PAPERS AND E-BRIEFS PRESENTED AT THE AES 152ND CONVENTION

MAY 2022, IN-PERSON AND ONLINE

Dynamic Range Controller Ear Training: Transfer of Learned Skills to a Related Task

A technical ear training program designed to teach participants to identify the audible effects of dynamic range compressors was developed and used to train graduate students in music production. A pre-/post-training listening test was used to determine if the students could transfer skills learned during the training program to a related listening task. The participants executed the task repeatedly and the participants' variance in their final compressor settings was measured. The pre-/post-tests were administered to the trained student group, an untrained student control group, and a group of recently graduated professional engineers. A reduction in variance between the pre-/post-test was measured in the trained group but not in the control group. The trained students also had less variance in their responses than the professional engineers.

Martin, Dennis; Massenburg, George; King, Richard
Schulich School of Music of McGill University, Montreal, Canada; Centre for Interdisciplinary Research in Music Media and Technology, Montreal, Canada

Loudness & Perception
Paper Number: 10545
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21658

Latency and Quality-of-Experience Analysis of a Networked Music Performance Framework for Realistic Interaction

Networked Music Performances (NMPs) are of increasing importance for creative and cultural professionals exploring new solutions and opportunities to perform at geographically distant locations. Although latency is one of the most important parameters for the transmission of audio information, it is rarely possible to obtain insights into the individual components of the transmission latency during NMP. In this publication a detailed evaluation of the latency budget of an ultra-low latency audio transmission in an NMP system for realistic interaction, build up between Hanover and Munich (˜500 km), is presented. To explore the Quality-of-Experience (QoE) of the NMP, objective results of a pop / rock piece played by five professional musicians are compared with their subjective perception. Measures of network performance, as well as the comparison of the objective musical results with the subjective user feedback suggest that the musicians had an experience close to a real performance.

Hupke, Robert; Dürre Jan; Werner, Norbert; Peissig, Jürgen
Leibniz University Hanover, Sennheiser electronic GmbH & Co. KG

Network Audio
Paper Number: 10546
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21659

"We'll Feel That in Post": Altering Emotional Perception With Different Piano Presentations In the Context of Lyrics

While much is known about how musical performers can communicate emotions through music, less research has been dedicated to determining if audio processing can affect the emotion of a musical work as it is perceived by the listener, particularly within the context of a narrative provided by sung lyrics. This paper presents a pilot experiment in which 8 participants were presented with two audio recordings of the same piano and vocal performance, with the piano presented with different timbral characteristics in each recording. Results demonstrate some limited correlation between change in timbre of the piano and change in perceived emotion by the listener.

Cann, Nathan
McGill University, Montreal, Canada

Audio Synthesis & Audio Effects
Paper Number: 10547
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21660

Geometrical Acoustics Approach to Cross Talk Cancellation

Crosstalk Cancellation (CTC) is a signal processing technique allowing for immersive sound reproduction from a limited number of loudspeakers. Pioneered in the sixties, CTC has lately gained much attraction due to upcoming Augmented Reality and Virtual Reality applications and generalization of 3D audio content. In this paper, we present a novel time-domain approach to CTC based on modeling of the system's geometrical acoustics. Our solution provides a simple processing model, as well as means to address robustness issues and adaptation to arbitrary listener positions.

Vancheri, Alberto; Leidi, Tiziano; Heeb, Thierry; Grossi, Loris; Spagnoli, Noah; Weiss, Daniel
University of Applied Sciences and Arts of Southern Switzerland; Weiss Engineering Ltd, Switzerland

Binaural Audio
Paper Number: 10548
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21661

Reduction of 3D Ambisonic to 2D using plane-wave decomposition

This articles deals with the reduction – or "projection" or "downscaling" – of a 3D-encoded Ambisonic sound field ("full-sphere" or "periphonic") into a 2D representation ("horizontal-only" or "planar" or "pantophonic"). We show that the reduction operation can be equivalently achieved by 1) applying conversion formula of the normalization factors, or 2) performing a plane-wave decomposition of the original sound field and re-encoding the resulting plane waves to 2D Ambisonic. The latter approach provides greater flexibility in adapting the content to horizontal-only reproduction.

Carpentier, Thibaut
STMS, IRCAM – CNRS – Sorbonne Université – Ministère de la Culture, Paris, France

Spatial Audio
Paper Number: 10549
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21662

Multi-diaphragm Micro-speaker vs. Single-diaphragm Micro-speaker

With wearable devices becoming increasingly popular, there is a foreseeable paradigm shift for small format or miniature audio transducers in the near future. Small single diaphragm micro-speaker transducer no longer meets the new expectations both in low frequency performance and overall sound pressure level. In this paper a new multi-diaphragm micro-speaker is presented. The acoustic performance of the transducer is simulated using finite element method. Furthermore, this paper analyses the low frequency performance between the single diaphragm micro-speaker and the multi-diaphragm micro-speaker.

Wei, George; Chen, Wendy; Wie, Tony
Guo Guang Electric Corporation, Guangzhou, P. R. China,

Loudspeakers and headphones
Paper Number: 10550
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21663

Spectral and spatial perceptions of comb-filtering for sound reinforcement applications.

Most sound reinforcement systems consist of multiple loudspeakers systems arranged strategically to cover the entire audience area. This study investigates the spectral and spatial perceptions of interferences that can be experienced in the shared coverage area between two full-range loudspeakers. A listening test was conducted to determine the effect of lag source delay, relative level, and angular separation, on the perception of spectral coloration and spatial impressions (width, localization shift, image separation). The results show that spectral coloration is considerably reduced when sources are spatially separated, even with a small azimuth angle (10°). It was also found that coloration audibility depends on the interaction between the audio track and the delay introduced. Finally, the type of perceived spatial degradation depends mainly on the spatial separation and on the relative level of the source arriving later in time (lag source).

Moulin, Samuel; Corteel, Etienne
L-Acoustics, Marcoussis, France

Room Acoustics
Paper Number: 10551
Available in the AES E-Library here: https: https://www.aes.org/e-lib/browse.cfm?elib=21664

Ambisonic Decoder Test Methodologies based on Loudspeaker Reproduction

The comparative evaluation of the quality of different Ambisonic decoding strategies presents a number of challenges, most notably the lack of a suitable reference signal other than the original, real-world audio scene. In a previous paper, a new test methodology for such evaluations was presented via a listening test conducted using binaural reproduction. In this paper, this methodology is further refined and the results of a new listening test using loudspeaker reproduction over a 7.0.4 array is presented. The results again indicate some significant differences between the decoders for certain attributes.

Bates, Enda; David, William; Dempsey, Daniel
ADAPT Centre, School of Engineering, Trinity College Dublin, Ireland

Spatial Audio
Paper Number: 10552

Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21665

Twang! A physically derived synthesis model for the sound of a vibrating bar

A physically derived synthesis model of the sound generated when a ruler is twanged while hanging over the edge of a solid surface is presented. This is a sound effect used in movies, TV, theatre performances and cartoons. The model is derived from the Euler-Bernoulli equation, offering the user a set of physical parameters to control ruler length as well as the material properties. Perceptual evaluation indicates that the model can be perceived as realistic as a recorded ruler twang as well as being able to replicate sounds of similar quality as an alternative synthesis model.

Selfridge, Rod; Andreasson, Mimmi; Bengtsson, Lina; Bengtsson, Bjarki Vidar; Lindborg, Emelie; Rydén, Matilda; Tez, Hazar; Reiss, Joshua
Edinburgh Napier University, Scotland; KTH Royal Institute of Technology, Sweden; Queen Mary University of London, UK

Audio Synthesis & Audio Effects
Paper Number: 10553
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21666

Velocity-Contolled Parameter Switching for Echo Cancellation in Immersive Telepresence with Continuously Changing Microphone Positions

The frequency-domain adaptive Kalman filter (FDAKF) is a popular choice for multichannel acoustic echo cancellation (AEC) due to its good initial convergence and robustness to double-talk. However, without additional measures its reconvergence and tracking capabilities are known to be suboptimal. Previous studies have particularly focused on abrupt echo path changes and have proposed different methods to optimize the filter's reconvergence. Motivated by our application of an acoustic echo cancellation system for immersive telepresence, this paper investigates continuous echo path changes caused by moving microphones. The echo cancellation performance of the FDAKF is studied for different parameters of the underlying model inside the Kalman filter. Experimental results show, that even in the very challenging scenario of a moving microphone, a small echo reduction can still be achieved with suitable parameters for the considered microphone velocities. Furthermore, a novel method is proposed, which includes a microphone motion-controlled online parameter switching for the FDAKF by means of external motion sensors. In this paper the method is studied within a proof-of-concept. Experiments show a behavior matched to static and dynamic phases and even an increased reconvergence speed in the transition from dynamic to more static phases.

Nophut, Marcel; Preihs, Stephan; Peissig, Jürgen
Institute of Communications Technology, Leibniz University Hannover, Germany

3D/Immersive/Spatial Audio
Paper Number: 10554
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21667

MMAD – Designing for Height – Practical Configurations

Although the basic philosophy behind the design of Microphone arrays, for 3D audio recording and reproduction, has been described in previous AES papers[1][2][3][4], no

specific examples have been given with respect to various Surround Sound arrays and the corresponding height arrays (1st layer of height array microphones and the Zenith microphone). This paper gives four examples of complete 3D Audio arrays with perfect critical linking. Examples include same microphone directivity arrays, as well as hybrid arrays. Suitable steering functions are discussed, and specific values are given for each array combination, so as to obtain perfect critical linking.

