AES 148TH CONVENTION PAPER AND E-BRIEF ABSTRACTS

JUNE 2–5, 2020, ONLINE

PAPER SESSIONS: APPLICATIONS

- **Study on Comparison of Individuality of Ear Canal Shape—**
Riki Kimura,[1] ; *Shohei Yano,*[1] *Rui Fujitsuka,*[1] *Naoki Wakui,*[1] *Takayuki Arakawa,*[2] *Takafumi Koshinaka*[2]
[1] NIT, Nagaoka College, Nagoka, Niigata, Japan
[2] NEC Corporation, Japan

Ear acoustic authentication, a type of biometric authentication, uses the acoustic characteristics of the ear canal as a feature. Because ear acoustic authentication acquires features using earphones, the process of authentication is easy, and the method has attracted much attention recently. However, the mechanism of the acoustic characteristics of the ear canal has not been sufficiently studied. In this study we verified two methods—the image matching method and Slicing method. In conclusion, Slicing method was found to outperform the image matching method, based on the results of this study.
Convention Paper 10348

- **Visualization of Differences in Ear Acoustic Characteristics Using t-SNE—**
Rei Ominato,[1] *Shohei Yano,*[1] *Naoki Wakui,*[1] *Shinnosuke Takamichi*[2]
[1] NIT, Nagaoka College, Nagoka, Niigata, Japan
[2] University of Tokyo, Tokyo, Japan

Ear acoustic authentication is a biometric authentication technology that recognizes the acoustic characteristics of the ear canal to authenticate users. However, compared to fingerprints, ear acoustic authentication has not been studied sufficiently with regards to the individuality of the acoustic characteristics of the ear canal. Therefore, a study on the visualization of ear canal acoustic characteristic differences using t-distributed stochastic neighbor embedding (t-SNE) which expresses the similarity in high-dimensional space and estimates the similarity in low-dimensional space, was conducted.
Convention Paper 10350

- **Content Matching for Sound Generating Objects within a Visual Scene Using a Computer Vision Approach—**
Daniel Turner,[1] *Chris Pike,*[2] *Damian Murphy*[1]
[1] University of York, York, UK
[2] BBC R&D, Salford, UK

The increase in and demand for immersive audio content production and consumption, particularly in VR, is driving the need for tools to facilitate creation. Immersive productions place additional demands on sound design teams, specifically around the increased complexity of scenes, increased number of sound producing objects, and the need to spatialize sound in 360 degrees. This paper presents an initial feasibility study for a methodology utilizing visual object detection in order to detect, track, and match content for sound generating objects, in this case based on a simple 2D visual scene. Results show that while successful for a single moving object there are limitations within the current computer vision system used which causes complications for scenes with multiple objects. Results also show that the recommendation of candidate sound effect files is heavily dependent on the accuracy of the visual object detection system and the labelling of the audio repository used.
Convention Paper 10375

PAPER SESSIONS: LOUDSPEAKERS

- **Upper Frequency Limit of Flat Panel Loudspeakers—Evaluation of the Voice Coil Break-Up Modes—**
Benjamin Zenker, Sebastian Merchel, M. Ercan Altinsoy, Dresden University of Technology, Dresden, Germany

During the monitoring of the behavior of different types of exciters attached to the same loudspeaker panel, significant deviations of the upper-frequency limit were discovered. These deviations depend on the resonance of the voice coil former and cannot be explained with the linear T/S parameters. This paper shows the indirect and direct measurement of the voice coil's break-up with two exemplary exciters. Furthermore, an FE-simulation model has been built to validate and visualize these break-up modes. Finally, a prototype with reinforced structure was constructed to increase the resonance frequency and to extend the frequency range of the loudspeaker panel.
Convention Paper 10324

- **The Effect of the Hand Position on Handheld Microphones' Frequency Response and Directivity—**
Eddy B. Brixen, DPA Microphones, A/S Allerød, Denmark, EBB-consult, Smorum, Denmark, The Danish National School of Performing Arts, Copenhagen, Denmark

When performing microphone measurements, all obstacles that may disturb the sound field around the microphone are removed. The microphone may even be suspended by thin wires to ensure that the influence of a mic stand is avoided. In real life, some microphones are managed somewhat differently, handheld microphones in particular. Some artists even prefer to cover most of the microphone grid with their hand ("cupping"). This habit affects the performance of the microphone. This paper presents measurements of the influence of five different hand positions on handheld vocal microphones. Both frequency response and directivity are measured.
Convention Paper 10328

• **Limitations of Single-Surface Horns for Directivity Control—** *Bjørn Kolbrek*, Celestion International Ltd., Ipswich, UK

From the early days of cinema sound and sound reinforcement, the directivity of loudspeakers has been an important problem. Early solutions to this problem was clusters of exponential horns and multicell horns. Multicell horns are expensive to produce, and therefore other solutions were developed, like the radial horn, the EV CE-horn, Altec's Manta-Ray horns, and the JBL Bi-radial horns. These horns are what can be called single-surface horns. However, such horns have severe restrictions on what is possible in terms of upper and lower break frequencies and acoustic loading. In fact, given the throat size, upper break frequency and coverage angles, the design is fixed, and only the mouth size can be varied. This paper will explain why.
Convention Paper 10334

• **Metamaterial Absorber for Loudspeaker Enclosures—** *Sebastien Degraeve, Jack Oclee-Brown*, GP Acoustics (UK), Ltd., Maidstone, UK

Acoustic metamaterial absorbers can realize previously unattainable absorption spectra with sub-wavelength dimensions approaching the theoretical minimum. Such an optimal metastructure is presented in this work and implemented in a loudspeaker drive unit. The strategy is discussed and the engineering challenges are highlighted. Special attention has been paid to optimize the driver-absorber coupling and preserve the unique properties of the metamaterial absorber by using a one-parameter horn and an exact impedance match at the interfaces. The results are finally compared to exponentially tapered tubes, demonstrating the superiority of the metamaterial approach, not only in terms of performance but also versatility, size and cost.
Convention Paper 10341

• **Predicting Loudspeaker Current Distortion with FEA—** *Marco Baratelli*,[1] Oliver Munroe[2]
[1] Bowers & Wilkins Group Ltd., Southwater, UK
[2] Laboratoire d'Acoustic de l'Université du Mans, Le Mans Cedex, France

In this work COMSOL Multiphysics is used to simulate the current distortion produced in a blocked coil loudspeaker using the sweep method. First, the normal magnetization curve and hysteresis loops of the soft magnetic material involved are measured. Then, current distortion measurements are performed on the loudspeaker motor under test. Finally, time domain models are developed first using a lossless soft magnetic material (BH curve) and then using a hysteretic model (Jiles-Atherton). The simulations are then compared to the measurements and the results discussed. The use of lossless magnetic material models strongly underestimates odd harmonics, proving that if a correct prediction of current distortion is desired it is necessary to accurately model magnetic hysteresis.
Convention Paper 10343

• **Reducing Condenser Microphone Distortion Using a Single Chip Phantom Power Circuit—***Axel Germann*,[1] *Joost Kist*[2]
[1] Prema Semiconductor GmbH, Mainz, Germany
[2] Phantom Sound B.V., Amsterdam, The Netherlands

A phantom power supply increases the load on the output circuit of a condenser microphone and produces harmonic distortion. This effect can be cancelled by substituting two controlled current sources for the phantom power supply that reduce the load on the microphone output. This article presents measurements obtained with the crucial parts of a single chip solution. The circuit was produced in an 80 V bipolar process. Distortion artefacts were reduced by up to 4 dB. A theoretical derivation for predicting the distortion of a phantom power circuit is given. Calculations based on the nonlinearity of transistors used in microphone output stages are compared with measurements and simulations.
Convention Paper 10347

• **The Modeling of a Finite Thickness Short-Ring for Lossy Blocked Voice Coil Impedance—***Isao Ginn Anazawa*, Ny Works, North Vancouver, BC, Canada

When a voice coil is surrounded by a conductive medium, such as a short-ring or pole piece in close proximity, the impedance becomes lossy. The conductive medium impedance itself derived from the well-known skin depth – impedance relationship that describes the impedance has the frequency-dependent of $\sqrt{\omega}$. This frequency dependence holds when the effective short-ring thickness is sufficiently thicker than the skin depth. In practice, many short-rings do not have sufficient thickness for mid to low frequency. Therefore the lossy voice coil impedance characteristics deviate from the idealized lossy impedance model. The short-ring thickness is implemented in the lossy impedance modelling. Then it is applied in a transformer-based blocked voice coil impedance model.
Convention Paper 10351

• **Development and Analysis of a MEMS Based Multi-Driver Headphone to Introduce Individualized Pinnae Frequency Shaping—***Alexander Vilkaitis, Michele Lucchi*, USound GmbH, Vienna, Austria

This paper discusses the process of developing and analyzing a multi-driver headphone for spatial audio using MEMS loudspeakers to reproduce pinnae dependent frequency shaping effects. MEMS Directional Transfer Functions from the headphone drivers are compared to the Kemar Head Directional Transfer Function using a broadband correlation function and in 1/3 octave band regions as per Blauert. Results show a statistically significant correlation between the MDTFs and HDTFs with correlation coefficients in the region above 0.6.The headphones do reproduce the individualized frequency shaping effects in the near field similar to those caused by the far field HRTF. The headphones are best able to reproduce the cues for front elevated and rear sources which correspond to MEMS speaker positions M2 and M4.
Convention Paper 10356

• **Non-Linear Acoustic Losses Prediction in Vented Loudspeaker Using Computational Fluid Dynamic Simulation—***Yves Pene, Yoachim Horyn, Christoph Combet*, L-Acoustics, Marcoussis, France

Bass-reflex designs can exhibit strong non-linear behavior around their resonant frequency with significant acoustic losses and parasite noise emission. These phenomena are mainly due to turbulences and flow separation at the port's inlet and outlet. This work proposes a method to predict the resulting non-linear acoustic losses for a given loudspeaker, enclosure volume, and port geometry. The approach consists of coupling computational fluid dynamics (CFD) simulation with loudspeaker non-linear motion modelization. Four different ports geometries mounted on one given loudspeaker enclosure are tested. The computed acoustic losses are compared with measurements and show a good agreement. The obtained results prove that the proposed method can predict non-linear losses with an average error less than 1 dB around the Helmholtz frequency.
Convention Paper 10359

• **Acoustic Vibration Analysis of the Shell-Structured Distributed Mode Loudspeaker—***Tasuku Kurosawa, Kan Okubo*, Tokyo Metroolitan University, Hino-city, Tokyo, Japan

Since the 1990s, Distributed Mode Loudspeaker (DML) has been widely researched and developed. The DML can provide wide-

band-characteristics and wide directivity in spite of a single diaphragm by using bending waves based vibration on the plate for acoustic radiation. This property is significantly different from traditional loudspeakers. However, the DML has structural problems (e.g., maximum dimension becomes large), because it uses a two-dimensionally spread flat plate. In this paper the authors examine the characteristics of shell-structured loudspeakers whose diaphragm is formed with an arbitrary curvature and compare them with conventional DML. Our results strongly suggest that the problems of the DML using a flat plate can be improved by employing a shell structure for the diaphragm.
Convention Paper 10363

PAPER SESSIONS: NETWORK

• **Generative Adversarial Networks for Audio Equalization: An Evaluation Study**—*Giovanni Pepe*,[1,2] *Leonardo Gabrielli*,[2] *Stefano Squartini*,[2] *Luca Cattani*,[1] *Carlo Tripodi*[1]
[1] ASK Industries Spa, Reggio Emilia,, Italy
[2] Università Politecnica Delle Marche, Ancona, Italy

In this paper we propose a neural network-based approach for audio equalization inside a car cabin. We consider the Generative Adversarial approach to generate FIR filters for binaural equalization at the driver listening position of the sound produced by multiple loudspeakers. The neural network is optimized to generate equalizing filters able to achieve a flat frequency response at one control position in a time-invariant scenario. Results are analyzed in the frequency domain, comparing the achieved frequency response with the desired one. Compared to previous works, the proposed approach provides better results with a very low error compared to the target response.
Convention Paper 10367

PAPER SESSIONS: PERCEPTION

• **Mix Clarity Prediction Based on a Multi-Resolution Inter-Band Relationship Measure**—*Andrew Parker, Steven Fenton*, University of Huddersfeld, Huddersfield, UK

