Furthermore, the effects of spatial sampling and truncation on the pre-equalization are discussed.

*Convention Paper 8121*

10:30

**P21-4 Discrete Wave Field Synthesis Using Fractional Order Filters and Fractional Delays**—*César D. Salvador*, Universidad de San Martín de Porres, Lima, Peru

A discretization of the generalized 2.5D Wave Field Synthesis driving functions is proposed in this paper. Time discretization is applied with special attention to the prefiltering that involves half-order systems and to the delaying that involves fractional-sample delays. Space discretization uses uniformly distributed loudspeakers along arbitrarily shaped contours: visual and numerical comparisons between lines and convex arcs, and between squares and circles, are shown. An immersive soundscape composed of nature sounds is reported as an example. Modeling uses MATLAB and real-time reproduction uses Pure Data. Simulations of synthesized plane and spherical wave fields, in the whole listening area, report a discretization percentage error of less than 1%, using 16 loudspeakers and 5th order IIR prefilters.

*Convention Paper 8122*

11:00

**P21-5 Immersive Virtual Sound Beyond 5.1 Channel Audio**—*Kangeun Lee, Changyong Son, Dohyung Kim*, Samsung Advanced Institute of Technology, Suwon, Korea

In this paper a virtual sound system is introduced for the next generation multichannel audio. The sound system provides a 9.1 channel surround sound via a conventional 5.1 loudspeaker layout and contents. In order to deliver 9.1 sound, the system includes channel upmixing and vertical sound localization that can create virtually localized sound at any spherical surface around human head. An amplitude panning coefficient is used to channel upmixing that includes a smoothing technique to reduce musical noise occurred by upmixing. The proposed vertical rendering is based on VBAP (vector based amplitude panning) using three loudspeakers among the 5.1. For quality test, our upmixing and virtual rendering method is evaluated in real 9.1 and 5.1 loudspeaker respectively and compared with Dolby Pro Logic IIz. The demonstrated performance is superior to the references.

*Convention Paper 8117*

11:30

**P21-6 Acoustical Zooming Based on a Parametric Sound Field Representation**—*Richard Schultz-Amling, Fabian Kuech, Oliver Thiergart,*

*Markus Kallinger*, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

Directional audio coding (DirAC) is a parametric approach to the analysis and reproduction of spatial sound. The DirAC parameters, namely direction-of-arrival and diffuseness of sound can be further exploited in modern teleconferencing systems. Based on the directional parameters, we can control a video camera to automatically steer on the active talker. In order to provide consistency between the visual and acoustical cues, the virtual recording position should match the visual movement. In this paper we present an approach for an acoustical zoom, which provides audio rendering that follows the movement of the visual scene. The algorithm does not rely on a priori information regarding the sound reproduction system as it operates directly in the DirAC parameter domain.

*Convention Paper 8120*

12:00

**P21-7 SoundDelta: A Study of Audio Augmented Reality Using WiFi-Distributed Ambisonic Cell Rendering**—*Nicholas Mariette,<sup>1</sup> Brian F. G. Katz,<sup>1</sup> Khaled Boussetta,<sup>2</sup> Olivier Guillerminet<sup>3</sup>*

<sup>1</sup>LIMSI-CNRS, Orsay, France

<sup>2</sup>Université Paris 13, Paris, France

<sup>3</sup>REMU, Paris, France

SoundDelta is an art/research project that produced several public audio augmented reality art-works. These spatial soundscapes were comprised of virtual sound sources located in a designated terrain such as a town square. Pedestrian users experienced the result as interactive binaural audio by walking through the augmented terrain, using headphones and the SoundDelta mobile device. SoundDelta uses a distributed "Ambisonic cell" architecture that scales efficiently for many users. A server renders Ambisonic audio for fixed user positions, which is streamed wirelessly to mobile users that render a custom, individualized binaural mix for their present position. A spatial cognition mapping experiment was conducted to validate the soundscape perception and compare with an individual rendering system.

*Convention Paper 8123*

12:30

**P21-8 Surround Sound Panning Technique Based on a Virtual Microphone Array**—*Filippo M. Fazi,<sup>1</sup> Toshiro Yamada,<sup>2</sup> Suketu Kamdar,<sup>2</sup> Philip A. Nelson,<sup>1</sup> Peter Otto<sup>2</sup>*

<sup>1</sup>University of Southampton, Southampton, UK

<sup>2</sup>University of California, San Diego, La Jolla, CA, USA

A multichannel panning technique is presented, which aims at reproducing a plane wave with an array of loudspeakers. The loudspeaker gains are computed by solving an acoustical inverse problem. The latter involves the inversion of a matrix of transfer functions between the loudspeakers and the elements of a virtual microphone array, the center of which corresponds to the location of the listener. The radius of the virtual microphone array is varied consistently with the frequency, in such a way that the transfer

function matrix is independent of the frequency. As a consequence, the inverse problem is solved for one frequency only, and the loudspeaker coefficients obtained can be implemented using simple gains.  
*Convention Paper 8119*

**Session P22**  
**09:00 – 12:00**

**Tuesday, May 25**  
**Room C5**

## MICROPHONES, CONVERTERS, AND AMPLIFIERS

Chair: **Mark Sandler**, Queen Mary University of London, London, UK

**09:00**

**P22-1 A Comparison of Phase-Shift Self-Oscillating and Carrier-Based PWM Modulation for Embedded Audio Amplifiers—Alexandre Huffenus,<sup>1</sup> Gaël Pillonnet,<sup>1</sup> Nacer Abouchi,<sup>1</sup> Frédéric Goutt<sup>2</sup>**

<sup>1</sup>Lyon Institute of Nanotechnology, Villeurbanne, France

<sup>2</sup>STMicroelectronics, Grenoble, France

This paper compares two modulation schemes for class-D amplifiers: phase-shift self-oscillating (PSSO) and carrier-based pulse width modulation (PWM). Theoretical analysis (modulation, frequency of oscillation, bandwidth, etc.), design procedure, and IC silicon evaluation will be shown for mono and stereo operation (on the same silicon die) on both structures. The design of both architectures will use as many identical building blocks as possible, to provide a fair, “all else being equal,” comparison. THD+N performance and idle consumption went from 0.02% and 5.6mA in PWM to 0.007% and 5.2mA in self-oscillating. Other advantages and drawbacks of the self-oscillating structure will be explained and compared to the classical carrier-based PWM one, with a focus on battery-powered applications.

