AES 140TH CONVENTION PROGRAM

JUNE 4–JUNE 7, 2016

PALAIS DES CONGRES, PARIS, FRANCE

The Winner of the 140th AES Convention Best Peer-Reviewed Paper Award is:

Phase Derivative Correction of Bandwidth-Extended Signals for Perceptual Audio Codecs— *Mikko-Ville Laitinen,*[1] *Sascha Disch,*[2] *Christopher Oates,*[2] *Ville Pulkki*[1]
[1]Aalto University, Espoo, Finland
[2]Fraunhofer IIS, Erlangen, Erlangen, Germany
Convention Paper 9490
To be presented on Saturday, June 4, in Session 2—Signal Processing—Part 1: Coding, Encoding, and Perception

* * * * *

The AES has launched an 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.A number of student-authored papers were nominated. The excellent quality of the submissions has made the selection process both challenging and exhilarating. The award-winning student paper will be honored during the Convention, and the student-authored manuscript will be considered for publication in a timely manner for 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 140th 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 Winner of the 140th AES Convention Student Paper Award is:

Plane Wave Identification with Circular Arrays by Means of a Finite Rate of Innovation Approach
—Falk-Martin Hoffmann, Filippo Maria Fazi, Philip Nelson, University of Southampton, Southampton, UK
Convention Paper 9521
To be presented on in Sessions 7: Audio Signal Processing—Part 2: Beamforming, Upmixing, HRTF and 12: Posters: Perception Part 1 and Audio Signal Processing Part 2

Session P1 **Saturday, June 4**
09:00 – 11:30 **Room 353**

AUDIO EQUIPMENT AND AUDIO FORMATS

Chair: **Menno van der Veen**, Ir. bureau Vanderveen, Hichtum, The Netherlands

09:00

P1-1 Linearization Technique of the Power Stage in Open-Loop Class D Amplifiers—*Federico Guanziroli, Pierangelo Confalonieri, Germano Nicollini*, STMicroelectronics, Milan, Italy

An efficient method to linearize the switching (power) stage of open-loop class D amplifiers is presented. This technique has been successfully applied to an open-loop fully-digital PWM class D amplifier designed in a 40 nm CMOS process leading to nearly 15 dB improvement in the Total Harmonic Distortion (THD). Simulated open-loop class D amplifier performance resulted to 105 dBA Signal-to-Noise Ratio (SNR), and 1W output power over 8 Ohm with 90% power efficiency and 0.014% THD.
Convention Paper 9484

09:30

P1-2 Physically-Based Large-Signal Modeling for Miniature Type Dual Triode Tube—*Shiori Oshimo, Kanako Takemoto, Toshihiko Hamasaki*, Hiroshima Institute of Technology, Hiroshima, Japan

A precise SPICE model for miniature (MT) triode tubes of high-μ 12AX7 and medium-μ 12AU7 is proposed, based on the physical analysis of the measurement results. Comparing the characteristics between these tubes, the grid current at lower plate voltage and positive grid bias condition is modeled successfully with novel perveance parameters for the first time, though it was known that the perveance depends on both grid and plate bias. It is shown that the modulation factor of the space charge for the MT triodes is different from the other classic tubes. The model is implemented in LTspice to result in a good replication for a variation of three-order magnitude of grid current and cathode current.
[Also a poster—see session P5-2]
Convention Paper 9485

10:00

P1-3 Analysis of Current MEMS Microphones for Cost-Effective Microphone Arrays—A Practical Approach—*Sven Kissner, Jörg Bitzer*, Jade Hochschule Oldenburg, Oldenburg, Germany

With this paper we present a practically relevant investigation of current, commercially available MEMS microphones (Micro-ElectroMechanical Systems). We compared the static noise floor exhibited by single and various parallel MEMS microphone configurations and a conventional and commonly used electret capsule, as well as the directivity patterns of selected configurations. The results suggest that while current types are exhibiting an already acceptable static noise floor, a direct parallel circuit of MEMS microphones allows further reductions of the noise floor close to the theoretical value of 3 dB SPL per doubling of number of microphones while maintaining omnidirectionality below 5 kHz.
Convention Paper 9486

10:30

P1-4 Matching the Amplifier to the Audio for Highly Efficient Linear Amplifiers—J*amie Angus*, University of Salford, Salford, Greater Manchester, UK

"Class-D" switching amplifiers are considered to be the most efficient amplifiers available on the market. However, designers must deal with supply rail, and radio frequency interference, as well as the need to switch power devices at high frequencies. Because of these, and other problems, not everyone wishes to use switching based technologies for their amplifiers. Unfortunately, linear amplifiers are significantly more inefficient than switching amplifiers, under sine wave testing. However real audio signals spend much more time at low amplitudes than a sine wave. By changing the switch points for "Class-G" or "Class-H" they can have efficiencies that rival "Class-D" amplifiers producing the same output. The paper develops optimum switch points for both single and multiple switching points, with respect to the expected amplitude distribution of the audio.
Convention Paper 9487

11:00

P1-5 Delay-Reduced Mode of MPEG-4 Enhanced Low Delay AAC (AAC-ELD)—*Markus Schnell, Wolfgang Jaegers, Pablo Delgado, Conrad Benndorf, Tobias Albert, Manfred Lutzky*, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

The MPEG-4 AAC Enhanced Low Delay (AAC-ELD) coder is well established in high quality communication applications, such as Apple's FaceTime, as well as in professional live broadcasting. Both applications require high interactivity, which typically demands an algorithmic codec delay between 15 ms and 35 ms. Recently, MPEG finalized a new delay-reduced mode for AAC-ELD featuring only a fraction of the regular algorithmic delay. This mode operates virtually at higher sampling rates while maintaining standard sampling rates for I/O. Supporting this feature, AAC-ELD can address even more delay critical applications, like wireless microphones or headsets for TV. In this paper main details of the delay-reduced mode of AAC-ELD are presented and application scenarios are outlined. Audio quality aspects are discussed and compared against other codecs with a delay below 10 ms.
[Also a poster—see session P5-3]
Convention Paper 9488

Session P2 **Saturday, June 4**
09:00 – 12:00 **Room 352B**

AUDIO SIGNAL PROCESSING—PART 1: CODING, ENCODING, AND PERCEPTION

Chair: **Dejan Todorovic**, Dirigent Acoustics, Belgrade, Serbia

09:00

P2-1 Low Complexity, Software Based, High Rate DSD Modulator Using Vector Quantification—*Thierry Heeb*,[1] *Tiziano Leidi*,[1] *Diego Frei*,[1] *Alexandre Lavanchy*[2]
[1]ISIN-SUPSI, Manno, Switzerland
[2]Engineered SA, Yverdon-les-Bains, Switzerland

High rate Direct Stream Digital (DSD) is emerging as a format of choice for distribution of high-definition audio content. However, real-time encoding of such streams requires considerable computing resources due to their high sampling rate, constraining implementations to hardware based platforms. In this paper we disclose a new modulator topology allowing for reduction in computational load and making real-time high rate DSD encoding suitable for software based implementation on off-the-shelf Digital Signal Processors (DSPs). We first present the architecture of the proposed modulator and then show results from a practical real-time implementation.
Convention Paper 9489

09:30

P2-2 Phase Derivative Correction of Bandwidth-Extended Signals for Perceptual Audio Codecs—*Mikko-Ville Laitinen*,[1] *Sascha Disch*,[2] *Christopher Oates*,[2] *Ville Pulkki*[1]
[1]Aalto University, Espoo, Finland
[2]Fraunhofer IIS, Erlangen, Erlangen, Germany

Bandwidth extension methods, such as spectral band replication (SBR), are often used in low-bit-rate codecs. They allow transmitting only a relatively narrow low-frequency region alongside with parametric information about the higher bands. The signal for the higher bands is obtained by simply copying it from the transmitted low-frequency region. The copied-up signal is processed by multiplying the magnitude spectrum with suitable gains based on the transmitted parameters to obtain a similar magnitude spectrum as that of the original signal. However, the phase spectrum of the copied-up signal is typically not processed but is directly used. In this paper we describe what are the perceptual consequences of using directly the copied-up phase spectrum. Based on the observed effects, two metrics for detecting the perceptually most significant effects are proposed. Based on these, methods how to correct the phase spectrum are proposed as well as strategies for minimizing the amount of transmitted additional parameter values for performing the correction. Finally, the results of formal listening tests are presented.
Convention Paper 9490

10:00

P2-3 AC-4 – The Next Generation Audio Codec—*Kristofer Kjörling*,[1] *Jonas Rödén*,[1] *Martin Wolters*,[2] *Jeff Riedmiller*,[3] *Arijit Biswas*,[2] *Per Ekstrand*,[1] *Alexander Gröschel*,[2] Per Hedelin,[1] *Toni Hirvonen*,[1] *Holger Hörich*,[2] *Janusz Klejsa*,[1] *Jeroen Koppens*,[1] *K. Krauss*,[2] *Heidi-Maria Lehtonen*,[1] *Karsten Linzmeier*,[2] *Hannes Muesch*,[3] *Harald Mundt*,[2] *Scott Norcross*,[3] *J. Popp*,[2] *Heiko Purnhagen*,[1]

Jonas Samuelsson,[1] Michael Schug,[2] L. Sehlström,[1] R. Thesing,[2] Lars Villemoes,[1] Mark Vinton[3]
[1]Dolby Sweden AB, Stockholm, Sweden
[2]Dolby Germany GmbH, Nuremberg, Germany
[3]Dolby Laboratories, San Francisco, CA USA

AC-4 is a state of the art audio codec standardized in ETSI (TS103 190 and TS103 190-2) and the TS103 190 is part of the DVB toolbox (TS101 154). AC-4 is an audio codec designed to address the current and future needs of video and audio entertainment services including broadcast and Internet streaming. As such, it incorporates a number of features beyond the traditional audio coding algorithms, such as capabilities to support immersive and personalized audio, support for advanced loudness management, video-frame synchronous coding, dialogue enhancement, etc. This paper will outline the thinking behind the design of the AC-4 codec, explain the different coding tools used, the systemic features included, and give an overview of performance and applications.
[Also a poster—see session P5-6]
Convention Paper 9491

10:30

P2-4 Using Phase Information to Improve the Reconstruction Accuracy in Sinusoidal Modeling—*Clara Hollomey,[1] David Moore,[1] Don Knox,[1] W. Owen Brimijoin,[2] William Whitmer[2]*
[1]Glasgow Caledonian University, Glasgow, Scotland, UK
[2]MRC Institute of Hearing Research, Glasgow, Scotland, UK

Sinusoidal modeling is one of the most common techniques for general purpose audio synthesis and analysis. Owing to the ever increasing amount of available computational resources, nowadays practically all types of sounds can be constructed up to a certain degree of perceptual accuracy. However, the method is computationally expensive and can for some cases, particularly for transient signals, still exceed the available computational resources. In this work methods derived from the realm of machine learning are exploited to provide a simple and efficient means to estimate the achievable reconstruction quality. The peculiarities of common classes of musical instruments are discussed and finally, the existing metrics are extended by information on the signal's phase propagation to allow for more accurate estimations.
[Also a poster—see session P5-8]
Convention Paper 9492

11:00

P2-5 Equalization of Spectral Dips Using Detection Thresholds—*Sunil G. Bharitkar,[1] Charles Q. Robinson,[2] Andrew Poulain[2]*
[1]HP Labs., Inc., San Francisco, CA, USA
[2]Dolby Laboratories, San Francisco, CA, USA

Frequency response equalization is often performed to improve audio reproduction. Variations from the target system response due to playback equipment or room acoustics can result in perceptible timbre distortion. In the first part of this paper we describe experiments conducted to determine the audibility of artificially introduced spectral dips. In particular, we measured notch depth detection threshold (independent variable) with respect to notch center frequency and Q-factor (independent variables). Listening tests were administered to 10 listeners in a small listening room and a screening room (small cinema with approximately 100 seats). Pink noise was used as the stimulus as it is perceptually flat (with roughly 3 dB/octave spectral tilt with frequency) and is known to be a reliable and discriminating signal for performing timbre judgments. The listeners gave consistent notch depth results with low variability around the mean value. The notch audibility data was then used to develop multiple candidate algorithms that generate equalization curves designed to perceptually match a desired target response, while minimizing the equalization gain applied. Informal subjective results validated the performance of the final algorithm.
Convention Paper 9493

11:30

P2-6 Single-Channel Audio Source Separation Using Deep Neural Network Ensembles—*Emad M. Grais, Gerard Roma, Andrew J. R. Simpson, Mark D. Plumbley,* University of Surrey, Guildford, Surrey, UK

Deep neural networks (DNNs) are often used to tackle the single channel source separation (SCSS) problem by predicting time-frequency masks. The predicted masks are then used to separate the sources from the mixed signal. Different types of masks produce separated sources with different levels of distortion and interference. Some types of masks produce separated sources with low distortion, while other masks produce low interference between the separated sources. In this paper a combination of different DNNs' predictions (masks) is used for SCSS to achieve better quality of the separated sources than using each DNN individually. We train four different DNNs by minimizing four different cost functions to predict four different masks. The first and second DNNs are trained to approximate reference binary and soft masks. The third DNN is trained to predict a mask from the reference sources directly. The last DNN is trained similarly to the third DNN but with an additional discriminative constraint to maximize the differences between the estimated sources. Our experimental results show that combining the predictions of different DNNs achieves separated sources with better quality than using each DNN individually.
[Also a poster—see session P5-7]
Convention Paper 9494

Session EB1 Saturday, June 4
09:00 – 10:15 Foyer

ENGINEERING BRIEFS—POSTERS

09:00

EB1-1 Sound Pressure Analysis For Closed-Box Loudspeaker Enclosures—*Charalampos Papadakos, Gavriil Kamaris, John Mourjopoulos,* University of Patras, Patras, Greece

This study employs a physical modeling method to explore the pressure distribution within typical closed-box loudspeaker enclosures of different shape and inner volume. The simulation results are compared to measurements in such enclosures. The results indicate that sound pressure within such enclosures often exceeds levels of 130 dB. The pressure profile is usually constant at lower frequencies and displays some strong resonances at higher frequencies due to normal modes. Such levels traditionally challenge enclosure air-tightness, box rigidity, but they can also provide useful acoustic energy for harvesting.
Engineering Brief 236

09:00

EB1-2 Mobile Platform Acoustical Noise Identification Using Internal and Reference Microphones—*Przemek Maziewski*, Intel Technology Poland, Gdansk, Poland

Proposed paper addresses the problem of microphone noise. The performance of built-in microphones in laptops and other mobile devices can suffer in the presence of noise. Identifying noisy components and separating the internal from external origins allows for the noise sources to be root caused and eliminated. This capability is crucial when developing new platforms. The proposed method employs a series of recordings, conducted using both built-in and reference microphones. The recordings are obtained under different operating conditions of the device, e.g., AC vs battery power. The recordings are then analyzed to identify different characteristics resulting from use of the internal versus the external microphone. Based on these results noise components can be separated and the potential noise source identified.
Engineering Brief 237

09:00

EB1-3 Automatically Generating VST Plugins from MATLAB Code—*Charles DeVane, Gabriele Bunkheila*, MathWorks, Natick, MA, USA

We describe the automatic generation of VST audio plugins from MATLAB code using the Audio System Toolbox from MathWorks. We provide MATLAB code for three complete example plugins, discuss problems that may be encountered, and describe a workflow to generate VST plugins as quickly and easily as possible.
Engineering Brief 238
[Presented by Gabriele Bunkheila]

09:00

EB1-4 Echo Thresholds for a 3-D Loudspeaker Configuration—*Lee Davis, Hyunkook Lee*, University of Huddersfield, Huddersfield, UK

Echo thresholds were examined with differing stimuli, lag sound directions, and decision criteria in a 3D loudspeaker reproduction environment. Two tests were undertaken to examine two different criteria: echo threshold with fusion and echo threshold with complete separation, each with three stimuli (orchestral, pink noise burst, and speech) and six lag sound directions in total. An adapted method of adjustment was used by subjects to control the delay between the lead and lag loudspeakers. Results showed that there were significant differences in echo threshold when the decision criteria differed. The orchestral stimulus was found to be significantly different from the pink noise burst and speech in both criteria. Few significant differences were noted between angles. In general, echo thresholds were higher with lag sources located in the median plane.
Engineering Brief 239

09:00

EB1-5 A New Response Method for Auditory Localization and Spread Tests—*Hyunkook Lee, Dale Johnson, Maksims Mironovs*, University of Huddersfield, Huddersfield, West Yorkshire, UK

This Engineering Brief presents a new response method developed for auditory localization and spread tests. The proposed method uses a flexible strip with a series of LEDs, which are powered by a microcontroller, for eliciting subjective responses. For the localization test, the position of an active LED is controlled and recorded in Max using a dial. For the spread test, multiple LEDs can be positioned on the strip to visually describe the lower and upper boundaries of the perceived image. The required system is easy to build and relatively inexpensive. Vertical stereophonic localization tests were conducted to compare between the LED method and a visual marker method. Results showed that the proposed method was more accurate, consistent, and time-efficient than the marker method.
Engineering Brief 240

09:00

EB1-6 Setting Up and Making an AES67 Network Coexist with Standard Network Traffic—*Mickaël Henry*,[1,2] *Lucas Rémond*,[3] *Nicolas Sturmel*[2]
[1]UVHC, Valenciennes, France
[2]Digigram S.A., Montbonnot, France
[3]CNSMDP, Paris, France

In this paper we will show how an AES67 network can coexist within a standard non-audio network. We will detail the difficulties usually encountered when setting up and using AES67 networks. We will analyze the utility of the network protocols required by AES67: (i) IGMP and its impact on devices features, (ii) PTP and the clock recovery performance when using PTP enabled switches, and (iii) QoS and the impact of non-audio traffic such as web and corporate traffic. We will use a set-up of 10 different AES67 compliant devices from many manufacturers and supporting various AoIP protocols all compliant to AES67. We will provide recommendations in order to provide proper quality of experience while making networks coexist.
Engineering Brief 241

09:00

EB1-7 Implementation of Faster than Real Time Audio Analysis for Use with Web Audio API: An FFT Case Study—*Luis Joglar-Ongay, Christopher Dewey, Jonathan Wakefield*, University of Huddersfield, Huddersfield, UK

There is significant interest in the audio community in developing web-based applications using HTML5 and Web Audio API. Whilst this newly emerging API goes some way to provide offline audio analysis in the web browser it is limited to a relatively basic FFT with fixed Blackman windowing and no overlap facility. Most previously documented solutions to this issue operate in real time. This paper demonstrates how to perform more sophisticated, faster than real time FFT analysis for use within Web Audio applications. It makes use of the Web Audio API and the dsp.js library. Academics and researchers can use this paper as a tutorial to develop similar solutions within their own web based audio applications.
Engineering Brief 242

09:00

EB1-8 Block-Sparse Fast Recursive Approximated Memory Improved Proportionate Affine Projection Algorithm—*Felix Albu*, Valahia University of Targoviste, Targoviste, Romania

A new approximated memory improved proportionate affine projection algorithm for block sparse echo cancellation is proposed. This contribution presents a fast recursive implementation combined with the use of dichotomous coordinate descent iterations. It is shown

that the proposed algorithm has superior convergence speed and tracking abilities for echo path changes in the context of acoustic and network echo cancellation applications. Also it is proved that these achievements are obtained while having a reduced numerical complexity than competing algorithms.
Engineering Brief 243

09:00

EB1-9 The Effect of Loop Length and Musical Material on Discrimination Between MP3 and WAV Files—*Denis Martin, Richard King*, McGill University, Montreal, Quebec, Canada; The Centre for Interdisciplinary Research in Music Media and Technology, Montreal, Quebec, Canada

Listening test results generally show that for bit-rates higher than 128 kbps listeners can rarely distinguish between MP3 and WAV files with any statistical significance while many audio professionals agree it is possible. This project attempts to explain some of the reasons why typical AB and ABX tests often fail by looking at the effects of loop length and music choice on listener success. An informal take-home AB listening test was used with varying musical material and music looping at different lengths. The results show that performance drops significantly with short loop lengths (<2sec, $p = .02$) and that the participants were able to discriminate between these two different file formats with great significance ($p < .001$).
Engineering Brief 244

09:00

EB1-10 Considerations When Calibrating Program Material Stimuli Using LUFS—*Malachy Ronan, Nicholas Ward, Robert Sazdov*, University of Limerick, Limerick, Ireland

While the LUFS standard was originally developed for broadcast applications, it offers a convenient means of calibrating program material stimuli to an equal loudness level, while remaining in a multichannel format. However, this calibration is based on an absolute sound pressure level of 60 dBA, the preferred listening level when watching television. Levels used in analytical listening and perceptual experiments tend to be significantly higher. This disparity may affect the accuracy of the Leq(RLB) weighting filter employed in LUFS meters. To address this issue, the development of the LUFS standard is examined to assess its suitability for the task. The findings suggest that a compromise between analytical listening and loudness matching in perceptual experiments requires careful consideration of experimental variables.
Engineering Brief 245

09:00

EB1-11 Selective Mixing Improves Reproduction Quaity with Portable Loudspeakers—*Piotr Kleczkowski, Tomasz Dziedzic*, AGH University of Science and Technology, Krakow, Poland

Selective mixing of sounds is an experimental method of mixing, first proposed in [1]. Further developments and listening experiments confirmed that inexperienced listeners more often than not prefer this type of processing over direct mixing, while it is the other way round with mixing engineers. It has been found lately, that besides the extent of the effect, there is another independent variable associated with this method—quality of the reproduction system. Experiments have shown, that the percentage of listeners choosing selective mixing versions is higher when the music is reproduced over small loudspeakers of portable devices, like notebook computers.
Engineering Brief 246
[eBrief will be presented by Bartlomiej Chojnacki]

Tutorial 1 **Saturday, June 4**
09:00 – 10:15 **Havane Amphitheatre**

MAIN MICROPHONE TECHNIQUES FOR 2.0 AND 5.1

Presenter: **Helmut Wittek**, SCHOEPS GmbH, Karlsruhe, Germany

This tutorial will explain step-by-step, using many practical examples, what a suitable stereophonic microphone array can look like. With 2.0 stereo setups as the starting point, multichannel setups will also be introduced. Many factors influence the choice of a stereophonic microphone setup, but the relevance of these factors can vary greatly depending on the application, such that there is never one single “correct” setup. Knowledge of various options gives a Tonmeister the ability to make optimal choices. In this session the free iPhone and browser App “Image Assistant” will be presented. It calculates the spatial characteristics of arbitrary stereophonic microphone arrays and auralizes the result. Moreover, the educational website www.hauptmikrofon.de is presented offering various comparative sound samples on the subject.
This session is presented in association with the AES Technical Committee on Spatial Audio

Session EB2 **Saturday, June 4**
10:30 – 12:00 **Foyer**

ENGINEERING BRIEFS—POSTERS

10:30

EB2-1 Investigation into the Perceptual Effects of Image Source Method Order—*Dale Johnson, Hyunkook Lee*, University of Huddersfield, Huddersfield, UK

This engineering brief explores the perceived effects and characteristics of impulse responses (IRs) generated using a custom, hybrid, geometric reverb algorithm. The algorithm makes use of a well known Image Source Method (ISM) and Ray Tracing methods. ISM is used to render the early reflections to a specified order while ray tracing renders the remaining reflections. IRs rendered at varying ISM orders appear to exhibit differences in perceptual characteristics, particularly in the early portion. To understand these characteristics, an elicitation test base was devised in order to acquire terms for the different characteristics. These terms were grouped in order to provide attributes for future grading tests.
Engineering Brief 247

10:30

EB2-2 The Influence of Discrete Arriving Reflections on Perceived Intelligibility and STI Measurements—*Ross Hammond,*[1] *Peter Mapp,*[2] *Adam J. Hill*[1]
[1]University of Derby, Derby, Derbyshire, UK
[2]Peter Mapp Associates, Colchester, Essex, UK

The most widely used objective intelligibility measurement method, the Speech Transmission Index (STI), does not completely match the highly complex auditory perception and human hearing system. Investigations were made into the impact of discrete reflections (with varying arrival times and amplitudes) on STI scores, subjective

intelligibility, and the subjective annoyance factor.' This allows the effect of comb filtering on the modulation transfer function matrix to be displayed, as well as demonstrates how the perceptual effects of a discrete delay cause subjective 'annoyance,' that is not necessarily mirrored by STI. This work provides evidence showing why STI should not be the sole verification method within public address and emergency announcement systems, where temporal properties also need thoughtful consideration.
Engineering Brief 248

10:30

EB2-3 Immersive Production Techniques in Cinematic Sound Design: Context and Spatialization—*Tom Downes, Malachy Ronan*, University of Limerick, Limerick, Ireland

Immersive formats are fast becoming a ubiquitous feature of film post-production workflow. However, little knowledge exists concerning production techniques addressing this increased spatial resolution. Questions therefore remain regarding their function in cinematic sound design. To address this issue, this paper evaluates the context required to prompt the use of elevated loudspeakers and examines the relevance of electroacoustic spatialization techniques to 3D cinematic formats. A contextually relevant scene from submarine classic Das Boot was selected to probe this question in a 9.1 loudspeaker configuration. It is hoped that this paper will prompt further discourse on the topic.
Engineering Brief 249

10:30

EB2-4 Perceptual Comparison of Localization with Soundman Binaural Microphones vs HRTF Post-Processing—*Blas Payri, Ramón Rodríguez Mariño*, Universitat Politècnica de València, Valencia, Spain

We realize a perceptual comparison of spatial sound localization using a synthetic pink noise and four recognizable sound sources: male and female speech, a closing door, and a sea-sound recording. Spatialization is created via binaural recording (Soundman OKM binaural microphones) or HRTF post-processing using filters available in Logic Pro, Protools, and Matlab. Eleven participants had to locate the source position of 72 stimuli combining different locations in azimuth. Results show that location recognition is generally low (36%). Although Soundman recordings show better results, no significant difference in localization accuracy is found between HRTF filtering systems and the binaural microphone recordings. We conclude that binaural 3D sound can easily be implemented with available commercial software, with no clear difference between systems.
Engineering Brief 250

10:30

EB2-5 VSV (Virtual Source Visualizer), A Practical Tool for 3D-Visualizing Acoustical Properties of Spatial Sounds—*Masataka Nakahara,[1] Akira Omoto,[1,2] Yasuhiko Nagatomo[3]*
[1]ONFUTURE Ltd., Tokyo, Japan
[2]Kyushu University, Fukuoka, Japan
[3]Evixer Inc., Tokyo, Japan

The authors have developed a practical tool that visualizes 3D acoustical properties of sound by using sound intensity information. The tool, VSV (Virtual Source Visualizer), consists of two main parts; analyzing software and measurement instruments. Since the goal is to provide a simple solution to 3D acoustic analysis, the authors have focused on the following items; obtain intuitively understandable results, and construct reliable system with inexpensive devices. In this paper usefulness and accuracy of our proposed method are discussed, and some examples of practical applications are also introduced.
Engineering Brief 251

10:30

EB2-6 Database of Binaural Room Impulse Responses of an Apartment-Like Environment—*Fiete Winter,[1] Hagen Wierstorf,[2] Ariel Podlubne,[3] Thomas Forgue,[3] Jérome Manhès,[3] Matthieu Herrb,[3] Sascha Spors,[1] Alexander Raake,[2] Patrick Danès[3]*
[1]Universtiat Rostock, Rostock, Germany
[2]Technische Universität Ilmenau, Ilmenau, Germany
[3]Université de Toulouse, Toulouse, France

We present a database of binaural room impulse responses (BRIRs) measured in an apartment-like environment. The BRIRs were captured at four different sound source positions, each combined with four listener positions. A head and torso simulator (HATS) with varying head-orientation in the range of ± 78 degrees with 2-degrees resolution was used. Additionally, BRIRs of 20 listener positions along a trajectory connecting two of the four positions were measured, each with a fixed head-orientation. The data is provided in the Spatially Oriented Format for Acoustics (SOFA) and it is freely available under the Creative Commons (CC-BY-4.0) license. It can be used to simulate complex acoustic scenes in order to study the process of auditory scene analysis for humans and machines.
Engineering Brief 252

10:30

EB2-7 Compatibility Study of Dolby Atmos Objects' Spacial Sound Localization Using a Visualization Method—*Takashi Mikami,[1] Masataka Nakahara,[1,2] Kazutaka Someya[3]*
[1]SONA Co., Tokyo, Japan
[2]ONFUTURE Ltd., Tokyo, Japan
[3]beBlue Co., Ltd., Tokyo, Japan

3D sound intensity measurement was carried out in two Dolby Atmos-compliant mixing rooms, and spatial sound localizations were visualized by using a newly developed visualizer, VSV (Virtual Source Visualizer), which locates sound directions on panoramic 4 pi space by using 3D sound intensity. Since in conventional channel-based sound design, sound localizations depend on loudspeakers' positions, there should be differences among mixing rooms. But in object-based sound design as is provided by Atmos, sound localization is rendered by RMU (Rendering and Mastering Unit) using metadata of azimuth and elevation angle and is expected not to depend on loudspeakers' positions. The session discusses inter-room compatibility / difference of sound direction using the visualization method between two mixing rooms, a small Home Atmos-compliant mixing room, and a Cinema Atmos-compliant large stage.
Engineering Brief 253

10:30

EB2-8 Controlling Program Loudness in Individualized Binaural Rendering of Multichannel Audio Contents—*Emmanuel Ponsot*,[1,2] *Hervé Dejardin*,[2] *Edwige Roncière*[2]
[1]STMS Lab (Ircam, CNRS, UPMC), Paris, France
[2]Radio France, Paris, France

For practical reasons, we often experience multichannel audio productions in a binaural context (e.g., headphones on mobile devices). To make listeners benefit from optimal binaural rendering ("BiLi project"), Radio France developed nouvOson (http://nouvoson.radiofrance.fr/), an online audio platform where listeners can select HRTFs and ITDs that best fit them. The goal of the present study was to control the program loudness (measured according to the ITU-R BS.1770 / R128 recommendations) after binauralization. To this end, we examined the influence of various parameters such as the audio content (synthetic vs. real broadcast audio), HRTFs, and ITDs on loudness. We propose a dynamic process, which adapts the gain in the binauralization chain so as to control the output loudness of virtual surround audio contents.
Engineering Brief 254

10:30

EB2-9 Presenting the S3A Object-Based Audio Drama Dataset—*James Woodcock*,[1] *Chris Pike*,[2] *Frank Melchior*,[2] *Philip Coleman*,[3] *Andreas Franck*,[4] *Adrian Hilton*[3]
[1]University of Salford, Salford, Greater Manchester, UK
[2]BBC Research and Development, Salford, Greater Manchester, UK
[3]University of Surrey, Guildford, Surrey, UK
[4]University of Southampton, Southampton, Hampshire, UK

This engineering brief reports on the production of three object-based audio drama scenes commissioned as part of the S3A project. 3D reproduction and an object-based workflow were considered and implemented from the initial script commissioning through to the final mix of the scenes. The scenes are being made available freely and without restriction as Broadcast Wave Format files containing all objects as separate tracks and all metadata necessary to render the scenes as an XML chunk in the header conforming to the Audio Definition Model specification (Recommendation ITU-R BS.2076 [1]). It is hoped that these scenes will find use in perceptual experiments and in the testing of 3D audio systems. The scenes are available via the following link: http://dx.doi.org/10.17866/rd.salford.3043921.
Engineering Brief 255

10:30

EB2-10 Installation of a Flexible 3D Audio Reproduction System into a Standardized Listening Room—*Russell Mason*, University of Surrey, Guildford, Surrey, UK

In order to undertake research into 3D audio reproduction systems, it was necessary to install a flexible loudspeaker rig into the ITU-R BS 1116 standard listening room at the University of Surrey. Using a mixture of aluminum truss and tube, a method for mounting loudspeakers in a manner that allows a wide range of layouts was created. As an example configuration, an installed 22.2 system is described. The method used to undertake bass management of this system, as well as methods to align the time of arrival, level, and frequency response of each channel are described. The resulting configuration is compared to the requirements of the ITU-R BS 1116 standard.
Engineering Brief 256

Workshop 1 **Saturday, June 4**
10:30 – 11:45 **Havane Amphitheatre**

MICROPHONES: WHAT CAN WE MEASURE AND WHAT DO WE HEAR?

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

Panelists: *Jürgen Breitlow*, Georg Neumann Berlin, Berlin, Germany; Sennheiser Electronic, Wedemark, Germany
David Josephson, Josephson Engineering, Inc., Santa Cruz, CA, USA
Martin Schneider, Georg Neumann GmbH, Berlin, Germany
Helmut Wittek, SCHOEPS GmbH, Karlsruhe, Germany

Comparison of microphones can be carried out by performing measurements and by listening. Both actions are equally important. The measured data make it easy to select the kind of microphones that suit your job. Measurements can take the microphone to its limits in all directions. Measurements help you to define the category of microphone you need for a job. However, listening can help you to define exactly the unit that you find sounds the best. In this workshop, the most important data are explained, and good practice regarding how to listen is presented.

