

2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2023)

**New Paltz, New York, USA
22-25 October 2023**



**IEEE Catalog Number: CFP23AUD-POD
ISBN: 979-8-3503-2373-3**

**Copyright © 2023 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP23AUD-POD
ISBN (Print-On-Demand):	979-8-3503-2373-3
ISBN (Online):	979-8-3503-2372-6
ISSN:	1931-1168

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

TABLE OF CONTENTS

SLMGAN: Exploiting Speech Language Model Representations for Unsupervised Zero-Shot Voice Conversion in GANs	1
<i>Yinghao Aaron Li, Cong Han, Nima Mesgarani</i>	
Yet Another Generative Model for Room Impulse Response Estimation	6
<i>Sungho Lee, Hyeong-Seok Choi, Kyogu Lee</i>	
An Improved Metric of Informational Masking for Perceptual Audio Quality Measurement	11
<i>Pablo M. Delgado, Jürgen Herre</i>	
Multichannel Subband-Fullband Gated Convolutional Recurrent Neural Network for Direction-Based Speech Enhancement with Head-Mounted Microphone Arrays	16
<i>Benjamin Stahl, Alois Sontacchi</i>	
A Novel Method to Detect Instrumental Music in a Large Scale Music Catalog	21
<i>Wo Jae Lee, Emanuele Coviello</i>	
Miipher: A Robust Speech Restoration Model Integrating Self-Supervised Speech and Text Representations	26
<i>Yuma Koizumi, Heiga Zen, Shigeki Karita, Yifan Ding, Kohei Yatabe, Nobuyuki Morioka, Yu Zhang, Wei Han, Ankur Bapna, Michiel Bacchiani</i>	
Neural Audio Decorrelation Using Generative Adversarial Networks	31
<i>Carlotta Anemüller, Oliver Thiergart, Emanuel A. P. Habets</i>	
Region-Of-Interest Oriented Constant-Beamwidth Beamforming with Rectangular Arrays	36
<i>Gal Itzhak, Israel Cohen</i>	
Deep Adaptation Control for Stereophonic Acoustic Echo Cancellation	41
<i>Amir Ivry, Israel Cohen, Baruch Berdugo</i>	
Histogram Layer Time Delay Neural Networks for Passive Sonar Classification	46
<i>Jarin Ritu, Ethan Barnes, Riley Martell, Alexandra Van Dine, Joshua Peeples</i>	
Music De-Limiter Networks Via Sample-Wise Gain Inversion	51
<i>Chang-Bin Jeon, Kyogu Lee</i>	
Convolutional Block-Matching Segmentation Algorithm with Application to Music Structure Analysis	56
<i>Axel Marmoret, Jérémy E. Cohen, Frédéric Bimbot</i>	
Efficient Deep Acoustic Echo Suppression with Condition-Aware Training	61
<i>Ernst Seidel, Pejman Mowlae, Tim Fingscheidt</i>	
Extending Audio Masked Autoencoders Toward Audio Restoration	66
<i>Zhi Zhong, Hao Shi, Masato Hirano, Kazuki Shimada, Kazuya Tateishi, Takashi Shibuya, Shusuke Takahashi, Yuki Mitsufuji</i>	
Hybrid Noise Shaping for Audio Coding Using Perfectly Overlapped Window	71
<i>Byeongho Jo, Seungkwon Beack</i>	