Previous work proposed a measure of Inter-Band Relationship (IBR) as a measure to predict punch and clarity in a musical piece. This paper proposes a multiresolution approach to increase the effectiveness of the measure when used to objectively measure mix clarity. Two listening tests were performed to elicit mix clarity scores for different mixes of the same song and songs of different styles. These scores were used to evaluate the IBR measures. The results indicate that the multiresolution IBR approach is more effective than the original measurement at predicting mix clarity. Spearman correlation values of $rho = 0.7697$ & $rho = 0.7882$ were achieved with respect to the two listening tests.
Convention Paper 10340

• **Evaluation of the Perceived Naturalness of Artificial Reverberation Algorithms**—*Stojan Djordjevic*,[1] *Hüseyin Hacihabiboglu*,[2] *Zoran Cvetkovic*,[3] *Enzo De Sena*[1]
[1] University of Surrey, Guildford, Surrey, UK
[2] Middle East Technical University (METU), Ankara,Turkey
[3] King's College London, London, UK

Listening tests were carried out using a modified MUSHRA method to compare the perceived naturalness of reverberation generated using scattering delay networks (SDNs), feedback delay networks (FDNs), CATT-Acoustic modelling, and convolution with recorded room impulse responses. The difference in naturalness ratings achieved by reverberation generated using FDNs and SDNs was statistically significant, with the mean rating being 12% higher for SDN stimuli than for FDN stimuli. It was also found that CATT-Acoustic models which had been simplified to a bare rectangular room received lower ratings than models that included furniture or irregular room shaping, suggesting that the scattering and mixing effects of irregularities cause improvements in perceived naturalness of the generated reverberation.
Convention Paper 10353

• **Are Full-Range Loudspeakers Necessary for the Top Layer of Three-Dimensional Audio?**—*Toru Kamekawa, Atsushi Marui*, Tokyo University of the Arts, Adachi-ku, Tokyo, Japan

When a human perceives a space by hearing, horizontal sound image localization and spaciousness are sensed based on ILD (Inter-aural level difference) and ITD (Inter-aural time difference). However, in the vertical direction, spectral cue and directional band caused by the difference in the frequency characteristic of the direction of arrival of sound caused by the shape of the ear are crucial. The authors investigated the difference between the original 22.2 multichannel sound and its filtered sound by limiting the playback frequency band of its top layer using various contents. The results demonstrated that there were no significant differences in spatial impression, even if the top layer does not have a band below approximately 400 Hz.
Convention Paper 10362

• **Influence of Horizontal Loudspeaker Layout Geometry on Sweet Area Shape for Widened/Diffuse Frontal Sound**—*Lukas Gölles*,[1,2], *Valerian Drack*,[1,2] *Franz Zotter*,[2] *Matthias Frank*[2]
[1] University of Technology, Graz, Austria
[2] University of Music and Performing Arts, Graz, Austria

The sweet area in which listeners perceive plausible images of virtual sound sources are known to improve with the Ambisonic rendering order and typically also with the radius of the loudspeaker layout. Partly, this knowledge stems from experiments using a rectangular loudspeaker layout, partly from experiments with a circular layout. This bears the question: Does the geometry (circle, square, wide or long rectangle layout) affect the sweet area shape and size? Our paper presents comparative listening experiments using different geometries to render a frontal sound through an Ambisonic widening/diffuseness effect. Although theory would assume the circular geometry as its ideal, a wide rectangular geometry tends to yield slightly more favorable properties.
Convention Paper 10369

• **Comparing Speech Identification under Degraded Acoustic Conditions between Native and Non-Native English Speakers**—*Filippo Ciarla*,[1]; *Eugenio Donati*,[2] *Christos Chousidis*[2]
[1] KP Acoustics, London, UK
[2] University of West London, London, UK

English became, without a doubt, the international language; and many people are facing the challenge to study and work in an English-speaking environment. However, the ability of understanding, when it comes to non-native English speakers, can be affected by the acoustic characteristics of the environment. The research presented in this paper is focusing on higher education. The project investigates the differences in understanding between the native and non-native English speakers and especially the way that this is affected by the acoustic characteristics of the classrooms and the general acoustic environment. Participants from both categories took part in this project by taking a number of diagnostic rhyme test under a controlled noise-changing environment. The outcome of the research shows that the acoustic characteristics and the noise conditions of the classroom affect the performance of understanding, for both native and non-native English speakers, in different ways. Therefore, new

intelligibility threshold needs to be defined for the design of these spaces when they are used for a diverse audience.
Convention Paper 10374

- **A New Approach to Predicting Listener's Preference Based on Acoustical Parameters—***Peter Critchell, Ludovico Ausiello*, Southampton Solent University, Southampton, UK

Since its conception, the study of room acoustics has explored the links between acoustical parameters and subjective preference. While there have been attempts to combine such metrics, e.g., Frick's combination of six acoustical parameters to predict "*acoustic quality*," no reliable method for prediction of listeners' preference has been univocally ascertained or included in any ISO standard. In this study an alternative perspective is presented—to derive a simple descriptor, "*Preference Rating*" (PR), through meta-analysis of metric-preference relationships, within the context of Rock and Pop venues. A statistical approach has been taken to determine the relative importance of a chosen set of factors in the form of mathematical weights. Results of this pilot study indicate that preference may be predicted by using eight acoustical parameters: Reverberation Time (RT), Bass Ratio (BR), Tonality (TN), Definition (D50), Early Decay Time (EDT), Bonello Distribution (MD), Background Noise, and Surface Diffusivity Index (SDI). Quantitative data and subjective evaluation data describing 20 venues (provided by Dr. Adelman-Larsen) were used to validate this new approach and showed strong correlation in 85% of the scenarios. This suggests that the rationale behind the presented method is meaningful and can be used to set a base upon which further testing and development can be conducted to improve the reliability of such empirical approach.
Convention Paper 10378

- **Comparing Training Effects Associated with Two Sets of HRTF Data on Auditory Localization Performance—***Sungyoung Kim,*[1] *Song Hui Chon,*[2] *Hiraku Okumura,*[3] *Shichi Sakamoto*[4]
[1] Rochester Institute of Technology, Rochester, NY, USA
[2] Belmont University, Nashville, TN, USA
[3] Yamaha Corporation, Hamamatsu, Shizuoka, Japan
[4] Tohoku University, Nashville, TN, USA

In this study we investigated the influence of specific generalized HRTF data on auditory localization in the context of augmented reality (AR). The localization training performance was compared over two weeks between two groups, each of which had received training using a different set of generalized HRTF data. The post-training results showed that training was more effective with one specific HRTF set. In particular, this HRTF set led to better performance in two following aspects: (1) its higher scores in the pre-training test enabling the first-time participants to be more accurate; and (2) its consistency over the entire training period, which demonstrates that the adaptation acquired with this particular set was easier to generalize in a more stable way.
Convention Paper 10379

PAPER SESSIONS: RECORDING, PRODUCTION, EDUCATION

- **Body-Controlled Sound Field Manipulation as a Performance Practice—***Damian Dziwis, Tim Lübeck, Christoph Pörschmann*, University of Applied Sciences Cologne, Cologne, Germany

Spatial composition represents a key aspect of contemporary acousmatic and computer music. The history of spatial composition practice has shown many different approaches of composition and performance tools, instruments, and interfaces. Furthermore, current developments and the increasing availability of virtual/augmented reality systems (XR) extend the possibilities in terms of sound rendering engines as well as environments and tools for creation and experience. In contrast to systems controlling parameters of simulated sound fields and virtual sound sources, we present an approach of XR-based and real-time body-controlled (motion and biofeedback sensors) sound field manipulation in the spatial domain. The approach can be applied not only to simulated sound fields but also to recorded ones and reproduced with various spatial rendering procedures.
Convention Paper 10358

- **Mixing with Intelligent Mixing Systems: Evolving Practices and Lessons from Computer Assisted Design—***M. Nyssim Lefford,*[1] *Gary Bromham,*[2] *David Moffat*[3]
[1] Luleå University of Technology. Pietå, Sweden
[2] Queen Mary University of London, London, UK
[3] University of Plymouth, Plymouth, UK

Intelligent Mixing Systems (IMS) are being integrated into mixing workflows, however, there is little discussion around how these technologies are impacting mixing practices. This study explores the possibilities and pitfalls of IMS by comparing to the use of Computer Assisted Design (CAD) tools in the wider design context. The aim of this paper is to take advice from the field of CAD about the potential benefits and known issues of computer-assistance in creative work thereby allowing audio engineers to take more informed decisions regarding the use of IMS within their workflows.
Convention Paper 10376

PAPER SESSIONS: ROOM ACOUSTICS

- **Visualization of Room Reflections Using a Linear Loudspeaker Array and a Single Microphone—***Cagdas Tuna,*[1] *Albert Prinn,*[1] *Felix Knauff,*[2] *Andreas Walther,*[2] *Emanuël Habets*[1]
[1] International Audio Laboratories Erlangen, Erlangen, Germany
[2] Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

A class of reflector localization techniques use multichannel room impulse responses (RIRs) for the visualization of room reflections on 2D maps. In this paper we focus on a setup comprising a linear loudspeaker array and a single microphone, which has become practically more relevant given the growing popularity of soundbars. First, the state-of-the-art visualization techniques developed for the generation of 2D time-domain direction-of-arrival (DOA) maps describing the temporal evolution of room reflections are reviewed, and then an alternative high-resolution mapping approach is introduced. The performances of the existing techniques and the proposed method are compared using simulated and experimental multichannel RIR data generated in various room conditions.
Convention Paper 10332

- **A Comparison between Plane Wave Cancellation and a Least-Squares Approach for the Zonal Control of Modes in a Rectangular Room—***Tom Bell,*[1,2] *Marco Baratelli*[2]
[1] University of Southampton, Southampton, UK
[2] Bowers & Wilkins, Southwater, UK

A least-squares approach for designing filters can provide an effective way of controlling sound pressure in specific sound zones. This paper compares two least-squares approaches (with two and four controlling loudspeakers) against the plane wave cancellation method, specifically chosen due to its effectiveness in rectangular room conditions. The methods are first compared in an ideal rectangular room scenario and then tested outside

of an ideal set-up to check which method performs best. Findings suggest that the a least-squares approaches consistently outperforms the plane wave cancellation method under all tested situations.
Convention Paper 10342

- **Quality of Musicians' Experience in Network Music Performance: A Subjective Evaluation**—*Konstantinos Tsioutas,[1] George Xylomenos,[1] Ioannis Doumanis,[2] Christos Angelou[1]*

[1] Athens University of Economics and Business, Athens, Greece
[2] University of Central Lancashire, Preston, UK

In Network Music Performance (NMP), audio quality and audio delay are considered to be the most critical variables affecting the Quality of Musicians' Experience (QoME). In order to quantify the extent to which these parameters affect QoME, we executed a pilot study where eight musicians performed music in pairs in a controlled NMP setting and were asked to evaluate eight variables related to perception, while the end-to-end delay and the quality of the exchanged audio were varied. We present the design and execution of this experiment and discuss its results and their implications for the tolerance of musicians to increased delay and degraded audio quality.
Convention Paper 10357

- **Towards Encoding Perceptually Salient Early Reflections for Parametric Spatial Audio Rendering**—*Fabian Brinkmann,[1] Hannes Gamper,[2] Nikunj Raghuvanshi,[2] Ivan Tashev[2]*

[1] TU Berlin, Berlin, Germany
[2] Microsoft Research, Redmond, WA, USA

Parametric spatial audio rendering promises fast and perceptually convincing audio cues that remain playback-system agnostic and enable aesthetic modifications of the acoustic experience within games and virtual reality. We propose a parametric encoder for spatial room impulse responses that is tested with 9 simulated rooms spanning a large range of sizes and reverberation times. A key component of the pipeline is a perceptually inspired model for determining a minimal set of salient early reflections to reduce computational complexity. The results of a listening study with 27 subjects suggest that rendering 6 early reflections is indiscernible from a fully-rendered reference for the tested speech content and frequency-independent room simulations based on the image source method. However, the proposed model requires further improvements with respect to detecting and selecting the most-salient early reflections.
Convention Paper 10380

PAPER SESSIONS: SIGNAL PROCESSING

- **Acoustic Source Localization and High Quality Beamforming Using Coincident Microphone Arrays**—*Jonathan D. Ziegler,[1,2] Hendrik Paukert,[1] Andreas Koch,[1] Andreas Schilling[2]*

[1] Stuttgart Media University, Stuttgart, Germany
[2] Eberhard Karls University Tuebingen, Germany