*Convention Paper 8127*

**09:30**

**P22-2 Digital-Input Class-D Audio Amplifier—Hassan Ihs, Christian Dufaza, Primachip SAS, Marseille, France**

Not only digital-input class-D audio amplifier directly converts digital PCM-coded audio signal into power, it also exhibits superior performances with respect to traditional PWM analog class-D amplifiers. In the latter, several analog techniques have to be deployed to combat the many side-effects inherent to analog blocks, resulting in complex circuitry. Digital domain provides many degrees of freedom that allow combating signal non-idealities with no or just little extra cost. As the output to the real world is analog, digital class-D amplifier requires Analog to Digital Conversion (ADC) in feedback to compensate for jitter, signal distortions, and power supply noise. With careful sigma delta modulation design and few digital techniques, the class-D loop is stabilized and achieves superior audio

performances. The audio signal cascade chain is tremendously simplified resulting in significant reduction in area cost and power consumption. For applications that require analog input processing, the digital class-D still accepts analog inputs with no extra cost.  
*Convention Paper 8128*

**10:00**

**P22-3 Digital PWM Amplifier Using Nonlinear Feedback and Predistortion—Peter Craven,<sup>1</sup> Larry Hand,<sup>2</sup> Brian Attwood,<sup>3</sup> Jack Andersen<sup>4</sup>**

<sup>1</sup>Algol Applications Ltd., Steyning, West Sussex, UK

<sup>2</sup>Intersil/D2Audio, Austin, TX, USA

<sup>3</sup>PWM Systems, Crawley, Sussex, UK

<sup>4</sup>D2Audio, Austin, TX, USA (deceased)

A nonlinear feedback topology is used to reduce the deviations of a practical PWM output stage from ideal theoretical behavior, the theoretical nonlinearity of the PWM process being corrected using predistortion. As the final output is analog, an ADC is needed if the feedback is to be digital, and several problems arise from the practical limitations of commercially-available ADCs, including delay and addition of ultrasonic noise. We show how these problems can be minimized and illustrate the performance of a digital PWM amplifier in which feedback results in a significant reduction of distortion throughout the audio range.

*Convention Paper 8129*

**10:30**

**P22-4 Microphone Choice: Large or Small, Single or Double?—Martin Schneider, Georg Neumann GmbH, Berlin, Germany**

How do large and small diaphragm condenser microphones differ? A common misapprehension is that large capsules necessarily become less directional at low frequencies. It is shown that this is not a question of large or small, but rather of single or double diaphragm design. The different behaviors have a direct impact on the sound engineers' choice and placement of microphone. Likewise, the much debated question of proximity effect with multi-pattern microphones and omnidirectional directivity is discussed.

*Convention Paper 8124*

**11:00**

**P22-5 Improvements on a Low-Cost Experimental Tetrahedral Ambisonic Microphone—Dan T. Hemingson, Mark J. Sarisky, University of Texas at Austin, Austin, TX, USA**

An earlier paper [Hemingson and Sarisky, “A Practical Comparison of Three Tetrahedral Ambisonic Microphones,” 126th AES Convention Munich May 2009] compared two low-cost tetrahedral ambisonic microphones, an experimental microphone and a Core Sound TetraMic, using a Soundfield MKV or SPS422B as a standard for comparison. This paper examines improvements to the experimental device, including that suggested in the “future work” section of the original paper. Modifications to the capsules and a redesign of the electronics package made signif-

icant improvements in the experimental microphone. Recordings were made in natural environments and of live performances, some simultaneously with the Soundfield standard. Of interest is the use of the low-cost surround microphone for student and experimental education.

*Convention Paper 8125*

**11:30**

**P22-6 Analysis of the Interaction between Ribbon Motor, Transformer, and Preamplifier and its Application in Ribbon Microphone Design—**  
*Julian David*, Audio Engineering Associates, Pasadena, CA, USA, University of Applied Sciences Düsseldorf, Düsseldorf, Germany

The transformer in ribbon microphones interacts with the complex impedances of both the ribbon motor and the subsequent preamplifier. This paper presents a test setup that takes the influences of source and load impedances into account for predicting the amplitude and phase response of a specific ribbon microphone design in combination with different transformers. As a result, the effect of the transformer on the amplitude and phase response of the system can be simulated in good approximation by means of a generic low-impedance source. This allows for optimizing the ribbon/transformer/load circuit under laboratory conditions in order to achieve the desired microphone performance.

*Convention Paper 8126*

**Workshop 14**  
**09:00 – 10:45**

**Tuesday, May 25**  
**Room C6**

### **10 THINGS TO GET RIGHT IN PA AND SOUND REINFORCEMENT**

Chair: **Peter Mapp**, Peter Mapp and Associates, Colchester, UK

Panelists: *Mark Bailey*, JBL/Harman International, UK  
*Jason Baird*, Martin Audio, High Wycombe, UK  
*Chris Foreman*, Community Professional Loudspeakers, Chester, PA, USA  
*Ralph Heinz*, Renkus Heinz, Foothill Ranch, CA, USA  
*Glenn Leembruggen*, Acoustic Directions Leichhardt, NSW, Australia

The workshop will discuss the 10 most important things to get right when designing/operating sound reinforcement and PA systems. However, there are many more things to consider than just the ten golden rules, and that the order of importance of these often changes depending upon the venue and type of system. The workshop aims to provide a practical approach to sound systems design and operation and will be illustrated with many practical examples and case histories. Each workshop panelist has many years practical experience and between them can cover just about any aspect of sound reinforcement and PA systems design, operation, and technology. Come along to a workshop that aims to answer questions you never knew you had but of course to find out the ten most important ones you will need to attend the session.

**Workshop 15**  
**09:00 – 10:45**

**Tuesday, May 25**  
**Room C1**

### **SINGLE-UNIT SURROUND MICROPHONES**

Chair: **Eddy B. Brixen**, EBB-consult, Smørum, Denmark

Panelists: *Mikkel Nymand*, DPA Microphones, Allerød, Denmark  
*Mattias Strömberg*, Milad, Helsingborg, Sweden  
*Helmut Wittek*, SCHOEPS Mikrofone GmbH, Karlsruhe, Germany

The workshop will present available single-unit surround sound microphones in a kind of shoot out. There are a number of these microphones available and more units are on their way. These microphones are based on different principles. However, due to their compact sizes there may/may not be restrictions to the performance. This workshop will present the different products and the ideas and theories behind them.