This session is presented in association with the AES Technical Committee on Microphones and Applications

Workshop 2 **Saturday, June 4**
10:30 – 12:00 **Room 352A**

IMMERSIVE AUDIO FOR VIRTUAL REALITY

Chair: **Gavin Kearney**, University of York, York, UK

Panelists: *Jamieson Brettle*
Pedro Corvo, Sony Computer Entertainment
Marcin Gorzel, Google, Dublin, Ireland
Jelle Van Mourik, Sony Computer Entertainment

In recent years, major advances in gaming technologies, such as cost-effective head-tracking and immersive visual headsets have paved the way for commercially viable virtual reality to be delivered to the individual. Now the consumer finally has the opportunity to experience new gaming, cinematic and social media experiences with truly immersive and interactive 3-D audio and video content. For many sound designers, rendering a truly dynamic and spatially coherent mix for VR presents a new learning curve in soundtrack production. What spatial audio techniques should we be using to create engaging and interactive 3-D mixes? What audio workflows should we employ for similar immersive experiences on headphones, 5.1 loudspeakers and beyond? Are new VR production methods backwards compatible with existing game audio pipelines? Can binaural reproduction work for everyone? In this workshop our panel of experts will present practical workflows for mixing and rendering 3-D sound for VR. The workshop will explore different production techniques for creating immersive mixes such as Ambisonics processing and Head-Related Transfer Function rendering. It will also explore the importance of environmental rendering for VR as well as outlining workflow challenges and pipelines for dynamic spatial audio over a variety of VR technologies and applications.

Saturday, June 4 11:00 Room 362

Technical Committee Meeting on Hearing and Hearing Loss Prevention

Special Event
Awards Presentation and Keynote Address
Saturday, June 4, 12:00 – 13:30
Havane Amphitheatre

Opening Remarks:
- Executive Director Bob Moses
- President John Krivit
- Convention Chair Mike Williams

Program:
- AES Awards Presentation by Sean Olive, Awards Chair
- Introduction of Keynote Speaker by Convention Chair
- Keynote Address by Alex U. Case

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:

CITATION AWARD

- **Humberto Teran**

BOARD OF GOVERNORS AWARD

- **Dorte Hammershøi**
- **Bert Kraaijpoel**

FELLOWSHIP AWARD

- **Dave Fisher**
- **Ralph Kessler**
- **Brian C. J. Moore**
- **Rozenn Nicol**
- **Kazuho Ono**

Keynote Speaker

This year's Keynote Speaker is Alex U. Case. Case is Associate Professor of Sound Recording Technology at the University of Massachusetts Lowell. In addition to full time undergraduate and graduate teaching, Case has presented dozens of invited lectures and master classes on audio, acoustics, and education across the U.S. and internationally. His research and professional activities focus on the technical foundations, creative motivations, and aesthetic merits of recording and signal processing techniques used in multitrack production. Widely published, Case is an author, with over 100 articles in journals and industry trade publications, entries in the Grove Dictionary of American Music, and two titles with Focal Press. He is also President Elect of the AES. His keynote address is entitled "Intuition, Rebellion, Courage, and Chance – Historic Moments of Creative Signal Processing That Resonate to This Day."

Artists learn, in part, by imitating those who came before them, but they succeed by finding their own voice. Audio artistry is no different. Recording engineers are influenced by the recordings they admire, Keeping up with the advances in audio is a full time job. A glance at the rich schedule of events that fill the Technical Program here at the 140th AES International Convention confirms: we always have so much to learn, to discover, and to master. Technology and craft are always in flux, presenting audio engineers with opportunities for newer, faster, and better. Some advances are fleeting—doomed to fade—but others become industry touchstones. Learning from the ear-grabbing sonic inventions that came before us, we can be better prepared should audio fate drop-in on our recording session. Let your search for better recording and mixing techniques be informed by these key moments in sound recording history.

Session P3 Saturday, June 4
13:30 – 17:00 Room 353

INSTRUMENTATION AND MEASUREMENT

Chair: **Bert Kraaijpoel**, Dutch Film Academy (NFA), Amsterdam, Netherlands; Royal Conservatory, The Hague, Netherlands

13:30

P3-1 Characterization and Measurement of Wind Noise around Microphones—*Chris Woolf*, Broadcast Engineering Systemsn Cornwall, UK

Producing wind noise measurements for microphones that correlate well with practical use has always proved difficult. Characterizing the noise numerically, rather than spectrally, has proved even harder. This paper explores some novel approaches to both problems. The airflow patterns of a newly designed wind generator are mapped, and a simple method of producing turbulent flow from a laminar stream is demonstrated. In order to characterize the wind noise numerically a dual number approach is explored as a possibility. This takes the spectral curves for a protected and unprotected microphone in an airstream, and produces two numbers: one for the level of noise reduction, and a second one for the accuracy with which the two curves track, duly corrected for audibility.
Convention Paper 9495

14:00

P3-2 Rocking Modes (Part 2: Diagnostics)—*William Cardenas, Wolfgang Klippel*, Klippel GmbH, Dresden, Germany

The rocking behavior of the diaphragm is a severe problem in headphones, micro-speakers, and other kinds of loudspeakers causing voice coil rubbing that limits the maximum acoustical output at low frequencies. The root causes of this problem are small imbalances in the distribution of the stiffness, mass, and force factor in the gap. Based on lumped parameter modeling, modal decomposition and signal flow charts presented in a previous paper (Part 1) this paper focuses on the practical measurement using laser vibrometry, parameter identification, and root cause analysis. New characteristics are presented that simplify the interpretation of the identified parameters. The new technique has been validated by numerical simulations and systematic modifications of a real transducer. The diagnostic value of the new measurement technique has been illustrated on a transducer used in headphones.
Convention Paper 9496

14:30

P3-3 Harmonic Distortion Measurement for Nonlinear System Identification—*John Vanderkooy*,[1] *Sean Thomson*[2]
[1]University of Waterloo, Waterloo, ON, Canada

[2]B&W Group Ltd., Steyning, West Sussex, UK

In order to model nonlinearities in loudspeakers, accurate measurement of harmonic distortion is necessary with particular attention to the relative phases of fundamental and harmonics. This paper outlines several ways that logarithmic sweeps can be used to achieve this goal. It is shown that Novak's redesign of the logsweep is not strictly necessary, if proper account is taken of the phase relationships of the various harmonics. We study several other types of sweeps and methods to extract precise harmonic amplitudes and phases, using tracking filter concepts. The paper also deals with measurement systems that may have fractional-sample delays between excitation, reference, and data channels. Such details are important for accurate phase characterization of transfer functions. An intermodulation example is given for which sweeps with a single instantaneous frequency are inadequate.
[Also a poster—see session P8-5]
Convention Paper 9497

15:00

P3-4 Evaluation of a Fast HRTF Measurement System— *Jan-Gerrit Richter, Gottfried Behler, Janina Fels*, RWTH Aachen University, Aachen, Germany

This paper describes and evaluates a measurement setup for individual Head-Related Transfer Functions (HRTFs) in high spatial resolution in a short time period. The setup is constructed to have as little impact on the measurement as possible. It consists of a circular arc segment of approximately 160 degrees on which a large number of broadband loudspeakers are placed forming one continuous surface. By rotating the subject or the arc horizontally, HRTFs are acquired along a spherical surface. To evaluate the influence of the measurement setup a solid sphere and an artificial head are measured and are compared with both the presented system, simulation data using Boundary Element Method, and a traditional, well evaluated HRTF measurement system with only one loudspeaker.
Convention Paper 9498

15:30

P3-5 Efficiency of Switch-Mode Power Audio Amplifiers —Test Signals and Measurement Techniques—*Niels Elkjær Iversen, Arnold Knott, Michael A. E. Andersen*, Technical University of Denmark, Kgs. Lyngby, Denmark

Switch-mode technology is greatly used for audio amplification. This is mainly due to the great efficiency this technology offers. Normally the efficiency of a switch-mode audio amplifier is measured using a sine wave input. However this paper shows that sine waves represent real audio very poorly. An alternative signal is proposed for test purposes. The efficiency of a switch-mode power audio amplifier is modelled and measured with both sine wave and the proposed test signal as inputs. The results show that the choice of switching devices with low on resistances are unfairly favored when measuring the efficiency with sine waves. A 10% efficiency improvement was found for low power outputs. It is therefore of great importance to use proper test signals when measuring the efficiency.
Convention Paper 9499

16:00

P3-6 ITU-R BS.1770 Based Loudness for Immersive Audio —*Scott Norcross, Sachin Nanda, Zack Cohen*, Dolby Laboratories, San Francisco, CA, USA

With the adoption of ITU-R BS.1770 and the creation of numerous loudness recommendations, measuring and controlling the loudness of audio for broadcast is now a standard practice for legacy (5.1 and stereo) content. With new immersive and personalized audio content, the measurement and controlling of loudness is still in its infancy. While ITU-R BS.1770 has recently been revised to support an arbitrary number of audio channels. However dynamic object-based audio measurement is not explicitly covered in this revision, though the revision can be used to measured the rendered object-based audio. This paper summarizes the results of subjective loudness matching tests that were conducted using rendered dynamic object-based audio to verify the revision of ITU-R BS.1770.
Convention Paper 9500

16:30

P3-7 Metrics for Constant Directivity—*Rahulram Sridhar, Joseph G. Tylka, Edgar Choueiri*, Princeton University, Princeton, NJ, USA

It is often desired that a transducer have a polar radiation pattern that is invariant with frequency, but there is currently no way of quantifying the extent to which a transducer possesses this quality (often called "constant directivity" or "controlled directivity"). To address the problem, commonly-accepted criteria are used to propose two definitions of constant directivity. The first, stricter definition, is that the polar radiation pattern of a transducer should be invariant over a specified frequency range, whereas the second definition is that the directivity factor (i.e., the ratio between the on-axis power spectrum and the average power spectrum over all directions), or index when expressed in dB, should be invariant with frequency. Furthermore, to quantify each criterion, five metrics are derived: (1) Fourier analysis of contour lines (i.e., lines of constant sensitivity over frequency and angle), (2) directional average of frequency response distortions, (3) distortion thresholding of polar responses, (4) standard deviation of directivity index, and (5) cross-correlation of polar responses. Measured polar radiation data for four loudspeakers are used to compute all five metrics that are then evaluated based on their ability to quantify constant directivity. Results show that all five metrics are able to quantify constant directivity according to the criterion on which each is based, while only two of them, metrics 4 and 5, are able to adequately quantify both proposed definitions of constant directivity.
[Also a poster—see session P8-6]
Convention Paper 9501

Session P4 Saturday, June 4
13:30 – 16:30 Room 352B

ROOM ACOUSTICS

Chair: **Ben Kok**, BEN KOK - acoustic consulting, Uden, The Netherlands

13:30

P4-1 Small-Rooms Dedicated to Music: From Room Response Analysis to Acoustic Design—*Lorenzo Rizzi*,[1] *Gabriele Ghelfi*,[1] *Maurizio Santini*[2]
[1]Suono e Vita - Acoustic Engineering, Lecco, Italy
[2]Università degli Studi di Bergamo, Bergamo, Italy

Reviewing elements of on-field professional experience gained by the authors in the analysis of small-rooms dedi-

cated to music, case studies offered by the everyday working practice allow to deal with specific situations, these are seldom described by usual theoretical models and literature. Using the analysis procedure developed and refined by authors, it is possible to investigate the characteristics of the acoustic response of the small-rooms with more detail. In this paper case studies of particular interest will be described: different small-room phenomena will be shown in the reported measurements.
[Also a poster—see session P8-2]
Convention Paper 9502

14:00

P4-2 Direction of Late Reverberation and Envelopment in Two Reproduced Berlin Concert Halls—*Winfried Lachenmayr,*[1] *Aki Haapaniemi,*[2] *Tapio Lokki*[2]
[1]Mueller-BBM, Munich, Germany
[2]Aalto University School of Science, Aalto, Finland

Most studies on the influence of the direction of late reverberation on listener envelopment (LEV) in concert halls have been conducted in laboratory conditions, i.e., where synthetic sound fields and a relatively limited number of loudspeakers were used to approximate a real, spatially quite complex acoustic situation. This study approaches LEV from the real acoustics. The late part of the sound field of two measured concert halls Berlin Konzerthaus and Berlin Philharmonie, auralized with a state-of-the-art reproduction method, is altered virtually regarding its' direction. Results suggest that the figure-of-eight weighting applied in late lateral level LJ for predicting envelopment is underestimating the importance of reverberation from directions such as ceiling and rear.
Convention Paper 9503

14:30

P4-3 Electronic Shell—Improvement of Room Acoustics without Orchestra Shell Utilizing Active Field Control—*Takayuki Watanabe, Hideo Miyazaki, Masahiro Ikeda,*
Yamaha Corporation, Hamamatsu, Shizuoka, Japan

This paper introduces an example of Electronic Shell acoustic enhancement system that was installed in a multi-purpose hall without an orchestra shell. The system is based on the concept of Active Field Control using electroacoustic means. The three objectives of this system were (1) the enhancement of early reflection for performers, (2) the increase of the reverberation time and the total sound energy on stage, and (3) the enhancement of early reflection in the audience area. The application of this system showed an improvement of about 1 to 2 dB in STearly and more than 2 dB in G in the audience area, which is equivalent or better performance than simple mobile typed orchestra shell.
[Also a poster—see session P8-3]
Convention Paper 9504

15:00

P4-4 Experimental Assessment of Low-Frequency Electroacoustic Absorbers for Modal Equalization in Actual Listening Rooms—*Etienne Rivet,*[1] *Sami Karkar,*[1] *Hervé Lissek,*[1] *Torje Nikolai Thorsen*[2] *Véronique Adam*[2]
[1]Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[2]Goldmund International, Monaco, Monaco

In listening rooms, low-frequency modal resonances lead to uneven distributions in space and frequency of the acoustic energy, as well as an alteration of the temporal behavior of the original music content. While usual absorption techniques have severe limitations for reducing the negative impact of room modes, the authors have previously proposed the use of electroacoustic absorbers for room modal equalization. This device consists of a current-driven, closed-box loudspeaker associated to a hybrid sensor-/shunt-based impedance control. In this communication we assess the performance of these electroacoustic absorbers in actual listening rooms, by measuring frequency responses at different locations, as well as their modal decay times. The electroacoustic absorbers perform as expected and the room modal equalization is clearly improved in the low-frequency range.
Convention Paper 9505

15:30

P4-5 Modeling Non-Shoebox Shaped Rooms with the Mode Matching Method—*Bjørn Kolbrek, U. Peter Svensson,*
Norwegian University of Science and Technology, Trondheim, Norway

When a room is not shoebox shaped, usually no analytical expressions exist for the determination of resonance frequencies and mode shapes. One option is to employ the Finite Element Method (FEM). In this paper an alternative method, the Mode Matching Method (MMM), is used to compute the transfer function and sound field of a non-shoebox shaped room with rigid walls and is compared to an FEM solution. The two methods show excellent agreement.
[Also a poster—see session P8-7]
Convention Paper 9506

16:00

P4-6 Room Acoustic Measurements Using a High SPL Dodecahedron—*Dario D'Orazio,*[1] *Simona De Cesaris,*[1] *Paolo Guidorzi,*[1] *Luca Barbaresi,*[1] *Massimo Garai,*[1] *Roberto Magalotti*[2]
[1]University of Bologna, Bologna, Italy
[2]B&C Speakers S.p.A., Bagno a Ripoli (FI), Italy

In this paper a dodecahedron with high powered loudspeakers is presented. The source is designed to allow high SPL with very low distortion. By comparing the prototype with a reference sound source, the high SPL dodecahedron show a flat frequency response over the 80 ÷ 5000 Hz one third octave bands, enough to meet all the ISO 3382 criteria. Laboratory measurements have been performed to test the performances and the robustness of the dodecahedron using different techniques at different sound pressure levels and background noises. The prototype allows a good signal-to-noise ratio of the impulse response also when 75 dB of stationary noise is added during the measurements.
Convention Paper 9507

Session P5 Saturday, June 4
13:30 – 15:30 Foyer

POSTERS: AUDIO EQUIPMENT, AUDIO FORMATS, AND AUDIO SIGNAL PROCESSING PART 1

13:30

P5-1 A Comparison of Optimization Methods for Compression Driver Design—*Michele Gasparini,*[1] *Emiliano Capucci,*[2] *Stefania Cecchi,*[1] *Romolo Toppi,*[2] *Francesco Piazza*[1]
[1]Universitá Politecnica della Marche, Ancona, Italy

[2]Faital S.P.A., Milan, Italy

Finite element analysis is a powerful and widespread mathematical technique capable of modeling even very complex physical systems. The use of this method is quite common in loudspeaker design processes, although simulations may often become time consuming. In order to reduce the number of simulations needed to define an optimal design, some advanced metaheuristic algorithms can be employed. The use of these techniques is well known in many optimization tasks when an analytical description of the system is not available a priori. In this paper a comparison among three different optimization procedures in the design of a compression driver is presented. The algorithms will be evaluated in terms of both convergence time and residual error.
Convention Paper 9508

13:30

P5-2 **Physically-Based Large-Signal Modeling for Miniature Type Dual Triode Tube—***Shiori Oshimo, Kanako Takemoto, Toshihiko Hamasaki*, Hiroshima Institute of Technology, Hiroshima, Japan

A precise SPICE model for miniature (MT) triode tubes of high-µ 12AX7 and medium-µ 12AU7 is proposed, based on the physical analysis of the measurement results. Comparing the characteristics between these tubes, the grid current at lower plate voltage and positive grid bias condition is modeled successfully with novel perveance parameters for the first time, though it was known that the perveance depends on both grid and plate bias. It is shown that the modulation factor of the space charge for the MT triodes is different from the other classic tubes. The model is implemented in LTspice to result in a good replication for a variation of three-order magnitude of grid current and cathode current.
[Also a lecture—see session P1-2]
Convention Paper 9485

13:30

P5-3 **Delay-Reduced Mode of MPEG-4 Enhanced Low Delay AAC (AAC-ELD)—***Markus Schnell, Wolfgang Jaegers, Pablo Delgado, Conrad Benndorf, Tobias Albert, Manfred Lutzky*, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

The MPEG-4 AAC Enhanced Low Delay (AAC-ELD) coder is well established in high quality communication applications, such as Apple's FaceTime, as well as in professional live broadcasting. Both applications require high interactivity, which typically demands an algorithmic codec delay between 15 ms and 35 ms. Recently, MPEG finalized a new delay-reduced mode for AAC-ELD featuring only a fraction of the regular algorithmic delay. This mode operates virtually at higher sampling rates while maintaining standard sampling rates for I/O. Supporting this feature, AAC-ELD can address even more delay critical applications, like wireless microphones or headsets for TV. In this paper main details of the delay-reduced mode of AAC-ELD are presented and application scenarios are outlined. Audio quality aspects are discussed and compared against other codecs with a delay below 10 ms.
[Also a lecture—see session P1-5]
Convention Paper 9488

13:30

P5-4 **Advances to a Frequency-Domain Parametric Coder of Wideband Speech—***Aníbal Ferreira,*[1] *Deepen Sinha*[2]
[1]University of Porto, Porto, Portugal
[2]ATC Labs, Newark, NJ, USA

In recent years, tools in perceptual coding of high-quality audio have been tailored to capture highly detailed information regarding signal components so that they gained an intrinsic ability to represent audio parametrically. In a recent paper we described a first validation model to such an approach applied to parametric coding of wideband speech. In this paper we describe specific advances to such an approach that improve coding efficiency and signal quality. A special focus is devoted to the fact that transmission to the decoder of any phase information is avoided, and that direct synthesis in the time-domain of the periodic content of speech is allowed in order to cope with fast F0 changes. A few examples of signal coding and transformation illustrate the impact of those improvements.
Convention Paper 9509

13:30

P5-5 **Visual Information Search in Digital Audio Workstations—***Joshua Mycroft, Tony Stockman, Joshua D. Reiss*, Queen Mary University of London, London, UK

As the amount of visual information within Digital Audio Workstations increases, the interface potentially becomes more cluttered and time consuming to navigate. The increased graphical information may tax available display space requirements and potentially overload visual perceptual and attentional bandwidth. This study investigates the extent to which Dynamic Query filters (sliders, buttons, and other filters) can be used in audio mixing interfaces to improve both visual search times and concurrent critical listening tasks (identifying subtle attenuation of named instruments in a multichannel mix). The results of the study suggest that the inclusion of Dynamic Query filters results in a higher amount of correctly completed visual and aural tasks.
Convention Paper 9510

13:30

P5-6 **AC-4 – The Next Generation Audio Codec—***Kristofer Kjörling,*[1] *Jonas Rödén*[1] *Martin Wolters,*[2] *Jeff Riedmiller,*[3] *Arijit Biswas,*[2] *Per Ekstrand,*[1] *Alexander Gröschel,*[2] *Per Hedelin,*[1] *Toni Hirvonen,*[1] *Holger Hörich,*[2] *Janusz Klejsa,*[1] *Jeroen Koppens,*[1] *K. Krauss,*[2] *Heidi-Maria Lehtonen,*[1] *Karsten Linzmeier,*[2] *Hannes Muesch,*[3] *Harald Mundt,*[2] *Scott Norcross,*[3] *J. Popp,*[2] *Heiko Purnhagen,*[1] *Jonas Samuelsson,*[1] *Michael Schug,*[2] *L. Sehlström,*[1] *R. Thesing,*[2] *Lars Villemoes,*[1] *Mark Vinton*[3]
[1]Dolby Sweden AB, Stockholm, Sweden
[2]Dolby Germany GmbH, Nuremberg, Germany
[3]Dolby Laboratories, San Francisco, CA USA

AC-4 is a state of the art audio codec standardized in ETSI (TS103 190 and TS103 190-2) and the TS103 190 is part of the DVB toolbox (TS101 154). AC-4 is an audio codec designed to address the current and future needs of video and audio entertainment services including broadcast and Internet streaming. As such, it incorporates a number of features beyond the traditional audio coding algorithms, such as capabilities to support immersive and personalized audio, support for advanced loudness management, video-frame synchronous coding, dialogue enhancement, etc. This paper will outline the thinking behind the design of the AC-4 codec, explain the different coding tools used, the systemic features included, and give an overview of performance and applications.
[Also a lecture—see session P2-3]
Convention Paper 9491

13:30

P5-7 **Single-Channel Audio Source Separation Using Deep Neural Network Ensembles**—*Emad M. Grais, Gerard Roma, Andrew J. R. Simpson, Mark D. Plumbley*, University of Surrey, Guildford, Surrey, UK

Deep neural networks (DNNs) are often used to tackle the single channel source separation (SCSS) problem by predicting time-frequency masks. The predicted masks are then used to separate the sources from the mixed signal. Different types of masks produce separated sources with different levels of distortion and interference. Some types of masks produce separated sources with low distortion, while other masks produce low interference between the separated sources. In this paper a combination of different DNNs' predictions (masks) is used for SCSS to achieve better quality of the separated sources than using each DNN individually. We train four different DNNs by minimizing four different cost functions to predict four different masks. The first and second DNNs are trained to approximate reference binary and soft masks. The third DNN is trained to predict a mask from the reference sources directly. The last DNN is trained similarly to the third DNN but with an additional discriminative constraint to maximize the differences between the estimated sources. Our experimental results show that combining the predictions of different DNNs achieves separated sources with better quality than using each DNN individually.
[Also a lecture—see session P2-6]
Convention Paper 9494

13:30

P5-8 **Using Phase Information to Improve the Reconstruction Accuracy in Sinusoidal Modeling**—
Clara Hollomey,[1] David Moore,[1] Don Knox,[1] W. Owen Brimijoin,[2] William Whitmer[2]
[1]Glasgow Caledonian University, Glasgow, Scotland, UK
[2]MRC/CSO Institute of Hearing Research, Glasgow, Scotland, UK

Sinusoidal modeling is one of the most common techniques for general purpose audio synthesis and analysis. Owing to the ever increasing amount of available computational resources, nowadays practically all types of sounds can be constructed up to a certain degree of perceptual accuracy. However, the method is computationally expensive and can for some cases, particularly for transient signals, still exceed the available computational resources. In this work methods derived from the realm of machine learning are exploited to provide a simple and efficient means to estimate the achievable reconstruction quality. The peculiarities of common classes of musical instruments are discussed and finally, the existing metrics are extended by information on the signal's phase propagation to allow for more accurate estimations.
[Also a lecture—see session P2-4]
Convention Paper 9492

13:30

P5-9 **Just Noticeable Difference of Interaural Level Difference to Frequency and Interaural Level Difference**
—*Heng Wang,[1] Cong Zhang,[1] Yafei Wu[2]*
[1]Wuhan Polytechnic University, Wuhan, Hubei, China
[2]Wuhan University, Wuhan, Hubei, China

In order to explore the perceptual mechanism of Interaural Level Difference (ILD) and research the relationship of ILD limen to frequency and ILD, this article selected eight values of ILD according to the qualitative analysis of ILD sensitivity by human ear. It was divided into 24 frequency bands as critical band and selected the center frequency of each band to test. This experiment adopted the traditional test methods (1 up/2 down and 2AFC). The results showed that: the thresholds of ILD are more significant with frequency, they are smaller at 500 Hz and 4000 Hz, a maximum value especially when it reaches about 1000 Hz; the thresholds increase as the reference values of ILD increase. This work will provide basic data for comprehensive exploring perceptual characteristics of the human ear and theoretical support for audio efficient compression.
Convention Paper 9511

Tutorial 2 **Saturday, June 4**
13:30 – 15:00 **Room 352A**

BINAURAL APPLICATIONS WITH AURO-3D IMMERSIVE SOUND

Presenters: **Wilfried Van Baelen**, Auro Technologies N.V., Mol, Belgium
Bert Van Daele, Auro Technologies N.V, Mol, Belgium

Auro-3D, and Immersive Sound in general, bring a new exciting listening experience to the audience on various reproduction systems. In many applications the audio will be reproduced using multiple speakers, but for many people headphones are actually the main listening device. Binaural processing allows to playback immersive sound on headphones by applying the necessary filters to recreate the directional cues related to the various speaker and/or object positions in the 3D auditory space. In this tutorial, a short explanation will be given about Auro Technologies' own technology to reproduce Auro-3D content over standard headphones. In a second part of this session, Auro's unique upmixing technology for headphones will be introduced, which provides a three-dimensional experience on headphones from standard stereo and surround audio tracks.

Workshop 3 **Saturday, June 4**
13:30 – 15:00 **Room 351**

PERCEPTUAL EVALUATION OF HIGH RESOLUTION AUDIO

Chair: **Joshua D. Reiss**, Queen Mary University of London, London, UK

Panelists: *George Massenburg*, Schulich School of Music, McGill University, Montreal, Quebec, Canada; Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Montreal, Quebec, Canada
Bob Schulein, RBS Consultants / ImmersAV Technology, Schaumburg, IL, USA
Thomas Sporer, Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany
J. Robert Stuart, Meridian Audio Ltd., Huntingdon, UK

This workshop focuses on past measurements and future potential in perceptual evaluation of high resolution audio. Past attempts to assess the audibility of higher resolutions (beyond 44.1 kHz, 16-bit) will be summarized with an overview of results, but the focus of the workshop is on testing and methodology itself. Discussion will include the problems and pitfalls of listening tests and demos and how they might be overcome. We shed some light on the psychoacoustic justifications behind the results of previous experiments including what is known and what is not. We discuss issues in evaluating quality and perception, the structuring of tests, configuration of the testing environment, and analysis of results. Attention will be paid to the choice of test material, the choice of test methodology, and the training of participants. We intend to engage the audience with lively discussion.

This session is presented in association with the AES Technical Committee on Hi-Resolution Audio

Saturday, June 4 13:30 Room 362

Technical Committee Meeting on Semantic Audio Analysis

Student Event and Career Development
Saturday, June 4, 13:45 – 15:30
Havane Amphitheatre

OPENING AND STUDENT DELEGATE ASSEMBLY MEETING – PART 1

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 and International Regions Vice Chair, announce the finalists in the Student Recording Competition categories and the Student Design Competition, and announce all upcoming student/education related 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. The SDA leaders will then lead a dialog to discuss important issues significant to all audio students. All students and educators are invited to participate in this meeting. Election results and Recording Competition and Design Competition Awards will be given at the Student Delegate Assembly Meeting–Part 2 on Tuesday, June 7.

Saturday, June 4 14:30 Room 362

Technical Committee Meeting on Perception and Subjective Evaluation of Audio Signals

Special Event
Saturday, June 4, 15:15–16:15
Room 352A

AUDIO PROJECTIONS 1—BINAURAL AUDIO FROM AURO-3D

Presenters: **Wilfried Van Baelen**, Auro Technologies N.V., Mol, Belgium
Bert Van Daele, Auro Technologies N.V., Mol, Belgium

Headphone Demos

Headphones are more and more becoming the main reproduction device for all kinds of content, including immersive sound through the use of binaural processing. In this session, Auro-3D content will be demonstrated using the Auro-Headphones real-time binaural processing, as used in mobile devices and consumer devices. The demonstration will not only include the reproduction of 'native' (recorded and mixed in) Auro-3D, but also the unique upmixing capabilities of Auro-Matic for Headphones, delivering an immersive listening experience from regular stereo and surround sources.

Special Event
Saturday, June 4, 15:15–16:15
Room 351

AUDIO PROJECTIONS 2—3D AUDIO PROJECTIONS

Presenter: **Daniel Shores**, Sono Luminus, Boyce, VA, USA; Shenandoah Conservatory Music Production and Recording Technology, Winchester, VA, USA

Crossing over Into Immersive Audio

Since its inception in 1996, Sono Luminus has operated with the singular focus of providing the home consumer with a unique listening experience, which we describe as Performance Fidelity. In 2006, Sono Luminus took the leap from studio to record label and have since created albums focusing in classical and acoustic music that have resulted in eighteen GRAMMY nominations, including nominations in the categories of Surround Sound and Best Engineered classical, as well as one Latin GRAMMY award winning album. In 2011, Sono Luminus began releasing it's recordings using the Pure Audio Blu-ray. Sono Luminus felt this would be the perfect delivery method for their surround recordings to the largest consumer base. In 2015, Sono Luminus began its experimentation and implementation of 9.1 Auro-3D recording. To date, Sono Luminus has created numerous 9.1 recordings, five of which are now commercially available with more on the way. Immersive audio quite literally takes the listening experience of the home consumer to the next level. It allows Sono Luminus the opportunity to deliver an even more in-depth, intriguing, and unique listening experience. Recording both on location, and in the 100 year old converted church in the Virginia countryside that is now the home of Sono Luminus Studios, Sono Luminus focuses on techniques for capturing native immersive audio, rather than mixing for the format. In the end though, it is all about the serving the music, and we have taken the opportunity to work with the musicians and composers to develop recordings that bring out the all of the musical nuances in a way not possible before. Examples in this projection will demonstrate various styles of music including choral, early music, Celtic, percussion, electronics, experimental music just to name a few demonstrating vastly different instrumentation and sonic textures.

Samples will include selections from:

- Peter Gregson - From the Album "Touch"
- Skylark Vocal Ensemble - From the Album "Crossing Over"
- Ensemble Galilei - From the Album "From Whence we Came"
- International Contemporary Ensemble - From the Album " In the Light of Air"
- Inscape - From the Album "Petrushka"
- Ayreheart - From the Album "Barley Moon"
- Los Angeles Percussion Quartet – From their upcoming untitled album.

Special Event
Saturday, June 4, 15:45–17:15
Havane Amphitheatre

PODCASTING—TELLING YOUR STORY WITH SOUND

Presenter: **Jim Anderson**, New York University, New York, NY, USA

A podcast can be more than a monologue or an interview; it can be a rich environment for using sound to tell your story. One can use sound in many ways, whether it's to set the scene, illustrate a concept, or enliven a journalistic endeavor. The talk will take the audience on an international aural journey from the "hollars" of Kentucky to the streets of Grenada in search of sounds. With Jim

Anderson's deep background in broadcasting, he will demonstrate the power of sound to illustrate and enrich a podcast.

Special Event
Saturday, June 4, 16:15–17:15
Room 352A

AUDIO PROJECTIONS 3—BINAURAL AUDIO PROJECTIONS FROM THE BBC

Presenter: **Tom Parnell**, BBC Research and Development, Salford, Greater Manchester, UK

BBC Research & Development has been working with program makers in recent years to create binaural productions. State-of-the-art techniques in recording, object-based mixing, and processing of audio have been applied to create spatial listening experiences for BBC audiences listening on headphones. This listening session will present clips of binaural program material produced by the BBC.

Program:

• Unearthed - An interactive natural history story with binaural sound, originally presented on the BBC Taster website.

• Ring - A binaural horror play, part of Radio 4's Fright Night on Halloween 2015.

• The Stone Tape - A binaural ghost story, part of Radio 4's Fright Night on Halloween 2015.

• The Turning Forest - A 3D sound fairy tale, shown as a virtual reality piece with dynamic binaural sound at TriBeCa Film Festival 2016.