This paper presents an application-oriented approach to Acoustic Source Localization using a coincident microphone array. Multiple processing blocks are presented to generate a reactive, yet stable Direction of Arrival estimation tuned toward speaker tracking. Building on an energy based scanning method, individual characteristics, such as sound field directivity and static sound source positions are used for adaptive smoothing of the detected angle. The methods and resulting performance gain are discussed for the individual components of the algorithm. Objective performance is evaluated using simulated and recorded data. Audio quality is assessed using listening tests, which show a significant increase in subjective sound quality, noise suppression, and speech intelligibility when combining the tracker with a beamforming algorithm for coincident microphone arrays.
Convention Paper 10321

- **A Deep Learning Approach to Sound Classification for Film Audio Post-Production**—*Guillermo G. Peeters, Joshua D. Reiss*, Queen Mary University of London, London, UK

Audio post-production for film involves, among other things, the manipulation of large amounts of audio data. There is a clear need for the automation of many organization and classification tasks that are currently performed manually and repeatedly by sound engineers, such as grouping and renaming multiple audio recordings. Here, we present a method to classify such sound files in two categories, ambient recordings and single-source sounds or sound effects. Automating these organization tasks requires a deep learning model capable of answering questions about the nature of each sound recording based on specific stereo and monaural features. This study focuses on identifying these features and on the design of one possible model. The relevant features for this type of audio classification and the model specifications are discussed. In addition, an evaluation of the model is presented, resulting in high accuracy, precision and recall values for audio classification.
Convention Paper 10322

- **Analysis of Non-Linear Transfer Functions of a Guitar-Effect-Pedal with a Starving Circuit**—*Masaki Inui,[1] Toshihiko Hamasaki,[1] Menno van der Veen[2]*

[1] Hiroshima Institute of Technology, Hiroshima, Japan
[2] Ir. Bureau Vanderveen, Hichtum, The Netherlands

Despite the recent trend of digital transformation, the popularity of Guitar-Effect-Pedal (GEP) has not declined. This paper describes the complexity of the nonlinear characteristics of analog circuitry in a GEP, which originates from not only clipping diodes but also the operational amplifier. It is well known that variation in the supply voltage of a GEP influences its sound. Based on this fact, we designed a bias starving circuit to control various transfer functions. Particular attention is paid to the difference between the odd and even intermodulation distortion. Those transfer functions were analyzed, using each 9th-order polynomial approximation. The analysis indicates that the analog specific properties beyond the approximation exist under a dynamic range of approximately minus 50 dB.
Convention Paper 10335

- **Prediction of Hearing Loss through Application of Deep Neural Network**— *Samuele Calabrese, Eugenio Donati, Christos Chousidis*, University of West London, London, UK

This paper describes a neural network designed to provide aid in the preventive diagnosis of hearing loss issues. Hearing loss is a widely widespread disability affecting millions of people worldwide. An ananymous dataset is used to train a neural network to evaluate hearing loss in prevention and early diagnosis with the aim of supporting health care by optimizing time and cost. The system is tested using a second set of data and results in a correct evaluation of whether the patient is affected by hearing loss or not.
Convention Paper 10373

- **Investigating Timbral Differences of Varied Velocity Snare Drum Strikes**—*Matthew Cheshire, Ryan Stables, Jason Hockman*, Birmingham City University, Birmingham, UK

Adjusting striking excitation velocity for percussion instru-

ments changes characteristics of the sound output, most notably in loudness and timbre. In this study a listening test is carried out to assess participant abilities in distinguishing between varied velocity snare strikes when the loudness disparity had been removed from recordings made with four common studio microphones. Results indicate that all participants are able to identify different velocities based on timbral differences alone. Temporal and spectral features were then extracted from the recordings to gain insight into which quantifiable differences are present between varied velocity recordings. Analysis revealed various features such as attack and decay time, fundamental frequency, and brightness to have significant differences for the varied velocity snare strikes.
Convention Paper 10382

PAPER SESSIONS: SPATIAL AUDIO

• **Accuracy of Photogrammetric Extraction of the Head and Torse Shape for Personal Acoustic HRTF Modeling—***Aki Mäkivirta,*[1] *Matti Malinen,*[2] *Jaan Johansson,*[2] *Ville Saari,*[1] *A. Karjalainen,*[2] *P. Vosough*[1]
[1] Genelec, Iisalmi, Finland
[2] Kuava, Kuopio, Finland

Photogrammetric computational methods can acquire precise personal head, external ear, and upper torso shapes using video captured with a mobile phone. We analyze the accuracy and repeatability of generating such 3D geometry information. A known head and torso simulator (Kemar) as well as a 3D printed plastic head-and-torso dummy of a real person are considered. The resulting 3D geometry data is compared to the manufacturer's or the actual 3D geometry. Shape differences between geometries are analyzed. The computer-modeled head-related transfer functions implied by the geometries are compared. The impacts of differences in geometries are discussed. Photogrammetric determination of the 3D head-and-torso geometry can have sufficient accuracy to allow the computation of realistic personal head-related transfer function.
Convention Paper 10323

• **Optimized Binaural Rendering of Next Generation Audio Using Virtual Loudspeaker Setups—***Felix Lau, Michael Meier*, Institut für Rundfunktechnik, Munich, Germany

A binaural rendering add-on based on the EBU ADM Renderer has been developed, which renders Next Generation Audio scenes using virtual loudspeaker systems. During development, two optimization approaches emerged regarding rendering quality and efficiency. The first approach concerns the rendering of coherent signals from different emitter positions, which is particularly relevant for virtual speakers since object positioning is based on amplitude panning. The second approach concerns the reduction of computational costs when rendering the binaural room response by using different virtual speaker layouts for the rendering of the direct path and of the room response. To evaluate both approaches, a listening test was conducted. The results of this test showed that each approach positively influenced either rendering quality or performance.
Convention Paper 10325

• **Effects of Rigid Spherical Scatterer on Spatial Audio Reproduction Quality—***Lauros Pajunen,*[1] *Archontis Politis,*[2] *Markus Vaalgamaa,*[3] *Stefan Strömmer,*[3] *Ville Pulkki*[1]
[1] Aalto University, Aalto, Finland
[2] Tampere University, Tampere, Finland
[3] Huawei Technologies, Finland

The spectral quality reproduced by multichannel spatial audio reproduction methods are studied in two regular and one irregular loudspeaker layouts when the human head is analytically modeled as a spherical scatterer. The layouts have such a number of loudspeakers that is found in some high-end laboratory conditions. The results verify that the reproduction quality drops above a certain frequency limit, and the introduced spectral errors are more severe with Ambisonics than with VBAP. The head-like scatterer increases errors on the contralateral side of a sound source, and the errors on the ipsilateral side are mostly unaffected. In addition, the frequency limit of the errors may be lower depending on the source direction.
Convention Paper 10333

• **Dataset Augmentation and Dimensionality Reduction of Pinna-Related Transfer Functions—***Corentin Guezenoc, Renaud Séguier*, IETR UMR CNRS 6164, CentraleSupélec, Cesson-Sévigné, France

Efficient modeling of the inter-individual variations of head-related transfer functions (HRTF) is a key matter to the individualization of binaural synthesis. In previous work we augmented a dataset of 119 pairs of ear shapes and pinna-related transfer functions (PRTFs), thus creating a wide dataset of 1005 ear shapes and PRTFs generated by random ear drawings (WiDESPREaD) and acoustical simulations. In this article we investigate the dimensionality reduction capacity of two principal component analysis (PCA) models of magnitude PRTFs, trained on WiDESPREaD and on the original dataset, respectively. We find that the model trained on the WiDESPREaD dataset performs best, regardless of the number of retained principal components.
Convention Paper 10337

• **Influence of Individual HRTF Preference on Localization Accuracy—A Comparison between Regular and Bone Conducting Headphones—***Tray Minh Voong,*[1] *Christoph Reuter,*[2] *Michael Oehler*[1]
[1] University Osnabrück, Osnabrück, Germany
[2] University of Vienna, Vianna, Austria

In this paper we investigate whether the selection of an individual preferred HRTF via a simple time-saving perceptive evaluation method based on a tournament mode also leads to a good localization accuracy. For many everyday application scenarios for which such a selection process could be useful (e.g., navigation solutions for visually impaired people) it is also helpful to perceive the real acoustic environment at the same time. Therefore, it is also examined whether the use of bone conducting headphones differs from that of regular headphones. The results show that the perceptive judgment criteria used in the tournament task are suitable for selecting an individual HRTF that also allows a good localization of the acoustic stimuli at the same time. No difference could be found between bone conducting headphones and regular headphones.
Convention paper 10345

• **Generalized Low Frequency 3D Audio—***Jacob Hollebon, Filippo Maria Fazi*, University of Southampton, Southampton, UK

There exist many different techniques to reproduce 3D audio over loudspeakers, each derived from different models and motivations. However, at low frequencies these varying reproduction methods appear more similar than previously thought. This paper produces an analytical analysis of the stereo sine law, head-tracked stereo sine law, stereo tangent law/vector base amplitude panning, crosstalk cancellation and first order Ambisonics. Many of these techniques are shown mathematically to be equal or subsets of each other, resulting in a more generalized theory for low frequency audio reproduction over loudspeakers. Finally, the performance of each of the reproduction methods is considered under a low frequency analysis framework derived from a

soundfield reproduction perspective.
Convention Paper 10346

- **Predicting Directional Sound-Localization of Human Listeners in both Horizontal and Vertical Dimensions—***Roberto Barumerli,*[1] *Piotr Majdak,*[2] *Jonas Reijniers,*[3] *Robert Baungartner,*[2] *Michele Geronazzo,*[4] *Federico Avanzini*[5]

[1] University of Padova, Padova, Italy
[2] Austrian Academy of Sciences, Vienna, Austria
[3] University of Antwerp, Antwerp, Belgium
[4] Aalborg University, Aalborg, Denmark
[5] University of Milan, Milan, Italy

Measuring and understanding spatial hearing is a fundamental step to create effective virtual auditory displays (VADs). The evaluation of such auralization systems often requires psychoacoustic experiments. This process can be time consuming and error prone resulting in a bottleneck for the evaluation complexity. In this work we evaluated a probabilistic auditory model for sound localization intended as a tool to assess VAD's abilities to provide static sound-localization cues to listeners. The outcome of the model, compared with actual results of psychoacoustic experiments, shows the advantages and limitations of this systematic evaluation.
Convention Paper 10360

- **Towards Mobile 3D HRTF Measurement—***Song Li, Aly Tobbala, Jürgen Peissig*, Leibniz Universität Hannover, Hannover, Germany

Virtual reality (VR)-based measurement systems have recently been developed to quickly measure individual head-related transfer functions (HRTFs) with arbitrary head movements. However, these systems are usually limited to estimate 2D HRTFs with a fixed distance between the loudspeaker and the subject. In the present study a mixed reality (MR)-based mobile measurement system for fast estimation of distance-dependent HRTFs (3D HRTFs) is proposed. During the measurement, subjects cannot only rotate their heads but also move their bodies towards or away from the loudspeaker. Additionally, the qualities of estimated HRTFs are calculated and displayed to subjects in real-time. Furthermore, the influences of the MR head mounted display (HMD) on the far- and near-field HRTFs are objectively evaluated.
Convention Paper 10364

- **Direction of Arrival Estimation Based on Transfer Function Learning Using Autoencoder Network—***Yiwen Wang, Xihong Wu, Tianshu Qu*, Peking University, Beijing, China

Direction-of-arrival (DOA) estimation based on microphone arrays has been a hot research topic in recent years. Transfer function (TF) based DOA method performs well because it considers both time difference and intensity difference. However, obtaining transfer function is a difficult task and transfer function based method is susceptible to noise. In this paper an autoencoder network structure is proposed for DOA estimation task. The network is used to learn the characteristics of the transfer function, which considers both time difference information and intensity difference information for DOA estimation. The proposed unsupervised training method helps minimize the burden for labeling training data. The evaluation experiments show that our method performs better than TF-based method in the noisy environment.
Convention Paper 10370

- **The Ambisonic Partially Matching Projection Decoding Method for Near-Field Sound Sources—***Zhongshu Ge, Liang Li, Tianshu Qu*, Peking University, Beijing, China

Recently, a new Ambisonic decoding method, the partially matching projection decoding (PMPD) method, has been proposed for irregular loudspeaker layouts. This method, however, works only with real-valued Ambisonic signals. This shortage limits the usage of this method with near-field reproduction problems. In this paper we propose two modifications, the change of criteria to choose loudspeakers when projecting the target and the change of termination condition, to the previous method to make it works for the complex-valued near-field compensated Ambisonic signals. Two evaluation experiments were conducted to test the performance of the modified PMPD method with different loudspeaker layouts. The results showed that the modified PMPD method is available for reproducing near-field sound sources with both irregular and regular loudspeaker layouts.
Convention Paper 10372