**Tutorial 9**  
**09:00 – 10:45**

**Tuesday, May 25**  
**Room C2**

### **COMPRESSION FX— USE YOUR POWER FOR GOOD, NOT EVIL**

Presenter: **Alex U. Case**, University of Massachusetts Lowell, MA, USA

Dynamic range compression, so often avoided by the purist and damned by the press, is most enthusiastically embraced by pop music creators. As an audio effect, it can be easily overused. Reined-in it can be difficult to perceive. It is always difficult to describe. As a tool, its controls can be counterintuitive, and its meters and flashing lights uninformative. In this tutorial—rich with audio examples—Case organizes the broad range of iconic effects created by audio compressors as they are used to reduce and control dynamic range, increase perceived loudness, improve intelligibility and articulation, reshape the amplitude envelope, add creative doses of distortion, and extract ambience cues, breaths, squeaks, and rattles. Learn when pop engineers reach for compression, know what sort of parameter settings are used (ratio, threshold, attack, and release), and advance your understanding of what to listen for and which way to tweak.

**Tuesday, May 25 09:00 Room Saint Julien**  
**Standards Committee Meeting SC-02-12 Audio**  
**Applications of Networks**

**Session P23**  
**10:30 – 12:00**

**Tuesday, May 25**  
**C4-Foyer**

### **POSTERS: AUDIO PROCESSING— MUSIC AND SPEECH SIGNAL PROCESSING**

**10:30**

**P23-1 Beta Divergence for Clustering in Monaural Blind Source Separation—**  
*Martin Spiertz*,  
*Volker Gnann*, RWTH Aachen University,  
Aachen, Germany

General purpose audio blind source separation algorithms have to deal with a large dynamic



range for the different sources to be separated. In the used algorithm the mixture is separated into single notes. These notes are clustered to construct the melodies played by the active sources. The non-negative matrix factorization (NMF) leads to good results in clustering the notes according to spectral features. The cost function for the NMF is controlled by the parameter beta. Beta should be adjusted properly depending on the dynamic difference of the sources. The novelty of this paper is to propose a simple unsupervised decision scheme that estimates the optimal parameter beta for increasing the separation quality over a large range of dynamic differences.

*Convention Paper 8130*

10:30

**P23-2 On the Effects of Room Reverberation in 3-D DOA Estimation Using a Tetrahedral**

**Microphone Array**—*Maximo Cobos, Jose J. Lopez, Amparo Marti*, Universidad Politécnic de Valencia, Valencia, Spain

This paper studies the accuracy in the estimation of the Direction-Of-Arrival (DOA) of multiple sound sources using a small microphone array. As other sparsity-based algorithms, the proposed method is able to work in undetermined scenarios, where the number of sound sources exceeds the number of microphones. Moreover, the tetrahedral shape of the array allows estimation of DOAs in the three-dimensional space easily, which is an advantage over other existing approaches. However, since the proposed processing is based on an anechoic signal model, the estimated DOA vectors are severely affected by room reflections. Experiments to analyze the resultant DOA distribution under different room conditions and source arrangements are discussed using both simulations and real recordings.

*Convention Paper 8131*

10:30

**P23-3 Long Term Cepstral Coefficients for Violin Identification**—*Ewa Lukasik*, Poznan University of Technology, Poznan, Poland

Cepstral coefficients in mel scale proved to be efficient features for speaker and musical instrument recognition. In this paper Long Term Cepstral Coefficients—LTCCs—of solo musical phrases are used as features for identification of individual violins. LTCC represents the envelope of LTAS—Long Term Average Spectrum—in linear scale useful to characterize the subtleties of violin sound in frequency domain. Results of the classification of 60 instruments are presented and discussed. It was shown, that if the experts' knowledge is applied to analyze violin sound, the results may be promising.

*Convention Paper 8132*

10:30

**P23-4 Adaptive Source Separation Based on Reliability of Spatial Feature Using**

**Multichannel Acoustic Observations**—*Mitsunori Mizumachi*, Kyushu Institute of Technology, Kitakyushu, Fukuoka, Japan

Separation of sound source can be achieved by spatial filtering with multichannel acoustic observations. However, the right algorithm should be prepared in each condition of acoustic scene. It is difficult to provide the suitable algorithm under real acoustic environments. In this paper an adaptive source separation scheme is proposed based on the reliability of a spatial feature, which gives an estimate of direction of arrival (DOA). As confidence measures for DOA estimates, the third and fourth moments for spatial features are employed to measure how sharp the main-lobes of spatial features are. This paper proposes to selectively use either spatial filters or frequency-selective filters without spatial filtering depending on the reliability of each DOA estimate.

*Convention Paper 8133*

10:30

**P23-5 A Heuristic Text-Driven Approach for Applied Phoneme Alignment**—*Konstantinos Avdelidis,*

*Charalampos Dimoulas, George Kalliris, George Papanikolaou*, Aristotle University of Thessaloniki, Thessaloniki, Greece

The paper introduces a phoneme matching algorithm considering a novel concept of functional strategy. In contrast to the classic methodologies that are focusing on the convergence to a fixed expected phonemic sequence (EPS), the presented method follows a more realistic approach. Based on text input, a soft EPS is populated taking into consideration the structural and linguistic deviations that may appear in a naturally spoken sequence. The results of the matching process is evaluated using fuzzy inference and is consisted of both the phoneme transition positions as well as the actual utterance phonemic content. An overview of convergence quality performance through a series of runs for the Greek language is presented.

*Convention Paper 8134*

10:30

**P23-6 Speech Enhancement with Hybrid Gain Functions**—*Xuejing Sun, Kuan-Chieh Yen,*

*Cambridge Silicon Radio, Auburn Hills, MI, USA*

This paper describes a hybrid gain function for single-channel acoustic noise suppression systems. The proposed gain function consists of Wiener filter and Minimum Mean Square Error—Log Spectral Amplitude estimator (MMSE-LSA) gain functions and selects respected gain values accordingly. Objective evaluation using a composite measure shows the hybrid gain function yields better results over using either of the two functions alone.