Representation Learning for Audio Privacy Preservation Using Source Separation and Robust Adversarial Learning	76
<i>Diep Luong, Minh Tran, Shayan Gharib, Konstantinos Drossos, Tuomas Virtanen</i>	
Neural Networks for Interference Reduction in Multi-Track Recordings	81
<i>Rajesh R, Padmanabhan Rajan</i>	
Directional Target Speaker Extraction Under Noisy Underdetermined Conditions Through Conditional Variational Autoencoder with Global Style Tokens	86
<i>Rui Wang, Tomoki Toda</i>	
Single-Channel Speaker Distance Estimation in Reverberant Environments.....	91
<i>Michael Neri, Archontis Politis, Daniel Krause, Marco Carli, Tuomas Virtanen</i>	
Signal Reconstruction from Mel-Spectrogram Based on Bi-Level Consistency of Full-Band Magnitude and Phase.....	96
<i>Yoshiki Masuyama, Natsuki Ueno, Nobutaka Ono</i>	
Blind Room Acoustic Parameters Estimation Using Mobile Audio Transformer	101
<i>Shivam Saini, Jürgen Peissig</i>	
Diff-Pitcher: Diffusion-Based Singing Voice Pitch Correction	106
<i>Jiarui Hai, Mounya Elhilali</i>	
Perceptual Musical Similarity Metric Learning with Graph Neural Networks	111
<i>Cyrus Vahidi, Shubhr Singh, Emmanouil Benetos, Huy Phan, Dan Stowell, György Fazekas, Mathieu Lagrange</i>	
Design of Frequency-Invariant Beamformers with Sparse Concentric Circular Arrays.....	116
<i>Yaakov Buchris, Israel Cohen, Alon Amar</i>	
Single Channel Speech Presence Probability Estimation Based on Hybrid Global-Local Information.....	121
<i>Shuai Tao, Yang Xiang, Himavanth Reddy, Jesper Rindom Jensen, Mads Græsbøll Christensen</i>	
Lace: A Light-Weight, Causal Model for Enhancing Coded Speech Through Adaptive Convolutions	126
<i>Jan Bütthe, Jean-Marc Valin, Ahmed Mustafa</i>	
Consolidating Compression and Revisiting Expansion: An Alternative Amplification Rule for Wide Dynamic Range Compression	131
<i>Alice Sokolova, Baris Aksanli, Fredric Harris, Harinath Garudadri</i>	
Low Bit Rate Binaural Link for Improved Ultra Low-Latency Low-Complexity Multichannel Speech Enhancement in Hearing Aids.....	136
<i>Nils L. Westhausen, Bernd T. Meyer</i>	
Correlation Based Glimpse Proportion Index.....	141
<i>Ahmed Alghamdi, Leonard Moen, Wai-Yip Chan, Daniel Fogerty, Jesper Jensen</i>	
Time-Domain Audio Source Separation Based on Gaussian Processes with Deep Kernel Learning.....	146
<i>Aditya Arie Nugraha, Diego Di Carlo, Yoshiaki Bando, Mathieu Fontaine, Kazuyoshi Yoshii</i>	
Perceptual Quality Enhancement of Sound Field Synthesis Based on Combination of Pressure and Amplitude Matching.....	151
<i>Keisuke Kimura, Shoichi Koyama, Hiroshi Saruwatari</i>	

Kernel Interpolation of Incident Sound Field in Region Including Scattering Objects	156
<i>Shoichi Koyama, Masaki Nakada, Juliano G. C. Ribeiro, Hiroshi Saruwatari</i>	
Multi-Source Direction-of-Arrival Estimation Using Group-Sparse Fitting of Steered Response Power Maps	161
<i>Elisa Tengan, Thomas Dietzen, Filip Elvander, Toon Van Waterschoot</i>	
The Effect of Spoken Language on Speech Enhancement Using Self-Supervised Speech Representation Loss Functions	166
<i>George Close, Thomas Hain, Stefan Goetze</i>	
Differentiable Representation of Warping Based on Lie Group Theory.....	171
<i>Atsushi Miyashita, Tomoki Toda</i>	
Leveraging Synthetic Data for Improving Chamber Ensemble Separation.....	176
<i>Saurjya Sarkar, Louise Thorpe, Emmanouil Benetos, Mark Sandler</i>	
CLIPSONIC: Text-To-Audio Synthesis with Unlabeled Videos and Pretrained Language-Vision Models	181
<i>Hao-Wen Dong, Xiaoyu Liu, Jordi Pons, Gautam Bhattacharya, Santiago Pascual, Joan Serrà, Taylor Berg-Kirkpatrick, Julian McAuley</i>	
Analysis of XLS-R for Speech Quality Assessment.....	186
<i>Bastiaan Tamm, Rik Vandenbergh, Hugo Van Hamme</i>	
Audio Inputs for Active Speaker Detection and Localization Via Microphone Array	191
<i>Davide Berghi, Philip J. B. Jackson</i>	
Temporal Noise Shaping on MDCT Subband Signals for Transform Audio Coding	196
<i>Richard Füg, Bernd Edler</i>	
Adaptive Sparse Linear Prediction in Fixed-Filter ANC Headphone Applications for Multi-Speaker Speech Reduction	201
<i>Yurii Iotov, Sidsel Marie Nørholm, Valiantsin Belyi, Mads Græsbøll Christensen</i>	
Distribution of Modal Damping in Absorptive Shoebox Rooms.....	206
<i>Maximilian Schäfer, Karolina Prawda, Rudolf Rabenstein, Sebastian J. Schlecht</i>	
Predicting Thresholds in an Auditory Overshoot Paradigm Using a Computational Subcortical Model with Efferent Feedback	211
<i>Afagh Farhadi, Laurel H. Carney</i>	
Exploring the Integration of Speech Separation and Recognition with Self-Supervised Learning Representation	216
<i>Yoshiki Masuyama, Xuankai Chang, Wangyou Zhang, Samuele Cornell, Zhong-Qiu Wang, Nobutaka Ono, Yanmin Qian, Shinji Watanabe</i>	
Wide-Area 6DOF Rendering of Multi-Point Ambisonic Recordings Based on Interpolation of Spatial Parameters	221
<i>Archontis Politis, Lauros Pajunen, Jussi Leppänen, Sujeet Mate, Antti Eronen</i>	
Compressing Audio CNNs with Graph Centrality Based Filter Pruning	226
<i>James A King, Arshdeep Singh, Mark D. Plumbley</i>	
All-In-One Metrical and Functional Structure Analysis with Neighborhood Attentions on Demixed Audio.....	231
<i>Taejun Kim, Juhan Nam</i>	