- **Multichannel Acoustic Echo Cancellation for Ambisonics-Based Immersive Distributed Performances—***Marcel Nophut, Robert Hupke, Stephan Preihs), Jürgen Peissig*, Leibniz Universität Hannover, Hanover, Germany

Distributed performances of musicians at distant locations are recently enjoying increasing interest due to the availability of larger bandwidths in network and mobile communication. Modern spatial audio capturing and multichannel reproduction techniques could make these performances an immersive and more realistic experience. But a bidirectional acoustic coupling of rooms introduces disturbing echo loops, which calls for Acoustic Echo Cancellation (AEC) methods. This contribution investigates the Frequency Domain Adaptive Kalman Filter, a state-of-the-art AEC algorithm, in a novel and practical context of a distributed music performance including an Ambisonics audio rendering. In particular, the possibility of using Ambisonic channels as reference signals for the echo canceling algorithm is investigated, which allows a significant reduction of the algorithm's computational load.
Convention Paper 10381

POSTERS: APPLICATIONS

- **SpiegeLib: An Automatic Synthesizer Programming Library—***Jordie Shier, George Tzanetakis, Kirk McNally*, University of Victoria, Victoria, Canada

Automatic synthesizer programming is the field of research focused on using algorithmic techniques to generate parameter settings and patch connections for a sound synthesizer. In this paper we present the Synthesizer Programming with Intelligent Exploration, Generation, and Evaluation Library (spiegelib), an open-source, object oriented software library to support continued development, collaboration, and reproducibility within this field. spiegelib is designed to be extensible, providing an API with classes for conducting automatic synthesizer programming research. The name spiegelib was chosen to pay homage to Laurie Spiegel, an early pioneer in electronic music. In this paper we review the algorithms currently implemented in spiegelib and provide an example case to illustrate an application of spiegelib in automatic synthesizer programming research.
Convention Paper 10377

- **An Advanced Audio System for Stereo Reproduction Enhancement—***Domenico Giliberti,*[1] *Festim Iseini,*[2] *Nicola Pelagalli,*[1] *Alessandro Terenzi,*[1] *Stefania Cecchi*[1]

[1] Universitá Politecnica delle Marche, Ancona, Italy
[2] IHP, Frankfurt (Oder), Germany

Stereo audio systems are widely used in different scenarios, espe-

cially for portable devices. For their usability, it is important to develop algorithms capable of improving audio performance. In this context, the proposed system aim to enhance spatial sound reproduction taking advantages of three psychoacoustic effects. Starting from a virtual bass algorithm used to increase the perception of lowest frequencies exploiting the missing fundamental phenomenon, a stereo enhancer is used to add a spatialization effect to sound reproduction in the medium frequency range. Finally, a decorrelator algorithm operating in the high frequency range and capable of improving the apparent source width of the audio signal is applied. The combination of these three effects generates a pleasant perception of the reproduced sound.
Convention Paper 10386

POSTERS: LOUDSPEAKERS

- **2D Sound Field Reproduction with Elliptical Loudspeaker Array Based on Circular Microphone Array Signals—***Yi Ren,*[1] *Kenta Imaizuma,*[2] *Kimitaka Tsutsumi,*[2] *Yoichi Haneda*[1]
[1] The University of Electro-Communications, Tokyo, Japan
[2] NTT Service Evolution Laboratories, Tokyo, Japan

In this paper we propose a method for sound field reproduction using an elliptical loudspeaker array and a circular microphone array. This study aims to reproduce the interior sound field of the elliptical loudspeaker array in a 2-dimensional sound field. The solutions of the Helmholtz equation in an elliptical coordinate system, which are known as Mathieu functions, are used to expand the sound field in an elliptical-coordinate-system-based wave domain. We provide a method to transform the signals recorded by a circular microphone array that are generally expanded with circular harmonics to the Mathieu expansion coefficients. These coefficients can be applied to derive the driving functions of loudspeakers. Computer simulations show that the method is valid for sound field reproduction.
Convention Paper 10327

- **Hybrid Constant Directivity Horn—***Dario Cinanni,* Speaker-LAB, Senigallia, Italy

A new horns family is presented, the Hybrid Constant Directivity (HCD), investigating some practical aspects of constant directivity design through physical and FEA 3D prototypes. Horn driver SPL simulations are conducted using a method already presented to the scientific community and here improved, lead to a minimum mismatch between horn simulations and measurements. A detailed directivity and numerical match of the beam-width are examined with a direct SPL comparison among exponential, tractrix, and spherical expansions. Then, horns aspect ratio is changed obtaining HCD elliptical and rectangular mouth horns referenced and correlated to the circular one SPL simulation. Also wave-front shapes, mouth diffraction effects and radiation impedances are analyzed. Finally, the mathematical model for calculating HCD horns is disclosed.
Convention Paper 10336

- **Physical Characteristics of Analog Audio Cables and Their Effect on Sound Quality—***Akihiko Yoneya,* Nagoya Institute of Technology, Nagoya, Japan

This paper deals with the change in sound quality due to an analog audio cable under a high impedance load. It has been quantitatively shown that the signal amplitude dependency of the time constant affects the audible sound quality. Under the handling conditions, a change in the capacitance of the cable due to the signal causes a change in the sound quality. Then, the transient response and the sensitivity of the capacitance change due to the signal were measured for some cables. In addition to showing the characteristics of each cable, we propose conditions that hardly affect sound quality.
Convention Paper 10338

- **Miniature Omnidirectional Sound Sources for Measurements Applications—***Bartlomeij Chojnacki, Tadeusz Kamisinski, Artur Flach,* AGH University of Science and Technology, Cracow, Poland

Scale modeling and near field measurements require special type of acoustic sources. Despite the omnidirectional directivity patterns usually smallest possible size is required to limit the multi-scattering problem between the source and measured objects. In the given state of art there are only few solutions possible to apply in this kind of measurement. We will present constructions used in Department of Mechanics and Vibroacoustic in AGH, divided between specific functions like scale modeling reverberation time measurements or sound insulation measurements and near-field HRTF measurements. Different parameters used for assessment of miniature omnidirectional sources will be discussed, presenting new approach for design and verification of this type of measurement sources.
Convention Paper 10355

POSTERS: PERCEPTION

- **On the Importance of Impedance for Perceptual Relevance of HRTF—***Slim Ghorbal,*[1] *Xavier Bonjour,*[2] *Renaud Séguier*[1]
[1] CentraleSupelec/IETR, Rennes France
[2] 3D Sound Labs, Paris, France

Head-Related Transfer Functions (HRTFs) are commonly seen as a very promising way to achieve sound spatialization. For this reason, they have become a hot topic over the past decades. A lot of effort has been put to produce as many of them as possible and among the employed methods are the acoustical measure and the numerical simulation. However, noticeable differences exist between the HRTFs derived from acoustic measures of a given subject and the simulated ones that can be computed from his 3D scan. In this paper, keeping focused on one subject, we show that the acoustic impedance parameter used in simulation can explain the observed differences. We present two methods to get a frequency-dependent impedance that leads to perceptually relevant HRTFs. Those HRTFs are subjectively evaluated by localization tests and compared to results obtained with measured HRTF, state-of-the-art HRTF, and non-personalized HRTF. We also discuss the generalization possibility of such approaches.
Convention Paper 10354

- **Multi-Task Based Sound Localization Model—***Tao Song, Tianshu Qu, Jing Chen,* Peking University, Beijing, China

For machine hearing in complex scenes (i.e., reverberation, multi-sound sources), sound localization either serves as the front-end or is implicitly encoded in speech enhancing models. However, extracting binaural cues for sound localization is dependent on the clarity of the input speech signals, and speech enhancing (i.e., dereverberation or denoise) can benefit the processing of sound localization. Based on the idea above, a multi-task based sound localization model is proposed in this study. The proposed model takes waveform as input and simultaneously estimates the azimuth of the sound source and time-frequency (T-F) mask. Localization experiments were performed using binaural simulation in reverberant environments, and results show that compared to the single-task sound localization method, the presence of the speech enhancement task can improve the localization performance.
Convention Paper 10366

- **Minimum BRIR Grid Resolution forInteractive Position Changes in Dynamic Binaural Synthesis—***Annika Neidhart, Boris Reif,* Technische Universität Ilmenau, Ilmenau, Germany

This paper presents a psychoacoustic study on the minimum BRIR grid resolution required for a smooth transition in an interactive listener translation in virtual acoustic environments produced over headphones. A listening area with a size of 1 m × 2 m was set up with a positional resolution of up to 5 cm. 77490 BRIRs were simulated for this purpose. Two listening experiments were conducted to study the perception of discontinuities in the representation with various BRIR grid resolutions. Four different test signals were taken into account. The results show that the required resolution depends on the signal. While for white noise, only the highest resolution (5 cm) brought the best results, for sounds with limited bandwidth lower resolutions appeared to be just as fine.
Convention Paper 10371

POSTERS: RECORDING/PRODUCTION/EDUCATION

- **Augmented Reality for DAW-Based Spatial Audio Creation Using Smartphones—***Adrià M. Cassorla,*[1] *Gavin Kearney,*[1] *Andy Hunt,*[1] *Hashim Riaz,*[2] *Mirek Stiles,*[2] *Damian T. Murphy*[1]
[1] University of York, York, UK
[2] Abbey Road Studios, London, UK

Most tools used to create and manipulate 3D sound are based on a 2D screen GUI. Besides, they are usually controlled using consumer peripherals such as mouse and keyboard. The main concept of this paper is to develop and analyze AR capabilities on smartphones to augment spatial audio production in a DAW. This paper describes the development of an iPhone app that sends OSC control signals to a DAW for real-time spatial panning of sound sources. The research is based on finding different methods to control 3D audio, exploring Apple's ARKit. Results of user tests show that the use of AR to control spatial audio is a feasible option to consider in further research in the audio production industry.
Convention Paper 10344

POSTERS: ROOM ACOUSTICS

- **Round Robin to Determine Reverberation Time—***Markus Zehner,*[1] *Daniel Zurwerra,*[2] *Andrew Goldberg*[3]
[1] www.zehner.ch, Erlinsbach, Switzerland
[2] Virtually Audio GmbH, Erlinsbach, Switzerland
[3] Independent, Finland

An interlaboratory round robin test to measure a hall's reverberation time is performed using different equipment employed by 17 professional participants. All common measuring methods are represented: direct recording of impulse responses using impulse sound sources, indirect impulse response generation using software, and recording interrupted noise. Systematic differences are detected between these methods. Scatterings in measurement results are generally low, and, compared to previous studies, standard deviation is similar however some participants clearly differed in performance from others. The interrupted noise method shows the largest data scatter. Reverberation time specified as a single value shows large differences because the number is determined in a variety of ways. A revision of ISO 3382 is recommended.
Convention Paper 10320

- **Listener-Perspective Dependency of Perceived Auditory Attributes in a Concert Hall—***Bogdan Ioan Bacila, Hyunkook Lee*, University of Huddersfield, Huddersfield, UK

This paper presents a subjective study of spatial impression attributes and their perception in a 6 Degree-Of-Freedom context. For a plausible reproduction of virtual acoustics in a VR/AR/MR it is important to understand how different spatial impression attributes change perceptually for different positions in the room and for different head orientations. An elicitation test was carried out using the Repertory Grid Technique (RGT), in a virtual environment, using Binaural Room Impulse Responses (BRIRs) recorded in a reverberant concert hall. Results show an agreement with current literature and a previous elicitation test. New attributes such as "echo/reverb directionality" were found to also be perceived by participants, as a result of the different listening positions and head orientations.
Convention Paper 10326

POSTERS: SIGNAL PROCESSING

- **Experimenting with 1D CNN Architectures for Generic Audio Classification—***Lazaros Vrysis, Iordanis Thoidis, Charalampos Dimoulas, George Papanikolaou*, Aristotle University of Thessaloniki, Thessaloniki, Greece

During the recent years, convolutional neural networks have been the standard on audio semantics, surpassing traditional classification approaches that employed hand-crafted feature engineering as front-end and various classifiers as back-end. Early studies were based on prominent 2D convolutional topologies for image recognition, adapting them to audio classification tasks. After the surge of deep learning in the past decade, real end-to-end audio learning, employing algorithms that directly process waveforms are to become the standard. This paper attempts a comparison between deep neural setups on typical audio classification tasks, focusing on optimizing 1D convolutional neural networks that can be deployed on various audio in-formation retrieval tasks, such as general audio detection and classification, environmental sound or speech emotion recognition.
Convention Paper 10329

