

15th ITG-Conference on Speech Communication

ITG-Fachbericht 312

Aachen, Germany
20-22 September 2023

ISBN: 978-1-7138-8611-2

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2023) by VDE VERLAG GMBH
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

For permission requests, please contact VDE VERLAG GMBH
at the address below.

VDE VERLAG GMBH
Bismarckstr. 33
P.O.B. 12 01 43
10625 Berlin, Germany

Phone: +49 30 34 80 01 - 0
Fax: +49 30 34 80 01 - 9088

kundenservice@vde-verlag.de

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Contents

Keynote I:

Leap into the Industrial Metaverse: Where Wearables and Sensors Meet Reality

Dr. Hendrik Witt, *TeamViewer*

Poster Session A

Chair: Magnus Schäfer

01	Ad Hoc Distributed Microphones Clustering: A Comparative Analysis on Using Coherence and Signal-specific Features.	11
	Stijn Kindt, Martijn Meeldijk and Nilesh Madhu, <i>Ghent University, Belgium</i>	
02	Exploiting an External Microphone to Improve Time-difference-of-arrival Estimates for Euclidean Distance Matrix-based Source Localization	16
	Klaus Brümman and Simon Doclo, <i>University of Oldenburg</i>	
03	Hearing Impairment in Crowdsourced Speech Quality Assessments: Its Effect and Screening with Digit Triplet Hearing Test	21
	Benedikt Schuh, Wafaa Wardah, Babak Naderi, Thilo Michal and Sebastian Möller, <i>Quality and Usability Lab, Technische Universität, Berlin, Deutschland</i>	
04	Long-term Conversation Analysis: Exploring Utility and Privacy.	26
	Francesco Nespoli ^{1,2} , Jule Pohlhausen ³ , Patrick A. Naylor ² and Joerg Bitzer ^{3,4} <i>¹Microsoft, London, UK, ²Imperial College London, ³Jade University of Applied Sciences, Oldenburg, ⁴Fraunhofer IDMT Dept. HSA, Oldenburg</i>	
05	Towards a Natural Reproduction of Binaural Recordings: Combining Binaural Cue Adaptation and Adaptive Crosstalk Cancellation.	31
	Tobias Kabzinski, Sebastian Nagel and Peter Jax, <i>RWTH Aachen University</i>	
06	Screening of Alzheimer’s Dementia up to 12 Years ahead from Conversational Speech of ILSE Study	36
	Ayimmisagul Ablimit, Elisa Brauße and Tanja Schultz, <i>University of Bremen</i>	
07	Speaker’s Articulatory Strategy Analysis: Theoretical Framework and Preliminary Experiment	41
	Antoine Serrurier, <i>Uniklinik RWTH Aachen</i>	
08	Speech-based Age and Gender Prediction with Transformers	46
	Felix Burkhardt ¹ , Johannes Wagner ¹ , Hagen Wierstorf ¹ , Florian Eyben ¹ and Björn Schuller ^{1,2,3} <i>¹audEERING GmbH, ²University of Augsburg, ³Imperial College London</i>	
09	Transfer Learning using Musical/Non-musical Mixtures for Multi-instrument Recognition . . .	51
	Hannes Bradl ¹ , Markus Huber ¹ and Franz Pernkopf ³ <i>¹Joanneum Research, Graz, Austria, ²sonible GmbH, Graz, Austria, ³TU Graz, Austria</i>	
10	U-DiT TTS: U-Diffusion Vision Transformer for Text-to-speech.	56
	Xin Jing ¹ , Yi Chang ³ , Zijiang Yang ¹ , Jiangjian Xie ² , Andreas Triantafyllopoulos ¹ and Björn Schuller ^{1,2,3} , <i>¹University of Augsburg, ²Beijing Forestry University China, ³Imperial College London</i>	

11	Using Perceptual Evaluation of Speech Quality (PESQ) Loss for DNN-based Speech Enhancement.....	61
	Lars Thieling, Lars Nippert and Peter Jax, <i>RWTH Aachen University</i>	

Keynote II:

Alicia & Bernhard: A Communication Love-story Through the Ages

Prof. Dr. Nilesh Madhu, *IDLab, Department of Electronics and Information Systems, Ghent University*

Show & Tell

3PASS – Realistic Testing Matters

Ekaterina Malkova, Christian Schüring, *HEAD acoustics GmbH, Herzogenrath*

Exploring Advanced Hearing Trends: Live Demos of Occlusion Cancellation, Low Noise Wide-band Transparency, Wideband ANC and Spatial Audio Immersion