*Convention Paper 8135*

10:30

**P23-7 Human Voice Modification Using Instantaneous Complex Frequency**—

*Magdalena Kaniewska*, Gdansk University of Technology, Gdansk, Poland

The paper presents the possibilities of changing human voice by modifying instantaneous complex frequency (ICF) of the speech signal. The proposed method provides a flexible way of

altering voice without the necessity of finding fundamental frequency and formants' positions or detecting voiced and unvoiced fragments of speech. The algorithm is simple and fast. Apart from ICF it uses signal factorization into two factors: one fully characterized by its envelope and the other with positive instantaneous frequency. ICFs of the factors are modified individually for different sound effects.

*Convention Paper 8136*

10:30

**P23-8 Designing Optimal Phoneme-Wise Fuzzy Cluster Analysis**—*Konstantinos Avdelidis, Charalampos Dimoulas, George Kalliris, George Papanikolaou, Aristotle University of Thessaloniki, Thessaloniki, Greece*

A large number of pattern classification algorithms and methodologies have been proposed for the phoneme recognition task during the last decades. The current paper presents a prototype distance-based fuzzy classifier, optimized for the needs of phoneme recognition. This is accomplished by the specially designed objective function and a respective training strategy. Particularly, each phonemic class is represented by a number of arbitrary-shaped clusters that adaptively match the corresponding features space distribution. The formulation of the approach is capable of delivering a variety of related conclusions based on fuzzy logic arithmetic. An overview of the inference capability is presented in combination with performance results for the Greek language.

*Convention Paper 8137*

*Convention Paper 8138 was withdrawn*

**Workshop 16**  
11:00 – 13:00

**Tuesday, May 25**  
Room C6

### MPEG UNIFIED SPEECH AND AUDIO CODING

Chair: **Ralf Geiger**, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

Panelists: *Philippe Gournay*, University of Sherbrooke / VoiceAge, Quebec, Canada  
*Max Neuendorf*, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany  
*Lars Villemoes*, Dolby Sweden AB, Stockholm, Sweden

Recently the ISO/MPEG standardization group has launched an activity on unified speech and audio coding (USAC). This codec aims at achieving consistently high quality for speech, music, and mixed content over a broad range of bit rates. This outperforms current state-of-the-art coders. While low bit rate speech codecs focus on efficient representation of speech signals, they fail for music signals. On the other hand, generic audio codecs are designed for any kind of audio signals, but they tend to show unsatisfactory results for speech signals, especially at low bit rates. The new USAC codec unifies the best of both systems. This workshop provides an overview of architecture, performance, and applications of this new unified coding scheme. Experts in the fields

of speech coding and audio coding will present details on the technical solutions employed to top-notch unified speech and audio coding.

### Special Event AES/APRS/MPG PLATINUM ENGINEER BEHIND THE GLASS

Tuesday, May 25, 11:00 – 13:00  
Room C1

Moderator: **Howard Massey**

Panelists: *Chuck Ainlay*  
*Andy Bradfield*  
*Phil Harding*  
*George Massenburg*

A look behind the scenes with some of today's top British and American audio engineers that is sure to be lively and controversial. Topics will include new media and online distribution; the application of legacy (analog) technology in an increasingly digital world; and the critical role played by the engineer in driving both the technical and creative aspects of the recording process.

*Howard Massey* is a leading audio industry consultant, technical writer, and author of *Behind the Glass* and *Behind the Glass Volume II*, two collections of in-depth interviews with top record producers and audio engineers widely used in recording schools the world over. He also co-authored legendary Beatles engineer Geoff Emerick's acclaimed 2006 memoir, *Here, There, and Everywhere: My Life Recording the Music of the Beatles*.

**Tuesday, May 25 11:00 Room Saint Julien**  
**Standards Committee Meeting SC-04-03**  
**Loudspeaker Modeling and Measurement**

**Tutorial 10**  
11:15 – 12:15

**Tuesday, May 25**  
**Room C2**

### ADR

Presenter: **Dave Humphries**, Loopsync, Canterbury, UK

ADR is becoming more necessary than ever. Noisy locations, special effects, and dialog changes are a major part of any drama production. Wind noise, generators, lighting chokes, aircraft, traffic, and rain can all be eliminated by good ADR.

In this tutorial Dave Humphries will discuss the reasons why we need to put actors through this and how we can help minimize the need. What a Dialog Editor needs to know and how to help actors succeed at ADR. He will also demonstrate the art of recording ADR live.

**Tuesday, May 25 12:00 Room Mouton Cadet**  
**Technical Council Meeting**

**Tuesday, May 25 13:00 Room Saint Julien**  
**Standards Committee Meeting SC-03-02 Transfer**  
**Technologies**

**Session P24**  
14:00 – 17:30

**Tuesday, May 25**  
**Room C3**

### INNOVATIVE APPLICATIONS

Chair: **John Dawson**, Arcam, Cambridge, UK

14:00

- P24-1 The Serendiptichord: A Wearable Instrument for Contemporary Dance Performance—**  
*Tim Murray-Browne, Di Mainstone, Nick Bryan-Kinns, Mark D. Plumbley, Queen Mary University of London, London, UK*

We describe a novel musical instrument designed for use in contemporary dance performance. This instrument, the Serendiptichord, takes the form of a headpiece plus associated pods that sense movements of the dancer, together with associated audio processing software driven by the sensors. Movements such as translating the pods or shaking the trunk of the headpiece cause selection and modification of sampled sounds. We discuss how we have closely integrated physical form, sensor choice, and positioning and software to avoid issues that otherwise arise with disconnection of the innate physical link between action and sound, leading to an instrument that non-musicians (in this case, dancers) are able to enjoy using immediately.  
*Convention Paper 8139*

14:30

- P24-2 A Novel User Interface for Musical Timbre Design—***Allan Seago,<sup>1</sup> Simon Holland,<sup>2</sup> Paul Mulholland<sup>2</sup>*  
<sup>1</sup>London Metropolitan University, London, UK  
<sup>2</sup>Open University, UK

The complex and multidimensional nature of musical timbre is a problem for the design of intuitive interfaces for sound synthesis. A useful approach to the manipulation of timbre involves the creation and subsequent navigation or search of  $n$ -dimensional coordinate spaces or timbre spaces. A novel timbre space search strategy is proposed based on weighted centroid localization (WCL). The methodology and results of user testing of two versions of this strategy in three distinctly different timbre spaces are presented and discussed. The paper concludes that this search strategy offers a useful means of locating a desired sound within a suitably configured timbre space.  
*Convention Paper 8140*

