

2025 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2025)

**Tahoe City, California, USA
12-15 October 2025**



**IEEE Catalog Number: CFP25AUD-POD
ISBN: 979-8-3315-3746-3**

**Copyright © 2025 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:	CFP25AUD-POD
ISBN (Print-On-Demand):	979-8-3315-3746-3
ISBN (Online):	979-8-3315-3745-6
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

Kernel Ridge Regression Based Sound Field Estimation Using a Rigid Spherical Microphone Array	1
<i>Ryo Matsuda, Juliano G. C. Ribeiro, Hitoshi Akiyama, Jorge Trevino</i>	
Optimal Region-Of-Interest Beamforming for Audio Conferencing with Dual Perpendicular Sparse Circular Sectors	6
<i>Gal Itzhak, Simon Doclo, Israel Cohen</i>	
Physics-Informed Direction-Aware Neural Acoustic Fields.....	11
<i>Yoshiki Masuyama, François G. Germain, Gordon Wichern, Christopher Ick, Jonathan Le Roux</i>	
Source and Sensor Placement for Sound Field Control Based on Mean Square Error with Prior Spatial Covariance.....	16
<i>Shihori Kozuka, Shoichi Koyama, Hiroaki Itou, Noriyoshi Kamado</i>	
Towards Perception-Informed Latent HRTF Representations.....	21
<i>You Zhang, Andrew Franci, Ruohan Gao, Paul Calamia, Zhiyao Duan, Ishwarya Ananthabhotla</i>	
Room Impulse Response Generation Conditioned on Acoustic Parameters	26
<i>Silvia Arellano, Chunghsin Yeh, Gautam Bhattacharya, Daniel Arteaga</i>	
Online Incremental Learning for Audio Classification Using a Pretrained Audio Model	31
<i>Manjunath Mulimani, Annamaria Mesaros</i>	
Sound Event Detection with Audio-Text Models and Heterogeneous Temporal Annotations	36
<i>Manu Harju, Annamaria Mesaros</i>	
Latent Acoustic Mapping for Direction of Arrival Estimation: A Self-Supervised Approach.....	41
<i>Adrian S. Roman, Iran R. Roman, Juan P. Bello</i>	
Distributed Asynchronous Device Speech Enhancement Via Windowed Cross-Attention	46
<i>Gene-Ping Yang, Sebastian Braun</i>	
The Test of Auditory-Vocal Affect (TAVA) Dataset	51
<i>Juan S. Gómez-Cañón, Camille Noufi, Jonathan Berger, Karen J. Parker, Daniel L. Bowling</i>	
Unsupervised Multi-Channel Speech Dereverberation Via Diffusion.....	56
<i>Yulun Wu, Zhongweiyang Xu, Jianchong Chen, Zhong-Qiu Wang, Romit Roy Choudhury</i>	
Microphone Occlusion Mitigation for Own-Voice Enhancement in Head-Worn Microphone Arrays Using Switching-Adaptive Beamforming	61
<i>Wiebke Middelberg, Jung-Suk Lee, Saeed Bagheri Sereshki, Ali Aroudi, Vladimir Tourbabin, Daniel D. E. Wong</i>	
Multi-Utterance Speech Separation and Association Trained on Short Segments	66
<i>Yuzhu Wang, Archontis Politis, Konstantinos Drossos, Tuomas Virtanen</i>	
Task-Specific Audio Coding for Machines: Machine-Learned Latent Features Are Codes for that Machine.....	71
<i>Anastasia Kuznetsova, Inseon Jang, Wootae Lim, Minje Kim</i>	