Special Event
Saturday, June 4, 16:15–17:15
Room 351

AUDIO PROJECTIONS 4—3D AUDIO PROJECTIONS FROM 2L

Presenter: **Morten Lindberg**, 2L (Lindberg Lyd AS), Oslo, Norway; Westerdals, Oslo School of Arts, Communication and Technology

The music captured by 2L features Norwegian composers and performers and an international repertoire reflected in the Nordic atmosphere. The surround sound recordings of Lindberg Lyd not only transform the entire listening experience, but also - more radically - these innovative recordings overturn some very basic concepts regarding how music is played and even composed. 2L emphasize surround sound with Pure Audio Blu-ray and HiRes file distribution, and have garnered no less than 23 American GRAMMY nominations since 2006. Sixteen of these in categories Best Engineered Album, Best Surround Sound Album and Producer of the Year 2L record in spacious acoustic venues: large concert halls, churches and cathedrals. This is actually where we can make the most intimate recordings. The qualities we seek in large rooms are not necessarily a big reverb, but openness due to the absence of close reflecting walls. Making an ambient and beautiful recording is the way of least resistance. Searching the fine edge between direct contact and openness - that's the real challenge! A really good recording should be able to bodily move the listener. This core quality of audio production is made by choosing the right venue for the repertoire, and by balancing the image in the placement of microphones and musicians relative to each other in that venue.

There is no method available today to reproduce the exact perception of attending a live performance. That leaves us with the art of illusion when it comes to recording music. As recording engineers and producers we need to do exactly the same as any good musician: interpret the music and the composer's intentions and adapt to the media where we perform. Immersive audio is a completely new conception of the musical experience. Recorded music is no longer a matter of a fixed one-dimensional setting, but rather a three-dimensional enveloping situation. Stereo can be described as a flat canvas, while immersive audio is a sculpture that you can literally move around and relate to spatially; surrounded by music you can move about in the aural space and choose angles, vantage points and positions.

Special Event
Saturday, June 4, 17:30 – 19:00
Havane Amphitheatre

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

Lecturer: **Rozenn Nicol**, Orange Labs, France Telecom, Lannion, France

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 140th AES Convention is Rozenn Nicol, a research engineer on spatial audio at Orange Labs in France. She has worked on the development of innovative 3D sound technologies such as binaural, WFS and ambisonics, to enhance future telecommunication services. The title of her lecture is "The 3D Audio Revolution from Labs to Mass Market."

Over the past 25 years, there have been many major evolutions in spatial audio. After more than 50 years of stereophony, new technologies, such as Wave Field Synthesis, Higher Order Ambisonics or Vector Base Amplitude Panning, were introduced. They promise an enhanced 3D audio experience, where virtual sound sources can be accurately synthesized in any direction. Various formats of multichannel audio are now proposed: from 5.1, to 7.1, 10.2, Auro-3D (9.1, 11.1 or 13.1), 22.2, and Dolby Atmos (5.1.2, 5.1.4, 7.1.2, 7.1.4, 9.1.2). Not only the number of channels increases for a better sound immersion, but also sound spatialization is extended to elevation effects. In parallel, more and more tools are available for the capture, the editing, the coding and the reproduction of spatial audio. However, since these evolutions require more and more complex setups of loudspeakers, spatial audio is faced with the risk of being limited to movie theaters or amusement parks. Fortunately, a new step was recently reached with the binaural adaptation of any multichannel audio to headphone listening. Pioneer experiments from radio or television (BBC, Radio France, France Télévisions) show that spatial is very close to becoming a mass market product.

Saturday, June 4 17:30 Room 362

Technical Committee Meeting on Audio Forensics

Student Event and Career Development
Saturday, June 4, 20:00 – 24:00
Abbey Road Institute Paris
51 rue Merlin de Thionville
92150 Suresnes

AES STUDENT PARTY

The AES Student Party is open to any 140th Convention participant with an ALL ACCESS STUDENT BADGE. A great opportunity to meet fellow students from around the world. Check the SDA website/ blog for full details. It will be hosted by Abbey Road Institute Paris and AES thanks them for their participation and support.

Session P6 **Sunday, June 5**
09:00 – 12:00 **Room 353**

PERCEPTION—PART 1

Chair: **Dan Mapes-Riordan**, Etymotic Research, Elk Grove Village, IL, USA; DMR Consulting, Evanston, IL, USA

09:00

P6-1 Perception of Low Frequency Transient Acoustic Phenomena in Small Rooms for Music—*Lorenzo Rizzi*,[1] *Federico Ascari*,[2] *Gabriele Ghelfi*,[1] **Michele Ferroni**[2]
[1]Suono e Vita - Acoustic Engineering, Lecco, Italy
[2]Politecnico di Milano, Milan, Italy

Reducing the gap between analysis of low-frequency behavior of small rooms and actual perception, we introduce the importance of transient energetic phenomena besides classic FFT steady state analysis. After a frequency and temporal domain analysis of real-world impulse responses of critical listening rooms, headphone tests were performed. Results show that, for short musical sounds, a new curve called "Overshoot Response" can be more useful than classic frequency response regarding the level perception. Furthermore, the perceived loss of definition after the convolution with R.I.R. is correlated with decaying time and two metrics that were defined—"Room Slowness" and "Room Inertia."
Convention Paper 9512

09:30

P6-2 The Reduction of Vertical Interchannel Crosstalk: The Analysis of Localization Thresholds for Musical Sources —*Rory Wallis, Hyunkook Lee*, University of Huddersfield, Huddersfield, UK

Musical sources were presented to subjects as phantom images from vertically arranged stereophonic loudspeakers. Loudspeakers were arranged in two layers: "main" and "height." Subjects reduced the amplitude of the height layer until the resultant phantom image matched the position of the same source presented from the lower loudspeaker alone; this was referred to as the "localization threshold." Delays ranging from 0–10 ms were applied to the height layer. The localization threshold was only significantly affected by the ICTD. The median threshold for 0 ms was –9.5 dB, which was significantly lower than the –7 dB found for the stimuli in which the height layer was delayed. No evidence was found to support the existence of the precedence effect in the median plane.
Convention Paper 9513

10:00

P6-3 The Perception of Vertical Image Spread by Interchannel Decorrelation—*Christopher Gribben, Hyunkook Lee*, University of Huddersfield, Huddersfield, UK

Subjective listening tests were conducted to assess the general perception of decorrelation in the vertical domain. Interchannel decorrelation was performed between a pair of loudspeakers in the median plane; one at ear level and the other elevated 30° above. The test stimuli consisted of decorrelated octave-band pink noise samples (63–8000 Hz), generated using three decorrelation techniques—each method featured three degrees of the interchannel cross-correlation coefficient (ICCC): 0.1, 0.4, and 0.7. Thirteen subjects participated in the experiment, using a pairwise comparison method to grade the sample with the greater perceived vertical image spread (VIS). Results suggest there is broadly little difference of overall VIS between decorrelation methods, and changes to vertical interchannel decorrelation appear to be better perceived in the upper-middle-frequencies.
[Also a poster—see session P12-16]
Convention Paper 9514

10:30

P6-4 Measurements to Determine the Ranking Accuracy of Perceptual Models—*Andy Pearce, Tim Brookes, Russell Mason, Martin Dewhirst*, University of Surrey, Guildford, Surrey, UK

Linear regression is commonly used in the audio industry to create objective measurement models that predict subjective data. For any model development, the measure used to evaluate the accuracy of the prediction is important. The most common of these assume a linear relationship between the subjective data and the prediction, though in the early stages of model development this is not always the case. Measures based on rank ordering (such as Spearman's test), can alternatively be used. Spearman's test, however, does not consider the variance of the subjective results. This paper presents a method of incorporating the subjective variance in the Spearman's rank ordering test using Monte Carlo simulations and shows how this can be used to develop predictive models.
Convention Paper 9515

11:00

P6-5 Assessment of the Impact of Spatial Audiovisual Coherence on Source Unmasking—*Julian Palacino*,[1] *Mathieu Paquier*,[1] *Vincent Koehl*,[1] *Frédéric Changenet*,[2] *Etienne Corteel*[3]
[1]UBO - LabSTICC, Lorient, France
[2]Radio France, Paris, France
[3]Sonic Emotion Labs, Paris, France

The present study aims at evaluating the contribution of spatial audiovisual coherence for sound source unmasking for live music mixing. Sound engineers working with WFS technologies for live sound mixing have reported that their mixing methods have radically changed. Using conventional mixing methods, the audio spectrum is balanced in order to get each instrument intelligible inside the stereo mix. In contrast, when using WFS technologies, the source intelligibility can be achieved thanks to spatial audiovisual coherence and/or sound spatialization (and without using spectral modifications). The respective effects of spatial audiovisual coherence and sound spatialization should be perceptually evaluated. As a first step, the ability of naive and expert subjects to identify a spatialized mix was evaluated by a discrimination task. For this purpose, live performances (rock, jazz, and classic) were played back to subjects with and without stereoscopic video display and VBAP or WFS audio rendering. Two sound engineers realized the audio mixing for three pieces of music and for both audio technologies in the same room where the test have been carried out.
[Also a poster—see session P12-10]
Convention Paper 9516

11:30

P6-6 Auditory Perception of the Listening Position in Virtual Rooms Using Static and Dynamic Binaural Synthesis *—Annika Neidhardt, Bernhard Fiedler, Tobias Heinl*, University of Technology Ilmenau, Ilmenau, Germany

Virtual auditory environments (VAEs) can be explored by controlling the position and orientation of an avatar and listening to the scene from its changing perspective. Reverberation is essential for immersion and plausibility as well as for externalization and the distance perception of the sound sources. These days, room simulation algorithms provide a high degree of realism for static and dynamic binaural reproduction. In this investigation, the ability of people to discriminate listening positions within a virtual room is studied. This is interesting to find out whether the state of the art room simulation algorithms are perceptually appropriate, but also to learn more about people's capability of orientating themselves within a purely acoustical scene. New findings will help designing suitable VAEs.
[Also a poster—see session P12-1]
Convention Paper 9517

Session P7 Sunday, June 5
09:00 – 12:30 Room 352B

AUDIO SIGNAL PROCESSING—PART 2: BEAMFORMING, UPMIXING, HRTF

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

09:00

P7-1 Dual-Channel Beamformer Based on Hybrid Coherence and Frequency Domain Filter for Noise Reduction in Reverberant Environments—*Hong Liu, Miao Sun*, Shenzhen Graduate School, Peking University, Guangdong, China

As an effective technique for suppressing coherent noise, adaptive beamforming shows a strong decrease in reverberant rooms due to multipath room reflections of received signals. In this paper a dual-channel beamformer based on noise coherence and frequency domain filter is proposed. First, hybrid coherence based on coherent-to-diffuse energy ratio (CDR) is introduced to approximate the coherence of noise signals. Then the hybrid coherence is used to estimate the noise power spectral density (PSD), which is applied to the frequency domain filter to reduce noise and reverberation components in microphone signals. Finally, outputs of the filter are processed by a beamformer to suppress residual noise. Experiments demonstrate that the proposed system has noticeable improvements in SNR and quality of the output in reverberant environments.
Convention Paper 9518

Convention Paper 9519 was withdrawn

10:00

P7-3 Estimation of Individualized HRTF in Unsupervised Conditions—*Mounira Maazaoui, Olivier Warusfel*, UMR STMS, IRCAM-CNRS-UPMC Sorbonne Universités, Paris, France

Head Related Transfer Functions (HRTF) are the key features of binaural sound spatialization. Those filters are specific to each individual and generally measured in an anechoic room using a complex process. Although the use of non-individual filters can cause perceptual artifacts, the generalization of such measurements is hardly accessible for large public. Thus, many authors have proposed alternative individualization methods to prevent from measuring HRTFs. Examples of such methods are based on numerical modeling, adaptation of non-individual HRTFs or selection of non-individual HRTFs from a database. In this article we propose an individualization method where the best matching set of HRTFs is selected from a database on the basis of unsupervised binaural recordings of the listener in a real-life environment.
Convention Paper 9520

10:30

P7-4 Plane Wave Identification with Circular Arrays by Means of a Finite Rate of Innovation Approach— *Falk-Martin Hoffmann, Filippo Maria Fazi, Philip Nelson*, University of Southampton, Southampton, UK

Many problems in the field of acoustic measurements depend on the direction of incoming wave fronts w.r.t. a measurement device or aperture. This knowledge can be useful for signal processing purposes such as noise reduction, source separation, de-aliasing, and super-resolution strategies among others. This paper presents a signal processing technique for the identification of the directions of travel for the principal plane wave components in a sound field measured with a circular microphone array. The technique is derived from a finite rate of innovation data model and the performance is evaluated by means of a simulation study for different numbers of plane waves in the sound field.
[Also a poster—see session P12-13]
Convention Paper 9521

11:00

P7-5 Mismatch between Interaural Level Differences Derived from Human Heads and Spherical Models—*Ramona Bomhardt, Janina Fels*, RWTH Aachen University, Aachen, Germany

The individualization of head-related transfer functions (HRTFs) is important for binaural reproduction to reduce measurement efforts and localization errors. One common assumption of individualization for frequencies below 6 kHz is that the sound pressure field around a sphere is similar to the one of a human head. To investigate the accuracy of this approximation, this paper compares the frequency-dependent interaural level difference (ILD) from a spherical approximation, a simulation using magnetic resonance imaging and individually measured HRTFs of 23 adults' heads. With this database, it is possible to analyze the influence of the head shape and the pinna on ILD using the boundary element method and the measured HRTFs. Meanwhile the mismatch between the spherical and human ILD below 1.5 kHz in the horizontal plane is small, they differ above. In the frequency range of 1.5 and 3.5 kHz, ILD of one side of the head is dominated by two maxima. The offset of the ear canal entrance towards the back of the head and the depth of the head are the two major influencing factors. In general, it is observed that the maxima of a spherical ILD are much smaller and more widely spaced than in the human ILD. Above 4 kHz the difference between human and spherical ILDs is even stronger.
Convention Paper 9522

11:30

P7-6 Stereo Panning Law Remastering Algorithm Based on Spatial Analysis—*François Becker, Benjamin Bernard*, Medialab Consulting SNP, Monaco, Monaco

Changing the panning law of a stereo mixture is often impossible when the original multitrack session cannot be retrieved or used, or when the mixing desk uses a fixed panning law. Yet such a modification would be of interest during tape mastering sessions, among other applications. We present a frequency-based algorithm that computes the panorama power ratio from stereo signals and changes the panning law without altering the original panorama.
[Also a poster—see session P19-11]
Convention Paper 9523

12:00

P7-7 Non-Linear Extraction of a Common Signal for Upmixing Stereo Sources—*François Becker, Benjamin Bernard*, Medialab Consulting SNP, Monaco, Monaco

In the context of a two- to three-channel upmix, center channel derivations fall within the field of common signal extraction methods. In this paper we explore the pertinence of the performance criteria that can be obtained from a probabilistic approach to source extraction; we propose a new, non-linear method to extract a common signal from two sources that makes the implementation choice of deeper extraction with a criteria of information preservation; and we provide the results of preliminary listening tests made with real-world audio materials.
[Also a poster—see session P19-12]
Convention Paper 9524

Session P8 **Sunday, June 5**
09:00 – 11:00 **Foyer**

POSTERS: ROOM ACOUSTICS, INSTRUMENTATION AND MEASUREMENT

09:00

P8-1 Numerical Modeling of Sound Intensity Distributions around Acoustic Transducer—*Adam Kurowski, Józef Kotus, Bozena Kostek, Andrzej Czyzewski*, Gdansk University of Technology, Gdansk, Poland

The aim of this research study is to measure, simulate, and compare sound intensity distribution generated by the acoustic transducers of the loudspeaker. The comparison of the gathered data allows for validating the numerical model of the acoustic radiation. An accurate model of a sound source is necessary in mathematical modeling of the sound field distribution near the scattering obstacles. An example of such obstacle is a human head. Preparation of a robust mathematical model of the sound field generated by a loudspeaker is one of the important factors in simulation of sound waves scattering by the human head. The numerical model is developed for the purpose of this kind of research.
Convention Paper 9525

09:00

P8-2 Small-Rooms Dedicated to Music: From Room Response Analysis to Acoustic Design—*Lorenzo Rizzi,*[1] *Gabriele Ghelfi,*[1] *Maurizio Santini*[3]
[1]Suono e Vita - Acoustic Engineering, Lecco, Italy
[2]Università degli Studi di Bergamo, Bergamo, Italy

Reviewing elements of on-field professional experience gained by the authors in the analysis of small-rooms dedicated to music, case studies offered by the everyday working practice allow to deal with specific situations, these are seldom described by usual theoretical models and literature. Using the analysis procedure developed and refined by authors, it is possible to investigate the characteristics of the acoustic response of the small-rooms with more detail. In this paper case studies of particular interest will be described: different small-room phenomena will be shown in the reported measurements.
[Also a lecture—see Session P4-1]
Convention Paper 9502

09:00

P8-3 Electronic Shell—Improvement of Room Acoustics without Orchestra Shell Utilizing Active Field Control—*Takayuki Watanabe, Hideo Miyazaki, Masahiro Ikeda*, Yamaha Corporation, Hamamatsu, Shizuoka, Japan

This paper introduces an example of Electronic Shell acoustic enhancement system that was installed in a multi-purpose hall without an orchestra shell. The system is based on the concept of Active Field Control using electroacoustic means. The three objectives of this system were (1) the enhancement of early reflection for performers, (2) the increase of the reverberation time and the total sound energy on stage, and (3) the enhancement of early reflection in the audience area. The application of this system showed an improvement of about 1 to 2 dB in STearly and more than 2 dB in G in the audience area, which is equivalent or better performance than simple mobile typed orchestra shell.
[Also a lecture—see session P4-3]
Convention Paper 9504

09:00

P8-4 A Novel Approach of Multichannel and Stereo Control Room Acoustic Treatment, Second Edition—*Bogic Petrovic, Zorica Davidovic*, MyRoom Acoustics, Belgrade, Serbia

This paper describes additional development and improvement for all walls and ceiling diffusers, a new principle for multichannel or stereo control room setup/treatment, as was originally published at the 129th AES Convention (Paper Number 8295). The main effort focused on lowering the price of treatment, optimization of LF absorption, simplification of diffuser construction, solution for long diffusers without periodic repetition of diffusive sequence, and increasing room decay. All of these procedures and design principles will be described and attached to this paper, including theoretical analysis and room acoustical measurements from some of the first control rooms built following this new and improved principle.
Convention Paper 9526

09:00

P8-5 Harmonic Distortion Measurement for Nonlinear System Identification—*John Vanderkooy,*[1] *Sean Thomson*[2]
[1]University of Waterloo, Waterloo, ON, Canada
[2]B&W Group Ltd., Steyning, West Sussex, UK

In order to model nonlinearities in loudspeakers, accurate measurement of harmonic distortion is necessary with particular attention to the relative phases of fundamental and harmonics. This paper outlines several ways

that logarithmic sweeps can be used to achieve this goal. It is shown that Novak's redesign of the logsweep is not strictly necessary, if proper account is taken of the phase relationships of the various harmonics. We study several other types of sweeps and methods to extract precise harmonic amplitudes and phases, using tracking filter concepts. The paper also deals with measurement systems that may have fractional-sample delays between excitation, reference, and data channels. Such details are important for accurate phase characterization of transfer functions. An intermodulation example is given for which sweeps with a single instantaneous frequency are inadequate.
[Also a lecture—see session P3-3]
Convention Paper 9497

09:00

P8-6 Metrics for Constant Directivity—*Rahulram Sridhar, Joseph G. Tylka, Edgar Choueiri*, Princeton University, Princeton, NJ, USA

It is often desired that a transducer have a polar radiation pattern that is invariant with frequency, but there is currently no way of quantifying the extent to which a transducer possesses this quality (often called "constant directivity" or "controlled directivity"). To address the problem, commonly-accepted criteria are used to propose two definitions of constant directivity. The first, stricter definition, is that the polar radiation pattern of a transducer should be invariant over a specified frequency range, whereas the second definition is that the directivity factor (i.e., the ratio between the on-axis power spectrum and the average power spectrum over all directions), or index when expressed in dB, should be invariant with frequency. Furthermore, to quantify each criterion, five metrics are derived: (1) Fourier analysis of contour lines (i.e., lines of constant sensitivity over frequency and angle), (2) directional average of frequency response distortions, (3) distortion thresholding of polar responses, (4) standard deviation of directivity index, and (5) cross-correlation of polar responses. Measured polar radiation data for four loudspeakers are used to compute all five metrics that are then evaluated based on their ability to quantify constant directivity. Results show that all five metrics are able to quantify constant directivity according to the criterion on which each is based, while only two of them, metrics 4 and 5, are able to adequately quantify both proposed definitions of constant directivity.
[Also a lecture—see session P3-7]
Convention Paper 9501

09:00

P8-7 Modeling Non-Shoebox Shaped Rooms with the Mode Matching Method—*Bjørn Kolbrek, U. Peter Svensson*, Norwegian University of Science and Technology, Trondheim, Norway

When a room is not shoebox shaped, usually no analytical expressions exist for the determination of resonance frequencies and mode shapes. One option is to employ the Finite Element Method (FEM). In this paper an alternative method, the Mode Matching Method (MMM), is used to compute the transfer function and sound field of a non-shoebox shaped room with rigid walls and is compared to an FEM solution. The two methods show excellent agreement.
[Also a lecture—see session P4-5]
Convention Paper 9506

Tutorial 3 **Sunday, June 5**
09:00 – 10:00 **Room 352A**

HEADPHONE VIRTUALIZATIONS—PRODUCE A NEW IMMERSIVE / 3D SOUND EXPERIENCE FOR THE MAIN AUDIO APPLICATION OF TODAY

Presenter: **Tom Ammermann**, New Audio Technology GmbH, Hamburg, Germany

Headphones are the most common audio playback device to listen to music. Games and mobile film entertainment is a growing headphone application, too. Furthermore virtual and augmented reality is evolving where headphones certainly will have a basic role. So listen simply stereo on headphones shouldn't be the end but the beginning of involving audio experiences with the flexibility of headphone applications. The tutorial will show different applications, tools, and provide a lot of listening examples from the new "Major Band" production.

Workshop 4 **Sunday, June 5**
09:00 – 10:00 **Room 351**

EXPERT TRANSFER TECHNIQUES: A SPECIAL FOCUS ON MECHANICAL DISCS

Chair: **Nadja Wallaszkovits**, Phonogrammarchiv, Austrian Academy of Science, Vienna, Austria; NOA GmbH

Panelists: *Klaus Blasquiz*, Paris, France
Stefano S. Cavaglieri, Fonoteca Nazionale Svizzera, Lugano, Switzerland
Jean-Hugues Chenot, Institut National de l'Audiovisuel, Bry-sur-Marne, France

The workshop leads through the problems of transfer, digitization, and restoration of historical obsolete disc formats. Starting with the possibilities, advantages, and limitations of a conventional mechanical transfer, the discussion will outline some of the most proven and tested optical transfer methods and technologies and their special usability with broken/ delaminated/ damaged discs. The different approaches will be presented, including various audio examples.

This session is presented in association with the AES Technical Committee on Archiving and Restoration

Student Event and Career Development
Sunday, June 5, 09:00 – 10:00
Havane Amphitheatre

STUDENT RECORDING CRITIQUES

Moderator: **Ian Corbett**, Kansas City Kansas Community College, -Kansas City, KS, USA; off-beat-open-hats recording & sound reinforcement

The Student Recording Critiques are non-competitive listening sessions, designed for students to listen to their recordings and productions on a world-class playback system, and receive feedback on their work. Students are invited to bring along their mixes and have them critiqued by a panel of renowned industry professionals in order to get pointers as to how they can push their skills to the next level.

It is vital that interested students sign up at the student booth immediately after the first SDA meeting and bring their work on CD, DVD, memory stick, or hard disc as clearly labelled 44.1 kHz

WAVE or AIFF files.

Please note that finalists of the Student Recording Competition are excluded to submit their work to the Recording Critiques to give non-finalists a way of having their work assessed.

Sunday, June 5 09:00 Room 363

Technical Committee Meeting on High-Resolution Audio

Sunday, June 5 10:00 Room 363

Technical Committee Meeting on Broadcast and Online Delivery

Workshop 5 Sunday, June 5
10:15 – 11:15 Room 352A

PRACTICAL PHYSICS OF HANDLING AND WIND NOISE

Chair: *Chris Woolf*, Broadcast Engineering Systems, Cornwall, UK

Panelists: *David Josephson*, Josephson Engineering, Inc., Santa Cruz, CA, USA
Michael Williams, Sounds of Scotland, Le Perreux sur Marne, France

To be of any practical use a microphone needs to be supported in some fashion and must be protected from all air movements and pressure changes that do NOT constitute useful audio. Myth, anecdote, and old habits are often relied upon as solutions to these problems, but an understanding of the down-to-earth, kitchen-table physics involved is considerably more useful.

This session is presented in association with the AES Technical Committee on Microphones and Applications

Workshop 6 Sunday, June 5
10:15 – 11:15 Room 351

SCREAMING STREAMING—LOUDNESS PEACE TALKS ARE MANDATORY

Chair: **Florian Camerer**, ORF, EBU, Vienna, Austria

Panelists: *Eelco Grimm*, HKU, Utrecht, The Netherlands
Thomas Lund, Genelec Oy, Iilsami, Finland

Streaming is rapidly becoming the preferred way to consume not only music but any content relying on a timed presentation (like broadcasting). So it seems only natural to apply the successful loudness levelling and normalization paradigm developed in broadcasting to streaming. Unfortunately, the loudness war in popular music as well as the limited level playback capabilities of European mobile devices have led to de-facto levels for streaming that are way above the Target Level in broadcasting. Most of the levels used in streaming also enforce severe compromises regarding the technical quality of the content. The panel will introduce this topic, reflect on the AES recommendation for streaming and report on recent findings regarding the loudness levels of streaming services as well as developments in European legislative bodies trying to improve the situation.

Student Event and Career Development
Sunday, June 5, 10:15 – 12:30
Havane Amphitheatre

RECORDING COMPETITION—PART 1

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. The top three finalists in each category, as identified by our judges, present a short summary of their production intentions and the key recording and mix techniques used to realize their goals. They then play their projects for all who attend. Meritorious awards are determined here and will be presented at the closing Student Delegate Assembly Meeting (SDA-2) on Tuesday afternoon. The competition is a great chance to hear the work of your fellow students at other educational institutions. Everyone learns from the judges' comments, even those who don't make it to the finals, and it's a great chance to meet other students and faculty.

10:15: Category 1—Traditional Acoustic Recording

Judges: Malgorzata Albinski, Thor Legvold, Morten Lindberg, Dan Shores

11:15: Category 2–Traditional Studio Recording

Judges: Jim Anderson, Richard King, Barry Marshall, Mandy Parnell

Sunday, June 5 11:00 Room 363

Technical Committee Meeting on Loudspeakers and Headphones

Workshop 7 Sunday, June 5
11:30 – 12:30 Room 352A

AUDIO RECORDING AND PRODUCTIONS FOR VIRTUAL REALITY /3 60-DEGREE APPLICATIONS

Chair: **Matthieu Parmentier**, francetélévisions, Paris, France

Panelists: *Frank Melchior*, BBC Research & Development, Salford, UK
Nils Peters, Qualcomm, San Diege, CA, USA
Jan Plogsties, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany
Ville Pullki, Aalto Univerisity. Espoo, Finland
Frederick Umminger, Sony Computer Entertainment America

At IBC, AES 139th, CES, and other events Virtual Reality has been a huge topic. VR producers more and more realize the potential and need for spatial audio processing for VR applications. This workshop will discuss the following topics:

• How to record audio for 360° video? —Mics: Can we use the same techniques as for Movie/TV productions? Or does a B-format mic do the trick?

• How to mix audio for 360/VR?—Formats: What are the formats to store and deliver, Channels, object or ambisonics, a combination or binaural?; Processing: How do we combine the different tracks? What Plugins and production tools are there?; Monitoring: How can we monitor what the user will hear?

• How to deliver audio for different VR applications?—Creating a VR app with spatial audio (SDKs, tools); User cases: mobile, VR glass, streaming, browser-based; Codec and limitations

• How to render audio for 360/VR?—Headphone rendering for VR glasses; Speaker playback for TV; What are resolution and latency requirements

• What quality aspects are important?—Accuracy and plausibility - What is the interaction with video?

This session is presented in association with the AES Technical Committees on Broadcast and On-Line Delivery and Spatial Audio

Professional Sound Expo
Sunday, June 5, 11:00 — 11:45
PSE Stage

A HIGH QUALITY CROSSTALK CANCELLATION LOUDSPEAKER

Presenters: **Christof Faller**, Illusonic GmbH, Zurich, Switzerland; EPFL, Lausanne, Switzerland
Daniel Weiss, Weiss Engineering Ltd., Uster, Switzerland

The presenters will describe a high quality crosstalk cancellation 3-way loudspeaker with algorithms design, including discussion of loudspeaker driver positions and crossover parameters. Crossover order and frequency is optimized for achieving wideband cancellation with well conditioned crosstalk cancellation filters. Reproduction of binaural, stereo, and multichannel surround content is explained and demonstrated.

Workshop 8 **Sunday, June 5**
11:30 – 12:30 **Room 351**

LOW FREQUENCY BEHAVIOR IN SMALL HIGH ACCURACY LISTENING ENVIRONMENTS

Chair: **Dirk Noy**, WSDG, Basel, Switzerland

Panelists: *Ben Kok*, BEN KOK - acoustic consulting, Uden, The Netherlands
Roger Roschnik, PSI Audio, Yverdon-les-Bains, Switzerland

Low frequency prediction in large and medium-size venues has become a standard in the audio industry. However, acoustic modeling of small rooms has not yet evolved into a widely accepted concept mainly because of the unavailability of one accurate tool set. The workshop will explore currently available software-based approaches and real world applications to low frequency prediction as well as present various means of low frequency treatments. Specific studio examples will illustrate comparisons of these approaches and their success in the field. The workshop will also explore the limitations of current LF design modeling, and in specific the underlying mathematical and numerical algorithms, one of them being ray tracing, which is only valid in frequency ranges where lengths are small compared to the characteristic dimensions of the room, another one being FEM/BEM (Finite Element / Boundary Element Methods) that can be employed to study static pressure distribution in three dimensional spaces. The common dividing line is often identified with the so-called Schroeder frequency. The workshop will review these theoretical prediction limits as well as hope to create a dialogue concerning future prediction, design and remediation techniques.

This session is presented in association with the AES Technical Committee on Acoustics and Sound Reinforcement

Professional Sound Expo
Sunday, June 5, 12:00 — 12:45
PSE Stage

SOUND QUALITY IN DIGITAL AUDIO INTERFACES

Presenter: **Jody Thorn**e, Prism Sound

Sound Quality in Digital Audio Interfaces – What do we mean, how do we achieve it. Dispelling the common myths and misconceptions.

Sunday, June 5 **12:00** **Room 363**

Technical Committee Meeting on Spatial Audio

Student Event and Career Development
Sunday, June 5, 12:30 – 14:30
Foyer

STUDENT DESIGN EXHIBITION

All accepted entries to the AES Student Design Competition are given the opportunity to show off their designs at this poster/ tabletop exhibition. The session is free and open to all convention attendees and is an opportunity for aspiring student hardware and software engineers to have their projects seen by the AES design community. It is an invaluable career-building event and a great place for companies to identify their next employees. Students from both audio and non-audio backgrounds are encouraged to participate. Few restrictions are placed on the nature of the projects, which may include loudspeaker designs, DSP plug-ins, analog hardware, signal analysis tools, mobile applications, and sound synthesis devices. Attendees will observe new, original ideas implemented in working-model prototypes.