Relative Transfer Function Vector Estimation for Acoustic Sensor Networks Exploiting Covariance Matrix Structure	236
<i>Wiebke Middelberg, Henri Gode, Simon Doclo</i>	
Optimizing Higher-Order Directional Audio Coding with Adaptive Mixing and Energy Matching for Ambisonic Compression and Upmixing	241
<i>Christoph Hold, Leo McCormack, Archontis Politis, Ville Pulkki</i>	
Array Configuration Mismatch in Deep DOA Estimation: Towards Robust Training	246
<i>Ayal Schwartz, Elior Hadad, Sharon Gannot, Shlomo E. Chazan</i>	
An Objective Evaluation of Hearing AIDS and DNN-Based Binaural Speech Enhancement in Complex Acoustic Scenes	251
<i>Enric Gusó, Joanna Luberadzka, Martí Baig, Umut Sayin, Xavier Serra</i>	
Slim-Tasnet: A Slimmable Neural Network for Speech Separation	256
<i>Mohamed Elminshawi, Srikanth Raj Chetupalli, Emanuël A. P. Habets</i>	
Computing Acoustic Onsets Via an Eikonal Solver	261
<i>Samuel F. Potter, Monte Hoover, Dmitry Zotkin, Ramani Duraiswami</i>	
A Differentiable Image Source Model for Room Acoustics Optimization	266
<i>Bowen Zhi, Alisha Sharma, Dmitry N. Zotkin, Ramani Duraiswami</i>	
Hyperbolic Unsupervised Anomalous Sound Detection.....	271
<i>François G. Germain, Gordon Wichern, Jonathan Le Roux</i>	
A Differentiable Acoustic Guitar Model for String-Specific Polyphonic Synthesis.....	276
<i>Andrew Wiggins, Youngmoo Kim</i>	
Quaternion Anti-Transfer Learning for Speech Emotion Recognition	281
<i>Eric Guizzo, Tillman Weyde, Giacomo Tarroni, Danilo Comminiello</i>	
SEFGAN: Harvesting the Power of Normalizing Flows and GANs for Efficient High-Quality Speech Enhancement.....	286
<i>Martin Strauss, Nicola Pia, Nagashree K. S. Rao, Bernd Edler</i>	
Towards on-Device Keyword Spotting Using Low-Footprint Quaternion Neural Models	291
<i>Aryan Chaudhary, Vinayak Abrol</i>	
Pretraining Respiratory Sound Representations Using Metadata and Contrastive Learning.....	296
<i>Ilyass Moummad, Nicolas Farrugia</i>	
A High-Rate Extension to Soundstream	301
<i>Hong-Goo Kang, Jan Skoglund, W. Bastiaan Kleijn, Andrew Storos, Hengchin Yeh</i>	
Estimating the Direction of Arrival of a Spoken Wake Word Using a Single Sensor on an Elastic Panel.....	306
<i>Tre Dipassio, Michael C. Heilemann, Benjamin Thompson, Mark F. Bocko</i>	
General Purpose Audio Effect Removal.....	311
<i>Matthew Rice, Christian J. Steinmetz, George Fazekas, Joshua D. Reiss</i>	
Location as Supervision for Weakly Supervised Multi-Channel Source Separation of Machine Sounds	316
<i>Ricardo Falcon-Perez, Gordon Wichern, François G. Germain, Jonathan Le Roux</i>	