VoxATtack: A Multimodal Attack on Voice Anonymization Systems	76
<i>Ahmad Aloradi, Ünal Ege Gaznepoglu, Emanuël A. P. Habets, Daniel Tenbrinck</i>	
Learning Perceptually Relevant Temporal Envelope Morphing.....	81
<i>Satvik Dixit, Sungjoon Park, Chris Donahue, Laurie M. Heller</i>	
Multi-Class-Token Transformer for Multitask Self-Supervised Music Information Retrieval.....	86
<i>Yuexuan Kong, Vincent Lostanlen, Romain Hennequin, Mathieu Lagrange, Gabriel Meseguer-Brocal</i>	
Conditioned Wave-U-Net for Acoustic Matching of Speech in Shared XR Environments.....	91
<i>Joanna Luberadзка, Enric Gusó, Umut Sayin</i>	
Fast Text-To-Audio Generation with Adversarial Post-Training	96
<i>Zachary Novack, Zach Evans, Zack Zukowski, Josiah Taylor, Cj Carr, Julian Parker, Adnan Al-Sinan, Gian Marco Iodice, Julian McAuley, Taylor Berg-Kirkpatrick, Jordi Pons</i>	
Post-Training Quantization for Audio Diffusion Transformers	101
<i>Tanmay Khandelwal, Magdalena Fuentes</i>	
Diffused Responsibility: Analyzing the Energy Consumption of Generative Text-To-Audio Diffusion Models.....	106
<i>Riccardo Passoni, Francesca Ronchini, Luca Comanducci, Romain Serizel, Fabio Antonacci</i>	
Benchmarking Sub-Genre Classification for Mainstage Dance Music	111
<i>Hongzhi Shu, Xinglin Li, Hongyu Jiang, Minghao Fu, Xinyu Li</i>	
RUMAA: Repeat-Aware Unified Music Audio Analysis for Score-Performance Alignment, Transcription, and Mistake Detection.....	116
<i>Sungkyun Chang, Simon Dixon, Emmanouil Benetos</i>	
FlexSED: Towards Open-Vocabulary Sound Event Detection.....	121
<i>Jiarui Hai, Helin Wang, Weizhe Guo, Mounya Elhilali</i>	
TACOS: Temporally-Aligned Audio CaptiOnS for Language-Audio Pretraining	126
<i>Paul Primus, Florian Schmid, Gerhard Widmer</i>	
Bridging Ears and Eyes: Analyzing Audio and Visual Large Language Models to Humans in Visible Sound Recognition and Reducing Their Sensory Gap Via Cross-Modal Distillation	131
<i>Xilin Jiang, Junkai Wu, Vishal Choudhari, Nima Mesgarani</i>	
OpenBEATs: A Fully Open-Source General-Purpose Audio Encoder.....	136
<i>Shikhar Bharadwaj, Samuele Cornell, Kwanghee Choi, Satoru Fukayama, Hye-Jin Shim, Soham Deshmukh, Shinji Watanabe</i>	
Modulation Discovery with Differentiable Digital Signal Processing	141
<i>Christopher Mitcheltree, Hao Hao Tan, Joshua D. Reiss</i>	
Transcribing Rhythmic Patterns of the Guitar Track in Polyphonic Music.....	146
<i>Aleksandr Lukoianov, Anssi Klapuri</i>	
Rethinking Non-Negative Matrix Factorization with Implicit Neural Representations	151
<i>Krishna Subramani, Paris Smaragdis, Takuya Higuchi, Mehrez Souden</i>	
Self-Steering Deep Non-Linear Spatially Selective Filters for Efficient Extraction of Moving Speakers Under Weak Guidance.....	156
<i>Jakob Kienegger, Alina Mannanova, Huajian Fang, Timo Gerkmann</i>	