Session P9 **Sunday, June 5**
12:45 – 13:45 **Room 353**

LIVE SOUND PRODUCTION AND UPMIXING

Chair: **Mark Drews**, University of Stavanger, Stavanger, Norway

12:45

P9-1 **A Hybrid Approach to Live Spatial Sound Mixing—** *Etienne Corteel,*[1] *Raphael Foulon,*[1] *Frédéric Changenet*[2]
[1]Sonic Emotion Labs, Paris, France
[2]Radio France, Paris, France

In this paper, we present an approach for live sound mixing that combines object oriented mixing with Wave Field Synthesis rendering with more standard live mixing techniques. This approach combines a standard mixing desk and an external processing unit. We first describe the system and the controls available to the sound engineer. Such system enables to create extensive contrast in the mix working on spatial positioning (angle, depth) but also projection of sound. We then review the use of the system in various musical genre (classical, jazz, pop) describing concrete application installations.
Convention Paper 9527

13:15

P9-2 **Mono-to-Stereo Upmixing—***Christian Uhle, Patrick Gampp*, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

A method for upmixing of single-channel audio signals for stereophonic sound reproduction in real-time is presented. To this end, the input signal is decomposed into a foreground signal and a background signal. The background signal is decorrelated using a network of nested allpass filters. The intensity of the decorrelation is controlled using a computational model for the perceived intensity of decorrelation. The foreground sound sources like singers and soloists are reproduced in the center of the stereo image. The proposed method enables upmixing from mono to stereo signals (and can also be applied to enhance the

stereo image) with low latency, moderate computational load, and low memory requirements. It produces output signals with a high sound quality and is suitable for automotive and low-bitrate streaming applications.
Convention Paper 9528

Session P10 **Sunday, June 5**
12:45 – 16:15 **Room 352B**

AUDIO QUALITY

Chair: **Robin Reumers**, Galaxy, Mol, Belgium

12:45

P10-1 Subjective Evaluation of High Resolution Audio through Headphones—*Mitsunori Mizumachi,*[1] *Ryuta Yamamoto,*[2] *Katsuyuki Niyada*[3]
[1]Kyushu Institute of Technology, Kitakyushu, Fukuoka, Japan
[2]Digifusion Japan Co., Ltd., Hiroshima, Japan
[3]Hiroshima Cosmopolitan University, Hiroshima, Japan

Recently, high resolution audio (HRA) can be played back through portable devices and spreads across musical genres and generation. It means that most people listen to HRA through headphones and earphones. In this study perceptual discrimination among audio formats including HRA has been invested using a headphones. Thirty-six subjects, who have a variety of audio and musical experience in the wide age range from 20s to 70s, participated in listening tests. Headphone presentation is superior in discriminating the details to the loudspeaker presentation. It is, however, found that the headphone presentation is weak in reproducing presence and reality. Audio enthusiasts and musicians could significantly discriminate audio formats than ordinary listeners in both headphone and loudspeaker listening conditions.
[Also a poster—see session P15-5]
Convention Paper 9529

13:15

P10-2 A Headphone Measurement System Covers both Audible Frequency and Beyond 20 kHz (Part 2)—*Naotaka Tsunoda, Takeshi Hara, Koji Nageno*, Sony Corporation, Tokyo, Japan

A new scheme consists of measurement by wide range HATS, and the free-field HRTF correction was proposed to enable entire frequency response measurement from audible frequency and higher frequency area up to 140 kHz and for direct comparison with free field loud speaker frequency response. This report supplements the previous report that described system concept by adding ear simulator detail and tips to obtain reliable data with much improved reproducibility.
[Also a poster—see session P15-7]
Convention Paper 9530

13:45

P10-3 Methodologies for High-dimensional Objective Assessment of Spatial Audio Quality—*Daniel Darcy, Kent Terry, Grant Davidson, Rich Graff, Alex Brandmeyer, Poppy Crum*, Dolby Laboratories, San Francisco, CA, USA

Traditional methods of subjective assessment of sound, such as ratings scales and forced-choice tasks, can be limited and time intensive in their ability to reflect the depth of experiential qualities associated with spatial hearing. Attempts to report localization of sound can be challenging when confounds or noise are introduced by constrained motions of head turning or pointing, and these approaches do not all record higher-dimensional features of sound like dispersion and trajectory. We propose a structured method of testing to reliably capture the quality of experience of spatial sound. Feature extraction of the high-dimensional representation of reported experiences converts to robust metrics used to tune and drive system performance toward desired perceptual attributes and optimal experiential performance.
Convention Paper 9531

14:15

P10-4 Objective Measures of Voice Quality for Mobile Handsets—*Holly Francois,*[1] *Scott Isabelle,*[2] *Eunmi Oh*[3]
[1]Samsung Electronics R&D Institute UK, Staines-Upon Thames, Surrey, UK
[2]Knowles Inc, Mountain View, CA, USA
[3]Samsung Electronics Co., Ltd., Seoul, Korea

Mobile phones include noise suppression to facilitate use in noisy environments; therefore listening tests in accordance with ITU-T P.835 are appropriate for compari ng handset performance. Objective speech quality measures are an often used cheaper alternative; however the results can be misleading, as rank order compared to listening tests is not always preserved. We compare the outputs of PESQ, POLQA, and 3Quest with the results of P.835 listening tests. As expected, measures intended for use with noise suppression perform that task better than tools that were not initially designed to do so. However, improved measures, that aim to preserve rank order while minimizing both maximum error and RMSE, would improve the reliability of comparative evaluations in background noise.
Convention Paper 9532

14:45

P10-5 The Difference between Stereophony and Wave Field Synthesis in the Context of Popular Music—*Christoph Hold,*[1] *Hagen Wierstorf,*[2] *Alexander Raake*[2]
[1]Technische Universität Berlin, Berlin, Germany
[2]Technische Universität Ilmenau, Ilmenau, Germany

Stereophony and Wave Field Synthesis (WFS) are capable of providing the listener with a rich spatial audio experience. They both come with different advantages and challenges. Due to different requirements during the music production stage, a meaningful direct comparison of both methods has rarely been carried out in previous research. As stereophony relies on a channel- and WFS on a model-based approach, the same mix cannot be used for both systems. In this study mixes of different popular-music recordings have been generated, each for two-channel stereophony, surround stereophony, and WFS. The focus is on comparability between the reproduction systems in terms of the resulting sound quality. In a paired-comparison test listeners rated their preferred listening experience.
[Also a poster—see session P15-10]
Convention Paper 9533

15:15

P10-6 Accelerometer Based Motional Feedback Integrated in a 2 3/4" Loudspeaker—*Ruben Bjerregaard, Anders N.*

Madsen, Henrik Schneider, Finn T. Agerkvist, Michael A. E. Andersen, Technical University of Denmark, Kgs. Lyngby, Denmark

It is a well known fact that loudspeakers produce distortion when they are driven into large diaphragm displacements. Various methods exist to reduce distortion using forward compensation and feedback methods. Acceleration based motional feedback is one of these methods and was already thoroughly described in the 1960s showing good results at low frequencies. In spite of this, the technique has mainly been used for closed box subwoofers to a limited extent. In this paper design and experimental results for a 2 3/4 " acceleration based motional feedback loudspeaker are shown to extend this feedback method to a small full range loudspeaker. Furthermore, the audio quality from the system with feedback is discussed based on measurements of harmonic distortion, intermodulation distortion, and subjective evaluation.
[Also a poster session—see session P15-6]
Convention Paper 9534

Convention Paper 9535 was withdrawn

15:45

P10-7 Visualization Tools for Soundstage Tuning in Cars—
Delphine Devallez,[1] *Alexandre Fénières*,[1] *Vincent Couteaux*[2]
[1]Arkamys, Paris, France
[2]Telecom ParisTech, Paris, France

In order to improve the spatial fidelity of automotive audio systems by means of digital signal processing, the authors investigated means to objectively assess the spatial perception of reproduced stereophonic sound in car cabins. It implied choosing a convenient binaural microphonic system representative of real listening situations and metrics to analyze interaural time differences under 1.5~kHz in those binaural recordings. Frequency-dependent correlation correctly showed the frequencies at which the fidelity was improved and allowed to quantify the improvement. The time-domain correlation seemed to be a good indicator of the apparent source width, but failed at giving the perceived azimuth of the virtual sound source. Therefore that metric must be refined to be used efficiently during audio tunings.
[Also a poster—see session P15-9]
Convention Paper 9536
[Paper will be presented by Alexandre Fénières]

Tutorial 4 — Sunday, June 5
12:45 – 13:45 — Room 351

ACOUSTICS ON A BUDGET

Presenters: **Eddy B. Brixen**, EBB-consult, Smørum, Denmark; DPA Microphones
Ben Kok, BEN KOK - acoustic consulting, Uden, The Netherlands

Achieving good acoustics in the studio is one of the most challenging tasks for an audio engineer, particularly when working on a limited budget. This tutorial will give a brief review of room acoustic requirements for project studios. The presentation includes practical tips on how to improve acoustic performance using common sense and commonly available low-cost materials. Also, methods how to identify and localize room acoustic problems without the use of advanced test equipment will be discussed; and finally, some do's and don'ts regarding loudspeaker placement.
This session is presented in association with the AES Technical Committee on Acoustics and Sound Reinforcement

Workshop 9 — Sunday, June 5
12:45 — 14:30 — Havane Amphitheatre

MIXING MUSIC: PART 2

Chair: **Richard King**, McGill University, Montreal, Quebec, Canada; The Centre for Interdisciplinary Research in Music Media and Technology, Montreal, Quebec, Canada

Panelists: *Jonathan Allen*
Erdo Groot, Polyhymnia International, Hoogland, Netherlands
George Massenburg, Schulich School of Music, McGill University, Montreal, Quebec, Canada; Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Montreal, Quebec, Canada
Rob Toulson, Anglia Ruskin University, Cambridge, UK

Proposed as a continuation of the very successful "Mixing Music" workshop at AES 139 (New York Oct. 2015). A panel of award-winning expert practitioners from varying backgrounds within the industry will spark interesting discussion and debate. Topics will include the process of mixing, techniques used, and proven methodologies that have yielded successful results over many years. Focus will remain on real information such as different ways to approach a mix, how to improve an existing mix, how to best interpret and address mix comments from an artist or the client. Balancing, use of processing, and listening levels will be addressed. Ample time will be reserved for questions, so that the audience will have a chance to solicit specific and meaningful information from the panel members.
*This session is presented in association with the AES Technical Committee on Recording Technology and Practice*s

Workshop 10 — Sunday, June 5
12:45 – 13:45 — Room 352A

BINAURAL SOUND IN THE AGE OF RADIO AND TELEVISION BROADCAST: WHY AND HOW?

Chair: **Marc Emerit**, Orange Labs, France

Panelists: *Hervé Dejardin*, Radio France, Paris, France
Frank Melchior, BBC Research and Development, Salford, UK
Markus Noisternig, UMR STMS IRCAM-CNRS-UPMC, Paris, France
Matthieu Parmentier, francetélévisions, Paris, France
Jan Plogsties, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

This workshop proposes an overview of broadcasters' strategies and developments to bring spatial audio contents to their audience. Thanks to the digital era, several functions can be added in multimedia players to completely renew the listeners' experience, such as personalized binaural processing within the end-user device. Radio France, Orange, BBC, and France Televisions, together with the IRCAM research center will explain and demonstrate their works incorporating standards such as AES69 (SOFA) and MPEGH-3D.
This session is presented in association with the AES Technical Committee on Spatial Audio

Professional Sound Expo
Sunday, June 5, 13:00 – 13:45
PSE Stage

RIBBON MICROPHONES

Presenter: **Sammy Rothman**

What are ribbon microphones and how do you use them? Sammy Rothman answers these questions and more in "AEA Ribbon Mics: Fix It in the Mic," which delves into all things ribbons including best miking practices and how ribbons mics work. The discussion will address the differences between ribbon and condenser microphones, between active and passive ribbon microphones and between various approaches to ribbon microphone design.

Sunday, June 5 13:00 Room 363

Technical Committee Meeting on Coding of Audio Signals

Session P11 Sunday, June 5
14:00 – 16:30 Room 353

AUDIO CONTENT MANAGEMENT AND APPLICATIONS IN AUDIO

Chair: **Mark Drews**, University of Stavanger, Stavanger, Norway

14:00

P11-1 Development Tools for Modern Audio Codecs—*Jonas Larsen, Martin Wolters*, Dolby Germany GmbH, Nuremberg, Germany

The Dolby Bitstream Syntax Description Language (BSDL) is a generic, XML-based language for describing the syntactical structure of compressed audio-visual streams. This paper describes how the representation of a bitstream syntax in the BSDL is used to ease the development of serialization, deserialization, and editing tools. Additionally, the formal syntax description allows realizing a range of novel analysis methods including bitstream syntax coverage measurements, detailed bitrate profiles, and the automatic generation of rich specification documentation. The approach is exemplified using the AC-4 codec.
Convention Paper 9537

14:30

P11-2 Can Bluetooth ever Replace the Wire?—*Jonny McClintock*, Qualcomm Technology International Ltd., Belfast, Northern Ireland, UK

Bluetooth is widely used as a wireless connection for audio applications including mobile phones, media players, and wearables, removing the need for cables. The combination of the A2DP protocol and frame based codecs used in many Bluetooth stereo audio implementations have led to excessive latency and acoustic performance significantly below CD quality. This paper will cover the latest developments in Bluetooth audio connectivity that will deliver CD quality audio, or better, and low latency for video and gaming applications. These developments together with the increased battery life delivered by Bluetooth Smart could lead to the elimination of wires for many applications.
[Also a poster—see session P15-11]
Convention Paper 9538

15:00

P11-3 Deep Neural Networks for Dynamic Range Compression in Mastering Applications—*Stylianos Ioannis Mimilakis,[1] Konstantinos Drossos,[2] Tuomas Virtanen,[2] Gerald Schuller[1]*
[1]Fraunhofer Institute for Digital Media Technology (IDMT), Ilmenau, Germany
[2]Tampere University of Technology, Tampere, Finland

The process of audio mastering often, if not always, includes various audio signal processing techniques such as frequency equalization and dynamic range compression. With respect to the genre and style of the audio content, the parameters of these techniques are controlled by a mastering engineer, in order to process the original audio material. This operation relies on musical and perceptually pleasing facets of the perceived acoustic characteristics, transmitted from the audio material under the mastering process. Modeling such dynamic operations, which involve adaptation regarding the audio content, becomes vital in automated applications since it significantly affects the overall performance. In this work we present a system capable of modelling such behavior focusing on the automatic dynamic range compression. It predicts frequency coefficients that allow the dynamic range compression, via a trained deep neural network, and applies them to unmastered audio signal served as input. Both dynamic range compression and the prediction of the corresponding frequency coefficients take place inside the time-frequency domain, using magnitude spectra acquired from a critical band filter bank, similar to humans' peripheral auditory system. Results from conducted listening tests, incorporating professional music producers and audio mastering engineers, demonstrate on average an equivalent performance compared to professionally mastered audio content. Improvements were also observed when compared to relevant and commercial software.
[Also a poster—see session P15-8]
Convention Paper 9539

15:30

P11-4 Principles of Control Protocol Design and Implementation—*Andrew Eales, Richard Foss, Rhodes University, Grahamstown, Eastern Cape, South Africa*

Control protocols are used within audio networks to manage both audio streams and networked audio devices. A number of control protocols for audio devices have been recently developed, including the AES standards AES64-2012 and AES70-2015. Despite these developments, an ontology of control protocol design and implementation does not exist. This paper proposes design and implementation heuristics for control protocols. Different categories of control protocol design and implementation heuristics are presented and the implications of individual heuristics are discussed. These heuristics allow the features provided by different control protocols to be compared and evaluated and provide guidelines for future control protocol development.
Convention Paper 9540

16:00

P11-5 Absorption Materials in Reflex Loudspeakers—*Juha Backman*, Genelec Oy, Iisalmi, Finland

It is well known that the placement of absorbent material has an effect on the behavior of ported (reflex) enclosures, even if the acoustic solution of the field inside the enclo-

sure would predict that the pressure field is quite homogeneous and that the flow velocities in the acoustic field are small. A CFD model is used to study this phenomenon, and the results indicate that there is strong vortex formation inside an unlined enclosure even at small volume velocities, and that the presence and the distribution of porous material has a strong effect on these vortices.
Convention Paper 9541

Tutorial 5 **Sunday, June 5**
14:00 – 15:15 **Room 352A**

WFS OBJECT BASED MIXING FOR THE "CONCERT & DRAMA" ENVIRONMENT PRODUCED BY RADIO FRANCE

Presenter: **Frédéric Changenet**, Radio France, Paris, France

For a year and a half, the Orchestre National de France (ONF) and the "service des fictions" (France Culture) have been co-producing "Concerts & Drama" that are a subtle mix of a radio drama and a symphonic music concert from original text and music. Actors and "sound effects experts" are on the stage of the Studio 104 (856 seats) with the orchestra in an oratorio configuration for a public show. WFS is used for the PA and a live mix is produced for broadcasting. In this tutorial the presenter will describe the technical and artistic aspects of these productions, with special attention devoted to the issues of an immersive 3D Sound Reinforcement using object based mixing in WFS.

Workshop 11 **Sunday, June 5**
14:00 – 15:15 **Room 351**

"COACHING" THE MIX TO OBTAIN A GREAT MASTERING RESULT

Presenters: **Andres A. Mayo**, Andres Mayo Mastering & Audio Post, Buenos Aires, Argentina
Rob Toulson, Anglia Ruskin University, Cambridge, UK

If you are mixing for CD, DVD, Blu-Ray, Vinyl or digital download, you will need to use different approaches in order to get the most out of these formats. Also, depending on the type of music you are producing, certain parameters (i.e., dynamics) will be dramatically different. And of course the mastering job will be different. This presentation/discussion is about the best way to finish your mix, and how the process of Mix Coaching (mix supervision prior to mastering) makes it possible to take full advantage of the mastering job. Real examples will be shown and attendees are encouraged to bring their own mix (one per person) for a critical listening session, depending on the time available.

This session is presented in association with the AES Technical Committee on Recording Technology and Practices

Professional Sound Expo
Sunday, June 5, 14:00 – 14:45
PSE Stage

ACOUSTIC ROOM TREATMENT

Presenter: **Lukas Rimbach**

Why do you need room treatments? This session will discuss the science and practice of treating rooms to optimize acoustic performance.

Sunday, June 5 **14:00** **Room 363**

Technical Committee Meeting on Sound for Digital Cinema and Television

Session P12 **Sunday, June 5**
14:45 – 16:45 **Foyer**

POSTERS: PERCEPTION PART 1 AND AUDIO SIGNAL PROCESSING PART 2

14:45

P12-1 Auditory Perception of the Listening Position in Virtual Rooms Using Static and Dynamic Binaural Synthesis—*Annika Neidhardt, Bernhard Fiedler, Tobias Heinl*, University of Technology Ilmenau, Ilmenau, Germany

Virtual auditory environments (VAEs) can be explored by controlling the position and orientation of an avatar and listening to the scene from its changing perspective. Reverberation is essential for immersion and plausibility as well as for externalization and the distance perception of the sound sources. These days, room simulation algorithms provide a high degree of realism for static and dynamic binaural reproduction. In this investigation, the ability of people to discriminate listening positions within a virtual room is studied. This is interesting to find out whether the state of the art room simulation algorithms are perceptually appropriate, but also to learn more about people's capability of orientating themselves within a purely acoustical scene. New findings will help designing suitable VAEs.
[Also a lecture—see session P6-6]
Convention Paper 9517

14:45

P12-2 A Low Computational Complexity Beamforming Scheme Concatenated with Noise Cancellation—*Jin Xie, Sungyub Yoo, Kapil Jain*, Marvell Technology Group Ltd., Santa Clara, CA, USA

In this paper we present a microphone beamforming algorithm. This algorithm has been implemented in Marvell's proprietary digital signal processor embedded in Marvell's audio codec chip. This beamforming algorithm features (1) easy to implement; (2) sound source localization (SSL) and sound source tracking; and (3) single in single out frequency domain noise cancellation. Lab tests show that the performance is better than the reference existing codec.
[Also a lecture—see sesion P7-2]
Convention Paper 9519
This paper was withdrawn

14:45

P12-3 An Innovative Structure for the Approximation of Stereo Reverberation Effect Using Mixed FIR/IIR Filters—*Andrea Primavera, Stefania Cecchi, Laura Romoli, Massimo Garai, Francesco Piazza*, Universitá Politecnica della Marche, Ancona (AN), Italy

Reverberation is a well-known effect that has an important role in our listening experience. Focusing on hybrid reverberator structures, an innovative structure for the approximation of stereo impulse responses considering low complexity filters is presented in this paper. More in detail, the conventional hybrid reverberator structure has been modified using also a tone correction filter for the emula-

tion of high frequency air-absorbing effect and introducing a technique for the reproduction of the stereo perception. On this basis, the presented approach allows to obtain a better approximation of the impulse responses considering both time and frequency domain. Several results are reported considering different real impulse responses and comparing the results with previous techniques in terms of computational complexity and reverberation quality.
Convention Paper 9542
[Paper presented by Michele Gasparini]

14:45

P12-4 Improvement of DUET for Blind Source Separation in Closely Spaced Stereo Microphone Recording—*Chan Jun Chun, Hong Kook Kim*, Gwangju Institute of Science and Tech (GIST), Gwangju, Korea

This paper proposes a blind source separation (BSS) method to improve the performance of the degenerate unmixing estimation technique (DUET) when sound sources are recorded using closely spaced stereo microphones. In particular, the attenuation-delay-based discrimination analysis employed in DUET is replaced with a microphone spacing- and source direction-based discrimination analysis in order to remedy the problem of DUET when the attenuation factors between recorded stereo audio signals are not distinguishable. In other words, the proposed BSS method generates a histogram as a function of the microphone spacing and the directional difference between stereo signals. Next, the generated histogram is used to partition the time-frequency representations of the mixtures into that of each sound source. The performance of the proposed method is evaluated by means of both objective and subjective measures. Consequently, it is shown from the evaluation that the proposed BSS method outperforms the conventional DUET in a closely spaced stereo microphone recording environment.
Convention Paper 9543

14:45

P12-5 A Phase-Matched Exponential Harmonic Weighting for Improved Sensation of Virtual Bass—*Hyungi Moon,*[1] *Gyutae Park,*[1] *Young-cheol Park,*[2] *Dae Hee Youn*[1]
[1]Yonsei University, Seoul, Korea
[2]Yonsei University, Wonju, Kwangwon-do, Korea

Virtual Bass System (VBS) is based on the psychoacoustic phenomenon called "missing fundamental" is widely used to extend the lower frequency limit of the small loudspeakers. The perceptual quality of the VBS is highly dependent on the weighting strategy for the generated harmonics. There have been several weighting strategies for the generated harmonics including loudness matching, exponential attenuation, and timbre matching. To precisely convey the weighting strategy, however, it is essential to match the phases between the reproduced harmonics to the natural harmonics contained in the original signal. In this paper limitations of the previous harmonic weighting schemes are addressed and a new harmonic weighting scheme is proposed. In the proposed weighting scheme, the slope of the attenuation weighting is dynamically varied according to the frequency of the missing fundamental, and a phase matching between the original and generated harmonics is performed prior to the harmonic weighting. Subjective tests show that the proposed method provides more natural and effective bass sensation than the conventional schemes.
Convention Paper 9544

14:45

P12-6 Extraction of Interchannel Coherent Component from Multichannel Audio—*Akio Ando, Hiroki Tanaka, Hiro Furuya*, University of Toyama, Toyama, Japan

Three-dimensional audio recording usually involves a number of spatially distributed microphones to capture the spatial sound. The temporal differences in arrival of sound from a source to microphones make the recorded signal less coherent than that with coincident microphones. In this paper a new method that extracts the interchannel coherent component from multichannel audio signal is proposed. It estimates the component of one channel signal from the other channel signals based on the least squares estimation. The experimental result showed that the new method can extract the interchannel coherent component from multichannel audio signal regardless of the number of channels of the signal.
Convention Paper 9545

14:45

P12-7 The Difference in Perceptual Attributes for the Distortion Timbre of the Electric Guitar between Guitar Players and Non-Guitar Players—*Koji Tsumoto, Atsushi Marui, Toru Kamekawa*, Tokyo University of the Arts, Adachi-ku, Tokyo, Japan

Subjective evaluation experiments were performed to reveal the perceptual attributes for the distorted timbre of the electric guitar. The motivation was to gain smoothness in conversation over the distorted timbre between guitar players and non-guitar players at the recording sessions. The signals of three guitar performance were distorted in three different amount of distortions with three kinds of frequency characteristics. That bring the total to twenty-seven stimuli. Sixteen non-guitar players and sixteen electric guitar players participated in the rating experiments using semantic scales anchored by eight bipolar adjective pairs. The result indicated both had similar perceptual attributes for distorted guitar timbres. One latent factor was found and was correlated with the acoustic features. The alterations of frequency characteristics did not appear as the variable affecting the judgment of distortion timbres.
Convention Paper 9546

14:45

P12-8 The Effect of a Vertical Reflection on the Relationship between Preference and Perceived Change in Timbre and Spatial Attributes—*Thomas Robotham, Matthew Stephenson, Hyunkook Lee,* University of Huddersfield, Huddersfield, UK

This study aims to investigate a vertical reflection's beneficial or detrimental contribution to subjective preference compared with perceived change in timbral and spatial attributes. A vertical reflection was electro-acoustically simulated and evaluated through subjective tests using musical stimuli in the context of listening for entertainment. Results indicate that the majority of subjects preferred audio reproduction with the addition of a reflection. Furthermore, there is a potential relationship between positive preference and the perceived level of both timbral and spatial differences, although this relationship is dependent on the stimuli presented. Subjects also described perceived differences where the reflection was present. These descriptors provide evidence suggest-

ing a link between timbral descriptions and preference. However, this link was not observed between preference and spatial descriptions.
Convention Paper 9547

14:45

P12-9 Relative Contribution of Interaural Time and Level Differences to Selectivity for Sound Localization—*Si Wang, Heng Wang, Cong Zhang*, Wuhan Polytechnic University, Wuhan, Hubei, China

In the present study, we measured threshold of interaural level difference in standard stimulus (ILDs) through the interaural time difference in variable stimulus (ITDv) and tested just notice difference of interaural time difference in standard stimulus (ITDs) by the interaural level differences in variable stimulus (ILDv) for sine wave over a frequency ranging from 150 to 1500 Hz at some lateral positions of sound image. Two separate experiments were conducted based on two alternative forced-choice (2AFC) and 1 up/2 down adaptive procedure. We could explore the relative contribution of Interaural Level Difference (ILD) and Interaural Time Difference(ITD) to sound localization as a function of position and frequency from these experimental data. The results showed lateral discrimination between stimuli are not difficult at frequencies of 350, 450, 570, and 700 Hz when we tested JND of ILD in standard stimulus and the auditory system is easier to discriminate two sound images and is more sensitive to localize the lateral positions of standard stimulus as frequency is varied from 700 to 1500 Hz when we measured JND of ITD in standard stimulus.
Convention Paper 9548

14:45

P12-10 Assessment of the Impact of Spatial Audiovisual Coherence on Source Unmasking—*Julian Palacino,[1] Mathieu Paquier,[1] Vincent Koehl,[1] Frédéric Changenet,[2] Etienne Corteel[3]*
[1]UBO - LabSTICC, Lorient, France
[2]Radio France, Paris, France
[3]Sonic Emotion Labs, Paris, France

The present study aims at evaluating the contribution of spatial audiovisual coherence for sound source unmasking for live music mixing. Sound engineers working with WFS technologies for live sound mixing have reported that their mixing methods have radically changed. Using conventional mixing methods, the audio spectrum is balanced in order to get each instrument intelligible inside the stereo mix. In contrast, when using WFS technologies, the source intelligibility can be achieved thanks to spatial audiovisual coherence and/or sound spatialization (and without using spectral modifications). The respective effects of spatial audiovisual coherence and sound spatialization should be perceptually evaluated. As a first step, the ability of naive and expert subjects to identify a spatialized mix was evaluated by a discrimination task. For this purpose, live performances (rock, jazz, and classic) were played back to subjects with and without stereoscopic video display and VBAP or WFS audio rendering. Two sound engineers realized the audio mixing for three pieces of music and for both audio technologies in the same room where the test have been carried out.
[Also a lecture—see session P6-5]
Convention Paper 9516

14:45

P12-11 Modeling the Perceptual Components of Loudspeaker Distortion—*Sune L Olsen,[1,2] Finn T. Agerkvist,[1] Ewen MacDonald,[1] Tore Stegenborg-Andersen,[2] Christer P. Volk[2]*
[1]Technical University of Denmark, Lyngby, Denmark
[2]DELTA SenseLab, Hørsholm, Denmark

While non-linear distortion in loudspeakers decreases audio quality, the perceptual consequences can vary substantially. This paper investigates the metric Rnonlin [1] which was developed to predict subjective measurements of sound quality in nonlinear systems. The generalizability of the metric in a practical setting was explored across a range of different loudspeakers and signals. Overall, the correlation of Rnonlin predictions with subjective ratings was poor. Based on further investigation, an additional normalization step is proposed, which substantially improves the ability of Rnonlin to predict the perceptual consequences of non-linear distortion.
Convention Paper 9549

14:45

P12-12 Comparison of the Objective and the Subjective Parameters of the Different Types of Microphone Preamplifiers—*Michal Luczynski, Maciej Sabiniok*, Wroclaw University of Technology, Wroclaw, Poland

The aim of this paper is to compare different types of microphone preamplifiers. The authors designed six types of preamps using different technologies (f.ex. based on vacuum tube, transistors, operational amplifiers). Assumed parameters such as input signal, gain, power supply were the same for all circuits. Preamps were tested by objective and subjective methods. Then the authors tried to find out relations between different gain components, electroacoustic parameters, and subjective sensation. The authors did not mean to create commercial devices; just to compare and classify objective and subjective parameters depending on the different types of microphone preamplifier.
Convention Paper 9550

14:45

P12-13 Plane Wave Identification with Circular Arrays by Means of a Finite Rate of Innovation Approach—*Falk-Martin Hoffmann, Filippo Maria Fazi, Philip Nelson*, University of Southampton, Southampton, UK

Many problems in the field of acoustic measurements depend on the direction of incoming wave fronts w.r.t. a measurement device or aperture. This knowledge can be useful for signal processing purposes such as noise reduction, source separation, de-aliasing, and super-resolution strategies among others. This paper presents a signal processing technique for the identification of the directions of travel for the principal plane wave components in a sound field measured with a circular microphone array. The technique is derived from a finite rate of innovation data model and the performance is evaluated by means of a simulation study for different numbers of plane waves in the sound field.
[Also a lecture—see session P7-4]
Convention Paper 9521

14:45

P12-14 Automatic Localization of a Virtual Sound Image Generated by a Stereophonic Configuration—*Laura*

Romoli,[1] *Stefania Cecchi,*[1] *Ferruccio Bettarelli,*[2] *Francesco Piazza*1
[1]Universitá Politecnica della Marche, Ancona, Italy
[2]Leaff Engineering, Ancona, Italy

Sound localization systems aim at providing the position of a particular sound source as perceived by the human auditory system. Interaural level difference, interaural time difference, and spectral representations of the binaural signals are the main cues adopted for localization. When two sound sources are simultaneously active, a virtual source is created. In this paper a novel approach is presented to provide the human perception of a sound image created by two loudspeakers. The solution is based on both frequency-dependent binaural and monaural cues in order to consider the human auditory system sensitivity to spatial sound localization. Experimental results proved the effectiveness of the proposed approach in correctly estimating the horizontal and vertical position of the virtual source.
Convention Paper 9551
[Paper presented by Michele Gasparini]

14:45

P12-15 The Effect of Early Impulse Response Length and Visual Environment on Externalization of Binaural Virtual Sources—*Joseph Sinker, Ben Shirley*, University of Salford, Salford, Greater Manchester, UK

When designing an audio-augmented-reality (AAR) system capable of rendering acoustic "overlays" to real environments, it is advantageous to create externalized virtual sources with minimal computational complexity. This paper describes experiments designed to explore the relationships between early impulse response (EIR) length, visual environment and perceived externalization, and to identify if reduced IR data can effectively render a virtual source in matched and unmatched environments. In both environments a broadly linear trend is exhibited between EIR length and perceived externalization, and statistical analysis suggests a threshold at approximately 30-40 ms above which the extension of the EIR yields no significant increase in externalization.
Convention Paper 9552

14:45

P12-16 The Perception of Vertical Image Spread by Interchannel Decorrelation—*Christopher Gribben, Hyunkook Lee*, University of Huddersfield, Huddersfield, UK

Subjective listening tests were conducted to assess the general perception of decorrelation in the vertical domain. Interchannel decorrelation was performed between a pair of loudspeakers in the median plane; one at ear level and the other elevated 30° above. The test stimuli consisted of decorrelated octave-band pink noise samples (63–8000 Hz), generated using three decorrelation techniques—each method featured three degrees of the interchannel cross-correlation coefficient (ICCC): 0.1, 0.4, and 0.7. Thirteen subjects participated in the experiment, using a pairwise comparison method to grade the sample with the greater perceived vertical image spread (VIS). Results suggest there is broadly little difference of overall VIS between decorrelation methods, and changes to vertical interchannel decorrelation appear to be better perceived in the upper-middle-frequencies.
[Also a lecture—see session 6-3]
Convention Paper 9514

Student Event and Career Development
Sunday, June 5, 14:45 – 17:00
Havane Amphitheatre

RECORDING COMPETITION—PART 2

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. The top three finalists in each category, as identified by our judges, present a short summary of their production intentions and the key recording and mix techniques used to realize their goals. They then play their projects for all who attend. Meritorious awards are determined here and will be presented at the closing Student Delegate Assembly Meeting (SDA-2) on Tuesday afternoon. The competition is a great chance to hear the work of your fellow students at other educational institutions. Everyone learns from the judges' comments, even those who don't make it to the finals, and it's a great chance to meet other students and faculty.

14:45: Category 4—Sound for Visual Media
Judges: Kris Gorski, Scott Levine, Liz Teutsch, Cecile Tournesac

15:45: Category 3—Modern Studio Recording & Electronic Music
Judges: David Miles Huber, Ian Kagey, Andres Mayo

Professional Sound Expo
Sunday, June 5, 15:00 – 15:45
PSE Stage

MEASUREMENT UNCERTAINTY IN AUDIO TESTING

Presenter: **Bruce Hofer**, Audio Precision, Inc., Beaverton, Oregon, USA

All measurements are subject to some degree of uncertainty including such factors as readout resolution, short term stability and temperature effects, and estimates of error with regard to national standards. Reported measurement results should also contain an estimate of their respective uncertainties so that comparisons can be meaningful. In the world of metrology, this process of reporting measurement uncertainty is very strict requiring either some knowledge about the distributions of the various factors and the concept of standard deviation. This session will briefly discuss some of the uncertainty factors that can affect traditional audio measurements.