15:00

- P24-3 Bi-Directional Audio-Tactile Interface for Portable Electronic Devices—***Neil Harris, New Transducers Ltd. (NXT), Cambridge, UK*

When an audio system uses the screen or case-work vibrating as the loudspeaker, it can also provide haptic feedback. Just as a loudspeaker may be used reciprocally as a microphone, the haptic feedback aspect of the design may be operated as a touch sensor. This paper considers how to model a basic system embodying these aspects, including the electrical part, with a finite element package. For a piezoelectric exciter, full reciprocal modeling is possible, but for electromagnetic exciters it is not, unless multi-physics simulation is supported. For the latter, a model using only lumped parameter mechanical elements is developed.  
*Convention Paper 8141*

15:30

- P24-4 Tactile Music Instrument Recognition for Audio Mixers—***Sebastian Merchel, Ercan Altinsoy, Maik Stamm, Dresden University of Technology, Dresden, Germany*

To use touch screens for digital audio workstations, particularly audio mixing consoles, is not very common today. One reason is the ease of use and the intuitive tactile feedback that hardware faders, knobs, and buttons provide. Adding tactile feedback to touch screens will largely improve usability. In addition touch screens can reproduce innovative extra tactile information. This paper investigates several design parameters for the generation of the tactile feedback. The results indicate that music instruments can be distinguished if tactile feedback is rendered from the audio signal. This helps to improve recognition of an audio signal source that is assigned, e.g., to a specific mixing channel. Applying this knowledge, the use of touch screens in audio applications becomes more intuitive.  
*Convention Paper 8142*

16:00

- P24-5 Augmented Reality Audio Editing—***Jacques Lemordant, Yohan Lasorsa, INRIA, Rhône-Alpes, France*

The concept of augmented reality audio (ARA) characterizes techniques where a physically real sound and voice environment is extended with virtual, geolocalized sound objects. We show that the authoring of an ARA scene can be done through an iterative process composed of two stages: in the first one the author has to move in the rendering zone to apprehend the audio spatialization and the chronology of the audio events, and in the second one a textual editing of the sequencing of the sound sources and DSP acoustics parameters is done. This authoring process is based on the joint use of two XML languages, OpenStreetMap for maps and A2ML for Interactive 3-D audio. A2ML, being a format for a cue-oriented interactive audio system, requests for interactive audio services are done through TCDL, a tag-based cue dispatching language. This separation of modeling and audio rendering is similar to what is done for the web of documents with HTML and CSS style sheets.  
*Convention Paper 8143*

16:30

- P24-6 Evaluation of a Haptic/Audio System for 3-D Targeting Tasks—***Lorenzo Picinali,<sup>1</sup> Bob Menelas,<sup>2</sup> Brian F. G. Katz,<sup>2</sup> Patrick Bourdoz<sup>2</sup>*  
<sup>1</sup>De Montfort University, Leicester, UK  
<sup>2</sup>LIMSI-CNRS, Orsay, France

While common user interface designs tend to focus on visual feedback, other sensory channels may be used in order to reduce the cognitive load of the visual one. In this paper non-visual environments are presented in order to investigate how users exploit information delivered through haptic and audio channels. A first experiment is designed to explore the effectiveness of a haptic audio system evaluated in a



single target localization task; a virtual magnet metaphor is exploited for the haptic rendering, while a parameter mapping sonification of the distance to the source, combined with 3-D audio spatialization, is used for the audio one. An evaluation is carried out in terms of the effectiveness of separate haptic and auditory feedbacks versus the combined multimodal feedback.

*Convention Paper 8144*

17:00

**P24-7 Track Displays in DAW Software: Beyond Waveform Views**—*Kristian Gohlke,<sup>1</sup> Michael Hltaky,<sup>1</sup> Sebastian Heise,<sup>1</sup> David Black,<sup>1</sup> Jörn Loviscach<sup>2</sup>*

<sup>1</sup>Hochschule Bremen (University of Applied Sciences), Bremen, Germany

<sup>2</sup>Fachhochschule Bielefeld (University of Applied Sciences), Bielefeld, Germany

For decades, digital audio workstation software has displayed the content of audio tracks through bare waveforms. We argue that the same real estate on the computer screen can be used for far more expressive and goal-oriented visualizations. Starting from a range of requirements and use cases, this paper discusses existing techniques from such fields as music visualization and music notation. It presents a number of novel techniques, aimed at better fulfilling the needs of the human operator. To this end, the paper draws upon methods from signal processing and music information retrieval as well as computer graphics.

*Convention Paper 8145*

**Session P25**  
14:00 – 18:00

**Tuesday, May 25**  
Room C5

**LISTENING TESTS AND EVALUATION  
PSYCHOACOUSTICS**

Chair: **Natanya Ford**, Buckinghamshire New University, UK

14:00

**P25-1 Toward a Statistically Well-Grounded Evaluation of Listening Tests—Avoiding Pitfalls, Misuse, and Misconceptions**—*Frederik Nagel,<sup>1</sup> Thomas Sporer,<sup>2</sup> Peter Sedlmeier<sup>3</sup>*

<sup>1</sup>Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

<sup>2</sup>Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany

<sup>3</sup>Technical University of Dresden, Dresden, Germany

Many recent publications in audio research present subjective evaluations of audio quality based on the Recommendation ITU-R BS.1534-1 (MUSHRA, MULTiple Stimuli with Hidden Reference and Anchor). This is a very welcome trend because it enables researchers to assess the implications of their developments. The evaluation of listening tests, however, sometimes suffers from an incomplete understanding of the underlying statistics. The present paper aims at

identifying the causes for the pitfalls and misconceptions in MUSHRA evaluations. It exemplifies the impact of falsely used or even misused statistics. Subsequently, schemes for evaluating the listeners' judgments that are well-grounded on statistical considerations comprising an understanding of the concepts of statistical power and effect size are proposed.

*Convention Paper 8146*

14:30

**P25-2 Audibility of Headphone Positioning Variability**—*Mathieu Paquier, Vincent Koehl, Université de Brest, Plouzané, France*

This paper aims at evaluating the audibility of spectral modifications induced by slight but realistic changes in the headphone position over a listener's ears. Recordings have been performed on a dummy head on which four different headphone models were placed eight times each. Musical excerpts and pink noise were played over the headphones and recorded with microphones located at the entrance of the blocked ear canal. These recordings were then presented to listeners over a single test headphone. The subjects had to assess the recordings in a 3I3AFC task to discriminate between the different headphone positions. The results indicate that, whatever the headphone model or the excerpt, the modifications caused by different positions were always perceived.