A Lightweight and Robust Method for Blind Wideband-To-Fullband Extension of Speech	161
<i>Jan B��the, Jean-Marc Valin</i>	
Listening Intention Estimation for Hearables with Natural Behavior Cues	166
<i>Pascal Baumann, Seraina Glaus, Ludovic Amruthalingam, Fabian Gr��ger, Ruksana Giurda, Simone Lionetti</i>	
Unveiling the Best Practices for Applying Speech Foundation Models to Speech Intelligibility Prediction for Hearing-Impaired People.....	171
<i>Haoshuai Zhou, Boxuan Cao, Changgeng Mo, Linkai Li, Shan Xiang Wang</i>	
Can Large Language Models Predict Audio Effects Parameters from Natural Language?	176
<i>Seunghoon Doh, Junghyun Koo, Marco A. Mart��nez-Ram��rez, Wei-Hsiang Liao, Juhan Nam, Yuki Mitsu��fuji</i>	
Temporal Adaptation of Pre-Trained Foundation Models for Music Structure Analysis	181
<i>Yixiao Zhang, Haonan Chen, Ju-Chiang Wang, Jitong Chen</i>	
Learn from Virtual Guitar: A Comparative Analysis of Automatic Guitar Transcription Using Synthetic and Real Audio	186
<i>Yuta Kusaka, Akira Maezawa</i>	
SpecMaskFoley: Steering Pretrained Spectral Masked Generative Transformer Toward Synchronized Video-To-Audio Synthesis Via ControlNet	191
<i>Zhi Zhong, Akira Takahashi, Shuyang Cui, Keisuke Toyama, Shusuke Takahashi, Yuki Mitsu��fuji</i>	
Neural-Network Based Interpolation of Late Reverberation in Coupled Spaces Using the Common Slopes Model.....	196
<i>Orchisama Das, Gloria Dal Santo, Sebastian J. Schlecht, Zoran Cvetkovic</i>	
Two Views, One Truth: Spectral and Self-Supervised Features Fusion for Robust Speech Deepfake Detection	201
<i>Yassine El Kheir, Arnab Das, Enes Erdem Erdogan, Fabian Ritter-Gutierrez, Tim Polzehl, Sebastian M��ller</i>	
Schr��dinger Bridge Consistency Trajectory Models for Speech Enhancement	206
<i>Shuichiro Nishigori, Koichi Saito, Naoki Murata, Masato Hirano, Shusuke Takahashi, Yuki Mitsu��fuji</i>	
A Unified Framework for Evaluating DNN-Based Feedforward, Feedback, and Hybrid Active Noise Cancellation.....	211
<i>Huajian Fang, Buye Xu, Jacob Donley, Ashutosh Pandey, Deliang Wang, Daniel Wong</i>	
Robust Speech Activity Detection in the Presence of Singing Voice	216
<i>Philipp Grundhuber, Mhd Modar Halimeh, Martin Strau���, Emanu��l A. P. Habets</i>	
Frequency-Domain Signal-To-Noise Ratios Illuminate the Effects of the Spectral Consistency Constraint and Griffin-Lim Algorithms	221
<i>Stephen Voran, Jaden Pieper</i>	
Source Separation by Flow Matching.....	226
<i>Robin Scheibler, John R. Hershey, Arnaud Doucet, Henry Li</i>	
Musical Source Separation Bake-Off: Comparing Objective Metrics with Human Perception.....	231
<i>Noah Jaffe, John Ashley Burgoyne</i>	

TGIF: Talker Group-Informed Familiarization of Target Speaker Extraction	236
<i>Tsun-An Hsieh, Minje Kim</i>	
SynSonic: Augmenting Sound Event Detection Through Text-To-Audio Diffusion ControlNet and Effective Sample Filtering.....	241
<i>Jiarui Hai, Mounya Elhilali</i>	
Physically Informed Spatial Regularization for Sound Event Localization and Detection	246
<i>Haocheng Liu, Diego Di Carlo, Aditya Arie Nugraha, Kazuyoshi Yoshii, Gaël Richard, Mathieu Fontaine</i>	
Stereo Reproduction in the Presence of Sample Rate Offsets	251
<i>Srikanth Korse, Andreas Walther, Emanuel A. P. Habets</i>	
Physics-Informed Transfer Learning for Data-Driven Sound Source Reconstruction in Near-Field Acoustic Holography	256
<i>Xinmeng Luan, Mirco Pezzoli, Fabio Antonacci, Augusto Sarti</i>	
Scene-Wide Acoustic Parameter Estimation	261
<i>Ricardo Falcon-Perez, Ruohan Gao, Gregor Mueckl, Sebastia V. Amengual Gari, Ishwarya Ananthabhotla</i>	
Cyclic Multichannel Wiener Filter for Acoustic Beamforming.....	266
<i>Giovanni Bologni, Richard Heusdens, Richard C. Hendriks</i>	
EffDiffSE: Efficient Diffusion-Based Frequency-Domain Speech Enhancement with Hybrid Discriminative and Generative DNNs	271
<i>Yihui Fu, Renzheng Shi, Marvin Sach, Wouter Tirry, Tim Fingscheidt</i>	
Adaptive Slimming for Scalable and Efficient Speech Enhancement	276
<i>Riccardo Miccini, Minje Kim, Clément Laroche, Luca Pezzarossa, Paris Smaragdis</i>	
ReverbMiipher: Generative Speech Restoration Meets Reverberation Characteristics Controllability	281
<i>Wataru Nakata, Yuma Koizumi, Shigeki Karita, Robin Scheibler, Haruko Ishikawa, Adriana Guevara-Rukoz, Heiga Zen, Michiel Bacchiani</i>	
IS3 : Generic Impulsive–Stationary Sound Separation in Acoustic Scenes Using Deep Filtering.....	286
<i>Clémentine Berger, Paraskevas Stamatiadis, Roland Badeau, Slim Essid</i>	
Hybrid-Sep: Language-Queried Audio Source Separation Via Pre-Trained Model Fusion and Adversarial Consistent Training	291
<i>Jianyuan Feng, Guangzheng Li, Yangfei Xu</i>	
UBGAN: Enhancing Coded Speech with Blind and Guided Bandwidth Extension	296
<i>Kishan Gupta, Srikanth Korse, Andreas Brendel, Nicola Pia, Guillaume Fuchs</i>	
Contrastive Representation Learning for Privacy-Preserving Fine-Tuning of Audio-Visual Speech Recognition	301
<i>Luca Becker, Rainer Martin</i>	
Low-Complexity Individualized Noise Reduction for Real-Time Processing	306
<i>Chuan Wen, Sarah Verhulst</i>	
JSQA: Speech Quality Assessment with Perceptually-Inspired Contrastive Pretraining Based on JND Audio Pairs	311
<i>Junyi Fan, Donald Williamson</i>	