Sunday, June 5 15:00 Room 363

Technical Committee Meeting on Automotive Audio

Tutorial 6 Sunday, June 5
15:30 – 16:45 Room 352A

"SEQUENCES"—THE LIVE PRODUCTION OF ELECTRONIC MUSIC FOR SURROUND REPRODUCTION

Presenter: **Hervé Dejardin**, Radio France, Paris, France

Presentation of the Production "Séquences."

How were the audio tracks recorded, mixed, and distributed as a series of 13 live electronic music videos in 5.1 and binaural, filmed in 360°, and composed entirely on machines? We will present the problems that appeared at different stages of production. The recording, post production, spatial design, the reverb., etc., will be discussed; and during the tutorial, binaural examples will be played for everybody to hear on headphones.

This session is presented in association with the AES Technical Committee on Microphones and Applications and Recording Technology and Practices

Tutorial 7 **Sunday, June 5**
15:30 – 17:00 **Room 351**

AURO-3D FORMATS AND TECHNOLOGIES—AN OVERVIEW

Presenters: **Wilfried Van Baelen**, Auro Technologies N.V., Mol, Belgium
Bert Van Deale, Auro Technologies N.V, Mol, Belgium

The audio world has recently seen the advent of several new audio formats, each claiming to bring a new three dimensional, immersive sound or "Surround with Height" listening experience to the listener. As a result several organizations are working on a standard for Immersive Sound, incorporating channel-based and/or object-based audio capabilities.

The introduction of such new formats bring new options as well as challenges for content creators such as movie studios, post production, and music facilities. After having experienced the fun side of the new creative options, engineers are still faced with several challenges, as well as optimizing the object-based elements of the Immersive Sound standard for reproduction on a plethora of reproduction systems, each with a number of common as well as unique features. At the same time choices need to be made to maximize the preservation of the creative intent on all reproduction systems.

This tutorial will help establish the fundamentals of the Auro 3D approach to Immersive sound, which enables simultaneous creation of content in various audio formats for movies and music, including the Immersive Sound formats. Several points are highlighted such as compatibility between deliverables, creative possibilities and limitations, and playback in theaters and at home.

Professional Sound Expo
Sunday, June 5, 16:00 – 16:45
PSE Stage

FRENCH RECORDING STUDIOS—2016

Presenter: **Franck Ernould**, Freelance, Paris, France

There have been many changes in the world of French recording studios since the last Paris AES convention in 2006 – Who is among the Top 10? Are there newcomers on the scene? Born-again studios? How do "mid-size" studios survive? What about residential and project studios? Who is active in mobile recording? And mastering? How do studios manage to survive in a rapidly changing world? This discussion is the result of meetings with current studio managers, and features many studio photographs, old and new.

Sunday, June 5 **16:00** **Room 363**

Technical Committee Meeting on Microphones and Applications

Special Event
Sunday, June 5, 17:00–18:00
Room 352A

AUDIO PROJECTIONS 5—BINAURAL AUDIO FROM RADIO FRANCE

Presenter: **Hervé Dejardin**, Radio France, Paris, France

"Sequences"—The Live Production of Electronic Music for Surround Sound Reproduction

Come and listen to extracts from "Sequences" in Binaural Sound reproduction. "Sequences" takes advantage of the latest innovations in the field of multichannel mixing techniques. This immersion in the surround audio space offers the listener the opportunity to better understand and experience the specificity of electronic live music. Radio France, France Télévisions and Milgram Productions have united their unique sound and image expertise for this project, and propose a new experience in electronic music listening.

Special Event
Sunday, June 5, 17:00–18:00
Room 351

AUDIO PROJECTIONS 6—3D AUDIO FROM AURO-3D

Presenters: **Wilfried Van Baelen**, Auro Technologies N.V., Mol, Belgium
Bert Van Daele, Auro Technologies N.V., Mol, Belgium

3D Loudspeaker Demos—9.1, 11.1, 13.1

In this listening session, various examples of immersive content in the Auro-3D formats will be played. The demonstrated content will range from movie excerpts and trailers to music in various genres. The examples will show the added value of 3D immersive sound reproduction from an artistic as well as an emotional point of view.

Sunday, June 5 **17:00** **Room 363**

Technical Committee Meeting on Signal Processing

Special Event
Sunday, June 5, 18:00–19:00
Room 352A

AUDIO PROJECTIONS 7—3D AUDIO PROJECTIONS FROM RADIO FRANCE

Presenter: **Frédéric Changenet**, Radio France, Paris, France

Ever since November 2015, Radio France has been presenting "Du cinéma pour vos oreilles," public WFS audio projections of Dramas, Concerts and Documentaries in studio 105 equipped with 30 loudspeakers. We propose listening to samples of these productions in the Binaural version.

Special Event
Sunday, June 5, 18:00–19:00
Room 351

AUDIO PROJECTIONS 8—SURROUND SOUND PROJECTIONS FROM RADIO FRANCE

Presenter: **Hervé Dejardin**, Radio France, Paris, France

"Sequences"—The Live Production of Electronic Music for Surround Sound Reproduction

Come and listen to extracts from "Sequences" in 8.0 Surround Sound Reproduction. "Sequences" takes advantage of the latest innovations in the field of multichannel mixing techniques. This immersion in a surround sound audio space offers the listener the opportunity to better understand and experience the specificity of electronic live music. Radio France, France Télévisions and Milgram Productions have united their unique sound and image expertise for this project and propose a new experience in electronic music listening in 8.0 Surround Sound.

Sunday, June 5 17:00 Room 363

Technical Committee Meeting on Network Audio Systems

Special Event
Sunday, June 5, 19:30–22:00
Issy Les Moulineaux

BANQUET

For this year's banquet join us in the heart of the Ile Saint Germain at Issy Les Moulineaux. Originally used entirely for agriculture, the island has had a chequered history becoming largely urbanized and almost industrialized at one point before being saved and transformed through a major project into an island park and haven of greenery. We will enjoy French seasonal specialities in a Napoleon III pavilion. French gastronomy is of course world renown and in 2010 was added by UNESCO to its lists of the world's "intangible cultural heritage." Our banquet menu will reflect this with a classic aperitif and three course meal sure to delight your palate.

There will be a musical interlude on a highly original instrument guaranteed to arouse your aural and visual curiosity—played by Robert Hebrard, musicologist and pioneer in experimental instrument manufacture in France. The number of seats is limited, but you can reserve your place now, and pay for your ticket(s) on arrival at the Convention. There will be bus transport from the Convention to the restaurant and returning to the Palais des Congrès afterwards.

85 € for AES members and nonmembers
Tickets will be available at the Special Events desk.

Session P13 Monday, June 6
08:45 – 11:15 Room 353

PERCEPTION—PART 2

Chair: **Thomas Görne**, Hamburg University of Applied Sciences, Hamburg, Germany

08:45

P13-1 Exploiting Envelope Fluctuations to Enhance Binaural Perception—*G. Christopher Stecker*, Vanderbilt University School of Medicine, Nashville, TN, USA

A review of recent and classic studies of binaural perception leads to the conclusion that envelope fluctuations, such as sound onsets, play a critical role in the sampling of spatial information from auditory stimuli. Specifically, listeners' perception of sound location corresponds with the binaural cues (interaural time and level differences) that coincide with brief increases in sound amplitude, and disregards binaural cues occurring at other times. This discrete, envelope-triggered sampling of binaural information can be exploited to enhance spatial perception of synthesized sound mixtures, or to facilitate the localization of mixture components.
[Also a poster—see session P19-9]
Convention Paper 9553

09:15

P13-2 Two Alternative Minimum-Phase Filters Tested Perceptually—*Robert Mores, Ralf Hendrych*, University of Applied Sciences Hamburg, Hamburg, Germany

A widely used method for designing minimum phase filters is based on the real cepstrum (Oppenheim, 1975). An alternative method is proposed for symmetric FIR filters that flips the filter's "left side" around the central coefficient to the "right side" using a sinus ramp of perceptually irrelevant duration. The resulting phase is nearly minimal and nearly linear. The method is applied to impulse responses. Perception tests use original sound samples (A), samples processed by real-cepstrum-based minimum phase filters (B), and samples processed by the proposed method (C). The tests reveal that for impulsive sound samples the perceived dissimilarity between A and C is smaller than the dissimilarity between A and B suggesting that the alternative method has some potential for sound processing.
[Also a poster—see session P19-1]
Convention Paper 9554

09:45

P13-3 Subjective Listening Tests for Preferred Room Response in Cinemas—Part 2: Preference Test Results
—*Linda A. Gedemer*, University of Salford, Salford, UK; Harman International, Northridge, CA, USA

SMPTE and ISO have specified near identical in-room target response curves for cinemas and dubbing stages. However, to this author's knowledge, to date these standards have never been scientifically tested and validated with modern technology and measurement techniques. For this reason it is still not known if the current SMPTE and ISO in-room target response curves are optimal or if better solutions exist. Using a Binaural Room Scanning system for room capture and simulation, various seating positions in three cinemas were reproduced through headphones for the purpose of conducting controlled listening experiments. This system used a binaural mannequin equipped with a computer-controlled rotating head to accurately capture binaural impulse responses of the sound system and the listening space which are then reproduced via calibrated headphones equipped with a head-tracker. In this way controlled listening evaluations can be made among different cinema audio systems tuned to different in-room target responses. Results from a MUSHRA-style preference test are presented.
[Also a poster—see session P19-8]
Convention Paper 9555

10:15

P13-4 Binaural Spatialization over a Bone Conduction Headset: Minimum Discernable Angular Difference—*Amit Barde*,[1] *William S. Helton*,[1] *Gun Lee*,[1] *Mark Billinghurst*[2]
[1]University of Canterbury, Christchurch, Canterbury, New Zealand
[2]University of South Australia, Mawson Lakes, South Australia, Australia

Binaural spatialization in the horizontal plane over a bone conduction headset (BCH) was investigated using inexpensive and commercially available hardware and software components. The aim of this study was to determine the minimum discernable angular difference between two successively spatialized sound sources. Localization accuracy and externalization was also explored. Statistically significant results were observed for angular separations of 10° and above. Localization accuracy was found to be significantly poorer than that seen for previous loudspeaker and headphone based reproduction. Localization errors

between 30° – 35° were observed for stimuli presented in front, back, and sides and 92% of the participants reported externalization. The study demonstrates that an acceptable level of spatial resolution and externalization is achievable using an inexpensive bone conduction headset and software components.
Convention Paper 9556

10:45

P13-5 The Harmonic Centroid as a Predictor of String Instrument Timbral Clarity—*Kirsten Hermes, Tim Brookes, Chris Hummersone*, University of Surrey, Guildford, Surrey, UK

Spectrum is an important factor in determining timbral clarity. An experiment where listeners rate the changes in timbral clarity resulting from spectral equalization (EQ) can provide insight into the relationship between EQ and the clarity of string instruments. Overall, higher frequencies contribute to clarity more positively than lower ones, but the relationship is program-item-
dependent. Fundamental frequency and spectral slope both appear to be important. Change in harmonic centroid (or dimensionless spectral centroid) correlates well with change in clarity, more so than octave band boosted/cut, harmonic number boosted/cut, or other variations on the spectral centroid.
[Also a poster—see session P19-7]
Convention Paper 9557

Session P14 Monday, June 6
08:45 – 12:15 Room 352B

AUDIO SIGNAL PROCESSING—PART 3: AUDIO APPLICATIONS

Chair: **Iva Salom**, Institute Mihajlo Pupin, University of Belgrade, Belgrade, Serbia

08:45

P14-1 Ensemble Effect Using Gaussian Matrices—*Connor McCullough*, Bose Corporation, Boston, MA, USA

The purpose of this paper is to propose an algorithm to serve as an alternative to the chorus effect, the current standard for simulating an ensemble from a single track. Due to the deterministic nature of chorus, specifically the use of an LFO to modulate the delay, chorus often has audible oscillation and does not truly model the behavior of musicians playing simultaneously. The proposed alternative is the implementation of a Gaussian-based algorithm that attempts to model the actual process of musicians playing together. This modeling will be achieved by generating a Gaussian matrix ([# of instruments] x [# of notes]), with each index containing a resampling factor that will temporally and tonally shift each note in a recording. While the Gaussian distribution will serve as the basis for the algorithm, additional constraints will be applied to the resampling factor in order to properly model ensemble behavior.
Convention Paper 9558

09:15

P14-2 A Loudness Function for Analog and Digital Sound Systems Based on Equal Loudness Level Contours—*Sofus Birkedal Nielsen*, Aalborg University, Aalborg, Denmark

A new and better loudness compensation has been designed based on the differences between the Equal Loudness Level Contours (ELLC) in ISO 226:2003. Sound productions are normally being mixed at a high Mixing Level (ML) in dB but often played at a lower listening level, which means that the perceived frequency balance will been changed both for LL lower or higher than ML. The differences in ELLC ask for a level based equalization using fractional-order filters. A designing technique for both analog and digital fractional-order filters has been developed. The analog solution is based on OPAMs and the digital solution is realized in a 16/32 bit fixed point DSP and could be implemented in any sound producing system.
Convention Paper 9559

09:45

P14-3 Spatial Multi-Zone Sound Field Reproduction Using Higher-Order Loudspeakers in Reverberant Rooms—*Keigo Wakayama, Hideaki Takada*, NTT Service Evolution Laboratories, Kanagawa, Japan

We propose a method for reproducing multi-zone sound fields in a reverberant room using an array of higher-order loudspeakers. This method enables sparse arrangement of loudspeakers and reproduction of independent sound fields for multiple listeners without the need for headphones. For multi-zone reproduction, global sound field coefficients are obtained using translation operator. By using the coefficient of the room transfer function measured or simulated with an extension of the image-source method, the loudspeakers' coefficients are then calculated with the minimum norm method in the cylindrical harmonic domain. From experiments of two-zone and three-zone examples, we show that there was a 2N + 1-fold decrease in the number of Nth-order loudspeakers for accurate reproduction with the proposed method compared to conventional methods.
Also a poster—see poster session P19-11
Convention Paper 9560

10:15

P14-4 Active Equalization for Loudspeaker Protection—*Christopher Painter,*[1] *Erfan Soltanmohammadi,*[2] *Kapil Jain*[2]
[1]Marvell Semiconductor, Inc., Longmont, CO, USA
[2]Marvell Semiconductor, Inc., Santa Clara, CA, USA

We present a time-varying linear equalization algorithm whose purpose is to protect a loudspeaker from damage under high drive conditions. It is suitable for implementation on a low-cost digital signal processor, often integrated on the same die as a high-performance audio codec. A typical application is in a portable wireless (e.g., Bluetooth) loudspeaker. For a given driver and enclosure design, the algorithm allows the power output of the loudspeaker to be maximized while introducing only minimal coloration or distortion. During the loudspeaker design phase, the parameters of the algorithm can be easily tuned by the designer, further optimizing the overall design for power output, robustness, and low distortion.
Convention Paper 9561
[Paper not presented]

10:45

P14-5 Comparison of Simple Self-Oscillating PWM Modulators—*Nicolai Dahl, Niels Elkjær Iversen, Arnold Knott, Michael A. E. Andersen*, Technical University of

Denmark, Kgs. Lyngby, Denmark

Switch-mode power amplifiers has become the conventional choice for audio applications due to their superior efficiency and excellent audio performance. These amplifiers rely on high frequency modulation of the audio input. Conventional modulators use a fixed high frequency for modulation. Self-oscillating modulators do not have a fixed modulation frequency and can provide good audio performance with very simple circuitry. This paper proposes a new type of self-oscillating modulator. The proposed modulator is compared to an already existing modulator of similar type and their performances are compared both theoretically and experimentally. The result shows that the proposed modulator provides a higher degree of linearity resulting in around 2% lower Total Harmonic Distortion (THD).
[Also a Poster—see session P19-10]
Convention Paper 9562

11:15

P14-6 Low Energy Audio DSP Design: Going Beyond The Hardware Barrier—*Jamie Angus*, University of Salford, Salford, Greater Manchester, UK

Modern digital audio signal processors need to be energy efficient, both for mobile audio and environmental concerns. Improving technology has been reducing the power of these devices via better, smaller, transistors and reduced voltage swings between one and zero. However, there is a limit to how far this improvement can go. To further reduce processor energy consumption the number of transitions between one and zero must be reduced. This paper presents a method of doing this to, instructions, addresses, and data. By looking at the interaction between their usage statistics and their digital representation and modifying it to match the usage a reduction in energy consumption is achieved. The paper present both measured usage statistics, and bit allocation strategies to achieve this.
Convention Paper 9563

11:45

P14-7 Modeling and Adaptive Filtering for Systems with Output Nonlinearity—*Erfan Soltanmohammadi*,[1] *Christopher Painter*,[2] *Kapil Jain*[1]
[1]Marvell Semiconductor, Inc., Santa Clara, CA, USA
[2]Marvell Semiconductor, Inc., Longmont, CO, USA

Many practical systems are nonlinear in nature, and the Volterra series, also known as nonlinear convolution, is widely used to model these systems. For nonlinear systems with infinite memory, such a modeling approach is usually not feasible because of multiple infinite summations. In practice, the full Volterra series representation of such a system is either approximated by just a few terms, or is otherwise simplified. In an audio system, a useful approximation is to model all memoryless and dynamical nonlinear effects as a combined nonlinearity at its output. In this paper we propose a new Volterra-based structure that accommodates nonlinear systems with output nonlinearity and infinite memory. We then propose an adaptation approach to estimate the Volterra kernels based on the Least Mean Squares (LMS) approach.
Convention Paper 9564
[Paper not presented]

Session P15 Monday, June 6
08:45 – 10:45 Foyer

POSTERS: LIVE SOUND PRACTICE, RENDERING, HUMAN FACTORS AND INTERFACES

08:45

P15-1 Evaluation of Quality Features of Spatial Audio Signals in Non-Standardized Rooms: Two Mixed Method Studies—*Ulrike Sloma*, Technische Universität Ilmenau, Ilmenau, Germany

It is known that the propagation and the characteristics of a reproduced sound wave in a room is influenced by the room acoustics. So does the perceived sound quality, too? To answer this question it is indispensable to research the quality evaluation of reproduced spatial audio signals in non-standardized rooms and to compare the results with those from standardized listening rooms, in which quality evaluation is usually conducted. Beside the overall quality it is reasonable to assess which parameters of the room acoustics have influence on which quality features. To evaluate the principle influence of different listening rooms on the perception of audio signals, two listening tests are conducted in which three acoustical different rooms are examined. In the first study the aim was to find out if there is an influence on the basic audio quality and five given quality features. This was realized as a single stimulus test. Based on the results a second test was conducted. The approach was adapted from the Open Profiling of Quality. The results from the studies suggest that the influence of the room characteristics are of minor importance on the perception of spatial audio signals.
Convention Paper 9565

08:45

P15-2 Novel Designs for the Audio Mixing Interface Based on Data Visualization First Principles—*Christopher Dewey, Jonathan Wakefield*, University of Huddersfield, Huddersfield, UK

Given the shortcomings of current audio mixing interfaces (AMIs) this study focuses on the development of alternative AMIs based on data visualization first principles. The elementary perceptual tasks defined by Cleveland informed the design process. Two design ideas were considered for pan: using the elementary perceptual tasks "scale" to display pan on either a single or multiple horizontal lines. Four design ideas were considered for level: using "length," "area," "saturation," or "scalable icon" for visualization. Each level idea was prototyped with each pan idea, totaling eight novel interfaces. Seven subjects undertook a usability evaluation, replicating a 16 channel reference mix with each interface. Results showed that "scalable icons" especially on multiple horizontal lines appear to show potential.
Convention Paper 9566

08:45

P15-3 The Method for Generating Movable Sound Source—*Heng Wang*,[1] *Yafei Wu*,[2] *Cong Zhang*[1]
[1]Wuhan Polytechnic University, Wuhan, Hubei, China
[2]Wuhan University, Wuhan, Hubei, China

The rapid development of 3D video inspired the demand for 3D audio technology and products, but the products on the market currently are limited to follow the original stereo or surround sound technology; it is difficult to

produce a three-dimensional sound field audio effect synchronized with 3D video content. The method is based on VBAP principle, derived 3D space movable sound source generating principles and formulas, and implement a method of generate movable sound source in the build 3D audio system. The produced virtual sound source could customize different trajectories and speed in 3D space. On the maximum base of keeping the existing audio equipment, only needing to configure the equipment to the allocation model, we could make the audience feel truly ubiquitous shock audio-visual enjoyment. After field tests, the movement of the movable virtual sound source multitasks generated by this method is obvious in the 3D sound field, and not only provides a good method for generating 3D movable sound source for future research and experimental film-making, but also for 3D audio in home entertainment promotion.
Convention Paper 9567

08:45

P15-4 Graphical Interface Aimed for Organizing Music Based on Mood of Music—*Magdalena Plewa,*[1] *Bozena Kostek,*[1] *Mateusz Bień*[2]
[1]Gdansk University of Technology, Gdansk, Poland
[2]Academy of Music in Kraków, Kraków, Poland

Mood of music is one of the most intuitive criteria for listeners, thus it is used in automated systems for organizing music. This study is based on the emotional content of music and its automatic recognition and contains outcomes of a series of experiments related to building models and description of emotions in music. One-hundred-fifty-four excerpts from 10 music genres were evaluated in the listening experiments using a graphical model proposed by the authors, dedicated to the subjective evaluation of mood of music. The proposed model of mood of music was created in a Max MSP environment. Automatic mood recognition employing SOM and ANN was carried out and both methods returned results coherent with subjective evaluation.
Convention Paper 9568

08:45

P15-5 Subjective Evaluation of High Resolution Audio through Headphones—*Mitsunori Mizumachi,*[1] *Ryuta Yamamoto,*[2] *Katsuyuki Niyada*[3]
[1]Kyushu Institute of Technology, Kitakyushu, Fukuoka, Japan
[2]Digifusion Japan Co., Ltd., Hiroshima, Japan
[3]Hiroshima Cosmopolitan University, Hiroshima, Japan

Recently, high resolution audio (HRA) can be played back through portable devices and spreads across musical genres and generation. It means that most people listen to HRA through headphones and earphones. In this study perceptual discrimination among audio formats including HRA has been invested using a headphones. Thirty-six subjects, who have a variety of audio and musical experience in the wide age range from 20s to 70s, participated in listening tests. Headphone presentation is superior in discriminating the details to the loudspeaker presentation. It is, however, found that the headphone presentation is weak in reproducing presence and reality. Audio enthusiasts and musicians could significantly discriminate audio formats than ordinary listeners in both headphone and loudspeaker listening conditions.
[Also a lecture—see session P10-1]
Convention Paper 9529

08:45

P15-6 Accelerometer Based Motional Feedback Integrated in a 2 3/4" Loudspeaker—*Ruben Bjerregaard, Anders N. Madsen, Henrik Schneider, Finn T. Agerkvist, Michael A. E. Andersen*, Technical University of Denmark, Kgs. Lyngby, Denmark

It is a well known fact that loudspeakers produce distortion when they are driven into large diaphragm displacements. Various methods exist to reduce distortion using forward compensation and feedback methods. Acceleration based motional feedback is one of these methods and was already thoroughly described in the 1960s showing good results at low frequencies. In spite of this, the technique has mainly been used for closed box subwoofers to a limited extent. In this paper design and experimental results for a 2 3/4 " acceleration based motional feedback loudspeaker are shown to extend this feedback method to a small full range loudspeaker. Furthermore, the audio quality from the system with feedback is discussed based on measurements of harmonic distortion, intermodulation distortion, and subjective evaluation.
[Also a lecture—see session P10-6]
Convention Paper 9534

08:45

P15-7 A Headphone Measurement System Covers both Audible Frequency and beyond 20 kHz (Part 2)—*Naotaka Tsunoda, Takeshi Hara, Koji Nageno*, Sony Corporation, Tokyo, Japan

A new scheme consists of measurement by wide range HATS, and the free-field HRTF correction was proposed to enable entire frequency response measurement from audible frequency and higher frequency area up to 140 kHz and for direct comparison with free field loud speaker frequency response. This report supplements the previous report that described system concept by adding ear simulator detail and tips to obtain reliable data with much improved reproducibility.
[Also a lecture—see session P10-2]
Convention Paper 9530

08:45

P15-8 Deep Neural Networks for Dynamic Range Compression in Mastering Applications—*Stylianos Ioannis Mimilakis,*[1] *Konstantinos Drossos,*[2] *Tuomas Virtanen,*[2] *Gerald Schuller*[1]
[1]Fraunhofer Institute for Digital Media Technology (IDMT), Ilmenau, Germany
[2]Tampere University of Technology, Tampere, Finland

The process of audio mastering often, if not always, includes various audio signal processing techniques such as frequency equalization and dynamic range compression. With respect to the genre and style of the audio content, the parameters of these techniques are controlled by a mastering engineer, in order to process the original audio material. This operation relies on musical and perceptually pleasing facets of the perceived acoustic characteristics, transmitted from the audio material under the mastering process. Modeling such dynamic operations, which involve adaptation regarding the audio content, becomes vital in automated applications since it significantly affects the overall performance. In this work we present a system capable of modelling such behavior focusing on the automatic dynamic range compression. It predicts frequency coefficients that allow the dynamic

range compression, via a trained deep neural network, and applies them to unmastered audio signal served as input. Both dynamic range compression and the prediction of the corresponding frequency coefficients take place inside the time-frequency domain, using magnitude spectra acquired from a critical band filter bank, similar to humans' peripheral auditory system. Results from conducted listening tests, incorporating professional music producers and audio mastering engineers, demonstrate on average an equivalent performance compared to professionally mastered audio content.Improvements were also observed when compared to relevant and commercial software.
[Also a lecture—see session P11-3]
Convention Paper 9539

08:45

P15-9 Visualization Tools for Soundstage Tuning in Cars—
Delphine Devallez,[1] *Alexandre Fénières,*[1]
Vincent Couteaux[2]
[1]Arkamys, Paris, France
[2]Telecom ParisTech - Paris, France

In order to improve the spatial fidelity of automotive audio systems by means of digital signal processing, the authors investigated means to objectively assess the spatial perception of reproduced stereophonic sound in car cabins. It implied choosing a convenient binaural microphonic system representative of real listening situations and metrics to analyze interaural time differences under 1.5~kHz in those binaural recordings. Frequency-dependent correlation correctly showed the frequencies at which the fidelity was improved and allowed to quantify the improvement. The time-domain correlation seemed to be a good indicator of the apparent source width, but failed at giving the perceived azimuth of the virtual sound source. Therefore that metric must be refined to be used efficiently during audio tunings.
[Also a lecture—see session P10-7]
Convention Paper 9536

08:45

P15-10 The Difference between Stereophony and Wave Field Synthesis in the Context of Popular Music—*Christoph Hold,*[1] *Hagen Wierstorf,*[2] *Alexander Raake*[2]
[1]Technische Universität Berlin, Berlin, Germany
[2]Technische Universität Ilmenau, Ilmenau, Germany

Stereophony and Wave Field Synthesis (WFS) are capable of providing the listener with a rich spatial audio experience. They both come with different advantages and challenges. Due to different requirements during the music production stage, a meaningful direct comparison of both methods has rarely been carried out in previous research. As stereophony relies on a channel- and WFS on a model-based approach, the same mix cannot be used for both systems. In this study mixes of different popular-music recordings have been generated, each for two-channel stereophony, surround stereophony, and WFS. The focus is on comparability between the reproduction systems in terms of the resulting sound quality. In a paired-comparison test listeners rated their preferred listening experience.
[Also a lecture—see session P10-5]
Convention Paper 9533

08:45

P15-11 Can Bluetooth ever Replace the Wire?—*Jonny McClintock*, Qualcomm Technology International Ltd., Belfast, Northern Ireland, UK

Bluetooth is widely used as a wireless connection for audio applications including mobile phones, media players, and wearables, removing the need for cables. The combination of the A2DP protocol and frame based codecs used in many Bluetooth stereo audio implementations have led to excessive latency and acoustic performance significantly below CD quality. This paper will cover the latest developments in Bluetooth audio connectivity that will deliver CD quality audio, or better, and low latency for video and gaming applications. These developments together with the increased battery life delivered by Bluetooth Smart could lead to the elimination of wires for many applications.
[Also a lecture—see session P11-2]
Convention Paper 9538

Tutorial 8 **Monday, June 6**
08:45 – 09:45 **Room 352A**

USING BINAURAL HEAD RECORDING TECHNIQUES WITH RELATED HD VIDEO FOR MUSIC PRODUCTION

Presenter: **Bob Schulein**, RBS Consultants / ImmersAV Technology, Schaumburg, IL, USA

The presentation will focus on producing music with related video using binaural head capture. Excerpts will be presented from a variety of productions that our group has created. Presenter Bob Schulein will discuss the synergy between binaural audio and related video as related to the artistic goals of productions.

Tutorial 9 **Monday, June 6**
08:45 – 09:45 **Room 351**

IMMERSIVE SOUND DESIGN WITH PARTICLE SYSTEMS

Presenters: **Nuno Fonseca**, ESTG, Polytechnic Institute of Leiria; Sound Particles, Leiria, Portugal
Yohann Bernard, Sound Designer, France

This tutorial will show how to use particle systems to create immersive sound. Using the new "Sound Particles" software, a 3D CGI-based software for audio, particle systems can be used to trigger thousands of sounds over a virtual 3D space, capturing the obtained result in different immersive formats. Based on original audio files taken from sound libraries, highly complex audio scenes can be created, such as a battlefield or an immersive 3D fire. This solution was already used on several movies and TV series and is currently being tested in all the major Hollywood studios.

Student Event and Career Development
Monday, June 6, 08:45 – 09:45
Havane Amphitheatre

STUDENT RECORDING CRITIQUES

Moderator: **Ian Corbett**, Kansas City Kansas Community College, Kansas City, KS, USA; off-beat-open-hats recording & sound reinforcement

The Student Recording Critiques are non-competitive listening sessions, designed for students to listen to their recordings and productions on a world-class playback system, and receive feedback on their work. Students are invited to bring along their mixes and have them critiqued by a panel of renowned industry professionals in order to get pointers as to how they can push their skills to the next level.

It is vital that interested students sign up at the student booth immediately after the first SDA meeting and bring their work on CD, DVD, memory stick, or hard disc as clearly labelled 44.1 kHz WAVE or AIFF files.

Please note that finalists of the Student Recording Competition are excluded to submit their work to the Recording Critiques to give non-finalists a way of having their work assessed.

Tutorial 10 **Monday, June 6**
10:00 – 11:15 **Room 351**

MULTICHANNEL 3D FOR SYNTHESIZER MUSIC WITH VOCALS

Presenter: **Lasse Nipkow**, Silent Work LLC, Zurich, Switzerland

In synthesizer music, vocals are usually strongly compressed, so that they hold their own against the rest of the mix. In a multichannel 3D format the conditions are different: the sounds are no longer all coming from the same direction as the vocals, but from other directions around the audience. If the vocals are only being played back from the center channel, even though they will come through better, there is still the danger they will not merge timbrally with the rest of the music. Hence new ground must be broken at the conception and mixing of a song, in order to attain a satisfactory result. In this presentation, psychoacoustic considerations relating to the aforementioned cases are discussed and practically demonstrated on the basis of audio / video examples.

Tutorial 11 **Monday, June 5**
10:00 – 11:00 **Room 352A**

USER-CENTERED AGILE GAME SOUND DESIGN

Presenter: **Arne Nykanen**

Good game sound or interactive sound design is often determined by the work processes used. This tutorial will take a comprehensive view on how experience from psychoacoustics, industrial design, user-centered design, agile software development, and computer games design can be merged to arrive at good sound design practices. Based on this review, a user-centered agile sound design practice is proposed. The core idea is to organize the design work into short iterations (sprints), listening to the evolving product frequently, and repeatedly throughout the design process, and get users (players) into the loop early on. Demonstrations will be made of how DAWs, software synthesizers, audio programming languages, game engines, and game audio middleware can be used to facilitate such design processes.

This session is presented in association with the AES Technical Committee on Audio for Games

Workshop 12 **Monday, June 6**
10:00 – 11:15 **Room 352A**

PERCEPTUAL EVALUATION INTERFACE DESIGN

Chair: **Brecht De Man**, Queen Mary University of London, London, UK

Panelists: *Jan Berg*, Luleå University of Technology, Piteå, Sweden
Sean Olive, Harman International, Northridge, CA, USA

Listening tests are a key component in a wide range of audio research and development, from loudspeaker construction over audio codecs to emotion in music. In this workshop we zoom in on the interface design of listening test software, covering important but often controversial design considerations including whether to enforce a particular use of scale; the need for anchor and reference; when to assess “preference” instead of “quality”; under which circumstances an uncontrolled, online test is acceptable. Even authorities in the domain of perceptual evaluation often disagree on these key issues that could significantly alter results of perceptual studies—or whether there are any results to speak of. The tradeoffs will be discussed for different applications, stimuli, and subjects.

This session is presented in association with the AES Technical Committee on Perception and subjective Evaluation of Audio Signals

Professional Sound Expo
Monday, June 6, 10:00 – 10:45
PSE Stage

GENELEC SMART ACTIVE MONITORS

Presenter: **Christophe Anet**, Genelec, Finland

Built for today’s fast-paced studio environment, Genelec Smart Active Monitors (SAM) are designed to meet your workflow and help you improve your sound reproduction quality. As space becomes more limited, listening areas are more confined and room acoustic problems more prevalent.