Williams, Michael
Sounds of Scotland, Le Perreux sur Marne, France

3D/Immersive/Spatial Audio
Paper Number: 10555
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21668

Survey of User Perspectives on Headphone Technology

Headphones are widely used to consume media content at home and on the move. Developments in signal processing technology and object-based audio media formats have raised new opportunities to improve the user experience by tailoring the audio rendering depending on the characteristics of the listener's environment. However, little is known about what consumers consider to be the deficiencies in current headphone-based listening, and therefore how best to target new developments in headphone technology. More than 400 respondents worldwide took part in a headphone listening experience survey. They were asked about how headphones could be improved, considering various contexts (home, outside, and public transport) and content (music, spoken word, radio drama/tv/film/online content, and telecommunication). The responses were coded into themes covering technologies (e.g. noise cancellation and transparency) and features (e.g. 3D audio) that they would like to see in future headphones. These observations highlight that users' requirements differ depending on the listening environment, but also highlight that the majority are satisfied by their headphone listening experience at home. The type of programme material also caused differences in the users' requirements, indicating that there is most scope for improving users' headphone listening experience for music. The survey also presented evidence of users' desire for newer technologies and features including 3D audio and sharing of multiple audio streams.

Rane, Milap; Coleman, Philip; Mason, Russell; Bech, Søren
Institute of Sound Recording, University of Surrey, Guildford, UK; Aalborg University, Department of Electronic Systems, Aalborg, Denmark; Bang Olufsen a/s, Struer, Denmark

Loudspeakers and headphones
Paper Number: 10556
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21669

Controlling the Balance Between Early and Late Reflections of an Impulse Response Using the Modal Decomposition Paradigm

Previous work has outlined a method of decomposing an acoustic impulse response (IR) into decaying sine components, which allows for parametrization of a modelled IR. This enables control over a convolution based reverberator akin to an algorithmic reverberator. A common control found in an algorithmic reverb is the balance between early and late reflections. This work introduces a method of adjusting the balance between early and late reflections without direct processing of the impulse response, through a component domain transform. This work extends the creative applications of indirect IR Processing through the use of the modal decomposition paradigm.

Monedero, Patxi; Howard, Michael
NUGEN Audio, Leeds, UK

Audio Synthesis & Audio Effects
Paper Number: 10557
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21670

MP3 compression classification through audio analysis statistics

MP3 audio compression can be undesirable in circumstances where high-quality music presentation is required and there is a lack of automated, evidenced, and open-source methods to determine this. This study introduced a new and accessible approach to discriminate between compression levels and identify lossy audio transcoding. Machine learning classifiers were trained on feature sets of audio analysis statistics, derived from multiple step-wise re-encodings of compressed audio samples. Two classifiers, a stacked model and a XGBoost-based model, had comparable accuracies to previous examples in the literature and marketplace (Stacked: 0.947, XGBoost: 0.970, Literature reference: 0.965, Commercial reference: 0.980). For transcoded samples, which hide compression levels with post-processing, the new classifiers were less accurate than existing methods. However, all methods were inaccurate in identifying transcodes where artificial noise was added via the μ-law encoder. A command-line implementation is available at gitlab.com/jammcfar/kbps_detect_proto.

McFarlane, Jamie; Chakravarthi, Bharathi Raja
National University of Ireland

Sound Classification
Paper Number: 10558
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21671

Bitrate Requirements for Opus with First, Second and Third Order Ambisonics reproduced in 5.1 and 7.1.4

In this paper, we present a study on the Basic Audio Quality of first, second and third order native Ambisonics recordings compressed with the Opus audio codec at 24, 32 and 48 kbps bitrates per channel. Specifically, we present subjective test results for Ambisonics in Opus decoded to ITU-R BS.2051-2 [1] speaker layouts (viz., 5.1 and 7.1.4) using IEM AllRAD decoder [2]. Results revealed that a bitrate of 48 kbps/channel is transparent for Basic Audio Quality for second and third order Ambisonics, while larger bitrates are required for first order Ambisonics.

Souza-Blanes, Ema; Tejeda-Ocampo, Carlos; Wang, Carren; Bharitkar, Sunil
Samsung Research America; Samsung Research Tijuana

3D/Immersive/Spatial Audio
Paper Number: 10559
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21672

Localization of Direct Source and Early Reflections Using HOA Processing and DNN Model

This paper proposes a novel direct source and first-order reflections localization method by integrating the high order Ambisonics (HOA) algorithm and deep neural network. We use the covariance matrix of HOA signals in the time domain as the input feature of the network, which contains precise spatial

information of the sound sources under reverberant scenarios. Besides, we use the deconvolution-based neural network (DCNN) for the spatial pseudo-spectrum (SPS) reconstruction, based on which the spatial relationship between elevation and azimuth can be depicted. Considering that the first-order reflections of the sound source also contain spatial directivity like the direct source, we treat both of them as the sources in the learning process. We have carried out a series of experiments based on simulated and measured data under different reverberant scenarios, which prove the effectiveness and accuracy of the proposed DCNN model.

Gao, Shan; Wu, Xihong; Qu, Tianshu
Key Laboratory on Machine Perception (Ministry of Education), School of Artificial Intelligence, Peking University, Beijing, China

Spatial Audio
Paper Number: 10560
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21673

Towards the Global Standard for Safe Listening Venues and Events

The Global Standard for safe listening venues and events was published by the World Health Organization on March 2 of this year, ahead of World Hearing Day. The key intention of the Standard is a reduction of overall audience exposure to high sound levels at live events worldwide, however, its implementation is up to individual nation-states, smaller jurisdictions, or even individual music venues. The authors were involved in discussions informing the Standard from an early stage, supporting its development by creating an evidence base around the measurement and management of sound levels in music venues. This paper provides an overview of the authors' research informing this contribution and highlights necessary further research.

Mulder, Johannes Mulder; Hill, Adam; Kok, Marcel; Burton, Jonathan; Lawrence, Michael
The National University Australia, Canberra; University of Derby, UK; dBcontrol, The Netherlands; Rational Acoustics LLC, USA

Room Acoustics
Paper Number: 10561
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21674

Conversational Speech Separation: an Evaluation Study for Streaming Applications

Continuous speech separation (CSS) is a recently proposed framework which aims at separating each speaker from an input mixture signal in a streaming fashion. Hereafter we perform an evaluation study on practical design considerations for a CSS system, addressing important aspects which have been neglected in recent works. In particular, we focus on the trade-off between separation performance, computational requirements and output latency showing how an offline separation algorithm can be used to perform CSS with a desired latency. We carry out an extensive analysis on the choice of CSS processing window size and hop size on sparsely overlapped data. We find out that the best trade-off between computational burden and performance is obtained for a window of 5 s.

Morrone, Giovanni; Cornell, Samuele; Zovato, Enrico; Brutti, Alessio; Squartini, Stefano
Università Politecnica delle Marche, Ancona, Italy; PerVoice S.p.A., Trento, Italy; Fondazione Bruno Kessler, Trento, Italy

Television Audio
Paper Number: 10562
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21675

Recording Arts and Studio Engineering for High-fidelity Multitrack Network Music Production

This article offers a practical framework and effective alternative technological solutions for reimagining recording arts and music production through free, open-source network audio technology during the pandemic. Network audio technology has become a crucial vehicle for music ensembles and bands to remotely rehearse, record, perform, and produce concerts and albums, which offer innovative opportunities to transcend geographical distance and deepen human connections between cultures and communities. By elaborating and analysing detailed qualitative case studies in multichannel high-fidelity network audio recording, mixing, and postproduction using hardware-software configuration, JackTrip, and Netty-McNetface as evidence, the authors prove the significance of open-source network audio technology as well as provide effective and economical solutions for musicians, composers, music educators, audio engineers, media artists, curators, and performing arts institutions to reimagine their future practices that lead to revolutionary institutional change during and beyond the pandemic.

Wu, Jiayue Cecilia; Miller, Scott
University of Colorado Denver, College of Arts and Media, U.S.A.; St. Cloud State University, School of the Arts, Department of Music, St. Cloud, MN, U.S.A.

Studio Technology
Paper Number: 10563
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21676

Comparison of Audio Quality of Teleconferencing Applications Using Subjective Test

The human need to communicate and connect during the Covid-19 pandemic has led to the increasing use of teleconferencing applications. Users naturally pay attention to audio quality in choosing a teleconference application from many available and easily accessible applications. Audio quality is partly affected by the audio coding method used and developed on the application and the noise introduced within the network. This work evaluates the audio quality of 5 popular teleconferencing applications using the subjective test. For complement, objective tests and Signal-to-Noise Ratio (SNR) assessments are also carried out. The standard used for the subjective test is ITU-R BS.1116-3: Methods for the Subjective Assessment of Small Impairments in Audio Systems, which aims to identify small differences between audio quality. The assessment is conducted by comparing the original and compressed audio. The original audio is recorded from the speaker side, while the compressed audio is recorded from the receiver side. Both are assessed using the subjective test method by 20 subjects. The assessment results of each audio teleconferencing application are different, even though some applications use the same codec. We also found that one of the most popular applications tends to have the lowest average score among the tested audio applications.

Faadhilah, Avelia Fairuz; Elfitri, Ikhwana
Department of Electrical Engineering, Faculty of Engineering, Universitas Andalas, Padang, Indonesia

Sound Quality & Perception
Paper Number: 10564
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21677

Ocean Wave Sound Synthesis and Perceptual Evaluation

We present and evaluate the implementation of a real-time, procedural ocean wave sound effect synthesis model that works in a web environment. This model uses filtering of noise rather than a physical model of ocean waves. The ocean waves sound synthesis model was implemented using the Web Audio API. A modular approach was adopted to achieve versatility and to expand the model to more complex techniques if needed. In the listening test, real world ocean wave sounds were compared against our sound model as well as ocean wave sounds created by other few synthesis techniques. The results indicate that the current implementation can successfully represent real ocean waves and the procedural model can outperform the other proposed approaches in terms of believability of the generated sound.