- **Audio-Based Detection of Malfunctioning Machines Using Deep Convolutional Autoencoders—***Iordanis Thoidis, Marios Giiuvanakis, George Papanikolaou*, Aristotle University of Thessaloniki, Thessaloniki, Greece

In this pape, we develop a modular deep convolutional autoencoder with a dense bottleneck structure to perform the task of unsupervised anomaly detection in machine operating sounds. The proposed model consists of multiple sub-networks with identical encoder-decoder structures, trained to learn a mapping function between different mel-scaled frequency bands. Experiments were conducted on the recently introduced MIMII (Malfunctioning Industrial Machine Inspection and Investigation) open benchmark dataset. Experimental results demonstrate that the proposed model yields improved fault detection performance in terms of the Area Under Curve (AUC) metric compared to the baseline approach.
Convention Paper 10330

- **Shelving Filter Cascade with Adjustable Transition Slope and Bandwidth—***Frank Schultz,*[1] *Nara Hahn,*[2] *Sascha Spors*[1]
[1] University of Rostock, Rostock, Germany
[2] Chalmers University of Technology, Göteborg, Sweden

A shelving filter that exhibits an adjustable transition band is derived from a cascade of second order infinite impulse response shelving filters. Two of three parameters, i.e., shelving level, transition slope, and transition bandwidth, can be freely adjusted in order to describe the design specifications. The accuracy of the resulting response depends on the number of deployed biquads

per octave. If this is set too small, deviations in level and bandwidth as well as a rippled slope can occur. The shelving filter cascade might be used in applications, that require a fractional-order slope in a certain bandwidth, such as for sound reinforcement system equalization, sound field synthesis and audio production.
Convention Paper 10339

• **Design of Multichannel FIR Filter Using Gradient Descent Optimizer For Personal Audio—***Seunghun Kim, Tack-Sung Choi*, CTO E&M Center, LG Electronics, Seoul, Korea

Personal audio systems provide an acoustically private space by increasing the acoustic isolation between the bright zone and the dark zone. Multichannel FIR filter is designed for the acoustic isolation by suggesting a target optimization problem and obtaining an optimum solution. As a practical attempt to implement a personal audio system, this paper proposes use of a gradient descent optimizer to computationally find the optimal multichannel FIR filter. The gradient descent optimizer minimizes the defined cost function based on the gradient. The cost function described in this paper is induced from the target optimization problem of the acoustic contrast control method. Experimental results in a real environment demonstrate the effectiveness of this method as achieving more than 20 dB difference at most frequencies under 8 kHz.
Convention Paper 10349

• **Pruning of an Audio Enhancing Deep Generative Neural Network—***Simon Plain, Arijit Biswas*, Dolby Germany GmbH, Nürnberg, Germany

The performance of deep neural networks has been shown to be very effective in multiple arenas. As their use becomes more widespread, increased focus is being placed on how implementable they are on a wider variety of devices. These devices often include those with low enough processing power that the operation of the neural network has a significant impact on storage and computation resources. Here we investigate the impact of pruning weights from a deep generative convolutional auto-encoder with skip connections. The chosen model is trained in a Generative Adversarial Network (GAN) setting for audio post-processing of a low bitrate audio coder. We evaluate performance using a combination of objective scores and listening tests.
Convention Paper 10352

• **Comparison of LMS-Based Adaptive Audio Filters for Identification—***Kristóf Horváth, Balázs Bank*, Budapest University of Technology and Economics, Budapest, Hungary

In the field of audio signal processing, logarithmic frequency resolution IIR filters, such as fixed-pole parallel filters and Kautz filters, are often used. These proven structures can efficiently approximate the frequency resolution of hearing, which is a highly desired property in audio applications. In recursive adaptive filtering however, the FIR structure with LMS algorithm is the most common. Since the linear frequency resolution of FIR filters is not well-suited for audio applications, in this paper we explore the possibility of combining the logarithmic frequency resolution IIR filters with the LMS algorithm. To this end the LMS algorithm is applied to fixed-pole parallel and Kautz filters, and the resulting structures are compared in terms of convergence properties.
Convention Paper 10365

• **A Novel Source Filter Model Using LSTM/K-means Machine Learning Methods for the Synthesis of Bowed-String Musical Instruments—***Hung-Chih Yang, Yiju Lin, Alvin Su*, National Cheng Kung University, Tainan, Taiwan

Synthesis of realistic bowed-string instrument sound is a difficult task due to the diversified playing techniques and the ever-changing dynamics that cause rapidly varying characteristics. The noise part closely related to the dynamic bow-string interaction is also regarded as an indispensable part of the musical sound. Neural networks have been applied to sound synthesis for years. In this paper a source filter synthesis model combined with a Long-Short-Term-Memory (LSTM) RNN predictor and a self-organized granular wavetable is proposed. The synthesis sound can be close to the recorded tones of a target bowed-string instrument. The timbre and the noise are both well preserved. Changes of pitch and dynamics can be easily achieved in real time, too.
Convention Paper 10368

• **Evaluation of Sparse Sound Field Models for Compressed Sensing in Multiple Sound Zones—***Leny Vincesla, Hyun Lim, Ahmet Kondoz*, Loughborough University London, London, UK

To implement a number of sound field reconstruction methods, it is often necessary to get a measure of room impulse responses (RIRs) of a region of interest. However, in many cases this requires a time-consuming effort due to repeated measuring processes. Compressed sensing can provide an alternative solution to obtain RIRs at any location in a domain. In this article we evaluate two different sparse sound field models and a compressed sensing algorithm for the creation of multiple sound zones. RIR estimates are obtained from the sparse models and used to derive the optimal loudspeaker filters. The experimental study indicates a significant improvement of the sound zone system performance from 300 to 3000 Hz using a reduced amount of RIRs.
Convention Paper 10383

• **Comparison of Sound of Organ Pipes in Contemporary and Historical Instruments—***Marta Kalman, Damian Koszewski, Bartlomiej Mróz*, Gdansk University of Technology, Gdansk, Poland

The aim of this research is to examine the differences in the timbre of organ pipes' sound between a historical and a contemporary organ instrument. The historical instrument is the Oliwa organ from Gdansk, Poland, and the contemporary one is from Kartuzy, Poland. Recordings are made of single notes played by an open labial pipe that belongs to the Principal rank. The analyses and comparison of several sound features compatible with audio descriptors defined in MPEG-7 standard are performed in the MATLAB environment. The influence of the distance between the microphone and the sound source on sound features is also examined, in order to judge whether the comb filter appears in close distances from the sound source.
Convention Paper 10384

POSTERS: SPATIAL AUDIO

• **A High-Frequency-Band Timbre Equalization Method for Transaural Reproduction with Two Frontal Loudspeakers—***Lulu Liu, Bosun Xie*, South China University of Technology, Guangzhou, China

A high-frequency band equalization method is proposed to further reduce the timbre coloration in transaural reproduction with two frontal loudspeakers. The high-frequency responses of a pair of filters for transaural synthesis are equalized by a frequency-dependent factor so that the overall power spectra of the responses remains constant, and the low-frequency responses of transaural filers are kept intact. Psychoacoustic experiment validates that the proposed method reduces the timbre coloration in transaural reproduction without introducing obvious perceivable

directional distortion for virtual source in the frontal quadrants of the horizontal plane.
Convention Paper 10331

- **Computed HRIRs and Ears Database for Acoustic Research—** *Slim Ghorbal,*[1] *Xavier Bonjour,*[2] *Renaud Séguier*[1]
[1] CentraleSupelec/IETR, Rennes, France
[2] 3D Sound Labs, Paris, France

This paper describes the different stages of creation of the CHEDAR database. It is comprised of 3D meshes generated from a morphable model of ear, head, and torso and their associated diffuse-field equalized Head-Related Impulse Responses (HRIRs). Focused on the influence of ear, it provides 1253 different ear shapes, making it the largest available HRIR database so far. Moreover, 4 different evaluation grids are used for computation, with radii ranging from 20 cm to 2 m, thus allowing to study near field as well as far field. The frequency range of the corresponding HRTFs goes from 100 Hz to 16 kHz by 100 Hz steps. The HRIRs are provided as .sofa files. The entire database, the largest so far, and unique by the type of gathered data, the number of entries and their resolutions is publicly available for academic purposes.
Convention Paper 10361

- **An Innovative Method for Binaural Room Impulse Responses Interpolation—***Valeria Bruschi, Stefano Nobili, Stefania Cecchi, Francesco Piazza*, Universitá Politecnica delle Marche, Ancona, Italy

The interpolation of room acoustic impulse responses is a widespread technique that allows to reduce measurement sets. In this paper an innovative method for Binaural Room Impulse Responses (BRIRs) interpolation is presented and tested. In the proposed method the BRIRs are decomposed in time and then divided in two frequency bands and an innovative peak detection and matching algorithm is applied for the early reflections, combined with a linear interpolation. A real room binaural impulse responses database has been employed in order to evaluate the algorithm comparing it with the state of the art. Experimental results have proved the effectiveness of the proposed approach.
Convention Paper 10385

E-BRIEFS: APPLICATIONS

- **Reimagining Robb: The Sound of the World's First Sample-Based Electronic Musical Instrument circa 1927 —***Michael Murphy, Richard Anstey*, Ryerson University, Toronto, Canada

This paper follows up on 2013 and 2014 AES presentations on recreating the first successful electronic organ, the 1927 Robb Wave Organ. Most historical literature lists the Chamberlin/Mellotron (c1960s) as the first usable sample-based electronic musical instruments. This paper will situate the earlier 1927 Robb Wave Organ as part of the evolution of sample-based instruments. It will also demonstrate the sampling method used for carving Pulse Amplitude Modulation (PAM) equivalent waveforms into spinning tone wheels. The authors have completed a physical recreation of the instrument and will demonstrate the various sampled sounds at the conference.
Engineering Brief 585

- **Production Vehicle Audio System Validation, Diagnostics, and Calibration—***Steve Hutt*, Equity Sound Investments

Interest in in the topic of production vehicle audio system variance has been growing since the publication of the papers "Loudspeaker Production Variance" (AES 125th Convention 2008, Convention Paper 7530), "Audio System Variance in Production Vehicles" (AES 48th International Conference Automotive Audio, 2012, Conference paper 5-3), and the workshop "Diagnostics for Production Vehicle Audio Systems" (AES 147th Convention, 2019). In this paper the author continues the discussion with a review of current vehicle audio system development and documentation practices and presents a proposal for system level specifications based on a reference system. A procedure is introduced to address validation methods and diagnostic procedures by analysis of complex system data metrics. A process for calibrating production vehicle audio systems to match a reference system is discussed.
Engineering Brief 593

- **Investigating User Interface Preferences for Controlling Background Foreground Balance on Connected TV—***Lawrence Pardoe, Lauren Ward, Hannah Clawson, Aimee Moulson, Chris Pike*, BBC R&D, Salford, UK

Next Generation Audio codecs are gaining traction in consumer devices. Using object-based audio techniques, they can offer significant personalization and accessibility benefits. However, the type of interface and degree of control desired by audiences remains unexplored. This paper describes two investigations. First, a paper prototyping study to broadly understand the most desirable features of object-based audio (n = 11 normal and hard of hearing participants). The second study determined the preferred level of control granularity for a foreground-background control for a cohort of normal hearing participants (n=18). Participants trended towards the more granular controls, with the graduated slider significantly preferable for the documentary and drama content presented. Qualitative data highlights that ease of use and clarity of purpose in controls is key.
Engineering Brief 604

- **Trans-Europe Express Audio: Testing 1000 Mile Low-Latency Uncompressed Audio between Edinburgh and Berlin Using GPS-Derived Word Clock, First with jacktrip then with Dante—***Paul Ferguson,*[1] *Chris Chafe,*[2,3] *Simon Gapp*[3]
[1] Edinburgh Napier University, Edinburgh, Scotland
[2] CCRMA,Stanford University, Stanford, CA, USA
[3] Technische Universität Berlin, Berlin, Germany

For nearly two decades networked audio research using jacktrip has shown that multichannel uncompressed audio was possible over a National Research and Education Network (NREN) and was now becoming viable over some public network connections. There was however, a "Dirty Secret,", in the absence of any synchronization between transmitting and receiving word clocks, periodic audio loss due to data over-run or over-run was a certainty. The authors describe a low-cost GPS-derived clocking solution for jacktrip and then apply it to off-the-shelf Dante equipment and Dante Domain Manager for the world's first long-distance Dante audio over standard academic networks.
Engineering Brief 605