SAM Systems draw on Genelec’s decades of experience and expertise to create an intelligent, flexible network of monitors and subwoofers that can adapt to your requirements. Use Genelec Loudspeaker Manager (GLMTM) 2.0 software to create monitoring systems ranging from traditional stereo to immersive audio setups relying on the proprietary power of GLM AutoCal to create an optimal monitoring environment.

Sponsored Session

Professional Sound Expo
Monday, June 6, 11:00 – 11:45
PSE Stage

ACTIVE VELOCITY ACOUSTIC ABSORPTION FOR LOW FREQUENCIES

Presenter: **Roger Roschnik**, PSI Audio, Yverdon-les-Bains, Switzerland

In recording and listening rooms, low-frequency modal resonances lead to uneven distributions in space and frequency of the acoustic energy, as well as an alteration of the temporal behavior of the original music content. Passive low frequency absorption has critical limitations in terms of volume and/or bandwidth. Other solutions exist to mitigate some of the effects of low-frequency modal resonance but all have other disadvantages associated. This presentation will discuss active velocity acoustic absorption, said to be extremely effective in absorbing low frequencies by modifying the acoustic impedance of air surrounding it. It also has many other advantages such as being effective over a large bandwidth of frequencies, working independently from the sound source, requiring no settings, can be turn on and off and moved from room to room, etc. We will explain how active absorption works and discuss its limitations as well as its advantages.

Tutorial 12 **Monday, June 6**
11:30 – 12:45 **Room 351**

PERCEPTUAL SIGNAL PROCESSING FOR 3D SOUND RECORDING

Presenter: **Hyunkook Lee**, University of Huddersfield, Huddersfield, UK

This tutorial/demo session will explain some of the psychoacoustic principles for vertical stereophonic perception and discuss various conventional and new signal processing methods for 3D recording from a perceptual point of view. The topics will include 3D upmixing and downmixing, rendering of virtual height/overhead loudspeaker image, and binauralization of 3D recording. The talk will be accompanied with practical 9.1 3D sound examples.

Tutorial 13 **Monday, June 6**
11:30 – 12:45 **Havane Amphitheatre**

100 YEARS OF CONDENSER MICROPHONES

Presenter: **Martin Schneider**, Georg Neumann GmbH, Berlin, Germany

On Dec. 20, 1916, E. C. Wente's patent of a "telephone transmitter" based on the capacitive principle was filed. It soon proved its usefulness as the most linear transducer principle and developed into the measurement microphones we use today. It was also quickly adopted for broadcast, recording, and film work and became the standard recording microphone type worldwide in the 1950s. One hundred years after E. C. Wente's patent application, the condenser microphone must be considered as the most important microphone transducer type, being the most linear and conceptually simple at the same time. This tutorial gives a detailed overview of the major developments concerning audio applications in these first 100 years of condenser microphone development.

Workshop 13 **Monday, June 6**
11:30 – 12:45 **Room 352A**

TRENDS AND DEVELOPMENTS FOR AUTOMOTIVE AUDIO

Chair: **Alfred Svobodnik**, Konzept-X GmbH, Karlsruhe, Germany

Panelists: *Timo Esser*, Alpine Electronics R&D Europe GmbH, Stuttgart, Germany
Ulrich Fox
Wolfram Jähn, Audi, Germany
Martin Kreißig, Daimler, Germany
Armin Prommersberger, Harman, Karlsbad, Germany

This workshop will focus on recent changes in requirements for automotive sound systems. Due to the ever increasing importance of ride comfort, topics like NVH (Noise Vibration and Harshness) seem to drive the trends and developments in that specific area of the audio industry. Top experts from industry will discuss these new challenges and what impact it might have on the whole industry.

This session is presented in association with the AES Technical Committee on Automotive Audio

Student Event and Career Development
Monday, June 6, 11:30 – 14:00
Foyer to Technical Program Rooms

EDUCATION/CAREER FAIR

The combined AES 140th Education and Career Fair will match job seekers with companies and prospective students with schools.

Companies

Looking for the best and brightest minds in the audio world? No place will have more of them assembled than the 140th Convention of the Audio Engineering Society. Companies are invited to participate in our Education and Career Fair, free of charge. This is the perfect chance to identify your ideal new hires! All attendees of the convention, students and professionals alike, are welcome to come visit with representatives from participating companies to find out more about job and internship opportunities in the audio industry. Bring your resume!

Schools

One of the best reasons to attend AES conventions is the opportunity to make important connections with your fellow educators from around the globe. Academic 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/institutions to participate. Admission is free and open to all convention attendees.

Session EB3 **Monday, June 6**
11:45 – 14:00 **Room 353**

ENGINEERING BRIEFS—LECTURES

Chair: **Christian Uhle**, Fraunhofer IIS, Erlangen, Germany

11:45

EB3-1 The Aerodynamics Phenomena of a Particular Bass-Reflex Port—*Victor Manuel Garcia-Alcaide, Sr.,[1] Sergi Palleja-Cabre,1 R. Castilla,[1] P. J. Gamez-Montero,[1] Jordi Romeu,[1] Teresa Pamies,[1] Joan Amate,[2] Natalia Milan[2]*
[1]Universitat Politècnica de Catalunya, Barcelona, Spain
[2]Amate Audio S.L., Barcelona, Spain

The aim of this paper is to study the aerodynamics phenomena of a particular bass-reflex port that causes unwanted noise in the audible frequency range. After discarding structural and mechanical vibration issues, the hypothesis that vortex shedding could be the source of the noise has been considered. Experimental and numerical evidences of the vortex, an analysis of its noise and the similarities between the real performance and the simulated one are presented. The simulations have been performed with axisymmetric geometries with the open source OpenFOAM toolbox. Additionally, three different experiments were carried out. First, acoustic signal experiments were done to analyze the response of the bass-reflex ports. Second, a mechanical vibration was tested in order to discard this source of noise. A good agreement has been found between numerical and experimental results, especially in the frequency band of the detected noise, around 1200 Hz. The presented CFD approach has proved a useful and cost-effective tool to face this kind of phenomena.
Engineering Brief 257
[eBrief presented by Joan Amate]

12:00

EB3-2 A Novel 32-Speakers Spherical Source—*Angelo Farina, Lorenzo Chiesi*, University of Parma, Parma, Italy

The construction and test of a novel compact spherical source equipped with 32 individually driven 2” loudspeakers is presented. The new sound source is designed for making room acoustics measurements, emulating the directivity pattern of various music instruments or human talkers and singers. The 32 signals feeding the loudspeakers can be obtained by three different approaches: a set of High Order Ambisonics coefficients computed for emulating the polar pattern of a fixed directivity source a set of SPS (Spatial PCM Sampling) signals recorded around a real source, employing a corresponding set of 32 microphones placed on a sphere surrounding the real source, a matrix of FIR filters, designed employing a mathematical theory almost identical to the one developed for creating virtual microphones from a spherical microphone array [1]. The presentation will show details of the construction of the new loudspeaker array, and the results of the first tests performed for evaluating the capability of creating arbitrary polar radiation patterns.
Engineering Brief 258

12:15

EB3-3 Distracting Noise—*Thomas Sporer,*[1] *Tobias Clauß,*[1] *Nicolas Pachatz,*[2] *Clemens Müller,*[2] *Matthias-Fritz Melzer,*[2] *Judith Liebetrau*[1]
[1]Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany
[2]Technical University of Ilmenau, Ilmenau, Germany

Noise in domestic and work environments is usually measured based on noise power. This is not reflecting the fact that temporal and spectral structure of the noise, but also the activity of the test subject influences the annoyance. In addition there is a difference between artificial noise signals and noise signals that probably have a meaning to the listener. In this study 15 assessors evaluated the perception of 23 natural noise stimuli at four different levels in two different situations. The situations are spatial recordings of a library and a canteen. The test subjects are not focusing on listening but on tasks but told to indicate when noise is distorting their activities.
Engineering Brief 259

12:30

EB3-4 Noise-Robust Speech Emotion Recognition Using Denoising Autoencoder—*Hun Kyu Ha, Nam Kyun Kim, Woo Kyeong Seong, Hong Kook Kim*, Gwangju Institute of Science and Tech (GIST), Gwangju, Korea

In this paper, a method of noise-robust speech emotion recognition under music noises is proposed by using a denoising autoencoder (DAE) and a support vector machine (SVM). The proposed method first trains a DAE by using emotional speech signals corrupted by music noises. Then, the output values from a middle layer of the DAE are used as speech features. Next, an SVM is trained to classify emotions using the DAE features. The performance of the proposed method is compared with that of a conventional SVM classifier. Consequently, it is shown that the proposed method relatively improves the overall emotion recognition rate by 9.76% under music noise conditions, compared to the conventional method.
Engineering Brief 260

12:45

EB3-5 Non-Intrusive Rumble Filtering by VLF Crossfeed with High Filter Slopes—*Douglas Self*, The Signal Transfer Company, London, UK

Vinyl discs create subsonic anti-phase signals because they are never perfectly flat and cause vertical stylus movement. This is often made worse by cartridge-arm resonance, giving amplitudes peaking around 10 Hz and requiring 40 dB of attenuation to reduce them to the vinyl noise floor. A conventional rumble filter needs very steep slopes to do this without unduly affecting the bottom of the audio band at 20 Hz. L-R crossfeed at low frequencies cancels the anti-phase signals, converting bass information to mono. This is not a new idea but has never caught on, probably because in published implementations the anti-phase filtering slope always comes out as –6dB/octave, no matter what order of lowpass filter is used to control the crossfeed. It is demonstrated that time-correction of the lowpass filter group delay with simple allpass filtering gives a much steeper slope of –18dB/octave for 2nd, 3rd, and 4th-order Butterworth filters, and intrusion into the audio band is minimized; this is believed novel. A practical design using 2nd-order filters was built and measured and gave the desired results.
Engineering Brief 261

13:00

EB3-6 The Misunderstood Transformer: “The Answer Lies in the Flux!”—*Michael Turner*, Nidec Motor Corporation, Harrogate, UK

Whether used for power supply or signal interfacing, transformers are a key component of audio equipment. What is surprising is the extent to which these apparently simple devices are misunderstood. This engineering brief seeks to dispel some of the more common myths, to clarify the relationships between voltage, current, flux, and saturation and to thereby assist with proper design, selection, and application of (mainly) power transformers
Engineering Brief 262

13:15

EB3-7 Designing a Laboratory for Immersive Arts—*Christopher Keyes,* Hong Kong Baptist University, Kowloon, Hong Kong

This brief gives an overview of a facility dedicated to 3D sound and multi-screen video. It houses a control room and a theater with the region’s only 24.2 channel sound system and 5 permanent HD video screens. At roughly 200 m3 it is a relatively small facility but has many uses. In its construction we were afforded a wide range of possibilities for spatial configurations and equipment choice. It is hoped that presenting some detail on these design decisions, including choices available and ultimately implemented, may be of use for readers planning and budgeting their own facilities.
Engineering Brief 263

13:30

EB3-8 Design and Implementation of a Low-Latency, Lightweight, High-Performance Voice Interface Front-End—*Thierry Heeb,*[1] *Andrew Stanford-Jason,*[2] *Tiziano Leidi*[1]
[1]ISIN-SUPSI, Manno, Switzerland
[2]XMOS Ltd., Bristol, UK

Smart voice interfaces are the enabler of a new generation of consumer products such as network connected, voice enabled personal assistants. These are based on distributed architectures where voice is captured and pre-processed locally before being sent to remote servers for se-

mantic analysis and response generation. A key element to achieve lowest cost and the best natural speech user experience is to keep latency to a minimum. This eBrief presents a lightweight, high-performance voice interface front-end software framework capable of handling multiple PDM microphones and integrating PDM to PCM conversion, high-resolution inter-channel delay, decimation, signal correction, and optional output data framing. The software forms a complete smart voice interface front-end running on the XMOS xCORE-200 architecture and achieving very low latency.
Engineering Brief 264

13:45

EB3-9 Multiphysical Simulation Methods for Loudspeakers—Nonlinear CAE-Based Simulations—*Alfred Svobodnik,[1] Roger Shively,[2] Marc-Olivier Chauveau,[3] Tommaso Nizzoli[1]*
[1]Konzept-X GmbH, Karlsruhe, Germany
[2]JJR Acoustics, LLC, Seattle, WA, USA
[3]Moca Audio, Tours, France

This is the third in a series of papers on the details of loudspeaker design using multiphysical computer aided engineering simulation methods. In this paper the simulation methodology for accurately modeling the nonlinear electromagnetics and structural dynamics of a loudspeaker will be presented. Primarily, the calculation of nonlinear force factor Bl(x), nonlinear inductance Le(x), and stiffness Kms(x) in the virtual world will be demonstrated. Finally, results will be presented correlating the simulated model results to the measured physical parameters. From that, the important aspects of the modeling that determine its accuracy will be discussed.
Engineering Brief 265

Professional Sound Expo
Monday, June 6, 12:00 – 12:45
PSE Stage

SOUND QUALITY IN DIGITAL AUDIO INTERFACES

Presenter: **Jody Thorne**, Prism Sound

Sound Quality in Digital Audio Interfaces—What do we mean, how do we achieve it. Dispelling the common myths and misconceptions.

Session P16 **Monday, June 6**
12:30 – 14:30 **Room 352B**

RECORDING AND PRODUCTION TECHNIQUES

Chair: **Sonja Krstic**, School of Electrical Engineering and Computer Science, Belgrade, Serbia

12:30

P16-1 Microphone Array Design Applied to Complete Hemispherical Sound Reproduction—From Integral 3D to Comfort 3D—*Michael Williams*, Sounds of Scotland, Le Perreux sur Marne, France

This paper describes the parameters that need to be taken into account in the design of a 13 channel microphone array recording system for reproduction also in a 13 loudspeaker hemispherical configuration. Both the microphone array and the loudspeaker array use 8 channels in the horizontal reference plane, 4 channels in the 45° elevation plane, and a Zenith channel at the top (90° elevation). This paper will also describe the various stages of advancement to complete 3D coverage (Integral 3D), and the logical development of this type of array to a new format proposition—the 16 channel "Comfort 3D format."
Convention Paper 9569

13:00

P16-2 Object-Based Audio Recording Methods—*Jean-Christophe Messonnier,[1] Jean-Marc Lyzwa,[1] Delphine Devallez,[2] Catherine De Boisheraud[1]*
[1]CNSMDP Conservatoire de Paris, Paris, France
[2]Arkamys, Paris, France

The new ADM standard enables to define an audio file as object-based audio. Along with many other functionalities, the polar coordinates can be specified for each audio object. An audio scene can therefore be described independently of the reproduction system. This means that an object-based recording can be rendered on a 5.1 system, a binaural system, or any other system. In the case of a binaural system, it also gives the opportunity to interact with the audio content, as a headtracker can be used to follow the movements of the listener's head and change the binaural rendering accordingly. This paper describes how such an object-based recording can be achieved.
[Also a poster—see session P19-6]
Convention Paper 9570

13:30

P16-3 A Further Investigation of Echo Thresholds for the Optimization of Fattening Delays—*Michael Uwins,[1] Dan Livesey[2]*
[1]University of Huddersfield, Huddersfield, UK
[2]Confetti College, Nottingham Trent University, Nottingham, UK

Since the introduction of stereophonic sound systems, mix engineers have developed and employed numerous artificial methods in order to enhance their productions. A simple yet notable example is the effect commonly known as "fattening," where a mono signal is cloned, delayed, and then panned to the opposite side of the stereo field. The technique can improve a sound's prominence in the mix by increasing its overall amplitude while creating a pseudostereo image and is a consequence of a renowned psychoacoustic phenomenon, the "precedence effect." The aim of this investigation was to build upon previous accepted studies, conducting further experiments in order to produce refined estimates for echo thresholds for elements common to a multi-track music production. This investigation obtained new estimates of echo thresholds and fattening delay times, for a variety of isolated instrumental and vocal recordings, as perceived by a sample population of trained mix engineers. The study concludes that current recommendation for delay times used to create fattening effects should be refined, taking into account not only those features of the but also the consequences of temporal and spectral masking, when applied in the context of a multitrack
[Also a poster—see session P19-2]
Convention Paper 9571

14:00

P16-4 An Investigation into the Sonic Signature of Three Classic Dynamic Range Compressors—*Austin Moore, Rupert Till, Jonathan Wakefield*, University of Huddersfield, Huddersfield, UK

Dynamic range compression (DRC) is a much-used process in music production. Traditionally the process was implemented in order to control the dynamic range of program material to minimize the potential of overloading recording devices. However, over time DRC became a process that was applied more as a creative effect and less as a preventative measure. In a professional recording environment it is not uncommon for engineers to have access to several different types of DRC unit, each with their own purportedly unique sonic signature. This paper investigates the differences between three popular vintage dynamic range compressors by conducting a number of measurements on the devices. The compressors were tested using: THD measurements, tone bursts, and objective analysis of music-based material using spectrum analysis and audio feature extraction.
Convention Paper 9572

Tutorial 14 **Monday, June 6**
13:00 – 14:00 **Room 351**

IMMERSED BY EMOTIONS: FILM MUSIC IN AURO3D

Presenter: **Patrick Lemmens**, Galaxy Studios, Mol, Belgium

The emotions in a film are crafted to a large extent by the music; a film without music has a cold and sterile feeling to it, while the music that exclusively accompanies a silent film is able to affect the feelings of the audience in practically any way desired by the composer. The third dimension of sound reproduction—which is added to the conventional 5.1 or 7.1 surround sound in cinema by using the different Auro3D formats—provides a fully immersive experience to the audience and gives them the impression of being in the middle of the action; especially the music plays an important role in establishing this immersive feeling, as the 3D format offers extended possibilities for creating the desired emotions to the composer and the music engineer.

Workshop 14 **Monday, June 6**
13:00 – 14:30 **Havane Amphitheatre**

THIS IS A MIX! THIS IS A MASTER!

Chair: **Ian Corbett**, Kansas City Kansas Community College, Kansas City, KS, USA; off-beat-open-hats recording & sound reinforcement

Panelists: *Mandy Parnell*, Black Saloon Studios, London, UK
Darcy Proper, Darcy Proper Mastering, Eemnes, The Netherlands; Wisseloord Studios, Hilversum, The Netherlands
Sander Van der Heide, The Saint of Sound, Soest, The Netherlands

Whether you are a student, home studio or project studio user, or someone entering the professional industry, most of the music mixes you hear and try to emulate have been professionally mastered. Too many novices try to recreate a "mastered" sound in their mix. This is undesirable and limits what the mastering engineer can do. Join our panel of mastering engineers as they present some "off-the-console" mixes, discuss what they did to the mix, play the resulting master, and discuss some other common issues they see in some of the material sent to them to master.

This session is presented in association with the AES Technical Committee on Recording Technology and Practices

Workshop 15 **Monday, June 6**
13:00 — 14:30 **Room 352A**

PERCEPTUAL ASSESSMENT OF SPATIAL SOUND: THE CASE OF BINAURAL REPRODUCTION

Chair: **Rozenn Nicol**, Orange Labs, France Telecom, Lannion, France

Panelists: *Areti Andreopoulou*, LIMSI-CNRS, Orsay, France
Emmanuel Ponsot, STMS Lab (Ircam, CNRS, UPMC), Paris, France; Radio France, Paris, France
Hagen Wierstorf, Technische Universität Ilmenau, Ilmenau, Germany
James Woodcock, University of Salford, Salford, Greater Manchester, UK

3D audio is now available both in professional and mass market products. Our listening experience is thus enhanced by the spatial dimension, which calls for revisiting the methods and tools used to assess the perception of sound reproduction. The workshop will discuss this issue in the particular case of binaural rendering which raises some specific questions: interaction between timbre and spatial information, HRTF customization, headphone reproduction., etc. This will be the opportunity to cross methods used for binaural listening with conventional methods of perceptual assessment.

This session is presented in association with the AES Technical Committees on Spatial Audio and Perception and Subjective Evaluation of Audio Signals

Professional Sound Expo
Monday, June 6, 13:00 – 13:45
PSE Stage

RIBBON MICROPHONES

Presenter: **Sammy Rothman**

What are ribbon microphones and how do you use them? Sammy Rothman answers these questions and more in "AEA Ribbon Mics: Fix It in the Mic," which delves into all things ribbons including best miking practices and how ribbons mics work. The discussion will address the differences between ribbon and condenser microphones, between active and passive ribbon microphones, and between various approaches to ribbon microphone design.

Professional Sound Expo
Monday, June 6, 14:00 – 14:45
PSE Stage

ACOUSTIC ROOM TREATMENT

Presenter: **Lukas Rimbach**

Why do you need room treatments? This session will discuss the science and practice of treating rooms to optimize acoustic performance.

Session P17 **Monday, June 6**
14:45 – 16:45 **Room 353**

RENDERING SYSTEMS

Chair: **Libor Husnik**, Czech Technical University, Prague, Czech Republic

14:45

P17-1 Optimized Spatial Layout for Virtual Spatial Audio Conference—*Liyun Pang, Pablo Hoffmann*, Huawei Technologies Düsseldorf GmbH, Düsseldorf, Germany

This paper is related to three-dimensional (3D) audio signal processing for virtual spatial audio conference applications. The core idea is to provide users of a virtual spatial audio conference with recommendations for the optimal spatial layout (spatial arrangement) of participants where optimal implies maximal speech intelligibility. That is, the listener's ability to understand speech from an individual speaker in a multi-speaker scenario is enhanced. The idea combines information of the individual's voice together with directional audio information to estimate the speech intelligibility of candidate spatial layouts (spatial arrangement). The candidate layout that provides the best speech intelligibility estimate is then selected.
Convention Paper 9573

15:15

P17-2 A Listener Adaptive Optimal Source Distribution System for Virtual Sound Imaging—*Marcos F. Simon Galvez, Takeshi Takeuchi, Filippo Maria Fazi*, University of Southampton, Southampton, Hampshire, UK

This paper describes the use of an optimal source distribution loudspeaker array for binaural reproduction. In this paper the device is made adaptive to the listener's position. This is obtained by using a fixed set of crosstalk cancellation filters created for a central listening position plus a listener-position dependent delay network, which varies the delay of each loudspeaker unit to maximize the cross-talk cancellation response at the listener's position. The paper introduces the formulation required for the adaptive control and simulated results predicting the performance of the device for symmetrical and asymmetrical listening positions. It is also shown how the proposed formulation has been implemented on a Sherwood S7 OPSODIS soundbar.
Convention Paper 9574

15:45

P17-3 Multizone Soundfield Reproduction with Virtual Elevations Using a Linear Loudspeaker Array—*Wenyu Jin, Milos Markovic, Liyun Pang*, Huawei Technologies Duesseldorf GmbH, Düsseldorf, Germany; European Research Center, Munich, Germany

Multizone soundfield reproduction has drawn the researchers' attention recently. The paper introduces a soundfield rendering system for a simultaneous reproduction of sound sources with different elevations over multiple listening areas using a linear loudspeaker array. A novel method based on the usage of HRTF (Head Related Transfer Function) spectral elevation cues in conjunction with a horizontal multi-zone sound rendering system is proposed. The proposed method is in a dual-band manner and it aims to accurately reproduce the desired 3-D elevated sound with the consideration of HRTF cues within the selected listening zones, while also minimizing the sound leakage to the targeted quiet zones over the entire audio frequency range. A listening test is conducted and the results confirm the feasibility of simultaneous multizone soundfield rendering with different elevation using a single 2-D loudspeaker array.
Convention Paper 9575
[Paper presented by Liyun Pang]

16:15

P17-4 Characterization of the Acoustical Directivity of a Speaker on a Sound Bar Enclosure: A Comparison between Measurements, Boundary Element Method, and a Spheroidal Analytical Model—*Vincent Roggerone,*[1] *Xavier Boutillon,*[1] *Etienne Corteel*[2]
[1]Ecole Polytechnique, Palaiseau, France
[2]Sonic Emotion Labs, Paris, France

The directivity of a sound bar of slender shape is analyzed. Measurements are compared to the results of a boundary element method. A good agreement is obtained in the low-mid frequency range. In order to reduce the computing time, a geometrical approximation based on a spheroidal analytical model is also considered. This approximation holds up to a certain frequency. The spheroid shape produces a more regular sound field pattern.
Convention Paper 9576

Session P18 **Monday, June 6**
14:45 – 15:45 **Room 352B**

HUMAN FACTORS AND INTERFACES

Chair: **Sorgun Akkor**, STD Ltd, Istanbul, Turkey

14:45

P18-1 Robustness of Speaker Recognition from Noisy Speech Samples and Mismatched Languages—*Ahmed Al-Noori, Francis F. Li, Philip J. Duncan*, University of Salford, Salford, Greater Manchester, UK

Speaker recognition systems can typically attain high performance in ideal conditions. However, significant degradations in accuracy are found in channel-mismatched scenarios. Non-stationary environmental noises and their variations are listed at the top of speaker recognition challenges. Gammtone frequency cepstral coefficient method (GFCC) has been developed to improve the robustness of speaker recognition. This paper presents systematic comparisons between performance of GFCC and conventional MFCC-based speaker verification systems with a purposely collected noisy speech data set. Furthermore, the current work extends the experiments to include investigations into language independency features in recognition phases. The results show that GFCC has better verification performance in noisy environments than MFCC. However, the GFCC shows a higher sensitivity to language mismatch between enrollment and recognition phase.
Convention Paper 9577

15:15

P18-2 A Reliable Singing Voice-Driven MIDI Controller Using Electroglottographic Signal—*Kostas Kehrakos,*[1] *Christos Chousidis,*[2] *Spyros Kouzoupis*[1]
[1]TEI of Crete, E. Daskalaki, Greece
[2]Brunel University, Uxbridge, London, UK

Modern synthesizers and software created a need for encoding musical performance. However, human singing voice, which is the dominant means of musical expres-

sion, lacks a reliable encoding system. This is because of the difficulties we face to extract the necessary control information from its heavy harmonic content. In this paper a novel singing voice MIDI controller, based on the Electroglottography is proposed. Electroglottographic signal has lower harmonic content than audio, but it contains enough information to describe music expression. The system uses autocorrelation for pitch extraction and a set of supplementary algorithms to provide information for dynamics and duration. The results show that the proposed system can serve as a platform for the development of reliable singing voice MIDI controllers.
Convention Paper 9578

Session P19 **Monday, June 6**
14:45 – 16:45 **Foyer**

POSTERS: PERCEPTION PART 2, AUDIO SIGNAL PROCESSING PART 3, AND RECORDING AND PRODUCTION TECHNIQUES

14:45

P19-1 Two Alternative Minimum-Phase Filters Tested Perceptually—*Robert Mores, Ralf Hendrych*, University of Applied Sciences Hamburg, Hamburg, Germany

A widely used method for designing minimum phase filters is based on the real cepstrum (Oppenheim, 1975). An alternative method is proposed for symmetric FIR filters that flips the filter's "left side" around the central coefficient to the "right side" using a sinus ramp of perceptually irrelevant duration. The resulting phase is nearly minimal and nearly linear. The method is applied to impulse responses. Perception tests use original sound samples (A), samples processed by real-cepstrum-based minimum phase filters (B), and samples processed by the proposed method (C). The tests reveal that for impulsive sound samples the perceived dissimilarity between A and C is smaller than the dissimilarity between A and B suggesting that the alternative method has some potential for sound processing.
[Also a lecture—see session P13-2]
Convention Paper 9554

14:45

P19-2 A Further Investigation of Echo Thresholds for the Optimization of Fattening Delays—*Michael Uwins,[1] Dan Livesey[2]*
[1]University of Huddersfield, Huddersfield, UK
[2]Confetti College, Nottingham Trent University, Nottingham, UK

Since the introduction of stereophonic sound systems, mix engineers have developed and employed numerous artificial methods in order to enhance their productions. A simple yet notable example is the effect commonly known as "fattening," where a mono signal is cloned, delayed, and then panned to the opposite side of the stereo field. The technique can improve a sound's prominence in the mix by increasing its overall amplitude while creating a pseudostereo image and is a consequence of a renowned psychoacoustic phenomenon, the "precedence effect." The aim of this investigation was to build upon previous accepted studies, conducting further experiments in order to produce refined estimates for echo thresholds for elements common to a multi-track music production. This investigation obtained new estimates of echo thresholds and fattening delay times, for a variety of isolated instrumental and vocal recordings, as perceived by a sample population of trained mix engineers. The study concludes that current recommendation for delay times used to create fattening effects should be refined, taking into account not only those features of the but also the consequences of temporal and spectral masking, when applied in the context of a multitrack mix.
[Also a lecture—see session P16-3]
Convention Paper 9571

14:45

P19-3 Extraction of Anthropometric Measures from 3D-Meshes for the Individualization of Head-Related Transfer Functions—*Manoj Dinakaran,[1,2] Peter Grosche,[1] Fabian Brinkmann,[2] Stefan Weinzierl[2]*
[1]Huawei Technologies, European Research Center, Munich, Germany
[2]Technical University of Berlin, Berlin, Germany

Anthropometric measures are used for individualizing head-related transfer functions (HRTFs) for example, by selecting best match HRTFs from a large library or by manipulating HRTF with respect to anthropometrics. Within this process, an accurate extraction of anthropometric measures is crucial as small changes may already influence the individualization. Anthropometrics can be measured in many different ways, e.g., from pictures or anthropometers. However, these approaches tend to be inaccurate. Therefore, we propose to use Kinect for generating individual 3D head-and-shoulder meshes from which anthropometrics are automatically extracted. This is achieved by identifying and measuring distances between characteristics points on the outline of each mesh and was found to yield accurate and reliable estimates of corresponding features. In our experiment, a large set of anthropometric measures was automatically extracted for 61 subjects and evaluated in terms of a cross-validation against the manually extracted correspondent.
Convention Paper 9579

14:45

P19-4 Methods of Phase-Aligning Individual Instruments Recorded with Multiple Microphones during Post-Production—*Bartlomiej Kruk,[1] Aleksander Sobecki[2]*
[1]State Higher Vocational School in Nysa, Nysa, Poland
[2]Wroclaw University of Technology, Wroclaw, Poland

When recording any instrument, like a guitar cabinet or a drum set with a multi-microphone setup, phase plays a key role in shaping the sound. Despite the importance, phase is often overlooked during the recording process because of lack of time or experience. Then during mixing stage engineers tend to use equalizers and compressors to correct issues that might originate in signals not being well time-aligned. Phase measuring tools like goniometers are widely used by mastering engineers to diagnose any phase related issues in a mix, yet their usefulness in shaping sounds of individual instruments is vastly overlooked. The main aim of this paper is to present and analyze easy phase-aligning methods.
Convention Paper 9580

14:45

P19-5 Wireless Sensor Networks for Sound Design: A Summary on Possibilities and Challenges—*Felipe Reinoso Carvalho, Abdellah Touhafi, Kris Steenhaut*, Vrije Universiteit Brussel, Pleinlaan, Brussels

This article presents opportunities of using Wireless Sensor Networks (WSNs) equipped with acoustic sensors as tools for sound design. We introduce the technology, examples considered as State of the Art, and several potential applications involving different profiles of sound design. The importance of adding real-time audio-messages into sound design is considered a main issue in this proposal. Actual technological situation and challenges are here discussed. The usage of WSNs for sound design is plausible, although technological challenges demand strong interaction between sound designers and WSN developers.
Convention Paper 9581

14:45

P19-6 Object-Based Audio Recording Methods—
Jean-Christophe Messonnier,[1] *Jean-Marc Lyzwa,*[1] *Delphine Devallez,*[2] *Catherine De Boisheraud*[1]
[1]CNSMDP Conservatoire de Paris, Paris, France
[2]Arkamys, Paris, France

The new ADM standard enables to define an audio file as object-based audio. Along with many other functionalities, the polar coordinates can be specified for each audio object. An audio scene can therefore be described independently of the reproduction system. This means that an object-based recording can be rendered on a 5.1 system, a binaural system, or any other system. In the case of a binaural system, it also gives the opportunity to interact with the audio content, as a headtracker can be used to follow the movements of the listener's head and change the binaural rendering accordingly. This paper describes how such an object-based recording can be achieved.
[Also a paper—see session P16-2]
Convention Paper 9570

14:45

P19-7 The Harmonic Centroid as a Predictor of String Instrument Timbral Clarity—*Kirsten Hermes, Tim Brookes, Chris Hummersone*, University of Surrey, Guildford, Surrey, UK

Spectrum is an important factor in determining timbral clarity. An experiment where listeners rate the changes in timbral clarity resulting from spectral equalization (EQ) can provide insight into the relationship between EQ and the clarity of string instruments. Overall, higher frequencies contribute to clarity more positively than lower ones, but the relationship is program-item-dependent. Fundamental frequency and spectral slope both appear to be important. Change in harmonic centroid (or dimensionless spectral centroid) correlates well with change in clarity, more so than octave band boosted/cut, harmonic number boosted/cut, or other variations on the spectral centroid.
[Also a paper—see session P13-5]
Convention Paper 9557

14:45

P19-8 Subjective Listening Tests for Preferred Room Response in Cinemas—Part 2: Preference Test Results —*Linda A. Gedemer*, University of Salford, Salford, UK; Harman International, Northridge, CA, USA