*Convention Paper 8147*

15:00

**P25-3 Objectivization of Audio-Video Correlation Assessment Experiments**—*Bartosz Kunka, Bozena Kostek, Gdansk University of Technology, Gdansk, Poland*

The purpose of this paper is to present a new method of conducting audio-visual correlation analysis employing a head-motion-free gaze tracking system. First, a review of related works in the domain of sound and vision correlation is presented. Then assumptions concerning audio-visual scene creation are briefly described. The objectivization process of carrying out correlation tests employing gaze-tracking system is outlined. The gaze tracking system developed at the Multimedia Systems Department is described, and its use for carrying out subjective tests is given. The results of subjective tests examining the relationship between video and audio associated with the video material are presented. Conclusions concerning the new methodology, as well as future work direction, are provided.

*Convention Paper 8148*

15:30

**P25-4 A New Time and Intensity Trade-Off Function for Localization of Natural Sound Sources**—*Hyunkook Lee, LG Electronics, Seoul, Korea*

This paper introduces a new set of psychoacoustic values of interchannel time difference (ICTD) and interchannel intensity difference (ICID) required for 10°, 20°, and 30° localization in the conventional stereophonic reproduction,

which were obtained using natural sound sources of musical instruments and wideband speech representing different characteristics. It then discusses the new concept of ICID and the ICTD trade-off function developed based on the relationship of the psychoacoustic values. The result of the listening test is conducted to verify the performance of the proposed method is also presented.

*Convention Paper 8149*

16:00

**P25-5 Effect of Signal-to-Noise Ratio and Visual Context on Environmental Sound**

**Identification—***Tifanie Bouchara,<sup>1</sup> Bruno L. Giordano,<sup>2</sup> Ilja Frissen,<sup>2</sup> Brian F. G. Katz,<sup>1</sup> Catherine Guastavino<sup>2</sup>*

<sup>1</sup>LIMSI-CNRS, Orsay, France

<sup>2</sup>McGill University, Montreal, Quebec, Canada

The recognition of environmental sounds is of main interest for the perception of our environment. This paper investigates whether visual context can counterbalance the impairing effect of signal degradation (signal-to-noise ratio, SNR) on the identification of environmental sounds. SNRs and semantic congruency between sensory modalities, i.e., auditory and visual information, were manipulated. Two categories of sound sources, living and nonliving were used. The participants' task was to indicate the category of the sound as fast as possible. Increasing SNRs and congruent audiovisual contexts enhanced identification accuracy and shortened reaction times. The results further indicated that living sound sources were recognized more accurately and faster than nonliving sound sources. A preliminary analysis of the acoustical factors mediating participants' responses revealed that the harmonic-to-noise ratio (HNR) sound signals was significantly associated with the probability of identifying a sound as living. Further, the extent to which participants' identifications were sensitive to the HNR appeared to be modulated by both SNR and audiovisual congruence.

*Convention Paper 8150*

16:30

**P25-6 The Influence of Individual Audio Impairments on Perceived Video Quality—**

*Leslie Gaston,<sup>1</sup> Jon Boley,<sup>2</sup> Scott Selter,<sup>1</sup> Jeffrey Ratterman<sup>1</sup>*

<sup>1</sup>University of Colorado, Denver, Denver, CO, USA

<sup>2</sup>LSB Audio LLC, Lafayette, IN, USA

As the audio, video, and related industries work toward establishing standards for subjective measures of audio/video quality, more information is needed to understand subjective audio/video interactions. This paper reports a contribution to this effort that aims to extend previous studies, which show that audio and video quality influence each other and that some audio artifacts affect overall quality more than others. In the current study, these findings are combined in a new experiment designed to reveal how individual impairments of audio affect perceived video quality. Our results show that some audio artifacts enhance the ability to identify video

artifacts, while others make discrimination more difficult.

*Convention Paper 8151*

*Paper presented by Scott Selter*

17:00

**P25-7 Vertical Localization of Sounds with Frequencies Changing over Three Octaves—**

*Eiichi Miyasaka, Tokyo City University, Yokohama, Kanagawa, Japan*

Vertical localization was investigated for sounds consisting of 22 tone-bursts ascending/descending along the whole-tone scale between C4 (262 Hz) and C7 (2093 Hz). The kinds of the sounds used were pure tones, one-third octave band noises, and piano-tones. The sounds were presented through a fixed loudspeaker (SP-A) set up just in front of listeners with numbered cards set perpendicularly (case-1) or with seven dummy loudspeakers attached to the numbered cards (case-2). The results show that most observers perceived the locations of the sound images moved upward from a loudspeaker around SP-A for the ascending sounds or downward for the descending sounds in both cases, although the sounds were radiated through the fixed loudspeaker (SP-A).

*Convention Paper 8152*

17:30

**P25-8 Variability in Perceptual Evaluation of HRTFs**

*—David Schönstein,<sup>1</sup> Brian F. G. Katz<sup>2</sup>*

<sup>1</sup>Arkamys, Paris, France

<sup>2</sup>Université Paris XI, Orsay Cedex, France

The implementation of the head-related transfer function (HRTF) is key to binaural rendering applications. An HRTF evaluation and selection is often required when individual HRTFs are not available. This paper examines the variability in perceptual evaluations of HRTFs using a listening test. A set of six different HRTFs was selected and was then used in a listening test based on the standardized MUSHRA method for evaluating audio quality. A total of six subjects participated, each having their own recorded HRTFs available. Subjects performed five repetitions of the listening test. While conclusive HRTF judgments were evident, a significantly large degree of variance was found. The effect of listener expertise on variability in perceptual judgments was also analyzed.