Tez, Hazar Emre; Selfridge, Rod; Reiss, Joshua
Queen Mary University of London, UK; KTH Royal Institute of Technology, Sweden

Audio Synthesis & Audio Effects
Paper Number: 10565
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21678

On the Specification of Threshold Limits in Speech Transmission Index Assessment of Distributed Sound Systems

Speech Transmission Index (STI) is a well-established method to evaluate the performance of a transmission channel for speech communication. The STI assessment technique is regulated by the IEC 60268-16 standard and its usage is widely adopted both in standard bodies and in tender specifications. STI is a successful implementation of an evaluation method that, despite inherent limitations, correlates very well with perception. In engineering practice it is often necessary to compare STI evaluations with limit values. This also requires specifying boundary conditions of the evaluation. In practical applications where STI fails to reach the target values, specifically in cases where distributed sound systems are used, it has been found that there is a polarization of the effects affecting the STI value. This leads to a particular sensitivity of the STI to boundary conditions and in fact, makes it impractical to reach targets in several scenarios. In these cases the specification of the STI limit targets is not sufficient to design effective sound systems and should be supported by a detailed and clear specification of the boundary conditions such as, but not limited to, the background noise.

Ponteggia, Daniele
Studio Ponteggia, Terni, Italy

Room Acoustics
Paper Number: 10566
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21679

Influence of the spider on loudspeaker thermal behavior: a predictive model based on lumped parameters

In the pre-design phase of a loudspeaker, rapid prediction of the loudspeaker thermal behavior is crucial. Voice coil (VC) temperature is not only influenced by the magnetic circuit materials and geometry, but also by thermal dissipation due to other speaker components e.g., the spider. The aim of this study is to develop a thermal model of the loudspeaker magnetic circuit that provides fast predictions of the VC temperature accounting for convective dissipation due to spider movement. An empirical method to evaluate the spider air permeability was presented and the related effects were implemented in the model. Life tests were performed on different loudspeakers to verify the accuracy of model results. Differences between predicted and measured VC temperatures were acceptable,

suggesting possible application to the pre-design phase as the model can provide fast predictions without requiring high-performance hardware. 1 Introduction

Villa, Luca; Corsini, Chiara; Spatafora, Grazia; Mele, Davide; Toppi, Romolo
Faital S.p.A, Milan, Italy

Loudspeakers and headphones
Paper Number: 10567
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21680

Multi-Input Architecture and Disentangled Representation Learning for Multi-Dimensional Modeling of Music Similarity

In the context of music information retrieval, similarity-based approaches are useful for a variety of tasks that benefit from a query-by-example approach. Music however, naturally decomposes into a set of semantically meaningful factors of variation. Current representation learning strategies pursue the disentanglement of such factors from deep representations, and result in highly interpretable models. This allows to model the perception of music similarity, which is highly subjective and multi-dimensional. While the focus of prior work is on metadata driven similarity, we suggest to directly model the human notion of multi-dimensional music similarity. To achieve this, we propose a multi-input deep neural network architecture, which simultaneously processes mel-spectrogram, CENSchromagram and tempogram representations in order to extract informative features for different disentangled musical dimensions: genre, mood, instrument, era, tempo, and key. We evaluated the proposed music similarity approach using a triplet prediction task and found that the proposed multi-input architecture outperforms a state of the art method. Furthermore, we present a novel multi-dimensional analysis to evaluate the influence of each disentangled dimension on the perception of music similarity.

Ribecky, Sebastian; Abeßer, Jakob; Lukashevich, Hanna
Semantic Music Technologies Group, Fraunhofer IDMT, Ilmenau, Germany

Sound Classification
Paper Number: 10568
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21681

Acoustic Design Review for the Historical Aula Magna at the University of Parma. Measurement and Simulation Tools to Predict the Amount of Absorption to be in Place.

The Aula Magna (Great Hall) of the University of Parma, historically known as the Hall of the Philosophers, is part of a 16th-century palace located in the core of the city. The Aula Magna hosts official ceremonies and graduations. The geometrical composition of the room, provided with a wagon vault, and the high level of reflections due to the hard finish materials lead to a poor quality of speech intelligibility. This paper deals with an acoustic design review of the current furniture inside the auditorium, with the purpose of adjusting the acoustic parameters to be suitable for the current uses of the room. Acoustic measurements have been undertaken in accordance with the standard requirements outlined by ISO 3382-1, capturing the existing conditions of the hall. A highly accurate digital model of the room was obtained with photogrammetry and modelling, in order to carry out numerical simulations regarding the implementation of the acoustic treatments. The quantity and quality of the proposed absorbing panels improve the listening conditions to a degree of comfort assessed against the criteria set by UNI 11532-2:2020. 1

Farina, Angelo; Bevilacqua, Antonella; Farina, Adriano
University of Parma, Parma, Italy

Room Acoustics
Paper Number: 10569
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21682

Classifying Sounds in Polyphonic Urban Sound Scenes

The deployment of machine listening algorithms in real-world application scenarios is challenging. In this paper, we investigate how the superposition of multiple sound events within complex sound scenes affects their recognition. As a basis for our research, we introduce the Urban Sound Monitoring (USM) dataset, which is a novel public benchmark dataset for urban sound monitoring tasks. It includes 24,000 sound scenes that are mixed from isolated sounds using different loudness levels, sound polyphony levels, and stereo panorama placements. In a benchmark experiment, we evaluate three deep neural network architectures for sound event tagging (SET) on the USM dataset. In addition to counting the overall number of sounds in a sound scene, we introduce a local sound polyphony measure as well as a temporal and frequency coverage measure of sounds which allow to characterize complex sound scenes. The analysis of these measures confirms that SET performance decreases for higher sound polyphony levels and larger temporal coverage of sounds.

Abeßer, Jakob
Fraunhofer IDMT, Ilmenau, Germany

Sound Classification
Paper Number: 10570
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21683

Perceptual optimization of hybrid stereo width control method compared with loudspeakers reproduction

Legacy stereo music sources cause unnatural spatial impressions through the earphones and headphones reproduction due to the lack of crosstalk, which is naturally produced in loudspeakers reproduction. The amplitude-based and phase-based hybrid stereo width control method is proposed to generate the crosstalk components between channels in headphones reproduction. The control parameters for both the previously-proposed amplitude-based and hybrid methods are perceptually optimized compared with the loudspeaker reproduction. Suitable parameter values are successfully obtained for both methods with classical music sources. It is confirmed that the suitable parameter values have a different trend in those methods depending on the music sources.

Mizumachi, Mitsunori; Ueno, Yui; Horiuchi, Toshiharu
Kyushu Institute of Technology, Fukuoka, Japan; KDDI Research, Inc, Saitama,

Recording Technologies
Paper Number: 10571
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21684

The Inaudible World: Measuring Sounds Perceived by Non-Human Life

The deleterious effects of audible and inaudible sounds on human and non-human life are becoming self-evident. Surveying hearing ranges across species show perceptual capabilities that extend into the infrasonic <20Hz and ultrasonic >20kHz ranges. This is far beyond the limited bandwidth of sounds perceived by human hearing mechanisms. Human-centric metrics and standards are insufficient to measure sounds perceived by other living beings who co-inhabit earth's acoustic environments. Such measurements are band limited to the 20Hz-20kHz human frequency range and audibility thresholds. These include the dBA weighting curve which form the majority of current health and safety standards and policies. Explorations of this area are currently creating new knowledge systems, opening up new areas of research in audio perception and facilitating the creation of new measurement devices and scales.

Milinusic, Christina
University of Lethbridge, Lethbridge, AB, Canada

Loudness & Perception
Paper Number: 10572
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21685

Simulation-Based Acoustic Design for a Modern Urban Church Sanctuary

This paper describes the architectural and acoustical challenges of building a sound-critical worship space in a dense urban setting. The plans for Redeemer Presbyterian Church's new building construction on the Upper East Side of Manhattan called for a sanctuary with over 450 seats in a lot only 50 feet wide. To maximize the visual space, the church chose an asymmetric diagonal layout over a more traditional symmetric shoebox-style hall. The new acoustical issues posed by this design were modeled in CATT-Acoustic to examine the distribution of sound absorption, total reverberation time within the space, and Binaural Quality Index (BQI) at different listener positions.

Judy, Patrick; Morgan, Andy; Boren, Braxton
American University, Washington, DC, USA; Morgan Acoustics, New York, NY, USA

Room Acoustics
Paper Number: 10573
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21686

Advances in Thunder Sound Synthesis

A recent comparative study evaluated all known thunder synthesis techniques in terms of their perceptual realness. The findings concluded that none of the synthesised audio extracts seemed as realistic as the genuine phenomenon. The work presented herein is motivated by those findings, and attempts to create a synthesised sound effect of thunder indistinguishable from a real recording. The technique supplements an existing implementation with physics-inspired, signal-based design elements intended to simulate environmental occurrences. In a listening test conducted with over 50 participants, this new implementation was perceived as the most realistic synthesised sound, though still distinguishable from a real recording. Further improvements to the model, based on insights from the listening test, were also implemented and described herein.

Fineberg, Eva; Walters, Jack; Reiss, Joshua
Queen Mary University of London, UK; Nemisindo Ltd, UK; Native Instruments GmbH, Germany

Audio Synthesis & Audio Effects
Paper Number: 10574
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21687

Acquisition of Continuous-Distance Near-Field Head-Related Transfer Functions on KEMAR Using Adaptive Filtering

Near-field Head-Related Transfer Functions (HRTFs) depend on both source direction (azimuth/elevation) and distance. The acquisition procedure for near-field HRTF data on a dense spatial grid is time-consuming and prone to measurement errors. Therefore, existing databases only cover a few discrete source distances. Coming from the fact that continuous-azimuth acquisition of HRTFs has been made possible by applying the Normalized Least Mean Square (NLMS) adaptive filtering method, in this work we applied the NLMS algorithm in measuring near-field HRTFs under continuous variation of source distance. We developed and validated a novel measurement setup that allows the acquisition of near-field HRTFs for source distances ranging from 20 to 120 cm with one recording. We then evaluated the measurement accuracy by analyzing the estimation error from the adaptive filtering algorithm and the key characteristics of the measured HRTFs associated with near-field binaural rendering.