SMPTE and ISO have specified near identical in-room target response curves for cinemas and dubbing stages. However, to this author's knowledge, to date these standards have never been scientifically tested and validated with modern technology and measurement techniques. For this reason it is still not known if the current SMPTE and ISO in-room target response curves are optimal or if better solutions exist. Using a Binaural Room Scanning system for room capture and simulation, various seating positions in three cinemas were reproduced through headphones for the purpose of conducting controlled listening experiments. This system used a binaural mannequin equipped with a computer-controlled rotating head to accurately capture binaural impulse responses of the sound system and the listening space which are then reproduced via calibrated headphones equipped with a head-tracker. In this way controlled listening evaluations can be made among different cinema audio systems tuned to different in-room target responses. Results from a MUSHRA-style preference test are presented.
[Also a lecture—see session P13-3]
Convention Paper 9555

14:45

P19-9 Exploiting Envelope Fluctuations to Enhance Binaural Perception—*G. Christopher Stecker*, Vanderbilt University School of Medicine, Nashville, TN, USA

A review of recent and classic studies of binaural perception leads to the conclusion that envelope fluctuations, such as sound onsets, play a critical role in the sampling of spatial information from auditory stimuli. Specifically, listeners' perception of sound location corresponds with the binaural cues (interaural time and level differences) that coincide with brief increases in sound amplitude, and disregards binaural cues occurring at other times. This discrete, envelope-triggered sampling of binaural information can be exploited to enhance spatial perception of synthesized sound mixtures, or to facilitate the localization of mixture components.
[Also a lecture—see session P13-1]
Convention Paper 9553

14:45

P19-10 Comparison of Simple Self-Oscillating PWM Modulators—*Nicolai Dahl, Niels Elkjær Iversen, Arnold Knott, Michael A. E. Andersen*, Technical University of Denmark, Kgs. Lyngby, Denmark

Switch-mode power amplifiers has become the conventional choice for audio applications due to their superior efficiency and excellent audio performance. These amplifiers rely on high frequency modulation of the audio input. Conventional modulators use a fixed high frequency for modulation. Self-oscillating modulators do not have a fixed modulation frequency and can provide good audio performance with very simple circuitry. This paper proposes a new type of self-oscillating modulator. The proposed modulator is compared to an already existing modulator of similar type and their performances are compared both theoretically and experimentally. The result shows that the proposed modulator provides a higher degree of linearity resulting in around 2% lower Total Harmonic Distortion (THD).
[Also a lecture—see session P14-5]
Convention Paper 9562

14:45

P19-11 Spatial Multi-Zone Sound Field Reproduction Using Higher-Order Loudspeakers in Reverberant Rooms—
Keigo Wakayama, Hideaki Takada, NTT Service Evolu-

tion Laboratories, Kanagawa, Japan

We propose a method for reproducing multi-zone sound fields in a reverberant room using an array of higher-order loudspeakers. This method enables sparse arrangement of loudspeakers and reproduction of independent sound fields for multiple listeners without the need for headphones. For multi-zone reproduction, global sound field coefficients are obtained using translation operator. By using the coefficient of the room transfer function measured or simulated with an extension of the image-source method, the loudspeakers' coefficients are then calculated with the minimum norm method in the cylindrical harmonic domain. From experiments of two-zone and three-zone examples, we show that there was a 2N + 1-fold decrease in the number of Nth-order loudspeakers for accurate reproduction with the proposed method compared to conventional methods.
Also a poster—see lecture session P14-3
Convention Paper 9560

14:45

P19-12 Stereo Panning Law Remastering Algorithm Based on Spatial Analysis—*François Becker, Benjamin Bernard*, Medialab Consulting SNP, Monaco, Monaco

Changing the panning law of a stereo mixture is often impossible when the original multitrack session cannot be retrieved or used, or when the mixing desk uses a fixed panning law. Yet such a modification would be of interest during tape mastering sessions, among other applications. We present a frequency-based algorithm that computes the panorama power ratio from stereo signals and changes the panning law without altering the original panorama.
Also a lecture—see session P7-6
Convention Paper 9523

14:45

P19-13 Non-Linear Extraction of a Common Signal for Upmixing Stereo Sources—*François Becker, Benjamin Bernard*, Medialab Consulting SNP, Monaco, Monaco

In the context of a two- to three-channel upmix, center channel derivations fall within the field of common signal extraction methods. In this paper we explore the pertinence of the performance criteria that can be obtained from a probabilistic approach to source extraction; we propose a new, non-linear method to extract a common signal from two sources that makes the implementation choice of deeper extraction with a criteria of information preservation; and we provide the results of preliminary listening tests made with real-world audio materials.
Also a lecture—see session P7-7
Convention Paper 9524

Tutorial 15 **Monday, June 6**
14:45 – 15:30 **Room 352A**

CREATING A VIRTUAL ACOUSMONIUM IN UNITY5

Presenter: **Christine Webster**, Soundwebster, ENSAD-LAB-EN-ER, Paris, France

The tutorial will show how it is possible to create an immersive 64 channel acousmonium in a 3D graphic scene with Unity5. Examples will come from Christine Webster's last EA virtual 3D immersive project "Empty Room." http://spatialmedia.ensadlab.fr/projet-empty-room/

This session is presented in association with the AES Technical Committee on Spatial Audio

Tutorial 16 **Monday, June 6**
14:45 – 15:45 **Room 351**

THE EVOLUTION OF STEREO TO SURROUND SOUND—WHAT DO THE VARIOUS PRESENT DAY FORMATS REALLY BRING TO THE SURROUND SOUND EXPERIENCE

Presenter: **Michael Williams**, Sounds of Scotland, Le Perreux sur Marne, France

This presentation will explain the configuration of the different formats, their faults, and the possible improvements that are possible. This presentation will also allow the listener to compare different formats - stereo with triphony, quadriphony with multichannel, multichannel with Blue ray, quadriphony with Blue ray, etc. This comparison process is the result of simultaneous multiformat recordings and uses fast cross-fading between the different formats. This fast fade technique is used in order to improve perception of the advantages and faults of different format configurations. This presentation will use a basic eight channel horizontal loudspeaker configuration which is the basic Blue ray loudspeaker configuration plus a back channel. This experience has already been presented to the French Section of the AES in a meeting in Lyon. As the comparison point of each format is produced, a unique method of synchronization of powerpoint slides will inform the listener of the structure of each individual reproduction format.

Special Event
Monday, June 6, 14:45 – 16:15
Havane Amphitheatre

EUROPEAN AUDIO EDUCATION: POSSIBILITIES AND CHALLENGES FOR STUDENTS AND EDUCATORS

Moderator: **Kyle P. Snyder**, Ohio University, School of Media Arts & Studies, Athens, OH, USA

Panelists: *Mark Drews, Nyssim Lefford, Theresa Leonard, Ufuk Onen, Denis Vautrin, Nicolas Veastegu*

In this roundtable discussion featuring invited experts from across the continent, we will briefly discuss a variety of topics germane to European audio education including the following: • Educational opportunities that are unique to Europe with a focus on exchange / internship opportunities. • Challenges unique to both European educators and students including suggestions from both the panel and audience.

Professional Sound Expo
Monday, June 6, 15:00 – 15:45
PSE Stage

MICROPHONE BASICS

Presenter: **John Willett**, Sound-Link ProAudio, Bicester, Oxfordshire, UK; Circle Sound Services, Bicester, Oxfordshire, UK

Microphones are the very first link in the recording chain, so it's important to understand them to use them effectively. This presentation will explain the differences between different types of microphones; explain polar-patterns and directivity, proximity effect relative recording distances and a little about room acoustics. The author found that many of these "golden nuggets" helped him greatly when he first understood them and hopes they will help you too.

Tutorial 17 **Monday, June 6**
15:45 – 16:45 **Room 352A**

BINAURAL AND AUDIOVISUAL CONTENT

Presenters: **Delphine Devallez**, Arkamys, Paris, France
Lidwine Ho, France télévisions, Paris, France

Broadcasters are now facing a digital and wireless world of interactivity and personalization. Users demand transmedia contents that need new tools and give new sensations. Binaural audio allows the perception of a 3D sound space around the listener in a very realistic way. This paradigm is totally new in television production that proposes scene and action in front of the viewer. There is a real impact of the vision on the binaural hearing, how to deal with the multisensory perception in TV production? How to use it and what are the issues? The role of head movement, captured via a head tracking device, is also discussed as it greatly improves the spatial perception. This tutorial session is illustrated with contents already broadcasted on a non-linear web TV platform. It is based on listening, viewing and explaining filmed extracts.

Tutorial 18 **Monday, June 6**
15:45 – 17:00 **Room 351**

3D SOUND REPRODUCTION—A STUDY OF DIFFERENT FORMAT REPRODUCTION POSSIBILITIES

Presenter: **Michael Williams**, Sounds of Scotland, Le Perreux sur Marne, France

This presentation will analyze the different formats of 3D sound reproduction and the basic principles of sound localization around a 3D microphone/loudspeaker array. It also explains the psychoacoustics of perception in a 3D hemispherical space. The ˜Witches Hat" localization and "Top Hat" localization will be explained to illustrate the parameters for this 3D localization—these two localization structures were presented by the author at a recent Taktons Convention in Novi Sad, Serbia. This presentation will demonstrate the characteristics of the present 3D formats with a multitude of sound recordings from the sounds of nature, to various other musical formations. The proposed sound formats include: • 5 horizontal channels + 4 elevation channels; • 5, 7, 8 horizontal channels with 4 (or 8) elevation channels, plus 1 Zenith channel. The structure of these two types of configurations leads quite naturally to a number of very interesting compatible configurations, which are also demonstrated using the technique of immediate quick fades between tracks of a simultaneous microphone array recording, for comparison without any time lapse. A certain number of recordings will be presented from the natural environment and also with some musical groups. All this will lead up to a demonstration of a new format ˜Comfort 3D" using a triple five (plus one) microphone array recording system. Some of the experimental listening tests with an artificial sound source moving around the microphone array at various angles of elevation can also be demonstrated if time allows.

Professional Sound Expo
Monday, June 6, 16:00 – 16:45
PSE Stage

MEASUREMENT MICROPHONES

Presenter: **Udo Wagner**, Microtech Gefell

This session will address Measurement Microphones, their characteristics, behavior, and applications in car acoustics, noise and vibration testing, car test stand equipment, and sensor calibration.

Special Event
Monday, June 6, 16:45 – 18:00
Room 352A

AUDIO PROJECTIONS 9—BINAURAL AUDIO FROM FRANCE TÉLÉVISIONS

Presenters: **Daniel Khamdamov**
Gilles Porte
Gaël Segalen

New Productions Using Binaural in the French Media Landscape

Presentation by Daniel KHAMDAMOV

TOUCHE FRANCAISE is a documentary web series (12 x 7 min) written by Guillaume Fédou and Jean-François Tatin, directed by Jean-François Tatin. The series offers a travel back into the French Touch musical scene, through a playlist of twelve iconic tracks of French electronic music from 1995 until today. It revisits the emergence of a new popular culture, portraying chronologically this movement and offering a sensory and musical journey back in times through an original and curated soundtrack of the French Touch sound. Interviews and never-before-seen personal archives of the internationally renowned French artists who contributed to this new musical style tell of a generational and global phenomenon born with the Internet bubble. Each of the 12 episodes delves into one chosen aspect of this movement and focuses on one track and artist such as Laurent Garnier, Air, Justice, Daft Punk, Sébastien Tellier, Vitalic, Kavinsky etc. Lasting approximately 7 min, the episodes are tailored for a quality audio streaming and perfect sonic immersion : binaural recordings and a spatial sound design offer the spectator a multisensory innovative experience.

Each episode ends with an original recording of the title track by a selected up and coming musician or band. French musician Christophe Chassol has composed the score and designed musical textures following his "ultrascore" technique, i.e., capturing a particular location's color, music, and sounds.

Presentation by Gaël SEGALEN

Dancers are musicians: Symétric à la lune ("Symmetric to the moon") is a musical piece made exclusively from the sounds of dance (moves, steps, breaths, …), a sound library recorded by Gaël Segalen in conjunction with dancers, improvising in public spaces, listening to their own sounds while producing them (mixed recording technics: binaural, stereo AB, mono, for many grounds, densities and perspectives). Made originally for headphones (audiowalks organized in different cities), and for boat cruises with the geolocalized app Soundways (Cap Digital, Futur en Seine) created by "Collective Mu for Bande Originale," an artistic exploration along the Canal de l'Ourcq, Paris and northern suburbs, summer 2014. Composed, reworked and mastered by Gaël Segalen for vinyl Edition on Paris based label Erratum (May 2016).

Special Event
Monday, June 6, 17:00 – 18:00
Room 351

AUDIO PROJECTIONS 10—3D AUDIO FROM THE JURASSIC AND EARLY TRIASSIC

Presenter: **Michael Williams**, Sounds of Scotland, Le Perreux sur Marne, France

Le Monde des Dinosaurs

The publication of the natural sounds of Jurasic Dinosaurs in the collection of biological recordings published by 'Frémeaux & Associés' is based on the research work done by Professor David Weishampel from John Hopkins University in Maryland, who studied the characteristics of air passing through the empty comb of the Parasaurolophus. The 'natural' sound recording engineer Jean-Luc Hérelle has put together the sounds of the Iguanodon and the tyrannosaurs in a Giant Sequoia Forest during the early-Jurassic (a period going back some 200 million years). The creation and production of this sound material is based on some recent scientific findings on the relationshship between dinosaurs and today's birds, and of course dinosaur replicas produced by paleontologist using much that has been discovered about dinosaurs since the 19th Cantury. These sounds are a real insight into this prehistoric world. If you were not able to experience the sounds of these giant animals at first hand, then get in your time machine and come to this fantastic soundscape audio projection.

This series of audio projection soundscapes will include extracts from:

1) The Plateosaurus, Nothosaurus, Coelophysis, Dimorphodon, Barapasaurus, Plesiosaurus and Rhamphorhynchus from the triassic/early Jurassic period.

2) A small group of Archaeopteryx who are hunting insects, surrounded by a few Camptosaurus, a Megalosaurus and a Scaphognathus. These are all from the Jurassic era.

3) The soundscape of an open swamp forest with the Pterodaustro, an Iguanodon, the Euhelopus, the Stegosaurus, the Acrocanthosaurus, the mighty Deinonychus, the spectacular Parasaurolophus, and the final scream of the Tyrannosaurus, accompanied by the song of the Pteranodon.

Special Event
Monday, June 6, 18:00 – 19:00
Room 352A

AUDIO PROJECTIONS 11—THE FRENCH MEDIA LANDSCAPE

Presenters: *Silvain Gire,* ARTE Radio / ARTE France, Issy-les-Moulineaux, France
David Kleinman, BinauralCircus, Paris, France
Pascal Rueff

New Productions Using Binaural in the French Media Landscape

Presentation by David KLEINMAN:

"Dernier round" ("Last round") is the first episode of the series "Qu'est-ce que j'fous là?" (What in Hell am I doing here?). 3 minutes in a boxing ring: a journey in my head, ears wide open… For "Dernier round," I was on the ring, fighting with gloves and everything … and microphones in my ears. Images are the best enemy of sound!... I'm searching for a way to reconcile them.

Presentation by Silvain GIRE

In the head: The conductor of the orchestra By Charlotte Rouault and Benoît Bories Production: ARTE Radio (4' 48") Hear the classical orchestra from the "point of listening" of the conductor. Violins on the left, cellos in the back … ARTE Radio produces the ongoing series "Dans la tête" ("In the head") in binaural audio. 4 minutes in the head of a metro driver, a hockey player, a basketball player in a wheelchair, a gynecologist …

Special Event
Monday, June 6, 18:00 – 19:00
Room 351

AUDIO PROJECTIONS 12—CONTEMPORARY MUSIC IN 3D AUDIO

Presenter: **Hugo Romano Guimaräes**, Neu Records, Barcelona, Spain

N is a production company based in Barcelona devoted to surround and 3D audio recordings in natural acoustics. We specialize in the recording and reproduction of classical and contemporary music in immersive audio formats, with a realistic aesthetic approach. The way we "see" sound is inspired by the soundscape of natural acoustics, with its immersive perspective, richness of timbres and colossal dynamic contrasts. When the perception of these aspects are carefully taken into account in doing an immersive recording and the right techniques are used, the result is a natural and realistic sound reproduction that creates the illusion of being there. We want people to hear what they couldn't experience in a live performance and we do that by transporting them to a privileged place in the center of the event: a place where you are truly immersed in the musical experience. This privileged situation is not a naturalistic recreation of a musical event: it's rather a hyperrealistic sound illusion of a musical event, created by the manipulation of several factors:

- Our microphone array design, which combines an association of different distances, angles and polar patterns in such a way that it produces not only a good sense of localization, but also an extremely natural timbre, and a sensation of space that music needs in order to expand and breathe;
- The selection of venues with voluminous spaces we choose for our projects, where sound can develop freely with unfettered dynamics, free of physical compression, and where we can find the perfect balance between direct and reverberant sound in a real context;
- And an artistic direction in close collaboration with the composers, which allows them to write music considering all these possibilities and creating new ways of experiencing music.

We present the albums of our label Neu Records in 3D sound installations, a diffusion format that allows the creation of an immersive listening context and gives a compelling social dimension to our projects. Our productions go beyond mere recordings, becoming a hybrid between sound art exhibitions and live concerts. At the same time, the physical format of the albums has also changed. The booklets become books with download codes, facilitating the reading of the texts we edit, and it's only a matter of time before physical CDs are no longer produced.

We have prepared a selection of works that we recorded for our label Neu Records:

- Ramon Humet: Interludi Meditatiu 1 Neopercusión & Kakizakai Kaoru
- Ramon Humet: Interludi Meditatiu 7 Neopercusión
- Ramon Humet: Interludi Meditatiu 15 Neopercusión
- Ramon Humet: Interludi Meditatiu 17 Neopercusión
- Fritz Hauser: Rundum We Spoke: Percussion
- Bernat Vivancos: O Lux Beata Latvian Radio Choir, Sigvards Klava

Special Event
Monday, June 6, 19:00 – 21:15
Saint Etienne du Mont Church

Place Sainte-Geneviève, 75005 Paris

Buses will be leaving Palais des Congres at 19:00. Concert will be 19:45 to 21:15 followed by buses back to Palais des Congres.

ORGAN CONCERT

Organists: **Graham Blyth**
Francis Rumsey

Saint-Etienne du Mont Church stands on the site of an abbey founded by Clovis and later dedicated to Ste Geneviève, the patroness of Paris. Such was the fame of this popular saint that the abbey proved too small to accommodate the pilgrimage crowds. Now part of the Lycée Henri IV, the Tour de Clovis (Tower of Clovis) is all that remains of the ancient abbey—you can see the tower from rue Clovis. Today the task of keeping Ste Geneviève's cult alive has fallen on the Saint-Etienne du Mont Church which almost adjoins the Panthéon. Building began in 1492 and was plagued by delays until its final completion in 1626. The architectural style of the church is unique in Paris, the result of the transition period between flamboyant gothic and renaissance. Inside the intricate and ornate 16th century choirscreen is the only example still in a church in Paris.

The organ case is the oldest one in Paris and is still in its original state, it was carved and built in 1631 by Jehan Buron. The organ itself has seen several transformations: an earlier 1772 four manual Cliquot instrument was completely revised in 1873 by Aristide Cavaillé-Coll who added a 16' Bombarde to the Grand Orgue, reduced the number of manuals to 3 and built a new 42-note Récit. Most of the old pipework was retained, mainly all Positif pipework; it was a 39-stop instrument. On this occasion, the organ was inaugurated with the participation of César Franck. Ten years later, Cavaillé-Col came back to improve his work. A major alteration was completed in 1956 by Beuchet-Debierre based on instructions by Maurice Duruflé and it is now a 90 stop instrument. Titular organist from 1929 until his death in 1986, Maurice Duruflé was also a notable composer, the use of his Pie Jesu by Michael Jackson in a prelude to a song on his album HIStory in 1995 may, or may not, have amused him.

Program to include:
Marchand: Grand Dialogue in C
Bonnet: Variations de Concert
Lefébure-Wely: Sortie in E flat
Franck: Chorale no. 3 in E minor
Messiaen: Dieu Parmi Nous

Challow Park Studios, Oxfordshire, UK, is sponsoring the concert

Session P20 **Tuesday, June 7**
08:45 – 10:45 **Room 353**

PERCEPTION: PART 3

Chair: **John Mourjopoulos**, University of Patras, Patras, Greece

08:45

P20-1 Conflicting Dynamic and Spectral Directional Cues Form Separate Auditory Images—*Henri Pöntynen, Olli Santala, Ville Pulkki*, Aalto University, Espoo, Finland

Auditory localization under conflicting dynamic and spectral cues was investigated in a listening experiment where head-motion-coupled amplitude panning was used to create front-back confusions with moving free-field stimuli. Subjects reported whether stimuli of various spectra formed auditory images in the front, rear or both hemiplanes simultaneously. The results show that panned low-pass stimuli were consistently localized to the rear hemiplane while high-pass stimuli did not produce hemiplane reversals. The main result of the experiment is that broadband stimuli providing low-frequency ITD sequences that are inconsistent with the source directions implied by the spectral cues can lead to the formation of two segregated auditory images. This effect was observed with both continuous and discontinuous stimulus spectra.
Convention Paper 9582
[Paper presented by Ville Pulkki]

09:15

P20-2 Discrimination of Formant Frequency in Pink Noise
—*Tomira Rogala*, Fryderyk Chopin University of Music, Warsaw, Poland

The paper reports an experiment conducted to determine discrimination thresholds for timbre in tonmeister students and non-musicians. The variations of timbre were obtained through introducing a 1/3-octave wide formant into the spectrum of noise and shifting the formant's center frequency. Discrimination thresholds were measured using a 3AFC procedure. The results have shown that the threshold values determined for tonmeister students were considerably lower than those obtained for non-musicians. In both groups of listeners a learning effect was observed: the thresholds decreased in successive measurement series completed by a listener. It also was found that the formant frequency discrimination thresholds depended on the formant frequency and were much higher at 125 Hz than at 315 Hz and higher frequencies.
Convention Paper 9583

09:45

P20-3 The Influence of Room Acoustics on Musical Performance and Interpretation—A Pilot Study—
Jan Berg, Sverker Jullander, Petter Sundkvist, Helge Kjekshus, Luleå University of Technology, Piteå, Sweden

Concert hall acoustics is an important factor that influences musical performance. Different acoustics lead to different musical results. For a musical performer, the artistic impression of a performance is paramount. Therefore, it is essential to study the relation between concert hall acoustics and musical performance. Such studies might also be relevant for architects and acousticians. A pilot study was devised, enabled by a unique concert hall with mechanically variable acoustics. A musician played the grand piano at four trials, each having a distinctive acoustic condition. The trials were recorded for later analysis. The performances were assessed by experts and the pianist himself. The results show that clear as well as subtle differences in interpretation and performance between the trials existed.
Convention Paper 9584

10:15

P20-4 Timbre Preferences of Four Listener Groups and the Influence of their Cultural Backgrounds—*Sungyoung Kim,*[1] *Ron Bakker,*[2] *Masahiro Ikeda*[3]
[1]Rochester Institute of Technology, Rochester, NY, USA
[2]Yamaha Music Europe, Vianen, The Netherlands
[3]Yamaha Corporation, Hamamatsu, Shizuoka, Japan

The cultural influence on listeners' timbre preference was investigated using the magnitude estimation method. Four listener groups (Dutch, Japanese, Korean, and American) participated in a listening test in their own countries. The listeners manipulated the timbre of five stimuli

(Dutch, Japanese, Korean and English popular song, and orchestral music) by adjusting gains of three frequency bands according to their preferences. The statistical analysis (a mixed design ANOVA) showed that only interaction factor of the listener groups and the stimuli significantly differentiated the preferred spectral responses of four listener groups. This implies that a listener group from one country had unique timbre preference that appeared by listening to a song in its own language.
[Also a poster—see session P22-3]
Convention Paper 9585

Session P21 **Tuesday, June 7**
08:45 – 11:45 **Room 352B**

IMMERSIVE AUDIO—PART 1

Chair: **Bob Schulein**, RBS Consultants / ImmersAV Technology, Schaumburg, IL, USA

08:45

P21-1 Low-Complexity Stereo Signal Decomposition and Source Separation for Application in Stereo to 3D Upmixing—*Sebastian Kraft, Udo Zölzer*, Helmut-Schmidt-University, Hamburg, Germany

In this paper we present a general low-complexity stereo signal decomposition approach. Based on a common stereo signal model, the panning coefficients and azimuth positions of the sources in a stereo mix are estimated. In a next step, this information is used to separate direct and ambient signal components. The simple algorithm can be implemented at low computational cost and its application in a stereo to 3D upmix context is described. Particular focus is put on the generation of additional ambient channels by using decorrelation filters in a tree structure. Finally, the separation performance is evaluated with several standard measures and compared to other algorithms.
[Also a poster—see session P22-4]
Convention Paper 9586

09:15

P21-2 Immersive Audio Delivery Using Joint Object Coding — *Heiko Purnhagen, Toni Hirvonen, Lars Villemoes, Jonas Samuelsson, Janusz Klejsa*, Dolby Sweden AB, Stockholm, Sweden

Immersive audio experiences (3D audio) are an important element of next-generation audio entertainment systems. This paper presents joint object coding techniques that enable the delivery of object-based immersive audio content (e.g., Dolby Atmos) at low bit rates. This is achieved by conveying a multichannel downmix of the immersive content using perceptual audio coding algorithms together with parametric side information that enables the reconstruction of the audio objects from the downmix in the decoder. An advanced joint object coding tool is part of the AC-4 system recently standardized by ETSI. Joint object coding is also used in a backwards compatible extension of the Dolby Digital Plus system. Listening test results illustrate the performance of joint object coding in these two applications.
[Also a poster—see session P22-6]
Convention Paper 9587

09:45

P21-3 Design and Subjective Evaluation of a Perceptually-Optimized Headphone Virtualizer—*Grant Davidson,[1] Daniel Darcy,[1] Louis Fielder,[1] Zhiwei Schuang,[2] Rich Graff,[1] Jeroen Breebaart,[3]* Poppy Crum[1]
[1]Dolby Laboratories, San Francisco, CA, USA
[2]Dolby Laboratories Intl. Services Co. Ltd., Beijing, China
[3]Dolby Australia Pty. Ltd., McMahons Point, NSW, Australia

We describe a novel method for designing echoic headphone virtualizers based on a stochastic room model and a numerical optimization procedure. The method aims to maximize sound source externalization under a natural-timbre constraint. The stochastic room model generates a number of binaural room impulse response (BRIR) candidates for each virtual channel, each embodying essential perceptual cues. A perceptually-based distortion metric evaluates the timbre of each candidate, and the optimal candidate is selected for use in the virtualizer. We designed a 7.1.4 channel virtualizer and evaluated it relative to a LoRo stereo downmix using a single-interval A:B preference test. For a pool of 10 listeners, the test resulted in an overall virtualizer preference of 75%, with no stereo test item preferred over binaural.
Convention Paper 9588

10:15

P21-4 An Open 3D Audio Production Chain Proposed by the Edison 3D Project—*Etienne Corteel,[1] David Pesce,[2] Raphael Foulon,[1] Gregory Pallone,[2] Frédéric Changenet,[3] Hervé Dejardin[3]*
[1]Sonic Emotion Labs, Paris, France
[2]b<>com - Cesson-Sévigné, France
[3]Radio France, Paris, France

In this paper we present a production chain for Next Generation Audio formats that combines standard digital audio workstations with external 3D audio rendering software using an open communication protocol. We first describe the scope of the Edison 3D project. In a second part, we revisit existing 3D audio formats outlining the need for an open format for content creation and archiving. We then describe the tools developed in Edison 3D enabling user interaction, storage of object positions in the timeline, monitoring of audio content in various rendering formats (stereo, 5.1, binaural, WFS, HOA), and export into the open and recently standardized (ITU BS.2076) 3D audio format: Audio Definition Model. We finally provide an outlook into future work.
Convention Paper 9589

10:45

P21-5 Perceptual Evaluation of Transpan for 5.1 Mixing of Acoustic Recordings—*Gaëtan Juge,[1] Amandine Pras,[1,2] Ilja Frissen[3]*
[1]Paris Conservatory (CNSMDP), Paris, France
[2]Stetson University, DeLand, FL, USA
[3]McGill University, Montreal, Quebec, Canada

We evaluate the efficiency of a 3D spatialization software named Transpan in the context of mixing acoustic recordings on a 5.1 reproduction system. The study aims to investigate if the use of the binaural with cross-talk cancellation (XTC) processing implemented in Transpan can improve the localization of lateral sources and their stability through listeners' movements. We administered a listening test to 22 expert listeners in Paris and in Berlin. The test consisted of comparisons among two mixes with and without Transpan binaural/XTC panning, for four classical music excerpts under five listening conditions, i.e., at the sweet spot and while performing specific move-

ments. Quantitative analysis of multiple choice questions showed that Transpan can enlarge the 5.1 sweet spot area toward the rear speakers. From qualitative analysis of participants' feedback emerged five main categories of comments, namely Localization stability; Precise localization accuracy; Vague localization accuracy; Timbral and spectral artifacts; and Spatial differences. Together the results show that Transpan allows for better source lateralization in 5.1 mixing.
[Also a poster—see session P22-5]
Convention Paper 9590

11:15

P21-6 The Influence of Head Tracking Latency on Binaural Rendering in Simple and Complex Sound Scenes—
Peter Stitt,[1] *Etienne Hendrickx*,[2] *Jean-Christophe Messonnier*,[2] *Brian Katz*[1]
[1]LIMSI Université Paris-Saclay, Orsay, France
[2]Paris Conservatory (CNSMDP), Paris, France

Head tracking has been shown to improve the quality of multiple aspects of binaural rendering for single sound sources, such as reduced front-back confusions. This paper presents the results of an AB experiment to investigate the influence of tracker latency on the perceived stability of virtual sounds. The stimuli used are a single frontal sound source and a complex (5 source) sound scene. A comparison is performed between the results for the simple and complex sound scenes and the head motions of the subjects for various latencies. The perceptibility threshold was found to be 10 ms higher for the complex scene compared to the simple one. The subject head movement speeds were found to be 6 degrees/s faster for the complex scene.
Convention Paper 9591

Tutorial 19 **Tuesday, June 7**
08:45 – 10:00 **Room 351**

HOW TO DO AUDIO SAMPLING, AND SAMPLE RATE CONVERSION, PROPERLY!

Presenter: **Jamie Angus**, University of Salford, Salford, Greater Manchester, UK; JASA Consultancy, York, UK

Sampling, and sample rate conversion, are critical processes in digital audio. The analogue signal must be sampled, so that it can be quantized into a digital word. In addition digital audio signals are often converted between sample rates, either as part of the conversion process, for mastering to a particular audio format, or varispeed playback. How does sampling affect the audio signal? What conditions must be met in order to sample audio signals properly? Does the sampling rate affect the timing, or lose information, from the music? Does changing the sample rate do any of these things. Are there further pitfalls? This tutorial will unravel these questions and, with audio examples, will show how sampling, and sample rate conversion, when done properly, can be done in a transparent and lossless fashion that preserves all of the original signal's qualities.
This session is presented in association with the AES Technical Committee on Signal Processing

Workshop 16 **Tuesday, June 7**
08:45 — 09:45 **Room 352A**

CODEC FOR ENHANCED VOICE SERVICES (EVS)— THE NEW 3GPP CODEC FOR COMMUNICATION

Chair: **Markus Multrus**, Fraunhofer IIS, Nuernberg, Germany

Panelists: *Stefan Bruhn*
Vaclav Eksler
Guillaume Fuchs
Jon Gibbs

Recently, 3GPP has finalized the standardization of its new communication codec for Enhanced Voice Services (EVS). This codec offers significant improvements over existing low-delay communication codecs. For a broad range of bit rates it supports coding of not only of narrowband and wideband audio, but also of super-wideband and fullband content up to 20 kHz audio bandwidth. The codec is based on a switched speech/audio coding scheme and features various tools for better compression efficiency and higher quality for speech, music, and mixed material. It provides all necessary features for (mobile-) communication and aims at providing a new level of user experience for all transmission-channel conditions. The workshop provides an overview over the codec, its performance, and applications together with a number of sound demonstrations.
This session is presented in association with the AES Technical Committee on Coding of Audio Signals

Student Event and Career Development
Tuesday, June 7, 08:45 – 11:15
Havane Amphitheatre

STUDENT DELEGATE ASSEMBLY MEETING—PART 2

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 Europe and International Regions. Judges' comments and awards will be presented for the Recording Competitions and Design Competitions. Plans for future student activities at local, regional, and international levels will be summarized.

Tutorial 20 **Tuesday, June 7**
10:00 – 11:00 **Room 351**

ASR—AUTOMATIC SPEAKER RECOGNITION

Presenter: **Antonio Moreno**, Agnitio, Spain

This tutorial gives an overview of ASR, Automatic Speaker Recognition. Special attention on parameters applied, the kind of speech segmentation applied, redundancy of the systems, hardware requirements, implementation in portable systems, the hit-rate, etc. The intention of the presentation is to be educational.