*Convention Paper 8153*

**Session P26**

**14:00 – 15:30**

**Tuesday, May 25**

**C4-Foyer**

**POSTERS: RECORDING, PRODUCTION, AND REPRODUCTION—MULTICHANNEL AND SPATIAL AUDIO**

14:00

**P26-1 Evaluation of Virtual Source Localization Using 3-D Loudspeaker Setups—**

*Florian Keiler, Johann-Markus Batke, Technicolor, Research, and Innovation, Hannover, Germany*

This paper evaluates the localization accuracy of ➤

different playback methods for 3-D spatial sound using a listening test. The playback methods are characterized by their panning functions that define the gain for each loudspeaker to playback a sound source positioned at a distinct pair of azimuth and elevation angles. The tested methods are Ambisonics decoding using the mode matching approach, vector base amplitude panning (VBAP), and a newly proposed 3-D robust panning approach. For irregular 3-D loudspeaker setups, as found in home environments, the mode matching shows poor localization. The new 3-D robust panning leads to a better localization and can also outperform the VBAP technology dependent on the source position and the loudspeaker setup used.

*Convention Paper 8060*

14:00

**P26-2 Optimization of the Localization Performance of Irregular Ambisonic Decoders for Multiple Off-Center Listeners**—*David Moore,<sup>1</sup> Jonathan Wakefield<sup>2</sup>*

<sup>1</sup>Glasgow Caledonian University, Glasgow, Lanarkshire, UK

<sup>2</sup>University of Huddersfield, Huddersfield, West Yorkshire, UK

This paper presents a method for optimizing the performance of irregular Ambisonic decoders for multiple off-center listeners. New off-center evaluation criteria are added to a multi-objective fitness function, based on auditory localization theory, which guides a heuristic search algorithm to derive decoder parameter sets for the ITU 5-speaker layout. The new evaluation criteria are based upon Gerzon's Metatheory of Auditory

Localization and have been modified to take into account off-center listening positions. The derived decoders exhibit improved theoretical localization performance for off-center listeners. The theoretical results are supported by initial listening test results.

*Convention Paper 8061*

14:00

**P26-3 Vibrational Behavior of High Aspect Ratio Multiactuator Panels**—*Basilio Pueo,<sup>1</sup> Jorge A. López,<sup>1</sup> Javier Moralejo,<sup>1</sup> José Javier López<sup>2</sup>*

<sup>1</sup>University of Alicante, Alicante, Spain

<sup>2</sup>Technical University of Valencia, Valencia, Spain

Multiactuator Panels (MAPs) consist of a flat panel of a light and stiff material to which a number of mechanical exciters are attached, creating bending waves that are then radiated as sound fields. MAPs can substitute the traditional dynamic loudspeaker arrays for Wave Field Synthesis (WFS) with added benefits, such as the low visual profile or omnidirectional radiation. However, the exciter interaction with the panel, the panel material, and the panel edge boundary conditions are some of the critical points that need to be evaluated and improved. In this paper the structural acoustic behavior of a high aspect ratio MAP is analyzed for two classical edge boundary conditions: free and clamped. For that purpose, the surface velocity over the whole area of the MAP prototype has been measured with a Laser Doppler Vibrometer (LDV), which helped in understanding the sound-generating behavior of the panel.

*Convention Paper 8062*

# 128<sup>th</sup> Convention Papers and CD-ROM

Convention Papers of many of the presentations given at the 126<sup>th</sup> Convention and a CD-ROM containing the 128<sup>th</sup> Convention Papers are available from the Audio Engineering Society. Prices follow:

**128<sup>th</sup> CONVENTION PAPERS (single copy)**

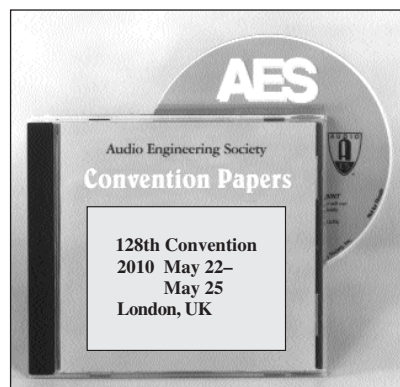
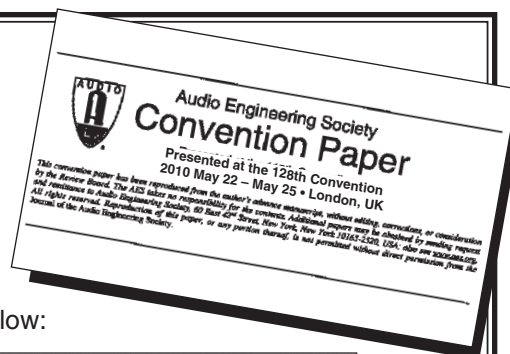
Member: \$ 5. (USA)

Nonmember: \$ 20. (USA)

**128<sup>th</sup> CD-ROM (193 papers)**

Member: \$190. (USA)

Nonmember: \$230. (USA)





14:00

**P26-4 Design of a Circular Microphone Array for Panoramic Audio Recording and Reproduction: Microphone Directivity—**

*Hüseyin Hacıhabiboglu, Enzo De Sena, Zoran Cvetkovic, King's College London, London, UK*

Design of a circularly symmetric multichannel recording and reproduction system is discussed in this paper. The system consists of an array of directional microphones evenly distributed on a circle and a matching array of loudspeakers. The relation between the microphone directivity and the radius of the circular array is established within the context of time-intensity stereophony. The microphone directivity design is identified as a constrained linear least squares optimization problem. Results of a subjective evaluation are presented that indicate the usefulness of the proposed microphone array design technique. *Convention Paper 8063*

14:00

**P26-5 Design of a Circular Microphone Array for Panoramic Audio Recording and Reproduction: Array Radius—**

*Enzo De Sena, Hüseyin Hacıhabiboglu, Zoran Cvetkovic, King's College London, London, UK*

A multichannel audio system proposed by Johnston and Lam aims at perceptual reconstruction of the sound field of an acoustic performance in its original venue. The system employs a circular microphone array, of 31 cm diameter, to capture relevant spatial cues. This design proved to be effective in the rendition of the auditory perspective, however other studies showed that there is still substantial room for improvement. This paper investigates the impact of the array diameter on the width and naturalness of the auditory images. To this end we propose a method for quantification and prediction of the perceived naturalness. Simulation results support array diameters close to that proposed by Johnston and Lam in the sense that they achieve optimal naturalness in the center of the listening area, but also suggest that larger arrays might provide a more graceful degradation of the naturalness for listening positions away from the center. *Convention Paper 8064*

14:00

**P26-6 MIAUDIO—Audio Mixture Digital Matrix—**

*David Pedrosa Branco, José Neto Vieira, Iouliia Skliarova, Universidade de Aveiro, Aveiro, Portugal*