Li, Yuqing; Preihs, Stephan; Peissig, Jürgen
Leibniz Universität Hannover, Germany

Binaural Audio
Paper Number: 10575
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21688

An End-to-End Binaural Sound Localization Model Based on the Equalization and Cancellation Theory

The end-to-end framework has been introduced into the binaural localization modeling and achieved higher localization accuracy than the other models, however, the reasonability and interpretability for applying the related neural networks remain unclear. It has been well documented that the auditory system relies on binaural cues for sound localization, and the equalization and cancellation (EC) theory describes how the binaural cues are extracted. In this paper, an end-to-end binaural localization model is proposed based on the EC theory. In the proposed model, a convolution neural network(CNN) with a specifically designed activation function is used to implement the EC theory. The proposed model was trained in synthesized rooms and evaluated in real rooms. Experiment results show that CNN kernels learned by the proposed model are corresponding to binaural cues, and the proposed model outperforms the current end-to-end model by a 10.73% improvement in localization accuracy and a 12.91%improvement in root mean square error(RMSE).

Song, Tao; Zhang, Wenwen; Chen, Jing
Peking University; Beijing University of Posts and Telecommunications

Binaural Audio
Paper Number: 10576
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21689

UP-WGAN: Upscaling Ambisonic Sound Scenes Using Wasserstein Generative Adversarial Networks

Sound field reconstruction using spherical harmonics (SH) has been widely used. However, order-limited summation leads to an inaccurate reconstruction of sound pressure when the reconstructed region is large. The reconstruction performance also degrades when it comes to high frequency. Upscaling ambisonic sound scenes is used to overcome the limitations. In this work, a deep-learning-based method for upscaling is proposed. Specifically, the generative adversarial network (GAN) is introduced. Instead of estimating the SH coefficients, a U-Net-based fully convolutional generator is introduced, which directly outputs the two-dimensional sound pressure. Results show that the proposed method significantly improves the upscaling results compared with the previous deep-learning-based method.

Wang, Yiwen; Wu, Xihong; Qu, Tianshu
Peking University, Beijing, China

Spatial Audio
Paper Number: 10577
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21690

Low Complexity Methods for Robust Stereo-to-Mono Down-mixing

Stereo to mono down-mix is a key component of parametric stereo coding to drastically reduce the bit rate, but at the same time it is also an irreversible process that is a potential source of undesirable artifacts. This paper aims to reduce typical distortions induced by down-mixing, such as signal cancellation, comb filtering or unnatural instabilities. Two down-mixing methods are designed with different trade-offs between natural timbre and energy preservation based on simple rules that ensure low complexity. The results of a listening test show that both the proposed methods have a substantial advantage over the passive down-mix, while being very competitive compared to more computationally demanding active down-mixing approaches. The proposed methods are, therefore, particularly well suited to low complexity stereo coding schemes, such as those required for communication applications.

Maben, Pallavi Maben; Borß, Christian; Edler, Bernd; Fuchs, Guillaume
Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany; Ittiam Systems Pvt. Ltd., Bengaluru, India

Audio Synthesis & Audio Effects
Paper Number: 10578
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21691

The Performance of A Personal Sound Zone System with Generic and Individualized Binaural Room Transfer Functions

The performance of a two-listener personal sound zone (PSZ) system consisting of eight frontal mid-range loud-speakers in a listening room was evaluated for the case where the PSZ filters were designed with the individualized BRTFs of a human listener, and compared to the case where the filters were designed using the generic BRTFs of a dummy head. The PSZ filters were designed using the pressure matching method and the PSZ performance was quantified in terms of measured Acoustic Contrast (AC) and robustness against slight head misalignments. It was found that, compared to the generic PSZ filters, the individualized ones significantly improve AC at all frequencies (200-7000 Hz) by an average of 5.3 dB and a maximum of 9.4 dB, but are less robust against head misalignments above 2 kHz with a maximum degradation of 3.6 dB in average AC. Even with this degradation, the AC spectrum of the individualized filters remains above that of their generic counterparts. Furthermore, using

generic BRTFs for one listener was found to be enough to degrade the AC for both listeners, implicating a coupling effect between the listeners' BRTFs.

Qiao, Yue; Choueiri, Edgar
Princeton University, Princeton, NJ, USA

Binaural Audio
Paper Number: 10579
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21692

Assessing the relevance of perceptually driven objective metrics in the presence of handling noise

This paper examines how perceptually driven objective metrics found in the speech enhancement and separation literature react when adding handling noise to speech corrupted with environmental noise. Identifying sensitive metrics will inform us which metrics are appropriate for the development or evaluation of speech enhancement techniques when dealing with handling noise. Using an in-house synthetic dataset and paired sample tests, we examine how nine different perceptual metrics behave on audio mixtures containing both handling and background noise. We show that eight of them react to handling noise but only when the handling to background noise power ratio is over a specific threshold which we identify using logistic regression.

Angonin, Céline; Theofanis Chourdakis, Emmanouil; Åeng, Ruben Andre
Nomono AS

Sound Quality & Perception
Paper Number: 10580
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21693

Comparison of Stereo Microphone Configuration and Playback Modes of Optimal Source Distribution Technique

Using an Optimal Source Distribution (OSD) implementation which is a crosstalk cancellation method for synthesizing virtual auditory space using horizontally placed loudspeakers, comparison of stereo microphone configurations and the influence of playback modes of OSD were investigated. A musical material was recorded in a studio using three AB-stereo pairs and a dummyhead. Three AB pairs were then processed with stereo loudspeaker playback simulating filter, resulting in seven stimuli. Attributes elicitation, dissimilarity rating, and attribute ratings were conducted and their results were analyzed. Participants to the listening test responded to microphone configuration differences in a similar manner with and without loudspeaker filter applied, while the two playback modes differentiated.

Marui, Atsushi; Yairi, Motoki; Hoshino, Tsuguto; Kamekawa,Toru
Tokyo University of the Arts, Japan; Kajima Technical Research Institute, Japan

Recording Technologies
Paper Number: 10581
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21694

Spatial extrapolation of early room impulse responses with source radiation model based on equivalent source method

The measurement of room impulse responses (RIRs) at multiple points is useful in most acoustic applications, such as sound field control. Recently, several methods have been proposed to estimate multiple RIRs. However, when using a small number of closely located microphones, the estimation accuracy degrades owing to the source directivity. In this study, we propose an RIR estimation method using a source radiation model based on the sparse equivalent source method (ESM). First, based on the sparse ESM, the source radiation was modeled in advance by the microphone array enclosing the sound source. Subsequently, the sound field, including the sound reflections, was modeled using the source radiation model based on the sparse ESM and the image source method. As observed from the simulation experiments, the estimation accuracy was improved at higher frequencies compared with the sparse ESM without the source radiation model.

Tsunokuni, Izumi; Matsuhashi, Haruka; Ikeda, Yusuke
Tokyo Denki University, Japan

Extended Reality Audio
Paper Number: 10582
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21695

Neural Synthesis of Footsteps Sound Effects with Generative Adversarial Networks

Footsteps are among the most ubiquitous sound effects in multimedia applications. There is substantial research into understanding the acoustic features and developing synthesis models for footstep sound effects. In this paper, we present a first attempt at adopting neural synthesis for this task. We implemented two GAN-based architectures and compared the results with real recordings as well as six traditional sound synthesis methods. Our architectures reached realism scores as high as recorded samples, showing encouraging results for the task at hand.

Comunità, Marco; Phan, Huy; Reiss, Joshua D.
Centre for Digital Music, Queen Mary University of London, UK

Audio Synthesis & Audio Effects
Paper Number: 10583
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21696

Localization of auditory events in scenarios with lateral, projected sound sources and frontal, attenuated direct sound

Loudspeaker arrays enable sound reproduction from different directions by taking advantage of wall reflections. Ideally, a listener only perceives sound coming from the reflective surface. Since sound arriving at the listener directly from the device is usually not attenuated completely it interferes with the reflected sound. Previous research has investigated perceptual influences of level differences between direct sound and reflected sound as well as the influence of type of stimulus in such scenarios. The current paper expands on this topic by including the influence of the relative delay between reflected sound and direct sound. Results suggest that the direct-to-reflection delay does not influence the localization of a projected sound, while endorsing the importance of the direct sound attenuation, as well as the strong influence of the signal type.

Männer, Johannes; Stenzel, Hanne; Walther, Andreas; Melchior, Frank
Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany; Hochschule der Medien, Stuttgart, Germany

Sound Quality & Perception
Paper Number: 10584
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21697

Synthesis of Wind Instruments using BiLSTM and Gaussian Mixture Model

In this work, a source filter model incorporating the BiLSTM and the Gaussian Mixture Model (GMM) for the synthesis of woodwind instruments is presented. Magnitude-modulated and pitch-synchronous impulse signals are the sources. The filter is converted by the correspondent DCT coefficients, divided into the low-frequency and high-frequency parts. The high-frequency part is modeled by the GMM. The BiLSTM recurrent networks are used to predict the low-frequency part DCT coefficients of the filters. The proposed method can synthesize realistic and expressive tones and breath noise as well when compared to the conventional Digital Waveguide Filter-based method.

