Table of Contents

Error Analysis in a Modular Meeting Transcription System 3
Peter Vieting, Simon Berger, Thilo von Neumann, Christoph Boeddeker, Ralf Schlüter, Reinhold Haeb-Umbach

Building a German-centric SpeechLLM Using Limited Data 8
Manas Maurya, Thomas Dethmann, Oliver Walter, Christoph Andreas Schmidt, Joachim Köhler

Investigation of Speech and Noise Latent Representations in Single-channel VAE-based Speech Enhancement 13
Jiatong Li, Simon Doclo

Unveiling Deep Speech Embeddings: Acoustic Insights into Hatespeech Detection 18
Rathi Adarshi Rammohan, Zhao Ren, Aleksandra Świderska, Dennis Küster, Tanja Schultz

The Role of Surprisal and Entropy in Spoken Intercomprehension: An Experiment on Translation of Cognates with Varied Predictability 23
Wei Xue, Julius Steuer, Dietrich Klakow, Bernd Möbius

Extending Manifold-Based MIMO System Identification to Adaptive Crosstalk Cancellation 28
Johannes Hahn, Tobias Kabzinski, Peter Jax

Speaker vs Noise Conditioning for Adaptive Speech Enhancement 33
Andreas Triantafyllopoulos, Iosif Tsangko, Michael Müller, Hendrik Schröter, Björn Schuller

YOLO-based Signal Detection of Amplitude Modulated Audio Transmissions in Realistic HF Scenarios 38
Lukas Henneke, Sebastian Urrigshardt, Fabian Fritz, Frank Kurth

Effectiveness of Acceleration Sensors on the Thorax and Abdomen for Speech Breathing Analysis 43
Dani Kazzy, Christian Kleiner, Susanne Fuchs, Peter Birkholz

Exploring In-Context Learning Capabilities of ChatGPT for Pathological Speech Detection 48
Mahdi Amiri, Hatef Otroshi Shahreza, Ina Kodrasi

Navigating PESQ: Up-to-Date Versions and Open Implementations 53
Matteo Torcoli, Mhd Modar Halimeh, Emanuël A. P. Habets

Mixed-Effects Models Neural Networks for Improved Speech-Based Predictions of Cognitive Decline 58
Jordan Behrendt, jiumend Zhang, Claudia Börnhorst, Tanja Schultz

Personalized Speech Synthesis for Zero-Shot Keyword Spotting 63
Fahrettin Gökgöz, Alessia Cornaggia-Urrigshardt, Kevin Wilkinghoff

Target Speaker Extraction: the Importance of a Powerful Extractor and Content-Informed Embeddings 68
Elias De Souter, Stijn Kindt, Kaixuan Yang, Haixin Zhao, Siyuan Song, Yanjue Song, Nilesh Madhu

Enhancement of Neural Embeddings for Speaker Identification in Ad-hoc Acoustic Sensor Networks and Multi-Speaker Scenarios 73
Philipp Intek, Luca Becker, Timm Koppelmann, Rainer Martin

A fully Zero-shot Approach to Obtaining Specialized and Compact Audio Tagging Models 78
Alexander Werning, Reinhold Häb-Umbach

Evaluating the Impact of Crowdsourced Audio Data on Speech Quality Assessment 83
Kirill Shchegelskiy, Malek El-Tannir, Wafaa Wardah, Tuğçe Melike Koçak Büyüktaş, Sebastian Möller

A Comparative Analysis on ASR System Combination for Attention, CTC, Factored Hybrid, and Transducer Models 88
Noureldin Bayoumi, Robin Schmitt, Tina Raissi, Albert Zeyer, Ralf Schlüter, Hermann Ney

Neural Prosody Prediction for German Articulatory Speech Synthesis 93
Peter Steiner, Zihao Huang, Arne-Lukas Fietkau, Peter Birkholz

Adapting the Fréchet Audio Distance as an Objective Metric for Text-to-Speech Quality Evaluation 98
Laurymas Zavistanavicius, Frank Zalkow, Christian Dittmar, Robert L. Stevenson

Towards Complex-Valued VAE-Based Distillation for Representation Learning in Speech Enhancement 103
Haixin Zhao, Kaixuan Yang, Nilesh Madhu

Room Reverberation Effectively Masks Deepfake Traces 108
Sophie Hoppe, Anabell Hacker, Markus Brückl

Early and Late Reflections in Aucoustic Echo Control: An Experimental Study on (Neural) Kalman Filters and DNN Methods 113
Ernst Seidel, Tim Fingscheidt

On the Application of Diffusion Models for Simultaneous Denoising and Dereverberation 118
Adrian Meise, Tobias Cord-Landwehr, Reinhold Haeb-Umbach

Binaural Distance Estimation Using a Joint Latent Representation of Acoustic Distance and Direct Path Response 123
Daniel Neudek, Benjamin Stodt, Stephan Getzmann, Rainer Martin

Detecting COPD Exacerbations Before Onset Using Vocal Biomarkers 128
Lars Nippert, Sami O. Simons, Julia Hoxha

Comparison of Knowledge Distillation Methods for Low-complexity Multi- microphone Speech Enhancement using the FT-JNF Architecture 133
Robert Metzger, Mattes Ohlenbusch, Christian Rollwage, Simon Doclo

Acoustical characterization and perceptual comparison of four types of 3D- printed vocal tract models for the German and Japanese vowels /a,e,i,o,u/ 138
Christian Kleiner, Peter Birkholz, Dominik Schäfer, Takayuki Arai

Low-Complexity Neural Wind Noise Reduction for Audio Recordings 143
Hesam Eftekhari, Srikanth Raj Chetupalli, Shrishti Saha Shetu, Emanuël A. P. Habets, Oliver Thiergart

An Improved Neural Network Architecture for Target Speech Extraction 148
David Joos, Friedrich Faubel, Jonas Jungclaussen, Markus Buck, Wolfgang Minker

A Very-Low Delay High-Performance Speech Vocoder Based on the Encodec Speech Decoder 153
Renzheng Shi, Tim Fingscheidt

Evaluating the Recognition Performance of the RehaLingo Speech Training System with Aphasic Speech 158
Hans-Günther Hirsch, Yannic Tiggelkamp, Christian Neumann, Tobias Bolten

Optimization of Feature and Loss Exponents for Lightweight DNN-based Binaural Speech Enhancement 163
Aleksej Chinaev, Gerald Enzner, Stefan Thaleiser

Unified Learnable 2D Convolutional Feature Extraction for ASR 168
Peter Vieting, Benedikt Hilmes, Ralf Schlüter, Hermann Ney

Blind Estimation of Head Rotations From Binaural Recordings 173
Erik Fleischhauer, Peter Jax

ReverbFX: A Dataset of Room Impulse Responses Derived from Reverb Effect Plugins for Singing Voice Dereverberation 178
Julius Richter, Till Svajda, Timo Gerkmann