Electroacoustic music is turning more and more to the sound diffusion techniques. Multichannel sound systems like BEAST and SARC are built so that the musician can independently control the intensity of several audio channels. This feature provides the possibility of creating several sound diffusion scenarios, i.e., immersion and the possibility of movement around the audience. The developed system (MIAUDIO) is a real-time sound diffusion system currently able to mix up to 8 audio input channels through 32 output channels. A hardware solution was adopt-

ed using a Field Programmable Gate Array (FPGA) to perform the mixture. The analog audio signals are conditioned, converted to digital format by several analog-to-digital converters, and then sent to the FPGA that is responsible to perform the mixing algorithm. The host computer connects to the FPGA via USB and is responsible for supplying the parameters that define the audio mixture. Being so, the user has control over the input levels through the output channels independently. MIAUDIO was successfully implemented with a low-cost solution when compared with similar systems. All the channels were tested using a Precision One system with very good results.

*Convention Paper 8065*

14:00

**P26-7 The Perception of Focused Sources in Wave Field Synthesis as a Function of Listener Angle—**

*Robert Oldfield, Ian Drumm, Jos Hirst, University of Salford, Salford, UK*

Wave field synthesis (WFS) is a volumetric sound reproduction technique that allows virtual sources to be positioned anywhere in space. The reproduction of the wave field of these sources means they can be accurately localized even when placed in front of the secondary sources/loudspeakers. Such "focused sources" are very important in WFS as they greatly add to the realism of an auditory scene, however the perception and localization-ability changes with listener and virtual source position. In this paper we present subjective tests to determine the localization accuracy as a function of angle, defining the subjective "view angle." We also show how improvements can be made through the addition of the first order image sources.

*Convention Paper 8066*

14:00

**P26-8 Gram-Schmidt-Based Downmixer and Decorrelator in the MPEG Surround Coding—**

*Der-Pei Chen, Hsu-Feng Hsiao, Han-Wen Hsu, Chi-Min Liu, National Chiao Tung University, Hsinchu, Taiwan*

MPEG Surround (MPS) coding is an efficient method for multichannel audio coding. In MPS coding, downmixing from multichannel signals into a fewer number of channels is an efficient way to achieve a high compression rate in an encoder. In decoder, an upmixing module combining with the decorrelator is the key module to reconstruct the multichannel signals. This paper considers the design of the downmixer and the decorrelator through the Gram-Schmidt orthogonal process. The individual and joint effects from the downmixer and decorrelator are verified through intensive subjective and objective quality measure.

*Convention Paper 8067*

**Workshop 17**  
14:00 – 15:30

**Tuesday, May 25**  
Room C1

**5.1 INTO 2 WON'T GO—THE PERILS OF FOLD-DOWN IN GAME AUDIO**

Chair: **Michael Kelly**, Sony Computer Entertainment Europe, London, UK

**Tuesday, May 25 14:30 Room Saint Julien**  
**AESSC Plenary Meeting**

Panelists: *Richard Furse*, Blue Ripple Sound Limited, London, UK  
*Simon Goodwin*, Codemasters, Southam, UK  
*Jean-Marc Jot*, DTS, Inc., Scotts Valley, CA, USA  
*Dave Malham*, University of York, York, UK

**Tutorial 11** **Tuesday, May 25**  
**16:00 – 17:30** **Room C6**

### LOUDSPEAKERS AND HEADPHONES

Presenter: **Wolfgang Klippel**, Klippel GmbH, Dresden, Germany

One mixing solution cannot suit mono, stereo, headphone, and various surround configurations. However games mix and position dozens of sounds on the fly, so they can readily make a custom mix, rather than rely upon downmixing or upmixing that penalizes listeners who do not use the default configuration. This workshop explains practical solutions to problems of stereo speaker, headphone, and mono compatibility (including Dolby ProLogic and 2.1 set-ups) without detriment to surround. It notes differences between the demands of games and cinema for surround and the challenges of reconciling de facto (quad+2) and theoretical (ITU 5.1) standard loudspeaker layouts and playing 5.1 channel content on a 7.1 loudspeaker system.

Distributed mechanical parameters describe the vibration and geometry of the sound radiating surface of loudspeaker drive units. This data is the basis for predicting the sound pressure output and a decomposition of the total vibration into modal and sound pressure related components. This analysis separates acoustical from mechanical problems, shows the relationship to the geometry and material properties, and gives indications for practical improvement. The tutorial combines the theoretical background with practical loudspeaker diagnostics illustrated on various kinds of transducers such as woofer, tweeter, compression driver, microspeaker, and headphones.

### Special Event

#### **AES/APRS—LIFE IN THE OLD DOGS YET— PART THREE: AFTER THE BALL— PROTECTING THE CROWN JEWELS**

Tuesday, May 25, 14:00 – 15:45  
Room C2

Moderator: **John Spencer**, BMS CHACE

Panelists: *Chris Clark*, British Library Sound Archive  
*Tommy D*, Producer  
*Simon Drake*, Naim Label Records  
*Tony Dunne*, A&R Coordinator, DECCA Records and UMTV/UMR, UK  
*Simon Hutchinson*, PPL  
*Paul Jessop*, Consultant IFI/RIAA  
*George Massenburg*, P&E Wing, NARAS  
*Crispin Murray*, Metropolis Mastering

A fascinating peek into the unspoken worlds of archiving and asset protection. It examines the issues surrounding retrievable formats that promise to future-proof recorded assets and the increasing importance of accurate recordings information (metadata). A unique group of experts from archiving and royalty distribution communities will hear a presentation from John Spencer, from BMS CHACE in Nashville, explaining his work with NARAS and the U.S. Library of Congress to establish an information schema for sound recording and Film and TV audio and then engage in a group discussion. The discussion then moves onto probably the most important topic to impact on the future of the sound and music economies—*how to keep what we've got and reward those who made it.*

### Student Event/Career Development

#### **STUDENT DELEGATE ASSEMBLY MEETING —PART 2**

Tuesday, May 25, 14:00 – 15:30  
Room C6

At this meeting the SDA will elect a new vice chair. One vote will be cast by the designated representative from each recognized AES student section in the European and International Regions. Judges' comments and awards will be presented for the Recording Competitions. Plans for future student activities at local, regional, and international levels will be summarized.