Lee, Yu-Wen; Yang, Hung-Chih; Su, Alvin W.Y.
SCREAM Lab., Department of CSIE, National Cheng-Kung University Tainan, Taiwan

Audio Synthesis & Audio Effects
Paper Number: 10585
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21698

On Loudspeakers as Recording Devices

Moving-coil electrodynamic loudspeakers and dynamic microphones use the same linear actuator technology at the core of their operation. Utilising this similarity, loudspeakers have a possible use as recording devices in cases where using dedicated microphones is not feasible. Such a use case exists in public address and voice alarm systems. This paper evaluates the feasibility of using the loudspeakers already in place in these systems as recording devices to provide information back to the system. A system using a single loudspeaker as both a playback and recording device simultaneously is analysed, modelled and simulated. The results show that using a current measuring set-up with an analogue-to-digital converter capable of detecting a range of roughly 120 dB, a speech signal incident at 46 dBSPL in a cone of 150° from a loudspeaker can be successfully estimated in an of?ce room with an announcement playing at 88 dBSPL and background interference present at the same time. As the estimated signal is unknown to the system, the solution generalises to other signal types as well.

Roest, Tobias; Martinez, Jorge; Oosterom, Han; Hendriks, Richard C.
TU Delft; Bosch Security Systems B.V.

Recording Technologies
Paper Number: 10586
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21699

Overall listening experience for binaurally reproduced audio

This work formulates a simple experiment to examine listener preference for binaurally recorded music reproduced via headphones in three different formats: mono, stereo and surround. This

paradigm introduces a framework that can associate prior listener preferences to the music content (referred to as cognitive factor) to perceptual factors associated with its presentation format, here due to the different Envelopment perceived by the listener for each of the three reproduction systems. A test procedure was employed to register the individual listener preference and perceived Envelopment for different music samples, presented in these alternative spatial audio systems. A simple empirical perceptual model for predicting with high accuracy the Envelopment of each sample is proposed, using features extracted directly from the binaural signal. Utilizing the listener preference for the stereo reproduced samples and incorporating the output of the above perceptual model, a linear model is implemented which can accurately predict the individual listener preference for samples reproduced via the other two tested systems, i.e. for mono and surround. At present, additional affective factors which can also contribute towards listener preference have not been considered. However, the proposed model achieves high accuracy for predicting individualized preference, showing also that the cognitive factors are dominant on listener preference decisions whilst a lesser impact was registered by the perceptual factors introduced by the audio system.

Moiragias, George; Economou, Konstantza; Mourjopoulos, John
University of Patras, Greece

Binaural Audio
Paper Number: 10587
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21700

Capturing Spatial Room Information for Reproduction in XR Listening Environments

An expansion on previous work involving “holographic sound recording” (HSR), this research delves into how sound sources for directional ambience should be captured for reproduction in a 6-DOF listening environment. We propose and compare two systems of ambient capture for extended reality (XR) using studio-grade microphones and first-order soundfield microphones. Both systems are based on the Hamasaki-square ambience capture technique. The Twins-Hamasaki Array utilizes four Sennheiser MKH800 Twins while the Ambeo-Hamasaki Array uses four Sennheiser Ambeo microphones. In a preliminary musical recording and exploration of both techniques, the spatial capture from these arrays, along with additional holophonic spot systems, were reproduced using Steam Audio in Unity’s 3D engine. Preliminary analysis was conducted with expert listeners to examine these proposed systems using perceptual audio attributes.The systems were compared with each other as well as a virtual ambient space generated using Steam Audio as a reference point for auditory room reconstruction in XR. Initial analysis shows progress towards a methodology for capturing directional room reflections using Hamasaki-based arrays.

Matsakis, Michael; Songmuang, Parichat; Zhang, Kathleen
New York University, NY, USA; McGill University, Montreal, QC, Canada

Extended Reality Audio
Paper Number: 10588
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21701

Circular harmonic processing with an ad-hoc smartphone array for speech enhancement

A circular microphone array may be assembled on the spot by combining the built-in microphones of several smartphones in a circular layout. The present work investigates different designs of such ad-hoc smartphone arrays that would allow the implementation of spatial audio techniques formulated in the circular harmonic domain. The tested techniques include circular harmonic beamforming for

Direction-Of-Arrival estimation and noise reduction, as well as spatial post-filters for additional background noise suppression. The evaluation of the different designs is performed using a simulated dataset of a multi-speaker event. The comparisons are done with objective metrics, as well as with a subjective listening test.

Bountourakis, Vasileios; Vryzasn Nikolaos; Dimoulas, Charalampos
Aalto University, Espoo, Finland; Aristotle University of Thessaloniki, Greece *(See document for exact affiliation information.)*

3D/Immersive/Spatial Audio
Paper Number: 10589
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21702

Moved By Sound: How head-tracked spatial audio affects autonomic emotional state and immersion-driven auditory orienting response in VR Environments

This paper presents a narrative content-driven virtual reality (VR) experiment using novel biosensing technology to evaluate emotional response to a complex, layered soundscape that includes discrete and ambient sound events, music, and speech. Stimuli were presented in a spatialized vs mono audio format, to determine whether head-tracked spatial audio exerts an effect on physiologically measured emotional response. The extent to which a listener's sense of immersion in a VR environment can be increased based on the spatial characteristics of the audio is also examined, both through the analysis of self-reported immersion scores and physical movement data. Finally, the study explores the relationship between the creators' own intentions for emotion elicitation within the stimulus material, and the recorded emotional responses that matched those intentions in both the spatialized and non-spatialized case. The results of the study provide evidence that spatial audio can significantly affect emotional response in Immersive Virtual Environments (IVEs). In addition, self-reported immersion metrics favour a spatial audio experience as compared to a non-spatial version, while physical movement data shows increased user intention and focused localization in the spatial vs non-spatial audio case. Finally, strong correlations were found between the creators of the soun

Warp, Richard; Zhu, Michael: Kiprijanovska, Ivana; Wiesler, Jonathan; Stafford, Scot; Mavridou, Ifigeneia
Pollen Music Group, San Franciso, CA, USA; emteq labs, Sussex Innovation Centre, Brighton, UK

Extended Reality Audio
Paper Number: 10590
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21703

A citizen science approach to support joint air quality and noise monitoring in urban areas

In the present work, a crowdsourcing approach is designed, to investigate the correlation between air and noise pollution in urban areas. Citizens are requested to provide air quality measurements and audio recordings using a prototype mobile application specially designed to motivate them to undertake the task of audiovisual capturing. Different use case scenarios of the application are presented, along with the technical specifications and service-based architecture. The UrESC22 dataset is formed, a subset of the ESC50 benchmark dataset consisting of all classes related to polluting activities (vehicles, engines, etc.). The dataset is used to train a convolutional neural network classifier for the detection of audio events related to air pollution.

Stamatiadou, Marina Eirini; Vryzas, Nikolaos; Vrysis, Lazaros; Saridou, Theodora; Dimoulas, Charalampos
Aristotle University of Thessaloniki, Greece

Machine Learning / Artificial Intelligence
Paper Number: 10591
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21704

On the use of integral input devices and relative mappings for an immersive audio mixing task. Part I.

This generation of spatial audio systems are capable of 3D sound reproduction and multi-channel sound sources panning and mixing. Currently, these systems require the user to employ standard, low-level interaction tools such as knobs and sliders to handle audio sources and reproduction setups with higher complexity. Innovative tools and control methods designed to meet the requirements of immersive audio environments are essential to enable efficient implementation of these technologies and software. This work investigated three types of input devices as controllers for an immersive audio mixing task in a spatial audio system. This paper concentrates on the subjective evaluation of the devices while a subsequent paper will focus on the objective data captured by the apparatus of the experiment. Results suggest implementing input devices that match the structure of the task greatly improved participants' sense of motivation and stimulation while completing the trials.

Quiroz, Diego; Martin, Denis
McGill University, Montreal, QC, Canada; Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Montreal, QC, Canada

Studio Technology
Paper Number: 10592
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21705

Disentangled estimation of reverberation parameters using temporal convolutional networks

Reverberation is ubiquitous in everyday listening environments, from meeting rooms to concert halls and record-ing studios. While reverberation is usually described by the reverberation time, getting further insight concerning the characteristics of a room requires to conduct acoustic measurements and calculate each reverberation param-eter manually. In this study, we propose ReverbNet, an end-to-end deep learning-based system to non-intrusively estimate multiple reverberation parameters from a single speech utterance. The proposed approach is evaluated using simulated room reverberation by two popular effect processors. We show that the proposed approach can jointly estimate multiple reverberation parameters from speech signals and can generalise to unseen speakers and diverse simulated environments. The results also indicate that the use of multiple branches disentangles the embedding space from misalignments between input features and subtasks.

Thoidis, Iordanis; Vryzas, Nikolaos; Vrysis, Lazaros; Kotsakis, Rigas; Kalliris, George; Dimoulas, Charalampos
Aristotle University of Thessaloniki, Greece

Machine Learning / Artificial Intelligence
Paper Number: 10593
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21706

Sensory evaluation of spatially dynamic audiovisual sound scenes: a review

Spatial audio systems are capable of rendering moving sound sources. This is becoming more common in commercial and domestic environments, driven by interest in spatial audio and virtual reality, and underpinned by object-based audio delivery formats. When consulting standard methods for sound quality evaluation in terms of spatially dynamic sound scenes (SDSS), challenges emerge indicating the methods are inadequate for the application. This article presents an overview of state-of-the-art sound quality evaluation methods, particularly focusing on their appropriateness for evaluation of SDSS. Limitations of current methodologies are discussed, and research of temporal evaluation methodologies used in other sensory sciences are reviewed for their potential applicability to audio quality assessment.

Porysek Moreta, Pia Nancy; Bech, Søren, Francombe, Jon; Østergaard, Jon; van de Par, Steven; Kaplanis, Neofytos
Bang & Olufsen a/s, Struer, Denmark; Aalborg University, Department of Electronic Systems, Aalborg, Denmark; Carl von Ossietzky University, Oldenburg, Germany

Sound Quality & Perception
Paper Number: 10594
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21707

An Improved Approach for Binaural Room Impulse Responses Interpolation in Real Environments

The interpolation of binaural room impulse responses (BRIRs) is a very common technique in the field of immersive audio rendering because it aims at reducing the responses measurements guaranteeing a correct sound spatialization. These approaches are usually based on the interpolation of the responses starting from the division in early reflections and reverberant tail considering a fixed arbitrary mixing time. However, this aspect could influence the interpolation methodology when real BRIRs are considered. In this paper, a state-of-the-art interpolation algorithm is improved adding an automatic procedure for the calculation of the exact mixing time and applied it to a set of BRIRs, measured in a real reverberant environment. Several experiments have proved the effectiveness of the proposed approach.