ENGINEERING BRIEFS: LOUDSPEAKERS & MICROPHONES

- **Analysis of Nonlinear Distortions in a Digital MEMS Microphone—** *Przemek Maziewski, Jan Banas, Damian Koszewski, Dominik Stanczak, Andrzej Trella*, Intel Technology Poland, Gdansk, Poland

MEMS microphone's nonlinear distortions are exploited in so-called ultrasonic dolphin attacks. The paper presents an analysis of problematic nonlinearities. For this purpose the theoretical model was used. It allows to specify the frequencies and amplitudes of these distortions. The effectiveness of the model was

illustrated by computer simulations. Then acoustic tests were carried out in a quiet anechoic chamber using a MEMS microphone. During testing, an additional reference microphone was used, which allowed to reduce the impact of non-linearity of non-MEMS test configuration components. Tests illustrate the significant level of MEMS microphone's nonlinear distortions. Some convergence of the theoretical model with results of laboratory tests with potential development directions was also shown.
Engineering Brief 582

- **Development of an Open Source customizable High Order Rigid Sphere Microphone Array**—*Oscar Moschner, Damian Dziwis, Tim Lübeck, Christoph Pörschmann*, Technische Hochschule Köln, Köln, Germany

We developed an analog rigid sphere microphone array with 64 microphones arranged according to the Fliege sampling grid. The array is suited for spherical harmonics processing and allows resolving spatial resolutions up to an order of N = 7. Incorporating a rapid prototyping approach, most parts of the array are 3D printable and script-based modification allows for an easy adaption to other sampling grids and radii. All the parts of the project are freely available under the GNU v3 license.
Engineering Brief 583

- **Power-Based Thermal Limits for Micro-Speaker Protection Algorithms**—*Congcong Peng,[1] Yun Shi,[1] Bixiang Yan,[2] Lianjun Wu,[3] Zhipeng Chen,[1] César Salvador,[1] Deheng Liu[1]*
[1] Silicon Integrated Co., Ltd., Wuhan, China;
[2] Guang Dong, OPPO Mobile Telecommunications Corp., Ltd., Dongguan, China
[3] Shenzhen Realme Mobile Communication Co., Ltd., Shenzhen, China

The increasing demand for high-definition audio in mobile platforms is creating the necessity of adequate operation limits for smart amplifiers to safely maximize the power delivered to micro-speakers. The maximum voice-coil temperature (T_{max}) is an important parameter for protection algorithms that prevent damage. Standard tests used by micro-speaker suppliers to specify T_{max} provide a conservative estimate that depends on the voice-coil materials. Other tests consider the rated noise power, a specification obtained by rigorous lifetime tests. This paper investigates the thermal limits when micro-speakers operate at their specified rated power. These power-based thermal limits are contrasted with values of T_{max} specified in datasheets. The evidence found is used to recommend procedures to measure power-based thermal limits that avoid over-protection.
Engineering Brief 594

- **DIY Modifications for Acoustically Transparent Headphones**—*Nils Meyer-Kahlen,[1] Daniel Rudrich,[2] Manuel Brandner,[2] Stefan Wirler,[1] Simon Windtner,[2] Matthias Frank[2]*
[1] Aalto University, Espoo, Finland
[2] University of Music and Performing Arts, Graz, Austria

Acoustically transparent headphones are an important tool for Augmented Reality (AR) research. To supplement a real acoustic scene with virtual sound sources or virtual acoustics, previous studies have used the very open AKG K1000. Since this model was discontinued and is expensive on the second-hand market, we present modifications to a regular open headphone. Measurements show that cutting out pieces of the ear-pads can yield transparency characteristics comparable to the K1000. Fully surrounding transparency measurements of the original and the modified models are presented. Moreover, based on headphone transfer function measurements, minimum-phase filters are derived, to compensate for the loss of low frequencies after modification. The instructions and open measurement data make it easy to replicate this DIY solution.
Engineering Brief 603

- **A Personal, 3D Printable Compact Spherical Loudspeaker Array**—*Valerian Drack,[1,2] Franz Zotter,[2] Natasha Barrett[3]*
[1] University of Technology, Graz, Austria
[2] University of Music and Performing Arts, Graz, Austria
[3] Norwegian Academy for Music, Oslo, Norway

Compact spherical loudspeaker arrays exhibit controllable directivity, and electroacoustic music recently became interested in using their beamforming capacities to compose spatial music. In particular, directivities of 3rd-order or higher turned out to offer a sufficiently precise horizontal control. Our engineering brief proposes the 170 compact spherical loudspeaker design that is DIY-3D-printable and only uses 8 audio channels. The 170 array houses seven 2.5" broadband transducers that allow third-order horizontal beamforming and a 6" subwoofer. We make the CAD model, electroacoustic measurements, and control filters openly accessible and show it's beampatterns for verification, based on those measurements.
Engineering Brief 607

- **An Improved Loudspeaker Model to Study Coupling in an Active Noise Control Setting**—*Samira Mohamady,[1] Niclas Broemme,[2] Allahyar Montazeri[3]*
[1] IAV GmbH, Munich, Germany
[2] Technical University of Ilmenau, Ilmenau, Germany
[3] Lancaster University, Lancaster, UK

The previous numerical studies confirms that the effect of coupling among loudspeakers cannot be neglected in a multiple channel active noise control (ANC) system with several loudspeakers and microphones, placed in a confined space. Depending on the electromechanical properties of the loudspeakers, the coupling happens through the enclosure acoustic modes and this will change the radiation impedance of the loudspeakers. As a result, the performance of the ANC system influence due to the changes of the supplied voltage to the secondary loudspeakers. Nevertheless, the previous investigations rely on the simplified model of the loudspeakers as a solid piston in the ANC frequency range of interest. In this study we aim to generalize these results by including a detailed finite element model (FEM) of the loudspeaker in which the electromechanical impedance of a real loudspeaker is simulated numerically and validated experimentally afterwards. The validated loudspeaker is used for the numerical simulation of the ANC system inside an enclosure using the finite element method
Engineering Brief 610

ENGINEERING BRIEFS: PERCEPTION AND EVALUATION

- **Investigation on the Perceived Audio Quality of Higher Order Ambisonics Recordings with a Reduced Number of A-Format Audio Channels**—*Alan Pawlak,[1] Łukasz Januszkiewicz[2]*
[1] University of Huddersfield, Huddersfield, UK
[2] Zylia Sp. z o o., Poznan, Poland

This project examines the effect of the reduced number of A-format audio channels obtained from a spherical microphone array on the overall quality of the 1st and 3rd order Ambisonics. The listening experiment employed four test scenarios involving two recording environments and two microphone positions. The evaluation was carried out using MUSHRA methodology with additional implementation of the scene rotation. The results suggest potential degradation in the quality of the 3rd order Ambisonics when the number of A-format channels is reduced. That was

found especially for the recordings made in a controlled environment. No effect of a reduced number of A-format channels was observed for the 1st order Ambisonics in any test scenario.
Engineering Brief 575

ENGINEERING BRIEFS: RECORDING, PRODUCTION, EDUCATION

- **Design of the Acoustic Enhancement System for Crowd Enhancement in the Qatar University Sports & Events Complex—***Takayuki Watanabe*,[1] *Dai Hashimoto*,[1] *Hideo Miyazaki*,[1] *Mohanad Is-Hak*[2]
[1] Yamaha Corp., Hamamatsu, Shizuoka, Japan
[2] Techno Q, Doha, Qatar

The Qatar Univ. Sports & Events Complex, which includes a Multi-Purpose Hall and a Training Hall was completed in November 2019. In addition to sports events, various other kinds of events, including lectures and pop concerts, will be held at this venue. To realize high speech intelligibility of the PA system, the entire interior except the floor is covered with acoustically absorptive materials. However, the resulting highly absorptive spaces are still required to provide an exciting crowd experience during sports events, and thus the Active Field Control (AFC) system was adopted to enhance loudness and reverberance. This engineering brief gives an overview of that system and the benefits achieved by its introduction.
Engineering Brief 572

- **Barrier Effect at Open-air Concerts, Part 1—***Joan F. La Roda Mauro*, Loudspeakercorner

In 2010, as a result of personal experience in concerts and festivals, the question arose as to whether the audience that crowds in front of the subwoofers produces a barrier effect for the very low frequencies. Being aware of the difficulty of conducting an accurate field study it was thought of using simulations; first with the FDTD method and then with COMSOL, both made by professors and students of the Polytechnic University of Gandia, belonging to the UPV. The aim of this study and the simulations is to find indications that in certain circumstances this phenomenon, which began to be called "Barrier Effect," occurs. It is not sought to quantify it precisely since the possible situations and conditions are infinite.
Engineering Brief 573

- **Music and Space: A Case of Live Immersive Music Performance with the Norwegian Post-Rock Band *Spurv*—** *Zachary Bresler*, University of Agder, Kristiansand, Norway

Normative practices for live popular music performances are seldom challenged, in particular those norms related to the staging of artists and the relationship of the audience to the stage. During a run of three concerts held at the University of Agder in collaboration with two music festivals in September 2019, such norms were reconsidered in a series of experimental immersive performances with the Norwegian post-rock band Spurv and electronic music artists Zachary Bresler and Kristian Isachsen. In this brief the background, organization, and setup for these concerts is described, followed by reflections on the performances with the purpose that others might take interest in this experiment and learn from the successes and shortcomings of the project.
Engineering Brief 601

ENGINEERING BRIEFS: ROOM ACOUSTICS

- **Analysis of Acoustic Parameters of Churches in Krakow—** *Piotr Cieslik, Adriana Silwanowicz*, AGH UST, University of Science and Technology, Kraków, Poland

The purpose of the project was to analyze acoustic parameters of churches in Krakow. Buildings in the Romanesque, Gothic, Baroque, and so-called contemporary styles were measured and analyzed. The results were compared with the recommendations described in the PN-B-02151-4: 2016-06 [1] standard, which specifies the acoustic parameters of public spaces. The results were also used to validate the numerical models of the measured churches. During measurements, the receivers were placed in the audience. The technique used in the project was an impulse response of the room using single, omnidirectional impulses. The measurements were made in accordance with the PN-EN ISO 3382-2. [2] standard.
Engineering Brief 579

- **Impact of Room Acoustics on Perceived Vocal Fatigue of Staff-Members in Higher-Education Environments: A Pilot Study—** *Sebastian Duran, Ludovico Ausiello*, Solent University, Southampton, UK

Although it is demonstrated how vocal fatigue and discomfort of teachers can be correlated to room acoustics properties of teaching environments [1, 2] the BB93 English normative (2015) does not provide acoustic requirements for higher-education learning environments. The presented pilot study aims to expand previous research regarding acoustic design criteria for higher education learning environments [4]. A comparative analysis between subjective and objective data collected in two lecture theaters at Solent University suggests how both excessive background noise levels and too low reverberation time values might lead to vocal fatigue issues, confirming the necessity to extend the current UK regulations and include in its scope Higher education learning environments.
Engineering Brief 584

- **Deep Diffusorbers: Real World Application and Scalability of Amplitude Grating Diffusers Covering Ultra-Wideband Absorbers in Small to Medium Sized Venues—***Gernot Ebenlechner*, Gernot Ebenlechner e.U., Gebirge, Austria

Modern sound systems are capable of reproducing audio at substantial levels down to 20 Hz. It is therefore of major interest to control room decay rates down to the lowest octaves even in the most demanding situations. To overcome space requirements and performance limitations of velocity-based absorbers below 80 Hz and at the same time provide sufficient diffusion and tailor-made absorption characteristics at mid and high frequencies ultra-wideband diffusorbers have been emerged. Real world application together with before-after measurements of such devices individually constructed of layered amplitude grating diffuser panels, porous absorber elements, and metallic rectangular Kirchhoff plates with free boundaries on resilient sheet show their performance and scalability in a variety of rooms ranging from 900 to 280,000 cubic feet volume.
Engineering Brief 612

ENGINEERING BRIEFS: SIGNAL PROCESSING

- **Computational Efficient Real-Time Capable Constant-Q Spectrum Analyzer—***Felix Holzmüller*,[1,2] *Paul Bereuter*,[1,2] *Philipp Merz*,[1,2] *Daniel Rudrich*,[1] *Alois Sontacchi*[1]
[1] University of Music and Performing Arts Graz, Graz, Austria
[2] Graz University of Technology, Graz, Austria