Is MixIT Really Unsuitable for Correlated Sources? Exploring MixIT for Unsupervised Pre-Training in Music Source Separation	316
<i>Kohei Saijo, Yoshiaki Bando</i>	
Beyond Architecture: The Critical Impact of Inference Overlap on Music Source Separation Benchmarks	321
<i>Harnick Khera, Johan Pauwels, Alan W. Archer-Boyd, Mark B. Sandler</i>	
Towards Reliable Objective Evaluation Metrics for Generative Singing Voice Separation Models	326
<i>Paul A. Bereuter, Benjamin Stahl, Mark D. Plumbley, Alois Sontacchi</i>	
Generating Separated Singing Vocals Using a Diffusion Model Conditioned on Music Mixtures	331
<i>Genís Plaja-Roglans, Yun-Ning Hung, Xavier Serra, Igor Pereira</i>	
Balancing Information Preservation and Disentanglement in Self-Supervised Music Representation Learning	336
<i>Julia Wilkins, Sivan Ding, Magdalena Fuentes, Juan Pablo Bello</i>	
SILA: Signal-To-Language Augmentation for Enhanced Control in Text-To-Audio Generation	341
<i>Sonal Kumar, Prem Seetharaman, Justin Salamon, Dinesh Manocha, Oriol Nieto</i>	
Perceptually-Driven Panning for an Extended Listening Area.....	346
<i>Pedro Lladó, Annika Neidhardt, Antoine R. Souchaud, Zoran Cvetkovic, Enzo De Sena</i>	
Theoretical Analysis of Recursive Implementations of Multi-Channel Cross-Talk Cancellation Systems.....	351
<i>Filippo Maria Fazi, Francesco Veronesi, Marcos F. Simón-Gálvez, Andreas Franck</i>	
SFC-L1: Sound Field Control with Least Absolute Deviation Regression	356
<i>Takuma Okamoto</i>	
Beamforming with Interaural Time-To-Level Difference Conversion for Hearing Loss Compensation.....	360
<i>Johannes W. De Vries, Timm-Jonas Bäumer, Stephan Töpken, Richard Heusdens, Steven Van De Par, Richard C. Hendriks</i>	
Robust One-Step Speech Enhancement Via Consistency Distillation	365
<i>Liang Xu, Longfei Felix Yan, W. Bastiaan Kleijn</i>	
Miipher-2: A Universal Speech Restoration Model for Million-Hour Scale Data Restoration	370
<i>Shigeki Karita, Yuma Koizumi, Heiga Zen, Haruko Ishikawa, Robin Scheibler, Michiel Bacchiani</i>	
Context-Aware Query Refinement for Target Sound Extraction: Handling Partially Matched Queries	375
<i>Ryo Sato, Chiho Haruta, Nobuhiko Hiruma, Keisuke Imoto</i>	
Combolutional Neural Networks.....	380
<i>Cameron Churchwell, Minje Kim, Paris Smaragdis</i>	
Incremental Averaging Method to Improve Graph-Based Time-Difference-Of-Arrival Estimation	385
<i>Klaus Brümman, Kouei Yamaoka, Nobutaka Ono, Simon Doclo</i>	
Learning Robust Spatial Representations from Binaural Audio Through Feature Distillation.....	390
<i>Holger Severin Bovbjerg, Jan Østergaard, Jesper Jensen, Shinji Watanabe, Zheng-Hua Tan</i>	