Bruschi, Valeria; Nobili, Stefano; Terenzi, Alessandro; Cecchi, Stefania
DII - Università Politecnica delle Marche, Ancona, Italy

Binaural Audio
Paper Number: 10595
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21708

Efficient neural networks for real-time modeling of analog dynamic range compression

Deep learning approaches have demonstrated success in modeling analog audio effects. Nevertheless, challenges remain in modeling more complex effects that involve time-varying nonlinear elements, such as dynamic range compressors. Existing neural network approaches for modeling compression either ignore the device parameters, do not attain sufficient accuracy, or otherwise require large noncausal models prohibiting real-time operation. In this work, we propose a modification to temporal convolutional networks (TCNs) enabling greater efficiency without sacrificing performance. By utilizing very sparse convolutional kernels through rapidly growing dilations, our model attains a

significant receptive field using fewer layers, reducing computation. Through a detailed evaluation we demonstrate our efficient and causal approach achieves state-of-the-art performance in modeling the analog LA-2A, is capable of real-time operation on CPU, and only requires 10 minutes of training data.

Steinmetz, Christian J.; Reiss, Joshua D.
Centre for Digital Music, Queen Mary University of London, UK

Machine Learning / Artificial Intelligence
Paper Number: 10596
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21709

Design and implementation of a flexible and low-budget acoustic treatment for a reference listening room

Audio reproduction quality depends on system qualities as well as on acoustic properties of the room, which can be controlled by appropriate acoustic treatment. The goal of our project was to adapt a 9 x 6 x 3 m shoe box shaped room, not initially intended for audio purposes, so that it could be used as a reference listening room in a range of applications requiring various acoustic conditions from live to dead. The treatment was based on the extensive use of wideband porous absorbers. Initially the room was too reverberant for audio purposes and produced intolerable flutter echo. Finally, we obtained reverberation time of 0.45 s in 63 Hz third octave band and a nearly flat reverberation time curve above 125 Hz, at the value of 0.2 s.

Kleczkowski, Piotr; Czesak, Karol; Kmiecik, Michal; Król Nowak, Aleksandra; Makuch, Teresa
AGH University of Science and Technology, Kraków, Poland

Room Acoustics
Paper Number: 10597
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21710

The next generation of audio accessibility

Technological advances have enabled new approaches to broadcast audio accessibility, leveraging metadata generated in production and machine learning to improve blind source separation (BSS). This work presents two contributions to accessibility knowledge: first, a quantitative comparison of two audio accessibility methods, Narrative Importance (NI) and Dolby AC-4 BSS. Secondly, an evaluation of the audio access needs of neurodivergent audiences. The paper presents two comparative studies. The first study shows that the AC-4 BSS and NI methods are ranked consistently higher for clarity of dialogue (compared to the original mix) whilst improving, or retaining, perceived quality. A second study quantifies the effect of these methods on word recognition, quality and listening effort for a cohort including normal hearing, d/Deaf, hard of hearing and neurodivergent individuals, with NI showing a significant improvement in all metrics. Surveys of participants indicated some overlap between Neurodivergent and d/Deaf and hard of hearing participants' access needs, with similar levels of subtitle usage in both groups.

McClenaghan, Iain; Pardoe, Lawrence; Ward, Lauren
BBC R&D, London, UK; University of York, York, UK

Sound Quality & Perception
Paper Number: 10598
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21711

Effect of spatial audio processing on attention redirection and speech intelligibility in multi-talker communication

Most current communication systems for mission critical applications use devices with relatively simple audio presentation to the user, employing single channel communication. An increasing number of applications are requiring users to monitor and listen to multiple talkers concurrently. This paper describes results of an experiment investigating the differences in performance in a transcription task when listeners were listening to multi-talker communication using two different Head-Related Transfer Functions (HRTF) data sets in comparison to monaural listening. The goal of this research is two-fold: 1) to assess the impact on attention redirection and the accuracy of the NATO phonetic word transcription based on the pre-selected HRTFs used, as compared to a diotic presentation, as the number of talkers increases; 2) to investigate whether there is a difference in performance between younger and older listeners. Results indicate a significant improvement when using spatial audio, in the accuracy of the transcription and number of missed characters, especially for participants in the older age group.

Growney, Eric; Roginska, Agnieszka
Otto Engineering Inc, Carpentersville, IL, USA; New York University, New York, NY, USA

Binaural Audio
Paper Number: 10599
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21712

All-Pass Hilbert Filters

This paper presents a method to generate all-pass filters whose phase is defined by the magnitude response of traditional IIR filter prototypes. The purpose of this is to make the overall shape of the filter controllable in an intuitive way so that the end user can specify arbitrary, but well-behaved, phase responses. At the core of the algorithm is the Hilbert transformer, which has been thoroughly studied with different implementations. However, with a simple modification we obtain the "Phase Displacer" which can produce an arbitrary constant phase shift at all frequencies. Expanding this idea per frequency we obtain the "Phase Equalizer": an all-pass filter whose shape is the magnitude response of either a shelving or peaking IIR filter.

Sierra, Juan
New York University, NY, USA

Audio Synthesis & Audio Effects
Paper Number: 10600
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21713

Spatially Dynamic Sound Quality Evaluation: A Literature Study

In the context of quality evaluation of reproduced sound, it is unknown if existing sensory evaluation methods for stationary spatial sound conditions would also work when sound sources or listeners move during the formation of a quality judgement in the listener's mind. To investigate this problem, we review literature in the fields of sound quality evaluation and dynamic sensory evaluation. We also describe specific methodologies for the temporal assessment of food quality and present them as potential candidates for spatially dynamic sound quality evaluation. Finally, we discuss the results of

a pilot experiment where we implemented the Temporal-Check-All-That-Apply (TCATA) methodology to evaluate the sound quality of spatially dynamic sound stimuli.

Gil, Juan; Bech, Søren; Christensen, Flemming
Aalborg University, Aalborg, Denmark; Bang & Olufsen A/S, Struer

Sound Quality & Perception
Paper Number: 10601
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21714

Temporal Evaluation of Sound Quality using TCATA

In this paper, we present an experiment based on a method called Temporal-Check-All-That-Apply (TCATA). The experiment was carried out to study the temporal perception of various attributes of sound quality when sounds move around a stationary listener. The first goal of the experiment was to understand if we could capture any changes in the perceived quality that reflected the spatial characteristics of the sounds reproduced in the lab using loudspeakers. The second goal was to evaluate if the method could record perceptual differences between stimuli with non-expert listeners. We concluded that TCATA is a suitable method to capture changes in the perception of various sound quality attributes over time. Still, we observed that dynamic physical characteristics are not necessarily represented as dynamic perceptual characteristics in all cases. The experimental variables programme, reproduction system, trajectory of a moving sound source, and loop showed a statistically significant effect on the results. Based on our findings, we propose recommendations for future investigations.

Gil, Juan; Bech, Søren; Christensen, Flemming
Aalborg University, Aalborg, Denmark; Bang & Olufsen A/S, Struer,

Sound Quality & Perception
Paper Number: 10602
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21715

Integrated Architectural and Acoustical Design of Media Recording Studios

Communication and connection among people in the 21st century, especially in the post-Covid era are enhanced through broadcast and recorded media. There is a significant expansion of web-based media, traditional television, radio, film and immersive entertainment and recreational facilities that require highly specialized, technically sophisticated recording and broadcast facilities to facilitate electronic communication and live communication among people. This presentation explores psycho-acoustic perceptual criteria for the spaces so recorded sounds are reproduced for immersive playback experiences; computer-based design methods to evaluate the performance of the architectural systems used in these rooms; auralizations and other perceptual methods to study the integration of digital technical media systems with the architectural design; and new instrumentation systems and innovative technology used during the construction phase to determine and/or verify the acoustical performance of the architectural and acoustical systems. The core of this presentation is to allow the audience to apply the links between human aural criteria that are emerging in the technical literature in the architectural design of these highly specialized facilities that are now being built in just about every type of building. The unique shapes, materials and building systems employed in these facilities are shown to have a perceptual and technical basis rooted in the human perception of sounds. The intentional design of specific sound qualities in the recorded sounds is shown to be affected by the architectural design of the spaces. The perception of these sound qualities is essential for the people

doing the recording inside the studio and also for the listeners of the recorded or broadcasted materials who may be listening in a conference room in an office, a lecture hall, a classroom, a cinema, a theme park, a pair of earbuds, a living room or home theater and just about any other space where broadcast or recorded sounds are heard. The scientific and perceptual basis for the shape, materials, building systems and recording technology to achieve the specific sonic qualities desired by the production and creative staff will be illustrated through applied case studies and recorded examples of sounds so attendees can synthesize and apply these principals in their practice.

Siebein, Keely; Siebein, Gary, Roa, Marylin; Miller, Jennifer; Vetterick, Matthew; Siebein Jr, Gary
Siebein Associates, Inc., Gainesville, FL, USA

Room Acoustics
Paper Number: 10603
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21716

Discriminability of concurrent virtual and real sound sources in an augmented audio scenario

This exploratory study investigates peoples' ability to discriminate between real and virtual sound sources in a position-dynamic headphone based augmented audio scene. For this purpose, an acoustic scene was created consisting of two loudspeakers at different positions in a small seminar room. Considering the presence of headphones, non-individualized BRIRs measured along a line with a dummy head wearing AKG K1000 headphones were used to allow for head rotation and translation. In a psychoacoustic experiment, participants had to explore the acoustic scene and tell which sound source they believe is real or virtual. The test cases included a dialog scenario, stereo pop-music and one person speaking while the other speaker played mono-music simultaneously. Results show that the participants were on trend able to debunk individual virtual sources. However, for the cases where both sound sources reproduced sound simultaneously, lower distinguishability rates were observed.