Stefan Liebich, Henning Niendieck, *Elevear GmbH, Aachen*

DemVis: Modular System for Speech-based Dementia Screening

Jordan Behrendt, Ayimnisagul Ablimit, Elisa Brauße, Tanja Schultz, *University of Bremen*

Poster Session B

Chair: Heiner Löllmann

12	Advances In End-to-end Conversational Speech Quality Prediction.....	66
	Stefan Bleiholder ¹ , Christian Kettler ² , Nils Rohrer ¹ and Steffen Weyer ² ¹ <i>HEAD acoustics GmbH, Herzogenrath</i> , ² <i>FH Aachen</i>	
13	Comparative Study of LC3plus and Lyra Codec on DNN-based Source Localisation for Hearing Aids.....	71
	Siyuan Song, Stijn Kindt, Jasper Maes, Alexander Bohlender and Nilesh Madhu <i>Ghent University, Belgium</i>	
14	Comparison of Different Neural Network Architectures for Spoken Language Identification.....	76
	Tala Bazazo ^{1,2} , Mohammad Zeineldeen ¹ , Christian Plahl ² , Ralf Schlüter ¹ and Hermann Ney ¹ ¹ <i>RWTH Aachen University</i> , ² <i>ebay, Aachen</i>	
15	Exploring Shapely Values for Blood Glucose Level Prediction from Speech.....	81
	Simone Pompe ¹ , Adria MalloI-Ragolta ¹ , Nicolas Schauer ² and Björn Schuller ^{1,3} ¹ <i>University of Augsburg</i> , ² <i>Vindex GmbH</i> , ³ <i>Imperial College London</i>	
16	LibriWASN: A Data Set for Meeting Separation, Diarization, and Recognition with Asynchronous Recording Devices.....	86
	Joerg Schmalenstroeer, Tobias Gburrek and Reinhold Haeb-Umbach, <i>Paderborn University</i>	
17	Reduced-complexity Binaural Source Localization for Headphones and Hearing Aids using Low-rank DRTF Approximations.....	91
	Jonas Förster ¹ , Helena Janning ² , Sebastian Nagel ² and Peter Jax ² ¹ <i>beyerdynamic GmbH & Co. KG</i> , ² <i>RWTH Aachen University</i>	

18	Single Channel Source Separation in the Wild - Conversational Speech in Realistic Environments	96
	Emil Berger, Barbara Schuppler, Martin Hagsmüller and Franz Pernkopf, <i>TU Graz, Austria</i>	
19	Subjective Performance Evaluation of Single-channel Speaker-conditioned Target Speaker Extraction Algorithms for Complex Acoustic Scenes	101
	Ragini Sinha ¹ , Ann-Christin Scherer ¹ , Simon Doclo ^{1,2} , Christian Rollwage ¹ and Jan Rannies ¹ ¹ Fraunhofer IDMT Dept. HSA, Oldenburg, ² University of Oldenburg	
20	Toward Semi-supervised Transcription of NAKO+ILSE: Influence of Automatic Speech Recognition Performance on Manual Transcription Effort	106
	Elisa Brauße, Kevin Scheck and Tanja Schultz, <i>University of Bremen</i>	
21	Towards a Brain Computer Interface for Speech Perception	111
	Harald Höge, <i>University of the Bundeswehr Munich</i>	
22	Wind Noise Reduction with a Diffusion-based Stochastic Regeneration Model	116
	Jean-Marie Lemerancier ¹ , Joachim Thiemann ² , Raphael Koning ² and Timo Gerkmann ¹ ¹ Universität Hamburg, ² Advanced Bionics, Hannover	

Lecture Session: Best Paper Award Contest

Chair: Walter Kellermann

23	Investigating Speaker Embedding Disentanglement on Natural Read Speech	121
	Michael Kuhlmann ¹ , Adrian Meise ¹ , Fritz Seebauer ² , Petra Wagner ² and Reinhold Haeb-Umbach ¹ , ¹ Paderborn University, ² Bielefeld University	
24	BRUDEX Database: Binaural Room Impulse Responses with Uniformly Distributed External Microphones	126
	Daniel Fejgin, Wiebke Middelberg and Simon Doclo, <i>University of Oldenburg</i>	
25	Comparative Analysis of the wav2vec 2.0 Feature Extractor	131
	Peter Vieting ¹ , Ralf Schlüter ^{1,2} and Hermann Ney ^{1,2} ¹ RWTH Aachen University, ² AppTek GmbH, Aachen	
26	Design of Low-order IIR Filters Based on Hankel Nuclear Norm Regularization for Achieving Acoustic Transparency	136
	Florian Hilgemann, Christoph Weyer and Peter Jax, <i>RWTH Aachen University</i>	
27	Fast Tracking of Time-variant Systems Using Local Affine Subspaces	141
	Till Hardenbicker and Peter Jax, <i>RWTH Aachen University</i>	
28	Generalized Wiener Filter for Nonlinear Acoustic Echo Control	146
	Svantje Voit and Gerald Enzner, <i>University of Oldenburg</i>	