Professional Sound Expo
Tuesday, June 7, 10:00 – 10:45
PSE Stage

PROTOOLS IN THE CLOUD

Presenter: **Dave Tyler**, AVID

What's new in Pro Tools 12? How do you get ready for Pro Tools Cloud Collaboration? These questions and more will be answered.

Sponsored Event

Workshop 17 **Tuesday, June 7**

10:15 — 11:15 **Room 352A**

HUMAN PERCEPTION AND LISTENING BY MACHINES

Chair: **Cleopatra Pike**, University of Surrey, Guildford, Surrey, UK

Panelist: *Amy Beeston*, University of Sheffield, Sheffield, UK

Automatic audio recognition is used in a wide range of technology applications (e.g., smartphone voice activated commands, speech to text applications, automatic alarm systems). This workshop will give an overview of automatic speech recognition and recognition of other audio by machines, as well a conceptual introduction to machine listening principles for the lay person. It will focus on examining how research into human perception can inform the development of machine listeners (e.g., how human methods for listening in reverberant environments and of reducing channel coloration can be incorporated into machines). We will aim to answer the question: what things can machines do better than, and worse than, a human? Suggestions for future perceptual work to incorporate knowledge of human perception into machine listening will be made.

This session is presented in association with the AES Technical Committee on Perception and Subjective Evaluation of Audio Signals

Professional Sound Expo
Tuesday, June 7, 11:00 – 11:45
PSE Stage

THE IMPORTANCE OF CONTROLLING PHASE AND DAMPING MONITORS

Presenters: **Christian Martin,** PSI Audio, Yverdon-les-Bains, Switzerland
Roger Roschnik, PSI Audio, Yverdon-les-Bains, Switzerland

It is generally recognized that the mission of an audio monitor is to transform an electrical signal into an acoustic signal as accurately as possible. Among the main disturbances inherent to any electroacoustic device and that need to be controlled, are phase distortion and harmonic distortion. This presentation will discuss the importance having a controlled phase response as well as optimal transducer damping for faithful audio transmission.

Tutorial 21 **Tuesday, June 7**
11:15 – 12:15 **Room 352A**

SOUND DESIGN AND ACCESSIBILITY IN FILM AND TELEVISION

Presenter: **Mariana Lopez**, Anglia Ruskin University, Cambridge, UK

Since the 1970s Audio Description (AD) has been making visual content accessible to visually impaired people through sound. In film and television a pre-recorded audio commentary provides information that clarifies the narrative, such as descriptions on actions, gestures and places. Although, throughout the years, digital technologies have been used to improve the mode of delivery of AD, the notions behind its design have been mostly unchanged despite significant advancements in the field of digital sound production and post production. This tutorial introduces participants to the history of sound design and accessibility in film and TV and discusses new research on the use of surround sound rendering and interactive media systems to create more spatially accurate soundtracks, as well as the introduction of first person narration.

Tutorial 22 **Tuesday, June 7**
11:15 – 12:30 **Room 351**

SOUNDS ACROSS THE SEA—A JOURNEY IN 9.1 IMMERSIVE AUDIO

Presenters: **Morten Lindberg**, 2L (Lindberg Lyd AS), Oslo, Norway; Westerdals - Oslo School of Arts, Communication and Technology
Daniel Shores, Sono Luminus, Boyce, VA, USA; Shenandoah Conservatory Music Production and Recording Technology, Winchester, VA, USA

For years Morten Lindberg and Daniel Shores have inspired each other across the Atlantic to develop new techniques and implementing new recording formats; initially in surround sound and lately moving forward in immersive audio. In this workshop we'll meet together playing a wide range of 9.1 examples, discussing microphone techniques, production workflow and musical sound design.

Workshop 18 **Tuesday, June 7**
11:30 — 12:15 **Havane Amphitheatre**

FRENCH RECORDING STUDIOS—PRESENT AND FUTURE

Presenter: **Franck Ernould**, Freelance, Paris, France

Recording studios are an endangered species. What happened in France since 2006? Some facilities disappeared, but most survived—with additional activities (audio training, event hosting, video shooting...). Some new studios appeared, and the Honky Chateau is back. How do they work? How do they see their future?

Student Event and Career Development
Tuesday, June 7, 11:30 — 12:30
Room 311

STUDENT RECORDING CRITIQUES

Moderator: **Ian Corbett**, Kansas City Kansas Community College, Kansas City, KS, USA; off-beat-open-hats recording & sound reinforcement

The Student Recording Critiques are non-competitive listening sessions, designed for students to listen to their recordings and productions on a world-class playback system, and receive feedback on their work. Students are invited to bring along their mixes and have them critiqued by a panel of renowned industry professionals in order to get pointers as to how they can push their skills to the next level.

It is vital that interested students sign up at the student booth immediately after the first SDA meeting and bring their work on CD, DVD, memory stick, or hard disc as clearly labelled 44.1 kHz WAVE or AIFF files.

Please note that finalists of the Student Recording Competition are excluded to submit their work to the Recording Critiques to give non-finalists a way of having their work assessed.

Session P22 **Tuesday, June 7**

12:00 – 14:00 Foyer

POSTERS: RENDERING, HUMAN FACTORS AND INTERFACES

12:00

P22-1 A Subjective Comparison of Discrete Surround Sound and Soundbar Technology by Using Mixed Methods—*Tim Walton,*[1,2] *Michael Evans,*[2] *David Kirk,*[1] *Frank Melchior*[2]
[1]Newcastle University, Newcastle-Upon-Tyne, UK
[2]BBC Research and Development, Salford, UK

In recent years, soundbars have seen a rise in interest from consumers of home audio. Such technology offers an alternative means to experience surround sound content compared to conventional discrete multichannel systems. This paper presents a subjective comparison between two soundbars—a discrete 5 channel surround system and a discrete stereo system for a range of content and listener experience—in order to evaluate how soundbar technology compares to conventional discrete systems. A mixed methods approach, Open Profiling of Quality, was used in order to deeper understand preference ratings for the various reproduction systems. The results show that the discrete surround system was significantly preferred to the soundbars for all content due to a combination of timbral and spatial factors.
Convention Paper 9592

12:00

P22-2 Spatial Post-Processing of Hard Panned Music for Headphone Reproduction—*Katie Gilchrest*, New York University, New York, NY USA

This paper presents a stereo expansion method for hard panned music of the 1960s and 1970s over headphones. Previous studies have attempted to externalize music across all genres, with most subjective tests ranking the unprocessed original as the most preferred version. The vast amount of recording and mixing techniques indicates that there may be more than one method of externalizing music for headphone listening. The algorithm "earGoggles" is a head-related transfer function (HRTF) model combined with room reflection simulation, crafted to externalize music with a correlation of 0.5 and below, produced from the era of early stereo mixing techniques. A subjective test was conducted in which subjects submitted preference ratings of four versions of seven programs: earGoggles, crosstalk simulation from the HRTF model, a room model, and the original unprocessed version. Experiment results showed that the mean overall preference rating was for the earGoggles algorithm, and least preferred was the original unprocessed version.
Convention Paper 9593
[Paper not presented]

12:00

P22-3 Timbre Preferences of Four Listener Groups and the Influence of their Cultural Backgrounds—*Sungyoung Kim,*[1] *Ron Bakker,*[2] *Masahiro Ikeda*[3]
[1]Rochester Institute of Technology, Rochester, NY, USA
[2]Yamaha Music Europe, Vianen, The Netherlands
[3]Yamaha Corporation, Hamamatsu, Shizuoka, Japan

The cultural influence on listeners' timbre preference was investigated using the magnitude estimation method. Four listener groups (Dutch, Japanese, Korean, and American) participated in a listening test in their own countries. The listeners manipulated the timbre of five stimuli (Dutch, Japanese, Korean and English popular song, and orchestral music) by adjusting gains of three frequency bands according to their preferences. The statistical analysis (a mixed design ANOVA) showed that only interaction factor of the listener groups and the stimuli significantly differentiated the preferred spectral responses of four listener groups. This implies that a listener group from one country had unique timbre preference that appeared by listening to a song in its own language.
[Also a lecture—see session P20-4]
Convention Paper 9585

12:00

P22-4 Low-Complexity Stereo Signal Decomposition and Source Separation for Application in Stereo to 3D Upmixing—*Sebastian Kraft, Udo Zölzer***,** Helmut-Schmidt-University, Hamburg, Germany

In this paper we present a general low-complexity stereo signal decomposition approach. Based on a common stereo signal model, the panning coefficients and azimuth positions of the sources in a stereo mix are estimated. In a next step, this information is used to separate direct and ambient signal components. The simple algorithm can be implemented at low computational cost and its application in a stereo to 3D upmix context is described. Particular focus is put on the generation of additional ambient channels by using decorrelation filters in a tree structure. Finally, the separation performance is evaluated with several standard measures and compared to other algorithms.
[Also a lecture—see session P21-1]
Convention Paper 9586

12:00

P22-5 Perceptual Evaluation of Transpan for 5.1 Mixing of Acoustic Recordings—*Gaëtan Juge,*[1] *Amandine Pras,*[1,2] *Ilja Frissen*[3]
[1]Paris Conservatory (CNSMDP), Paris, France
[2]Stetson University, DeLand, FL, USA
[3]McGill University, Montreal, Quebec, Canada

We evaluate the efficiency of a 3D spatialization software named Transpan in the context of mixing acoustic recordings on a 5.1 reproduction system. The study aims to investigate if the use of the binaural with cross-talk cancellation (XTC) processing implemented in Transpan can improve the localization of lateral sources and their stability through listeners' movements. We administered a listening test to 22 expert listeners in Paris and in Berlin. The test consisted of comparisons among two mixes with and without Transpan binaural/XTC panning, for four classical music excerpts under five listening conditions, i.e., at the sweet spot and while performing specific movements. Quantitative analysis of multiple choice questions showed that Transpan can enlarge the 5.1 sweet spot area toward the rear speakers. From qualitative analysis of participants' feedback emerged five main categories of comments, namely Localization stability; Precise localization accuracy; Vague localization accuracy; Timbral and spectral artifacts; and Spatial differences. Together the results show that Transpan allows for better source lateralization in 5.1 mixing.
[Also a lecture—see session P21-5]
Convention Paper 9590

12:00

P22-6 Immersive Audio Delivery Using Joint Object Coding — *Heiko Purnhagen, Toni Hirvonen, Lars Villemoes, Jonas Samuelsson, Janusz Klejsa*, Dolby Sweden AB, Stockholm, Sweden

Immersive audio experiences (3D audio) are an important element of next-generation audio entertainment systems. This paper presents joint object coding techniques that enable the delivery of object-based immersive audio content (e.g., Dolby Atmos) at low bit rates. This is achieved by conveying a multichannel downmix of the immersive content using perceptual audio coding algorithms together with parametric side information that enables the reconstruction of the audio objects from the downmix in the decoder. An advanced joint object coding tool is part of the AC-4 system recently standardized by ETSI. Joint object coding is also used in a backwards compatible extension of the Dolby Digital Plus system. Listening test results illustrate the performance of joint object coding in these two applications.
[Also a lecture – see session P21-2]
Convention Paper 9587

Session EB4 Tuesday, June 7
12:00 – 13:45 Room 353

ENGINEERING BRIEFS—LECTURES

Chair: **Thomas Görne**, Hamburg University of Applied Sciences, Hamburg, Germany

12:00

EB4-1 A Survey of Suggested Techniques for Height Channel Capture in Multichannel Recording—*Richard King, Will Howie, Jack Kelly*, McGill University, Montreal, QC, Canada; Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Montreal, Quebec, Canada

Capturing audio in three dimensions is becoming a required skill for many recording engineers. Playback formats and systems now exist that take advantage of height channels, which introduce the aspect of elevation into the experience. In this engineering brief several exploratory techniques in height channel capture are reviewed and compared. Techniques optimized for conventional 5.1 surround sound are employed, and additional microphones are added to increase the immersive experience. Methods that have proven to be successful in 5.1 recordings are modified for 3D audio capture and the results are discussed. This case studywill show an overview of the groundwork currently underway.
Engineering Brief 266

12:15

EB4-2 Perceptually Significant Parameters in Stereo and Binaural Mixing with Logic Pro Binaural Panner—*Blas Payri, Juan-Manuel Sanchis-Rico*, Universitat Politècnica de València, Valencia, Spain

We conducted a perception experiment using organ chords recorded with 6 microphones and mixed in stereo and binaural, varying in maximum angle distribution (0º, 87º, 174º), and for binaural mixes, varying in elevation, front-rear distribution, and postprocessing. N=51 participants (audio-related students) listened to the 20 stimuli with headphones, classified them according to similarity, and rated their valence and immersion. Results show a high agreement on similarity (Cronbach's alpha=.93) but very low agreement on valence and immersion ratings. The parameters that were perceived are the difference binaural/stereo, the binaural postprocessing style, and in a lesser degree, the angle. Elevation and rear distribution of sources did not yield any significant response.
Engineering Brief 267

12:30

EB4-3 3D Tune-In: The Use of 3D Sound and Gamification to Aid Better Adoption of Hearing Aid Technologies—
Yuli Levtov,[1] Lorenzo Picinali,[1] Mirabelle D'Cruz,[1] Luca Simeone[2]
[1]Reactify Music LLP, London, UK
[2]On behalf of 3D Tune-In consortium

3D Tune-In is an EU-funded project with the primary aim of improving the quality of life of hearing aid users. This is an introductory paper outlining the project's innovative approach to achieving this goal, namely via the 3D Tune-In Toolkit, and a suite of accompanying games and applications. The 3D Tune-In Toolkit is a flexible, cross-platform library of code and guidelines that gives traditional game and software developers access to high-quality sound spatialization algorithms. The accompanying suite of games and applications will then make thorough use of the 3D Tune-In toolkit in order to address the problem of the under-exploitation of advanced hearing aid features, among others.
Engineering Brief 268

12:45

EB4-4 Binaural Auditory Feature Classification for Stereo Image Evaluation in Listening Rooms—*Gavriil Kamaris, Stamatis Karlos, Nikos Fazakis, Stergios Terpinas, John Mourjopoulos*, University of Patras, Patras, Greece

Two aspects of stereo imaging accuracy from audio system listening have been investigated: (i) panned phantom image localization accuracy at 5-degree steps and (ii) sweet spot spatial spread from the ideal anechoic reference. The simulated study used loudspeakers of different directivity under ideal anechoic or realistic varying reverberant room conditions and extracted binaural auditory features (ILDs, ITDs, and ICs) from the received audio signals. For evaluation, a Decision Tree classifier was used under a sparse data self-training achieving localization accuracy ranging from 92% (for ideal anechoic when training/test data were similar audio category), down to 55% (for high reverberation when training/test data were different music segments).Sweet spot accuracy was defined and evaluated as a spatial spread statistical distribution function.
Engineering Brief 269

13:00

EB4-5 Elevation Control in Binaural Rendering—*Aleksandr Karapetyan, Felix Fleischmann, Jan Plogsties*, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

Most binaural audio algorithms render the sound image solely on the horizontal plane. Recently, immersive and object-based audio applications like VR and games require control of the sound image position in the height dimension. However, measurements from elevated loudspeakers require a 3D loudspeaker setup. By analyzing early reflections of BRIRs with a fixed elevation, spectral cues for the height perception are extracted and applied to HRTFs. The

parametrization of these cues allows the control of height perception. The sound image can be moved to higher as well as lower positions. The performance of this method has been evaluated by means of a listening test.
Engineering Brief 270

13:15

EB4-6 Headphone Virtualization for Immersive Audio Monitoring—*Michael Smyth, Stephen Smyth*, Smyth Research Ltd., Bangor, UK

There are a number of competing immersive audio encoding formats, such as Dolby Atmos, Auro-3D, DTS-X, and MPEG-H, but, to date, there is no single loudspeaker format for monitoring them all. While it is argued that one of the benefits of using audio objects within immersive audio is that it allows rendering to different, and even competing, loudspeaker formats, nevertheless the documented native formats of each immersive codec must be considered the reference point when monitoring each immersive audio system. This implies that the ability to switch between different loudspeaker layouts will be important when monitoring different immersive audio formats. The solution outlined here is based on the generation of virtual loudspeakers within DSP hardware and their reproduction over normal stereo headphones. The integrated system is designed to allow the accurate monitoring of any immersive audio system of up to 32 loudspeaker sources, with the ability to switch almost instantly between formats.
Engineering Brief 271

13:30

EB4-7 Temporal Envelope for Audio Classification—*Ewa Lukasik,[1] Cong Yang,[2] Lukasz Kurzawski[3]*
[1]Poznan University of Technology, Poznan, Poland
[2]University of Siegen, Siegen, Germany
[3]RecArt, Poznan, Poland

The paper reviews some applications of temporal envelope of audio signal from the perspective of a sound engineer. It contrasts the parametric representation of the temporal envelope (e.g., temporal centroid, attack time, attack slope) with the more global representation based on envelope shape descriptors. Such an approach would mimic the sound engineer expertise and could be useful for such classification tasks, as music genre, speech/music, musical instruments classification, and others.
Engineering Brief 272

Professional Sound Expo
Tuesday, June 7, 12:00 – 12:45
PSE Stage

THE DREAM OF REMOTE PRODUCTION: GET MORE OF EVERYTHING FOR LESS

Presenter: **Henry Goodman**, Calrec

Remote production provides the opportunity to produce live programing in a control room hundreds of miles from an event, with no latency and with a more efficient utilization of equipment and personnel. Broadcasters are already reaping the benefits: San Francisco-based Pac-12, for example, provide live coverage of some 850 events every year, mixed on an audio console in a centralized production facility hundreds of miles away.

Calrec's Henry Goodman explains that while there is no substitute for a skilled OB team, for niche programming such as local sports or local music events, remote broadcasting provides an incentive for broadcasters to increase their range of local programming.

Session P23 **Tuesday, June 7**
12:15 – 14:15 **Room 352B**

IMMERSIVE AUDIO: PART 2

Chair: **Robin Reumers**, Galaxy, Mol, Belgium

12:15

P23-1 Exploring the Benefits of Surround Sound in Contemporary Live Music Performances—*John Crossley*, University of Derby, Derby, UK

Spatial audio utilizing 5.1 surround sound and newer developments such as object oriented audio has become well established in cinema and home theaters. The expansion of this into live musical performance is quite limited. This work explores the benefits of surround sound for contemporary music performance. A 20-channel Wavefield synthesis system was compared to a high quality stereo sound reinforcement system under identical experimental conditions. An original composition was used to avoid familiarity with program material and to encourage focus on spatial considerations. Data drawn from audiences at both performances is used to quantify the perceptual differences for the average audience and to draw conclusions as to the usefulness of using a system of this type in an "average" contemporary live music performance.
Convention Paper 9594

12:45

P23-2 Extended Bass Management Methods for Cost-Efficient Immersive Audio Reproduction in Digital Cinema—*Toni Hirvonen,[1] Charles Q. Robinson[2]*
[1]Dolby Laboratories, Stockholm, Sweden
[2]Dolby Laboratories, San Francisco, CA, USA

New, more sophisticated cinema audio formats have recently been developed and deployed that provide a more immersive sound field for the audience members. This paper discusses techniques that can reduce the economic costs associated with the installation and maintenance of the audio reproduction devices in contemporary immersive digital cinema, while best retaining the benefits of these new formats. The focus of this paper is novel bass management techniques that enable the use of cost-effective loudspeakers. These proposed bass processing methods take advantage of the reduced spatial saliency of low frequency audio and allow for a reduced spatial resolution for audio signals in that range. We present subjective tests conducted in a state of the art cinema installation that illustrate the effects of the proposed solutions on sound quality. Some of the these techniques have been incorporated to the future version of the Dolby Atmos cinema specification.
Convention Paper 9595

13:15

P23-3 Local Wave Field Synthesis by Spatial Band-Limitation in the Circular/Spherical Harmonics Domain—*Nara Hahn, Fiete Winter, Sascha Spors*, University of Rostock,

Rostock, Germany

The achievable accuracy of sound field synthesis (SFS) techniques, such as Wave Field Synthesis (WFS), is mainly limited in practice due to the limited loudspeaker density. Above the so called spatial aliasing frequency, considerable artifacts are introduced in the synthesized sound field. In local SFS, the accuracy within a local listening area is increased at the cost of degradations outside. In this paper a new approach for local WFS is proposed. The WFS driving functions are computed based on an order-limited harmonics expansion of the target sound field. A local listening area is created around the shifted expansion center where the synthesized sound field exhibits higher accuracy. The size of the local area is controlled by the expansion order of the driving function. The derivations of 2D, 3D, and 2.5D driving functions are given, and the synthesized sound fields are evaluated by numerical simulations.
Convention Paper 9596

13:45

P23-4 Investigation on Subjective HRTF Rating Repeatability —*Areti Andreopoulou, Brian Katz*, LIMSI-CNRS, Orsay, France

This paper investigates the repeatability of an HRTF evaluation protocol, assessing the spatial quality of binaural stimuli, moving along pre-defined trajectories on the horizontal and median planes, on a forced-choice 9-point rating scale. The protocol assessment was based on data simulations and subjective studies. Repeatability was evaluated as a function of the size and content of the HRTF corpus, the trajectories, and the resolution of the rating scale. Analysis of the data revealed that HRTF rating is a reliable, yet challenging task with low repeatability rates of [about] 50%. Therefore participant screening through pre-tests should be used to maximize reliability of the responses.
Convention Paper 9597

Tutorial 23 — Tuesday, June 7
12:30 – 13:45 — Room 352A

AUDIO FORENSICS—WHAT'S IT ALL ABOUT?

Presenter: **Eddy B. Brixen**, EBB-consult, Smørum, Denmark; DPA Microphones

Working with audio forensics is serious business. Depending on the work of the forensics engineer, people may eventually end up in prison. This tutorial will present the kind of work related to the field. This covers fields as acoustics, when audio analysis can be a part of the crime scene investigation. Voice acoustics: Who was speaking? Electro acoustics: Checking data on tapes, discs or other data storage media. Recording techniques: Is this recording an original production or is it a copy of others' work. Even building acoustics and psychoacoustics, when the question is raised: Who could hear what? However, the most important 'everyday work' of the audio forensics engineers is cleaning of audio recordings and providing transcriptions.

This session is presented in association with the AES Technical Committee on Microphones and Applications

Tutorial 24 — Tuesday, June 7
12:30 – 13:45 — Havane Amphitheatre

CREATIVE VOCAL PRODUCTION TECHNIQUES

Presenter: **Michelle Desachy**, Estudio19, Mexico City, Mexico

Tutorial where we will dwell on the treatment of the lead voice and the background voices to create atmospheres and sonic flavors using recording techniques, using different microphones and the reasons why, using typical an non typical effects, using plug ins, and most of all using creativity, we will explain how human creativity works, the historical reasons why creativity plays an essential role in human development of art and how this can be applied to music and specifically on vocal production.

Tutorial 25 — Tuesday, June 7
12:30 – 13:45 — Room 351

ZEN-MEETS-TECH EXPERIENCE

Presenter: **David Miles Huber**

Four-times Grammy-nominated electronic artist David Miles Huber will be presenting "Parallax Eden," which has been released on the Auro Technologies own Auro-3D Creative Label s a Pure Audio Blu-ray with an exceptional Auro-3D, stereo, and surround mixes.

The Auro-3D mix creates an exciting three-dimensional soundscape thanks to the addition of a Height layer in front, above and all around the listener. Auro-3D is an exciting move from two-dimensional surround sound formats to three-dimensional sound and is the only sound system on the market that has both 5.1 Surround and Auro 9.1 in one standard PCM delivery file with high resolution audio in each channel, making easy distribution on any system in the world possible.

"Parallax Eden" is a Grammy-nominated project that is unique in more ways than one. Composed, produced, and mixed by electronic music artist, David Miles Huber, this project has been layered in various multi-dimensional ways. Not only does it offer the listener mixes in various format versions (Stereo, 5.1 Surround, and last but not least Auro 9.1 immersive sound), Parallax Eden offers up to three completely different mix versions. All three formats will hold a different mix of the original compositions, ranging from the "Chill Mixes" up to the uptempo "Berlin Remixes." Musically, "Parallax Eden" transports the listener to a new place, where down-tempo beats are carefully crafted with up-tempo grooves in imaginative ways, taking you on a journey that's both tranquil and energizing.

Professional Sound Expo
Tuesday, June 7, 13:00 – 13:45
PSE Stage

RIBBON MICROPHONES

Presenter: **Sammy Rothman**

What are ribbon microphones and how do you use them? Sammy Rothman answers these questions and more in "AEA Ribbon Mics: Fix It in the Mic" which delves into all things ribbons including best miking practices and how ribbons mics work. The discussion will address the differences between ribbon and condenser microphones, between active and passive ribbon microphones and between various approaches to ribbon microphone design.

Session EB5 — Tuesday, June 7
14:00 – 15:30 — Room 353

ENGINEERING BRIEFS—LECTURES

Chair: **Emiliano Caballero Fraccaroli**, Electric Lady Studios, New York, NY, USA

14:00

EB5-1 An Investigation into Kinect and Middleware Error and Their Suitability for Academic Listening Tests—*Thomas Johnson, Ian Gibson, Ben Evans, Mark Wendl*, University of Huddersfield, Huddersfield, West Yorkshire, UK

This paper investigates the accuracy and error introduced by middleware applications when used with the Kinect. The middleware applications (Synapse and GMS v3.0) were tested to quantify the error they introduce compared to the error of the Kinect and assess their suitability for use in academic listening tests.
Engineering Brief 273

14:15

EB5-2 How Can Actor Network Theory and Ecological Approach to the Perception Be Used to Analyze the Creative Audio Mixing Practice?—*Yong Ju Lee*, University of West London, London College of Music, London, UK

In audio mixing, communication between the artist or producer and the mix engineer are crucial elements in creating a track that is authentic and aesthetically pleasing. Through a MA Record Production module "Performance in The Studio," the researcher explored the idea that mix engineers, artists, and producers develop and select appropriate sounds for a track through a process of negotiation. Furthermore, this negotiation occurs through both verbal and non-verbal communication. Specifically, the researcher aims to look at subjective, or as 'vague', metaphorical descriptions and moments where the engineer, producer, and artists agree on the sound by recommendation and by synchronizing their expectations. However, a metaphorical description cannot define an exact meaning as it is insufficient linguistic tools. The researcher uses Actor Network Theory to understand this negotiation between the technical and the creative and the role of this process of communication and cognition in understanding the interaction and synchronization of the participants' mental representation of the mix and mix process. Furthermore, the researcher uses the Ecological Approach to Perception to analyze specific behavior and response from participants in the mixing process.
Engineering Brief 274

14:30

EB5-3 On the Silver Globe Revisited—*Joanna Napieralska,*[1] *Dorota Nowocien*[2]
[1]Frederic Chopin University of Music, Warsaw, Poland
[2]The Felix Nowowiejski Academy of Music, Bydgoszcz, Poland

On the Silver Globe is known to be the preeminent expression of Zulawski's visionary ideas. Its shooting began in 1976, was halted in 1978 by the communist authorities, and then reconstituted in 1988. Digitally restored in 2016 – courtesy of the Polish Film Institute —it had its real first showing on the 20th of February at Lincoln Center in New York, just three days after the director's death and 28 years after its premiere at the Cannes Film Festival, when the mono loudspeaker went down. Sound restoration is an example of the capabilities of modern technology in pursuing compromise between fidelity to the mono 35 mm magnetic original and digital 5.1 cinema in terms of sync, loudness, timbre, and spatial standards.
Engineering Brief 275
[eBrief presented by Dorota Nowocien]

14:45

EB5-4 Analyzing Sonic Similarity in Hip-Hop through Critical Listening and Music Theory—*Denis Martin, Ben Duinker, David Benson*, McGill University, Montreal, Quebec, Canada; The Centre for Interdisciplinary Research in Music Media and Technology, Montreal, Quebec, Canada

The notion of a musical artist's or genre's sound is frequently evoked, but what sonic parameters define this sound? We address this question through an in-depth corpus analysis of 100 critically and commercially acclaimed hip-hop tracks from the genre's golden age that we define as 1986–1996. We operationalize the term sound as the sum of both musical and production parameters. The practices of music theory and critical listening are brought together to analyze each track across 75 musical and production parameters in 9 categories. Through statistical analysis of our data set, we demonstrate that from these parameters we are capable of assembling groups of songs that sound alike. These groups are then compared against pre-existing groupings such as geographical location and recording label.
Engineering Brief 276

15:00

EB5-5 "Space Explorations": Broadening Binaural Horizons with Directionally-Matched Impulse Response Convolution Reverb—*Matthew Lien*, Whispering Willows Records Inc., Whitehorse, Yukon, Canada; Universal Music Publishing, Taipei, Taiwan

More people are listening with earphones than in the history of recorded music. But earphones locate typical audio claustrophobically in-and-around the listener's head due to an absence of localizing information the brain requires to externalize sound. When combined with recent trends to highly compress music, the results are an unnatural and unhealthy listening experience—a dumbing-down of the auditory faculty. But the rise of earphones has also brought binaural technology onto the radar. While most binaural music productions have been limited to capturing live performances within a single space, the pioneering application of directional binaural impulse response convolution reverb paired with directionally-matched binaural studio recordings restores acoustically diverse spatialization to music.
Engineering Brief 277

15:15

EB5-6 An Automated Source Separation Technology and Its Practical Applications—*Alexandre Vaneph, Ellie McNeil, François Rigaud, Rick Silva*, Audionamix

Audio source separation, the process of un-mixing, has long been seen as unreachable, "the holy grail." Recent progress in coupling digital signal processing with machine deep learning puts this process within reach of the typical sound audio engineer. Using our technology, we will demonstrate a few examples of separations focused on isolating voice tracks from fully arranged mixes and the

opportunities that can be realized from this technology in a series of industry case studies.
Engineering Brief 278

Tutorial 26 **Tuesday, June 7**
14:00 – 15:00 **Room 352A**

APPLICATIONS OF BINAURAL PSYCHOACOUSTICS IN AUDIO—DESIGNING SPATIAL AUDIO TECHNIQUES FOR HUMAN LISTENERS

Presenter: **Ville Pulkki**, Aalto University, Espoo, Finland

The human spatial hearing mechanisms are based on signal analysis in the auditory system. The monaural attributes of, and binaural differences between ear canal signals form the time- and frequency-dependent spatial cues. The cues are subconsciously analyzed by the brain, which forms the auditory spatial image. This tutorial will review current knowledge of the psychoacoustics of spatial hearing, in particular the perception of directions and distances of sources in free field and in reverberant conditions. Non-linear signal-dependent spatial audio technologies analyze spatial metadata directly from sound field or spatial audio signals with time-frequency-space resolution matching with hearing mechanisms. Optimally, the technologies should designed to reproduce the spatial cues with just good enough resolution without overdoing. The trends in the technologies will be discussed in the tutorial.

This session is presented in association with the AES Technical Committee on Spatial Audio

Tutorial 27 **Tuesday, June 7**
14:00 – 15:45 **Havane Amphitheatre**

PARAMETRIC SPATIAL AUDIO PROCESSING: AN OVERVIEW AND RECENT ADVANCES

Presenters: **Emanuël A. P. Habets**, International Audio Laboratories Erlangen, Erlangen, Germany
Oliver Thiergart, International Audio Laboratories Erlangen, Erlangen, Germany

Parametric spatial processing is a promising and emerging technique that is fundamentally different from traditional spatial processing techniques. First, a relatively simple sound field model is adopted and the parameters of the model (such as for example the direction of arrival and diffuseness), are estimated in a time-frequency domain. Second, the estimated parameters are used to process the received microphone signals. The compact and efficient representation of the sound field can be used to develop algorithms for different applications. In this tutorial different sound field models and corresponding parameter estimation techniques will be presented. We will then focus on selected applications such as speech enhancement (directional filtering and dereverberation), acoustical zooming, spatial audio communication, surround sound recording, and reproduction.

This session is presented in association with the AES Technical Committee on Spatial Audio

Professional Sound Expo
Tuesday, June 7, 14:00 – 14:45
PSE Stage

DIGITAL MICROPHONES—AES42 AND ALL THAT

Presenter: **John Willett**, Sound-Link ProAudio, Bicester, Oxfordshire, UK; Circle Sound Services, Bicester, Oxfordshire, UK,

AES42 Digital Microphones have been around for over 15 years now and still many people seem to be in the dark about them and the advantages they offer. Over these years the number of microphones has risen to over 40 from three major manufacturers (four if you include the measurement microphone from Microtech Gefell) and an increasing number of interfaces and recorders is becoming available. In this session the presenter will look at what AES42 is all about and discuss the advantages of digital microphones, with a look at the extensions to the standard added in AES42-2010.

Workshop 19 **Tuesday, June 7**
15:15 — 15:45 **Room 352A**

TRUE PHANTOM POWER

Presenter: **Joost Kist**, Phantom Sound B.V., Amsterdam, The Netherlands

The standard for phantom power has existed for 50 years. Still there is room for improvement within the boundaries of the standard P48. The 6K8 ohm resistor acts as a load for the emitter-follower in the microphone. This causes distortion. Replacement of this resistor with an electronic circuit that acts not as a load improves the performance of the microphone considerably. The output is 0.3 dB louder, the THD is reduced, especially the odd harmonics. The noisefloor is lower. Subjective: less Sss sounds, more ambient sound.