Schneiderwind, Christian; Neidhardt, Annika
Technische Universität Ilmenau, Germany

Extended Reality Audio
Paper Number: 10604
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21717

Mixing for in-ear monitors: understanding the work of monitor mixing engineers

The monitor mixing engineer is a key function in a public music performance when in-ear monitoring (IEM) is utilized. This paper aims to expand our knowledge about monitor mixing in general and IEM mixing in particular by investigating monitor mixing engineers' reasoning, decisions and actions. Four experienced monitor mixing engineers were interviewed on monitor mixing in general, IEM mixing and hearing health. Among the results are found that the engineers seek to create a fruitful working relationship with the performers. The engineers also describe what creates a functional mix and they show a high awareness of their responsibility for the comfort and well-being of the artist, both in a psychological sense as well as in providing sound levels that are not harmful.

Berg, Jan; Johannesson, Tomas; Nykänen, Arne
Luleå University of Technology, Luleå, Sweden

Loudness & Perception
Paper Number: 10605
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21718

Production Approaches, Loudspeaker Configurations, and Signal Processing Architectures for Live Virtual Acoustic Performance

Production workflows and signal processing architectures are described for staging live virtual acoustic concerts at venues having different loudspeaker configurations. The idea is to provide the audience with an immersive experience, while giving the performers the monitoring and sense of space needed to occupy and interact with the simulated acoustic. The auralization processing and monitoring are described for loudspeakers located (a) centrally, radiating outward from near the stage; (b) around the venue perimeter, radiating inward toward the audience; and (c) throughout the hall. Detailed production procedures and signal processing architectures are described using virtual acoustics performances by Cappella Romana in the TivoliVredenburg Grote Zaal in Utrecht with centrally positioned loudspeaker arrays, the Ritz Carlton Ballroom in San Francisco with perimeter configured loudspeakers, and the Bing Concert Hall at Stanford University with loudspeakers mounted throughout the hall.

Estakhrian, Hassan; Abel, Jonathan
Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, CA, USA

Room Acoustics
Paper Number: 10606
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21719

Watching on the Small Screen: The Relationship Between the Perception of Audio and Video Resolutions

A new quality assessment test was carried out to examine the relationship between the perception of audio and video resolutions. Three video resolutions and four audio resolutions were used to answer the question: “Does lower resolution video influence the perceived quality of audio, or vice versa?” Subjects were asked to use their own equipment, which they would be likely to stream media with. They were asked to watch a short video clip of various qualities and to indicate the perceived audio and video qualities on separate 5-point Likert scales. Four unique 10-second video clips were presented in each of 12 experimental conditions. The perceived audio and video quality ratings data showed different effects of audio and video resolutions. The perceived video quality ratings showed a significant effect of audio resolutions, whereas the perceived audio quality did not show a significant effect of video resolutions. Subjects were divided into two groups based on the self-identification of whether they were visually or auditorily inclined. These groups showed slightly different response patterns in the perceived audio quality ratings.

Bartel, Nicholas; Hui Chon, Song
Belmont University, Nashville, TN, USA

Television Audio
Paper Number: 10607
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21720

Plausibility of an approaching motion towards a virtual sound source II: In a reverberant seminar room

This study investigates the plausibility of dynamic binaural audio scenarios wherein the listener interactively walks towards a virtual sound source. An originally measured BRIR set was manipulated and simplified systematically to challenge plausibility, explore its limits, and examine the relevance of selected acoustic properties. After the first investigation in a quite dry listening laboratory, this second exploratory study repeats and extends the experiment in a considerably more reverberant room. The participants had to rate externalization, continuity, stability of the apparent sound source, impression of walking towards the sound source and the plausibility of the virtual acoustic scene. The results confirm the observations of the first study in the different acoustic environment. Both studies indicate much room for simplifications, but certain modifications seriously affect plausibility. Even inexperienced listeners notice if the progress of the auditory distance change does not match their own walking motion. In addition, the meaning of context and expectation for the perception of binaural audio is highlighted.

Neidhardt, Annika; Kamandi, Samaneh
Technische Universität Ilmenau, Germany

Sound Quality & Perception
Paper Number: 10608
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21721

Study of the audibility of background music in TV programs: towards a normative proposal

In this work, the problem of the audibility of background music in television programs is studied. After reviewing the state of the art and verifying its incipient state, the problem is faced considering the 3 levels of audibility defined by the WIPO and beginning the study trying to find the threshold between inaudible and audible. It is considered that the music and voice tracks are available separately and a series of subjective tests are prepared, carried out with abundant realistic material and in controlled conditions as similar as possible to the television room of an average home. An analysis of the results reveals that the difference in integrated loudness between voice and music is the most defining factor in audibility, although the type of music also reveals a certain influence. To take this influence into account, various indicators related to the momentary loudness of the signal were tested, finally obtaining a highly correlated statistic. By means of a linear regression, an expression dependent on both parameters was obtained that provides a very stable final estimator and with a mean error with respect to the jury's mean of about 0.9 dB for the sound material tested. This result can serve as a basis for the elaboration of a recommendation in this field. For the case of broadcast analysis where voice and music are mixed, the new voice-music separation techniques based on deep learning neural networks allow resynthesizing both isolated tracks at the destination to apply the proposed algorithm.

López, Jose J.; Ramallo, Suso
Universitat Politècnica de València, Valencia, Spain

Television Audio
Paper Number: 10609:
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21722

Degradation in reproduction accuracy due to sound scattered by listener's head in local sound field synthesis

The purpose of this study is to investigate the phenomenon of degradation in the accuracy of local sound field synthesis (LSFS) due to the sound scattered by a listener’s head. In conventional sound field synthesis (SFS) methods, the degradation in accuracy due to a listener’s head is negligible,

because the degradation are smaller at the low reproducible frequencies than the discretization artifacts of synthesized sound field. As LSFS method synthesizes the sound field only to a narrow extent at higher frequencies which is not considered in the conventional methods, how degraded the reproduction accuracy due to scattered sound in LSFS must be investigated. We conducted simulation experiments, using a rigid sphere for modeling the sound scattered by the head, using two LSFS methods: local wave field synthesis with virtual secondary sources (LWFS-VSS), and the pressure-matching method. The following two points were investigated: (i) The dependency of degradation on the frequency of sound and reproduction position; and (ii) the relationship between the virtual source distance and reproduction accuracy. The results showed that the degradation in the accuracy at the position opposite to the virtual source became larger as the frequency increased. Regarding the distance of the virtual source, when the source was placed near the listener's head, the reproduction accuracy was significantly low. Specifically, in the case of LWFS-VSS, as the virtual source approached the head, the reproduction accuracy became more degraded compared with the no-scattering condition.

Tsunokuni, Izumi; Ikeda, Yusuke
Tokyo Denki University, Japan

Spatial Audio
Paper Number: 10610
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21723

An open dataset of measured HRTFs perturbed by headphones

An acoustic transfer function dataset is presented that contains the influence on the head-related transfer function (HRTF) of headphones worn by a dummy-head. The database contains the HRTFs of a KEMAR dummy head in the horizontal plane with an angular resolution of 5 degrees, as well as the perturbed HRTFs of 34 different headphones of all types and functionalities, such as over-ear, on-ear, in-ear, and earplug headphones, with and without active noise control and hear-through functionality. Each headphone was measured a total of 5 times after repositioning it. The data set allows partial inter- and intra-individual investigations, for example of the angle-dependent insertion loss, since some headphone

Schlieper, Roman; Preihs, Stephan; Peissig, Jürgen
Gottfried Wilhelm Leibniz Universität Hannover, Germany

Loudspeakers and headphones
Paper Number: 10611
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21724

Time-Frequency Adaptive Room Optimization of Audio Signals

Room equalization (REQ) is a common method to adapt audio signals to the room in which they are reproduced in. REQ for example attenuates the audio signal at the room resonance frequencies, to reduce negative effects at those frequencies, when the signal is played back. REQ is a time-invariant method. Recently a time-frequency adaptive method to adapt audio signals to rooms has been proposed [1]. The results of a subjective evaluation are presented in this paper. Amount of room reverb and quality are assessed in a blank room, same room with absorbers, and blank room with time-frequency adaptive processing.

Maurer, Samuel; Faller, Christof
Graz University of Technology, Graz, Austria; Illusonic GmbH, Uster, Switzerland

Room Acoustics
Paper Number: 10612
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21725

Application of the Multi-Objective Optimisation Framework to the Creation of Personal Sound Zone

The main methods used to create multiple personal sound zones are optimisation processes involving conflicting objective functions. The two most common methods, Acoustic Contrast Control (ACC) and Pressure Matching (PM), aim to simultaneously optimise various objectives: the system performance in terms of reproduction error, the amount of acoustic energy in the so-called bright zones and dark zones, and the amount of energy required by the system to obtain the desired performance (the array effort). In this work, the general framework of multi-objective optimisation and the main techniques employed to solve this class of problems are firstly introduced. Then, it is shown that the most renowned methods developed in the last 40 years in the field of personal sound zone systems are special cases of theses general multi-objective techniques.