Soft-Constrained Spatially Selective Active Noise Control for Open-Fitting Hearables	395
<i>Tong Xiao, Reinhild Roden, Matthias Blau, Simon Doclo</i>	
Low-Rank Adaptation of Deep Prior Neural Networks for Room Impulse Response Reconstruction	400
<i>Mirco Pezzoli, Federico Miotello, Shoichi Koyama, Fabio Antonacci</i>	
MB-RIRs: A Synthetic Room Impulse Response Dataset with Frequency-Dependent Absorption Coefficients	404
<i>Enric Gusó, Joanna Luberadзка, Umut Sayin, Xavier Serra</i>	
Controlling the Parameterized Multi-Channel Wiener Filter Using a Tiny Neural Network	409
<i>Eric Grinstein, Ashutosh Pandey, Cole Li, Shanmukha Srinivas, Juan Azcarreta, Jacob Donley, Sanha Lee, Ali Aroudi, Çağdas Bilen</i>	
Long-Context Modeling Networks for Monaural Speech Enhancement: A Comparative Study	414
<i>Qiquan Zhang, Moran Chen, Zeyang Song, Hexin Liu, Xiangyu Zhang, Haizhou Li</i>	
FasTUSS: Faster Task-Aware Unified Source Separation.....	419
<i>Francesco Paissan, Gordon Wichern, Yoshiki Masuyama, Ryo Aihara, François G. Germain, Kohei Saijo, Jonathan Le Roux</i>	
RADE: A Neural Codec for Transmitting Speech Over HF Radio Channels	424
<i>David Rowe, Jean-Marc Valin</i>	
The Perception of Phase Intercept Distortion and Its Application in Data Augmentation.....	429
<i>Venkatakrishnan Vaidyanathapuram Krishnan, Nathaniel Condit-Schultz</i>	
Device-Centric Room Impulse Response Augmentation Evaluated on Room Geometry Inference	434
<i>Çağdas Tuna, Andreas Walther, Emanuel A. P. Habets</i>	
Modeling Multi-Level Hearing Loss for Speech Intelligibility Prediction.....	439
<i>Xiajie Zhou, Candy Olivia Mawalim, Masashi Unoki</i>	
Moises-Light: Resource-Efficient Band-Split U-Net for Music Source Separation.....	444
<i>Yun-Ning Amy Hung, Igor Pereira, Filip Korzeniowski</i>	
Improving Inference-Time Optimisation for Vocal Effects Style Transfer with a Gaussian Prior.....	449
<i>Chin-Yun Yu, Marco A. Martínez-Ramírez, Junghyun Koo, Wei-Hsiang Liao, Yuki Mitsufuji, György Fazekas</i>	
Self-Supervised Representation Learning with a JEPA Framework for Multi-Instrument Music Transcription.....	454
<i>Mary Pilataki, Matthias Mauch, Simon Dixon</i>	
DiTVC: One-Shot Voice Conversion Via Diffusion Transformer with Environment and Speaking Rate Cloning.....	459
<i>Yunyun Wang, Jiaqi Su, Adam Finkelstein, Rithesh Kumar, Ke Chen, Zeyu Jin</i>	
Learning to Upsample and Upmix Audio in the Latent Domain	464
<i>Dimitrios Bralios, Paris Smaragdis, Jonah Casebeer</i>	
Head-Related Transfer Function Individualization Using Anthropometric Features and Spatially Independent Latent Representation	469
<i>Ryan Niu, Shoichi Koyama, Tomohiko Nakamura</i>	

Author Index