Poster Session C

Chair: Ralf Schlüter

29	Compression of End-to-end Non-autoregressive Image-to-speech System for Low-resourced Devices	151
	Gokul Srinivasagan, Michael Deisher and Munir Georges ¹ Saarland University, ² Intel Corporation, Hillsboro, USA, ³ Intel Labs, Munich, ⁴ TU Ingolstadt	

30	CRNN-based Multi-DOA Estimator: Comparing Classification and Regression	156
	Pieter Cooreman, Alexander Bohlender and Nilesh Madhu, <i>Ghent University, Belgium</i>	
31	Development of Hybrid ASR Systems for Low Resource Medical Domain Conversational Telephone Speech	161
	Christoph Lüscher ^{1,2} , Mohammad Zeineldeen ^{1,2} , Zijian Yang ^{1,2} , Tina Raissi ¹ , Peter Vieting ¹ , Khai Le-Duc ¹ , Weiyue Wang ² , Ralf Schlüter ^{1,2} and Hermann Ney ^{1,2} ¹ <i>RWTH Aachen University</i> , ² <i>AppTek GmbH, Aachen</i>	
32	Evaluation of HRTF Models for Binaural Cue Adaptation	166
	Sebastian Nagel and Peter Jax, <i>RWTH Aachen University</i>	
33	Global vs. Local Federated Learning in Heterogeneous Acoustic Environments	171
	Rene Glitza, Luca Becker and Rainer Martin, <i>Ruhr University Bochum</i>	
34	In-the-wild Speech Emotion Conversion Using Disentangled Self-supervised Representations and Neural Vocoder-based Resynthesis	176
	Navin Raj Prabhu, Nale Lehmann-Willenbrock and Timo Gerkmann, <i>Universität Hamburg</i>	
35	Low-complexity Real-time Single-channel Speech Enhancement Based on Skip-GRUs	181
	Ragini Sinha ¹ , Christian Rollwage ¹ and Simon Doclo ^{1,2} ¹ <i>Fraunhofer IDMT Dept. HSA, Oldenburg</i> , ² <i>University of Oldenburg</i>	
36	Multi-speaker Text-to-speech Using Forward Tacotron with Improved Duration Prediction	186
	Kishor Kayyar Lakshminarayana ¹ , Christian Dittmar ¹ , Nicola Pia ¹ and Emanuël A.P. Habets ² ¹ <i>Fraunhofer IIS, Erlangen</i> , ² <i>International Audio Laboratories Erlangen</i>	
37	On Feature Importance and Interpretability of Speaker Representations	191
	Frederik Rautenberg ¹ , Michael Kuhlmann ¹ , Jana Wiechmann ² , Fritz Seebauer ² , Petra Wagner ² and Reinhold Haeb-Umbach ¹ , ¹ <i>Paderborn University</i> , ² <i>Bielefeld University</i>	
38	Quantifying Harmonic Distortions in Audio Playback Systems	196
	Magnus Schäfer, <i>HEAD acoustics GmbH, Herzogenrath</i>	
39	Stream-ETS: Low-latency End-to-end Speech Synthesis from Electromyography Signals ...	200
	Kevin Scheck, Darius Ivucic, Zhao Ren and Tanja Schultz, <i>University of Bremen</i>	

Keynote III:

Information Extraction from Speech and Audio Sensor Data - Challenges for Applied Research in the Security Domain

Prof. Dr. Frank Kurth, Fraunhofer FKIE, Dept. Communication Systems

Poster Session D

Chair: Christiane Antweiler

40	Analyzing and Improving Neural Speaker Embeddings for ASR	205
	Christoph Lüscher ^{1,2} , Jingjing Xu ^{1,2} , Mohammad Zeineldeen ^{1,2} , Ralf Schlüter ^{1,2} and Hermann Ney ^{1,2} , ¹ <i>RWTH Aachen University</i> , ² <i>AppTek GmbH, Aachen</i>	
41	Distribution Mismatch Correction for Acoustic Scene Classification	210
	Lukas Maier, Alexander Fuchs and Franz Pernkopf, <i>TU Graz, Austria</i>	