The constant-Q transform (CQT) is a valuable tool for music information retrieval, e.g., for chroma calculation and harmonic analysis. In this e-Brief we propose a block based, real-time

capable, efficient analysis algorithm resting upon a subsampling technique performed with fast Fourier transform. In addition, advanced features such as time resolution enhancement towards lower frequencies and a robust CQT-based tuner are presented. Finally a reference-implementation in C++ in form of a VST3-Plugin is introduced. The plugin's source code will be available openly for further development.
Engineering Brief 567

- **Analysis of an Alternative Approach to Digital Domain Volume Control Claiming High Perceptual Audio Quality—***Thierry Heeb, Tiziano Leidi*, SUPSI, Institute for Information Systems and Networking, Manno, Switzerland

In an increasingly digital environment, digital domain volume control offers a cost-effective alternative to analogue domain implementations. However, it has not found widespread use in the upper segment of the consumer audio market. Even if advances in digital to analogue converters and signal processing technologies have addressed the shortcomings of early days' implementations, many audiophiles still complain about inferior sound quality of digital solutions. In this Engineering Brief we analyze an alternative approach to digital domain volume control said to have scored high in casual subjective listening tests. Looking at bit-level arithmetic and information propagation considerations, we present objective elements that distinguish the new volume control from traditional approaches. These may explain the reported superior perceived audio quality.
Engineering Brief 599

- **Tutorial on Scaling of the Discrete Fourier Transform and the Implied Physical Units of the Spectra of Time-Discrete Signals —** *Jens Ahrens, Carl Andersson, Patrik Höstmad, Wolfgang Kropp*, Chalmers University of Technology, Gothenburg, Sweden

The combination of the time-discrete property of digital signals together with the commonly employed definition of the discrete Fourier transform (DFT) can cause ambiguity when interpreting magnitude spectra with respect to the physical unit of the signal under consideration. Standardized scaling of spectra increases the comparability of frequency-domain data that are published in scientific articles or data sheets of commercial products. We present and discuss in this tutorial a collection of the most relevant scaling options for DFT spectra to yield amplitude spectra, power spectra, and power density spectra, and we illustrate how an implied physical unit of the underlying signal is reflected by the magnitude of the spectrum. The tutorial is accompanied by Matlab/Octave scripts that demonstrate the different cases.
Engineering Brief 600

- **A Novel and High Efficient Simulation Model of Loudspeaker—** *Huixian Cao, Zhiwen Chen*, Tymphany Acoustic Technology Co., Ltd., Huizhou, China

As is known to all, a loudspeaker is a multi-physics coupling model. A fully coupled multi-physics simulation for 3D drivers is time-consuming or even unavailable in practical applications. To overcome this, this paper presents a novel model for simulating the performance of a loudspeaker with high efficiency, which is very useful for 3D simulation. The model needs a mechanical finite element model to simulate the vibration of driver. At the same time, the electrical and acoustical properties are included by using formulas of lumped model and Rayleigh integral in the same software. This model is easy to be practiced with a higher efficiency and less resource of hardware, which can provide a powerful simulation tool for transducer engineers.
Engineering Brief 602

ENGINEERING BRIEFS: SPATIAL AUDIO

- **Audio Portraiture Sound Design—The Development and Creation of Audio Portraiture within Immersive and Binaural Audio Environments—***Maree Sheehan*, Auckland University of Technology, Auckland, New Zealand

This paper examines the sound design development and creation of audio-portraiture within immersive and binaural audio environments. Sound design as a creative inquiry into the development of audio-portraiture explores how sound has the ability to sonically represent the multi-dimensional facets of humans. This study particularly looks at Maori women from New Zealand as the participants. As a platform for gathering and designing significant sonic representations of these women, the utilization of both immersive and binaural technologies was used. Importantly, the sound design and audio technologies symbiotically support the manifestation of audio-portraiture that will be discussed in this paper.
Engineering Brief 566

- **Recording First-Order Ambisonics with a Differential Array of Two Dual-Diaphragm Microphones—***Thomas Deppisch*,[1] *Christoph Frank*[2]
[1] University of Technology and University of Music and Performing Arts, Graz, Austria
[2] Austrian Audio GmbH, Vienna, Austria

First-order Ambisonics consists of four signals corresponding to one omnidirectional and three figure-of-eight polar patterns aligned with the Cartesian axes. We obtain these signals from a matched pair of dual-diaphragm microphones, where each diaphragm output is accessible individually. While the signals for figure-of-eight patterns in x- and y-direction are calculated directly from the microphone outputs, the z-direction figure-of-eight signal is calculated as differential signal from omnidirectional patterns of both microphones, resulting in low-frequency attenuation and a phase shift. We equalize the frequency response using a combination of a model-based IIR and a measurement-based FIR filter and provide an open-source plugin performing the processing.
Engineering Brief 589

- **Real-Time Auralization while Having Prepared in Advance for Possible Head Movements—***Mantas Tamulionis*, Vilnius Gediminas Technical University, Vilnius, Lithuania

In real-time binaural rendering (or auralization), it is important to ensure that no artifacts that may result from CPU overload are heard by the listener. The research is based on the work of A. Lindau. The author states that human cannot detect a difference between signals processed with different HRTFs that represent less than 3 degrees of head position change. The method proposed performs pre-filtering and prepares three variants of the auralized signal: one corresponding to the current position of the listener's head and other two required when the head rotates more than 3 degrees to either side. The right signal can be played immediately. This method allows reducing the size of HRTF database, computation time and saving CPU labor.
Engineering Brief 591

- **Creating Virtual Height Loudspeakers for Dolby Atmos and Auro-3D Using VHAP—***Hyunkook Lee, Kacper Borzym*, University of Huddersfield, Huddersfield, UK

This engineering brief evaluates the effectiveness of VHAP (virtual hemispherical amplitude panning) as a tool to create virtual height loudspeakers for Dolby Atmos 5.1.4 and Auro-3D 9.1 content. Previous studies showed that VHAP can effectively render elevated phantom images using only three or four ear-level

loudspeakers. This study subjectively compared the 4-channel and 3-channel VHAP stimuli rendered against the original 5.1.4, in terms of sense of height, sense of width and overall quality of experience. Five excerpts from Dolby Atmos and Auro-3D content of different types (film, classical and EDM) were tested. Results suggest that VHAP is able to create virtual height loudspeakers with no significant level of degradation, and for certain stimuli it can be slightly preferred to the original 5.1.4.
Engineering Brief 592

• **Comparison of Perception of Spatial Localization between Channel and Object Based Audio—***Tomas Oramus, Petr Neubauer*, Academy of Performing Arts in Prague, Prague, Czech Republic

Surround sound has been in almost every cinema for several decades. In 2012 Dolby, Inc. announced a new spatial audio format – Dolby Atmos with object-based audio approach. Thanks to it, sound designers can take advantage of more precise positioning of sounds in space. This paper examines whether the listeners can perceive the position of sounds with higher precision compared to 5.1 and 7.1. Listening tests with 127 respondents were conducted to compare perceived position of six samples, each of them reproduced in 5.1, 7.1 and Atmos. Results do not show increased precision of spatial localization when using object audio; however, further analysis shows significantly different perception of sound positions among the examined formats based on the listener's position.
Engineering Brief 608

ENGINEERING BRIEFS POSTERS: APPLICATIONS

• **Exploring Audio Device Orchestration in Workshops with Young People—***Jon Francombe, Kristian Hentschel, Suzanne Clarke*, BBC Research & Development, Salford, UK

When considering development of a new technology, it is important to account for the wants and needs of the target audience. Two day-long workshops were performed to explore the concept of audio device orchestration—using multiple connected, synchronized devices to create or augment a media experience—with 16- to 18-year-olds. The workshops utilized a variety of ideation techniques including warm-up exercises, idea generation exercises, and co-creation of prototypes. A thematic analysis was performed on the outputs to explore the participants' attitudes to audio technology and device orchestration. The results suggested a strong desire for positive application of technology and content, focusing on issues such as wellbeing and togetherness. The results match closely with previous research on values for digital wellbeing.
Engineering Brief 570

• **Exploring Audio Device Orchestration in Workshops with Audio Professionals—***Kristian Hentschel, Jon Francombe*, BBC Research & Development, Salford, UK

Device orchestration is the concept of using multiple connected, synchronized devices to create or augment a media experience. It enables interactive and immersive experiences for multiple listeners, but there are challenges in producing content for such flexible reproduction. A series of co-creation workshops with audio professionals was conducted to develop ideas and discover workflow issues. Over the course of six workshops, a number of working prototypes were developed using a production tool beta version. A thematic analysis of the ideas generated revealed workflow issues as well as use cases and content ideas. Common themes included gamification and togetherness, correlating well with outcomes from workshops with potential audience members. Applications for augmenting sports, drama, entertainment, and educational formats were also suggested.
Engineering Brief 571

• **HOAST: A Higher-Order Ambisonics Streaming Platform—**
Thomas Deppisch,[1] Nils Meyer-Kahlen,[2] Benjamin Hofer,[3] Tomasz Łatka[4] Tomasz Zernicki[4]
[1] University of Music and Performing Arts, Graz, Graz, Austria
[2] Aalto University, Espoo, Finland
[3] Hofer Web Solutions, Graz, Austria
[4] Zylia Sp. z o. o., Poznan, Poland

The availability of free, user-friendly software tools as well as affordable hardware is boosting interest in higher-order Ambisonics productions not only in research communities but also in the fields of Pro Audio and Virtual Reality. However, there is no practical solution available for presenting such productions publicly in a web browser. The largest commercial platforms, for example, are limited to first- or second-order binaural playback. We introduce the higher-order Ambisonics streaming platform HOAST, a new 360° video-platform, which allows for up to fourth-order Ambisonics audio material. Apart from implementation details of state-of-the-art binaural decoding and acoustic zoom, this contribution describes the current state of multichannel web audio and related challenges.
Engineering Brief 590

ENGINEERING BRIEFS POSTERS: LOUDSPEAKERS AND MICROPHONES

• **MEMS Based Audio Speaker Module—***Markus Hänsler,[1] Giacomo Muraro,[2] Christian Novotny,[2] Richard Murphy,[2] Jakob Spötl,[2] Andrea Rusconi Clerici[2]*
[1] USound GmbH, Graz, Austria
[2] USound GmbH, Vienna, Austria

This paper introduces a novel MEMS technology-based audio module for in-ear headphone applications. The device has a small form factor with the whole audio module integrated inside the tip of an in-ear headphone. It is a complete audio solution containing a surface mounted MEMS speaker, an integrated audio amplifier with energy recovery, a digital MEMS microphone, as well as digital signal processing for equalization and open-loop predistortion. The module can be manufactured in an SMT line enabling fully automated high-volume mass production with low costs and high yield. The mechanical structure and key benefits of a MEMS based speaker module are discussed in detail, along with the fundamentals of the audio amplifier implementation and signal processing.
Engineering Brief 577

• **Voice Coil Temperature Monitoring: A New In-House Developed Measurement System—***Luca Villa, Chiara Corsini, Grazia Spatafora, Emiliano Capucci, Davide Mele, Romolo Toppi*, Faital S.p.A., Milan, Italy

The increase in temperature of a loudspeaker voice coil (VC) might damage the electrodynamic transducer and produces power compression effect. Usually, DC resistance is evaluated and related to VC thanks to standard thermal coefficients. Unfortunately, typical commercial tools are not customizable according to the user's needs. We have developed an instrument based on the same principle, allowing for user definition of the thermal coefficients, customized software interface, portable and able to increase our lab capabilities. DC resistance of a compression driver and two woofers is measured during life tests using our system and results are compared with Klippel Power Test module (PWT), considered the gold standard. Results showed differences lower than 3.5%, confirming the validity of the new instrument.
Engineering Brief 580

- **Omni-Directional Sound Source Using Facing Ultrasonic Transducer Arrays—***Kyoka Okamoto, Kan Okubo*, Tokyo Metropolitan University, Hino, Tokyo, Japan

We present a new omni-directional sound source using facing ultrasonic transducer arrays. Ultrasonic transducers have been studied for sensors in acoustic sensing, arrayed parametric loudspeakers, or non-contact manipulation. In this research ultrasonic transducer arrays are placed face-to-face on a straight line, and the arrays radiate sound with different ultrasonic frequency each other from both sides. Then the audible sound with differential frequency of the ultrasound signals is omni-directionally emitted. In this report we fabricated sound source employing a facing ultrasonic transducer array and evaluated the frequency characteristics and directivity of the sound source experimentally. This result suggests that it is possible to produce simpler, smaller, and light-weighted omni-directional loudspeakers with higher design like a ring and polyhedron.
Engineering Brief 597