Gallian, Wilfried; Fazi, Filippo Maria
ISSR, University of Southampton, UK

Loudspeakers and headphones
Paper Number: 10613
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21726

Non-Ideal Operational Amplifier Emulation in Digital Model of Analog Distortion Effect Pedal

Digital models of analog guitar effects pedals have largely ignored the impact of non-ideal components on the resulting timbre, though the physical limitations of analog components are sometimes key to achieving the intended effect. The signature sound of the Pro Co RAT is largely attributed to the non-ideal characteristics of the Motorola LM308 operational amplifier, particularly the slew-rate, gain-bandwidth product and supply voltage. Analysis of harmonic and spectral content shows that the inclusion of these non-ideal component characteristics results in a more accurate recreation of the Pro Co RAT distortion effect. In a comparison of real-time digital models, the additional computational cost of the non-ideal model was negligible.

Leete, Timothy; Tarr, Eric; Ko, Doyuen
Belmont University, Nashville, TN, USA

Audio Synthesis & Audio Effects
eBrief:666
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21727

A dataset of measured spatial room impulse responses in different rooms including visualization

In this contribution, an open-source dataset of captured spatial room impulse responses (SRIRs) is presented. The data was collected in different enclosed spaces at the Technische Universität Ilmenau using an open self-build microphone array design following the spatial decomposition method (SDM)

guidelines. The included rooms were selected based on their distinctive acoustical properties resulting from their general build and furnishing as required by their utility. Three different classes of spaces can be distinguished, including seminar rooms, offices, and classrooms. For each considered space different source-receiver positions were recorded, including 360? images for each condition. The dataset can be utilized for various augmented or virtual reality applications, using either a loudspeaker or headphone-based reproduction alongside the appropriate head-related transfer function sets. The complete database, including the measured impulse responses as well as the corresponding images, is publicly available.

Klein, Florian; Surdu, Tatiana; Aretz, Arthur; Birth, Kilian; Edelmann, Niklas; Seitelmann, Florian; Ziener, Christian; Werner, Stephan; Sporer, Thomas
Technische Universtät Ilmenau, Germany; Fraunhofer Institute for Digital Media Technology, Germany

Extended Reality Audio
eBrief:667
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21728

Developing a Binaural Renderer for Audio Definition Model Content

The Audio Definition Model (ADM) can be used to represent object- and scene-based audio programmes, and there is a standardised method for reproducing ADM content on loudspeakers, but not currently for headphones. We present the design and implementation of a binaural renderer for the ADM, which supports the features of the ADM while maintaining high-quality output. For rendering objects the system uses virtual loudspeaker rendering with windowed binaural room impulse responses (BRIRs). To reduce comb-filtering effects, delay is removed from the BRIRs and replaced with a per-ear and per-object variable fractional delay line. When rendering diffuse sources, the original delays are used, as the varied onset delays help create perceived extent. The overall gain of each source is adjusted dynamically to compensate for loudness changes caused by interactions between BRIRs of neighbouring loudspeakers. The system is available as an open-source C++ library based on the VISR framework, suitable for adding real-time head-tracked binaural output to applications, and is built into the EAR Production Suite.

Nixon, Thomas; Franck, Andreas; Pike, Chris; Reich, Galen:
British Broadcasting Corporation, UK; University of Southampton, UK; Sonos, Inc., USA
Binaural Audio
eBrief:668
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21729

Design of a lightweight acoustical measurement room

The paper presents the design principles of an acoustic test chamber, where the insulation requirements of typical measurement rooms are relaxed and so constructing the surfaces using very lightweight materials, consisting only of absorbents and a simple frame, is possible. The test chamber constructed according to these principles shows good absorption characteristics down to 200Hz and has a significantly larger free space for measurements than a conventional chamber designed using wedges and solid walls.

Backman, Juha
AAC Technologies Solutions Finland, Turku, Finland

Room Acoustics
eBrief:669

Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21730

Designing sound system in-band headroom based on expected difference between C- and A-weighted levels

Sound pressure level (SPL) is the standard metric for regulations regarding environmental noise exposure. Because performances are often regulated by their A-weighted sound level, it is tempting to think that A-weighted level should be the primary design consideration for sound system headroom. Because A-weighting disregards significant low-frequency energy, it is possible to create a wide variety of spectra with the same A-weighted level, but each having a different spectral shape and C-weighted level. While regulators correlate excessive A-weighted levels with hearing damage, A-weighted levels are less well correlated with community annoyance. The Netherlands has recognized this and created a permitting system incorporating the difference between C- and A-weighted sound levels (C-A) as a measure of low-frequency content. This Brief gives supporting evidence for the correlation between C-A levels and different musical genres and offers complementary design guidance corresponding to sound system headroom with emphasis on in-band levels.

van Veen, Merlijn, Schwenke, Roger; McCarthy, Bob
Meyer Sound Laboratories, Berkeley, CA, USA

Loudspeakers and headphones
eBrief:670
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21731

Overview of Evaluation Methods of Sound Field Reproduction Systems

Sound reproduction is a task where the goal is to accurately reproduce sound field of a previously captured audio scene in a defined controlled receiving space. According to the current state-of-the-art technology there is no system available which can perform such task perfectly. Unrestricted reproduction of a three-dimensional sound field would require a spatially continuous sound source, therefore any real sound field reproduction system, consisting of discrete sound sources, offers only an approximation of the original sound field. Each system can be objectively evaluated against various aspects of sound field reproduction accuracy like spectral and level matching, quality and distortions, time, and directivity cues, "sweet spot" size, the influence of obstacles in the receiving space (like equipment or people) or using perceptual objective metrics. The goal of this work is to compile these evaluation methods and provide mapping to the abovementioned key aspects of sound field reproduction. The authors contextualize the overview in the acoustic testing perspective, where specific aspects of a sound reproduction accuracy matter in evaluation of different audio-related feature of a device (e.g., for testing quality of Direction of Arrival – directivity ques are critical for evaluation, whereas for testing Acoustical Scene Classification – spectral and level accuracy is more relevant) but the evaluation methods and findings can be translated into other areas of audio research.

Banas, Jan; Grzywa, Michal; Jezierski, Ryszard; Klinke, Piotr; Koszewski, Damian; Kuklinowski, Maciej; Maziewski, Przemyslaw; Pach, Pawel; Stanczak, Dominik; Trella, Pawel
Intel Technology Poland, Gdansk, Poland

Spatial Audio
eBrief:671
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21732

Numerical and Experimental Analysis of a Metamaterial-based Acoustic Superlens

For many years, the engineering limitations in a single loudspeaker have offered no solution to the problem of delivering sound only to parts of an audience. Precise control on how sound is delivered to an audience has required multiple loudspeakers, either through their distribution or through DSP. The recent uptake of acoustic metamaterials, however, seem to offer different solutions. Using devices based on acoustic metamaterials, for instance, brings to acoustics design principles that come directly from optics, at a reasonable manufacturing cost. In this work, we design, numerically simulate, and characterise an acoustic converging superlens: a 3D-printed device capable of focusing an incoming plane wave at a distance less than one wavelength. We show how a loudspeaker at a fixed distance from the lens results in an "image" of the source at a distance prescribed by the thin-lens equation. Finally, we propose possible applications of such an acoustic superlens to future audio experiences.

Chisari, Letizia; Ricciardi, Enrico; Memoli, Gianluca
Metasonixx Ltd, London, UK; Labirinti Acustici, Milan, Italy; University of Sussex, Brighton, UK

Loudspeakers and headphones
eBrief:672
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21733

Phase Mitigation Through Filter Design

In both acoustic and digital systems, delays and the resulting phase interference are an innate feature of sound recording; traditionally, phase-interference mitigation is applied through temporal offset to attempt time coherence between multiple signal paths. Filter design presents an alternative solution to phase issues, wherein predictive modeling allows for a filter to apply corrective magnitude response. Such application of filter design presents its own set of problems and could further be explored in creative, rather than remedial, settings.

Bailey, SeanTemple University, Philadelphia PA, USA

Audio Synthesis & Audio Effects
eBrief:673
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21734

Virtual Studio Production Tools with Personalized Head Related Transfer Functions for Mixing and Monitoring Dolby Atmos and Multichannel Sound

With the increasing popularity of audiophile headphones in this decade, the need for mixing over headphones is on the rise. Studio engineers use headphones as a critical tool for checking their mixes over the headphones before publishing them. As dolby atmos music and surround sound music is currently regaining popularity, there is also an increasing need for having multi channel speaker setups and associated gear in the studio to produce music in such formats. Such systems are extremely expensive and time consuming to set up. In this engineering brief, we present virtual studio production tools for mixing and monitoring atmos and multichannel sound with personalized head-related transfer functions (HRTFs). This paper talks in detail how the acoustics of the studio, including speaker, and headphone responses are captured accurately for a truly immersive experience. The acoustic fingerprint of the studio is then integrated with the personalized HRTFs predicted using machine learning algorithms that use an ear image as an input. Such novel tools will bring the power of

personalized spatial audio and dolby atmos production in hands of millions of at-home mixing engineers and producers.

Sunder, Kaushik; Jain, Sunder
Embody, San Mateo, CA, USA

Binaural Audio
eBrief:674
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21735

Applause detection filter design for remote live-viewing with adaptive modeling filter

The COVID-19 pandemic prevents us from enjoying live performances. On the other hand, commercial audio-visual transmission systems, such as live viewing systems, have become more popular and have been increasing. The APRICOT: (APplause for Realistic Immersive Contents Transmission) system was developed and used in some trials to enhance the reality for live viewing. This paper describes an applause sound extraction method for automation of applause sound transmission and a simulation experiment using the sound source recorded live at the venue to assess the applause sound extraction performance. We used an adaptive filter to model the room transfer function. In addition, we designed the inverse filter to emphasize applause sounds and extracted them. The experimental evaluation showed that the system extracted the applause sounds almost correctly under various conditions from the performance sound source.

Kawahara, Kazuhiko; Karakawa, Masahiro; Omoto, Akira; Kamamoto, Yutaka
Kyushu University, Japan; Nippon Telegraph and Telephone Corp, Japan

Audio Synthesis & Audio Effects
eBrief:675
Available in the AES E-Library here: https://www.aes.org/e-lib/browse.cfm?elib=21736