42	Exploratory Evaluation of Speech Content Masking	215
	Jennifer Williams ¹ , Karla Pizzi ^{2,3} , Paul-Gauthier Noé ⁴ and Sneha Das ⁵	
	<i>¹University of Southampton, U.K. ²Fraunhofer AISEC, Garching, ³Technical University Munich</i>	
	<i>⁴University of Avignon, France, ⁵Technical University of Denmark</i>	
43	Exploring Visualization Techniques for Interpretable Learning in Speech Enhancement Deep Neural Networks	220
	Eike Nustede and Jörn Anemüller, <i>University of Oldenburg</i>	
44	Feedback-aware Design of an Occlusion Effect Reduction System Using an Earbud-mounted Vibration Sensor	225
	Christoph Weyer and Peter Jax, <i>RWTH Aachen University</i>	
45	Fuzzy-clustering-supported Assignment of Smart-speaker-based Microphone Arrays to Acoustic Sources in Reverberant Acoustic Environments	230
	Luca Becker ¹ , Stijn Kindt ² and Rainer Martin ¹	
	<i>¹Ruhr University Bochum, ²Ghent University, Belgium</i>	
46	Investigating Disentanglement of Speaker Identity and Characteristics through User Experience	235
	Sai Sirisha Rallabandi ¹ and Sebastian Möller ^{1,2}	
	<i>¹TU Berlin, ²German Research Center for Artificial Intelligence, Berlin</i>	
47	Language Recognition for SSB Modulated HF Radio Signals of Short Duration	240
	Alessia Cornaggia-Urrigshardt, Fabian Fritz, Lukas Henneke, Frank Kurth, Christian Schlich and Kevin Wilkinghoff, <i>Fraunhofer FKIE, Wachtberg</i>	
48	Self-learning and Active-learning for Electromyography-to-speech Conversion	245
	Zhao Ren, Kevin Scheck and Tanja Schultz, <i>University of Bremen</i>	
49	Target-speaker Voice Activity Detection in Multi-talker Scenarios: An Empirical Study	250
	Ahmad Aloradi, Mohamed Elminshawi, Srikanth Raj Chetupalli and Emanuël A.P. Habets <i>International Audio Laboratories Erlangen</i>	
50	Uncertainty-driven Hybrid Fusion for Audio-visual Phoneme Recognition	255
	Huajian Fang, Simone Frintrop and Timo Gerkmann, <i>Universität Hamburg</i>	

Lecture Session: Generative Methods for Speech Processing

Chairs: Tim Fingscheidt and Timo Gerkmann

51	On the Behavior of Intrusive and Non-intrusive Speech Enhancement Metrics in Predictive and Generative Settings	260
	Danilo de Oliveira, Julius Richter, Jean-Marie Lemerancier, Tal Peer and Timo Gerkmann, <i>Universität Hamburg</i>	
52	Evaluation Metrics for Generative Speech Enhancement Methods: Issues and Perspectives	265
	Jan Pirklbauer ¹ , Marvin Sach ¹ , Kristoff Fluyt ² , Wouter Tirry ² , Wafaa Wardah ³ , Sebastian Möller ³ and Tim Fingscheidt ¹	
	<i>¹TU Braunschweig, ²Goodix Technology (Belgium) BV, Leuven, ³TU Berlin</i>	

53	Improving the Naturalness of Synthesized Spectrograms for TTS Using GAN-Based Post-processing	270
	Paolo Sani ¹ , Judith Bauer ¹ , Frank Zalkow ¹ , Emanuël A.P. Habets ^{1,2} and Christian Dittmar ¹ <i>¹Fraunhofer IIS, Erlangen, ²International Audio Laboratories Erlangen</i>	
54	Audio-visual Speech Enhancement with Score-based Generative Models	275
	Julius Richter, Simone Frintrop and Timo Gerkmann, <i>Universität Hamburg</i>	
55	A Maximum Entropy Information Bottleneck (MEIB) Regularization for Generative Speech Enhancement with HiFi-GAN	280
	Marvin Sach ¹ , Jan Pirklbauer ¹ , Kristoff Fluyt ² , Wouter Tirry ² and Tim Fingscheidt ¹ <i>¹TU Braunschweig, ²Goodix Technology (Belgium) BV, Leuven</i>	