- **Cantilever Vibrating Element for Loudspeaker Breakup Modes Control—***Dario Cinanni*, ASK Industries Spa, Monte San Vito, Italy

Conventional loudspeaker membranes generally use cone or dome shapes, which permit the increase of the first frequency which presents a non-rigid body motion behavior, the so-called break-up frequency. A flat membrane, for example, has a minor geometrical stiffness if compared to standard shapes, bringing some vibration problems. On the contrary flat membranes have an interesting potential, which is a wide dispersion characteristic in the high-frequency range. A new mechanical device is presented to control loudspeaker flat membrane vibrations and it relies on both mechanical driving and damping. Both virtual and physical loudspeaker prototypes are developed and then compared using simulations and measurements.
Engineering Brief 606

ENGINEERING BRIEFS POSTERS: PERCEPTION & EVALUATION

- **Investigating Perceptual Tolerance of Audio-Tactile Synchrony—***Rai Sato*,[1] *Sungyoung Kim*,[2] *Atsushi Marui*[1]
[1] Tokyo University of the Arts, Tokyo, Japan
[2] Rochester Institute of Technology, Rochester, NY, USA

For a congruent perception of multimodal stimuli, a proper timing relation is required. In this study we investigated this timing relation between auditory and vibrotactile sensations, in the context of a hand-held device, when auditory parameters are manipulated. The parameters include (1) a frequency of a pure tone and (2) a spectral kurtosis of a band-passed pink noise. The results show that the frequency parameter is correlated with both the point of subjective simultaneity (PSS) and just noticeable difference (JND). However, we could not find an effect from the change of spectral kurtosis on PSS nor JND. This indicated that the timing of both modal stimuli may need to be changed depending on a frequency of sound stimuli.
Engineering Brief 574

- **The Influence of Loudspeaker-Listener Distance on the Detection of Low-Bitrate Audio Coding Artifacts—** *Alan Pawlak, Hyunkook Lee*, University of Huddersfield, Huddersfield, UK

This study investigated the influence of loudspeaker-listener distance on the detectability of low-bitrate coding artifacts. Two sets of ABX listening tests were conducted in an ITU-R BS.1116-compliant listening room: (i) headphone reproduction (reference anechoic condition) and (ii) loudspeaker reproduction in the listening room (at 1 m, 2 m, and 4 m listening distances). Results from the headphone test for each subject determined the bitrate to be tested for the same subject in the loudspeaker test. Results showed that only the subjects who passed the headphone ABX test at a higher bitrate were affected by listening distance in the loudspeaker test. No clear evidence was found to support the conventional recommendation of minimum 2 m listening distance for the critical listening test.
Engineering Brief 576

- **Investigating User Preference for Reverberation Plugins—** *Kevin Garland, Malachy Ronan*, Limerick Institute of Technology, Limerick, Ireland

Music producers claim strong allegiances to specific reverberation plugins despite these plugins being derived from similar algorithms. This preference may be due to confounding variables including interface design, control parameters provided to the user, adjectives used to describe these parameters, and preset design. This paper describes a listening experiment undertaken to determine whether user preferences exist when close-matched reverberation parameters are held constant on four plugins. The results are discussed within the context of reverberation plugin development.
Engineering Brief 578

- **Comparing the Perception of "Sense of Presence" between a Stereo Mix and a Binaural Mix in Immersive Music—***Andrea De Sotgiu, Mauro Coccoli, Gianni Vercelli*, DIBRIS - Università degli Studi di Genova, Genova, Italy

The aim of this paper is to investigate if binaural music involves listeners more than stereophonic music. A survey was conducted on differences in perception when listening to music with headphones. An experiment was carried-on on a sample of eight people to draw some preliminary results on the preferences of listeners between a stereo mix and a binaural mix. To understand the degree of "sense of presence" of the tested subjects, a questionnaire based on the ITC-SOPI model was used, modified for the specific needs of the experiment. The lack of standards in both the mix techniques and the software to be used to create a professional product in binaural format was one of the significant difficulties encountered.
Engineering Brief 588

- **Understanding Users' Choices and Constraints when Positioning Loudspeakers in Living Rooms—***Craig Cieciura*,[1] *Russell Mason*,[1] *Philip Coleman*,[1] *Jon Francombe*[2]
[1] University of Surrey, Guildford, Surrey
[2] BBC Research and Development, Salford, UK

A study was conducted in participants' homes to ascertain how they would position one to eight compact wireless loudspeakers, with the goal of enhancing their existing system. The eleven participants described three key themes, creating an arrangement that: was spatially balanced and evenly distributed; maintained the room's aesthetics; maintained the room's functionality. In practice, the results showed that participants prioritized aesthetics and functionality, while balance was not usually achieved. It was concluded that a hierarchy of preferred positions in each space exists, as the same positions were reused while positioning differing numbers of loudspeakers and by different participants in each location. Consistencies were observed between the locations that can be used to estimate loudspeaker positions for a given living room layout.
Engineering Brief 596

• **A Mixed-Methods Evaluation of Preferences between Binaural and Stereo Broadcast Audio with Experienced and Non-Experienced Listeners—***Alice Foster, Chris Pike, Jon Francombe*, BBC Research & Development, Salford, UK

An online experiment was conducted to determine preferences between binaural and stereo versions of the same audio material, as well as the reasons for these preferences. It was run with program producers who had knowledge and experience of binaural audio and with members of the general public, more typical of broadcast audiences. The participants performed paired comparisons using a six-point preference scale and described their reasons for that preference using a free-text response. There were six audio items, including classical and pop music, sports, and drama. Inexperienced listeners were less often able to hear differences between the two versions and used less specific justifications for preferences that existed. Both groups often identified positive spatial characteristics of binaural versions.
Engineering Brief 609

ENGINEERING BRIEFS POSTERS: SIGNAL PROCESSING

• **A Python Library for Multichannel Acoustic Signal Processing—** *Andres Perez-Lopez*,[1,2] *Archontis Politis*[3]
[1] Eurecat, Centre Tecnologic de Catalunya, Barcelons, Spain
[2] Pompeu Fabra University, Barcelona, Spain
[3] Tampere University, Tampere, Finland

The programming language Python is receiving increasing attention among the audio research community, partially motivated by the growth of data science and machine learning fields and its potential applications. In this work we present a Python library for acoustic simulation and microphone array processing with a special focus on spherical geometries. More specifically, the library provides a shoebox impulse response generator, a microphone array response simulator with arbitrary geometries and sensor directivities, and a set of methods for signal dependent and independent processing in the spherical harmonic domain.
Engineering Brief 569

• **Investigation on Balanced-Delay Filter-Bank for Encoding Multichannel Audio Signals—***Ikhwana Elfitri, Reski Yulian Fauzan, Amirul Luthfi*, Universitas Andalas Padang, Indonesia

Various MPEG audio standards, such as MPEG Surround, MPEG Advanced Audio Coding (AAC), MPEG Spatial Audio Object Coding (SAOC) as well as MPEG-H 3D Audio, employ a similar powerful hybrid filter-bank. Even though the filter-bank introduces an unbalanced delay, the overall filter distortion is very low. However, when MPEG Surround encodes multichannel audio, such as 22.2 channels, the whole filter-bank structure can form a multi-layer filtering. In this paper we investigate a case of balancing the delay for minimizing the overall filter distortion. The results of the experiments using 22.2 multi-channel audio signals show that the proposed balanced-delay filter-bank is capable of increasing the Signal-to-Noise Ratio up to 8 dB.
Engineering Brief 595

• **Development of a 4-pi Sampling Reverberator, VSVerb. - Phase Detection—***Masataka Nakahara*,[1,2] *Yasuhiko Nagatomo*,[3] Akira Omoto[1,4]
[1] ONFUTURE Ltd., Tokyo, Japan
[2] SONA Corporation, Tokyo, Japan
[3] Evixar Inc., Tokyo, Japan
[4] Kyushu University, Fukuoka, Japan

The authors developed a 4-pi sampling reverberator, named "VSVerb," which restores a 4-pi reverberant field by using information of virtual sound sources that are captured in a target space. Acoustical properties of virtual sound sources are detected from measured sound intensities, and they are translated into time responses. So far VSVerb has given either plus or minus signs to reflection sounds randomly in order to reduce DC bias of a reverberation. In this report the authors propose an alternative method that detects phase characteristics of reflection sounds, i.e., detected sound sources, by analyzing obtained source-intensities and their related wave forms. By using the method, VSVerb becomes to generate 4-pi reverberations by 100% of analytical procedures.
Engineering Brief 598

ENGINEERING BRIEFS POSTERS: SPATIAL AUDIO

• **pysofaconventions, a Python API for SOFA—***Andres Perez-Lopez*, Eurecat, Centre Tecnologic de Catalunya, Barcelona, Spain, Pompeu Fabra University, Barcelona, Spain

Spatial audio is a research field with an active development, motivated by the advances in Augmented and Virtual Reality. One of the main building blocks for spatial audio and acoustic research is the availability of real, measured impulse responses. The SOFA convention (AES69-2015) is a standardized file format for the storage of such data, with a widespread support among the research community. In this work we present pysofaconventions, a full implementation of the SOFA specification for the Python programming language.
Engineering Brief 568

• **Directional Dependency of Subjective Sound Pressure Perception on Three-Dimensional Sound—***Akihito Nakai, Minoru Tsuji, Toru Chinen*, Sony Corporation, Tokyo, Japan

Sony launched 360 Reality Audio in 2019 that provides a new music experience using object-based spatial audio technology. In this music experience sounds arrive from various directions. The direction from which the sound arrives affects the subjective sound pressure sensitivity [1, 2, 3]. Sivonen measured the subjective sound pressure sensitivity in seven directions on the left half of the horizontal plane and on the upper front quarter in the median plane [1]. In this e-brief the number of directions is increased to 31, which includes a whole horizontal plane and the elevations lower than the front center for measurements with three band-limited signals (i.e., 93 conditions in total). As a result, 74 conditions are observed with statistically significant differences.
Engineering Brief 581

• **Rendering a Negatively Elevated Image Based on Relative HRTF Cues—***Anamaria Madalina Nastasa, Hyunkook Lee*, University of Huddersfield, Huddersfield, UK

A study was carried out in order to determine major cues for the perception of negative elevation in the median plane. Extensive analysis focusing on the differences of negative elevation angles to 0° in the median plane was conducted using multiple individual and KEMAR HRTF datasets. Based on the analysis, a parametric equalization model using two notches and one peak is proposed. This model has been applied to a white noise stimulus in a listening test using headphones. The results suggest that the proposed method could effectively create the impression of a negatively elevated image.
Engineering Brief 586

• **Perceptual Evaluation of Binaural Rendering and Stereo Width Control in Headphone Reproduction—***Yui Ueno*,[1] *Mistunori Mizumachi*,[1] *Toshiharu Horiuchi*[2]
[1] Kyushu Institute of Technology, Fukuoka, Japan
[2] KDDI Research, Inc., Saitama, Japan

Two-channel stereophonic sound reproduction is one of the most

important issues in audio engineering. Most of stereo music sources, however, cannot naturally reproduce spatial impression through earphones and headphones. In this study stereo sources are converted by either binaural rendering or stereo width control. The former convolves head-related impulse responses into the original stereo sources, and the latter shrinks the perceptual spatial width by adding the monaural signal. Perceptual differences in between the above methods are investigated on several spatial attributes. It is confirmed that the binaural rendering is suited to reproduce the spatial impression related to the whole sound field and the stereo width control is suitable for naturally characterizing the spatial attributes for individual sound sources.
Engineering Brief 587

• **Emily's World: Behind the Scenes of a Binaural Synthesis Production**—*David Poirier-Quinot*,[1] *Lucie Hardoin*,[2] *Brian F.G. Katz*[1]
[1] Sorbonne Université, Paris, France
[2] Lucie Hardoin, www.sonbinaural.com, Finistère, France

Three main workflows exist to create binaural content: binaural recording, binaural decoding of Ambisonic recordings, and binaural synthesis. Binaural synthesis, as an object-oriented approach, offers the highest flexibility throughout the creation process (composition, mixing, etc.). Conversely, it typically requires separate takes for each audio source and animations to construct the final spatial mix. "Binaural Steps - Emily's World" is an 8 minute short radio-fiction commissioned to highlight the possibilities of binaural synthesis using studio and Foley recordings. Synthesis was achieved using the Anaglyph binaural audio engine, designed to facilitate the transition of spatial hearing research from laboratory to industry.
Engineering Brief 611