Covariance Blocking and Whitening Method for Successive Relative Transfer Function Vector Estimation in Multi-Speaker Scenarios	321
<i>Henri Gode, Simon Doclo</i>	
Bridging High-Quality Audio and Video Via Language for Sound Effects Retrieval from Visual Queries	326
<i>Julia Wilkins, Justin Salamon, Magdalena Fuentes, Juan Pablo Bello, Oriol Nieto</i>	
Low-Complexity Higher Order Scattering Delay Networks	331
<i>Leny Vinesclas, Matteo Scerbo, Hüseyin Hacıhabiboglu, Zoran Cvetkovic, Enzo De Sena</i>	
Annotating Jazz Recordings Using Lead Sheet Alignment with Deep Chroma Features.....	336
<i>Ivan Shanin, Simon Dixon</i>	
Robust Audio Anti-Spoofing System Based on Low-Frequency Sub-Band Information.....	341
<i>Menglu Li, Xiao-Ping Zhang</i>	
Flexible Multichannel Speech Enhancement for Noise-Robust Frontend.....	346
<i>Ante Jukic, Jagadeesh Balam, Boris Ginsburg</i>	
Sound Source Distance Estimation in Diverse and Dynamic Acoustic Conditions.....	351
<i>Saksham Singh Kushwaha, Iran R. Roman, Magdalena Fuentes, Juan Pablo Bello</i>	
Mixed-Delay Distributed Beamforming for Own-Speech Separation in Hearing Devices with Wireless Remote Microphones.....	356
<i>Ryan M. Corey</i>	
Inverted Cardioid Topology for Multi-Radius Spherical Microphone Arrays	361
<i>Mark R. P. Thomas, Jan-Hendrik Hanschke</i>	
Complete and Separate: Conditional Separation with Missing Target Source Attribute Completion.....	366
<i>Dimitrios Bralios, Efthymios Tzinis, Paris Smaragdis</i>	
Mitigating Cross-Database Differences for Learning Unified HRTF Representation	371
<i>Yutong Wen, You Zhang, Zhiyao Duan</i>	
Learning Sub-Dimensional HRTF Representations Towards Individualization Applications - Traditional and Deep Learning Approaches	376
<i>Devansh Zurale, Shlomo Dubnov</i>	
Masked Frequency Modeling for Improving Packet Loss Concealment in Speech Transmission Systems.....	381
<i>Da-Hee Yang, Donghyun Kim, Joon-Hyuk Chang</i>	
Automatic Detection of Poor Tone Quality in Classical Guitar Playing Using Deep Anomaly Detection Method	386
<i>Kenta Ogawa, Shun Sawada, Kouichi Katsurada, Hidehumi Ohmura</i>	
Class Activation Mapping-Driven Data Augmentation: Masking Significant Regions for Enhanced Acoustic Scene Classification	391
<i>Pil Moo Byun, Jeong-Hwan Choi, Joon-Hyuk Chang</i>	
Fitting Auditory Filterbanks with Multiresolution Neural Networks	396
<i>Vincent Lostanlen, Daniel Haider, Han Han, Mathieu Lagrange, Peter Balazs, Martin Ehler</i>	

AECSQI: Referenceless Acoustic Echo Cancellation Measures Using Speech Quality and Intelligibility Improvement	401
<i>Jin Woo Lee, Hyeong-Seok Choi, Kyogu Lee</i>	

Diffusion Posterior Sampling for Informed Single-Channel Dereverberation	406
<i>Jean-Marie Lemerrier, Simon Welker, Timo Gerkmann</i>	

Author Index