VDE ITG Voice Message OKT 2025
Newsletter des ITG-Fachausschusses Sprachkommunikation (AT 3), 28. Ausgabe

Editorial

Liebe Leserinnen und Leser,

wieder gibt es viele gute und interessante Neuigkeiten aus unserer Runde – eine neue Professur, Preise, eine neue Ausgabe der CHiME-Challenge und neben mehreren Journalartikeln diesmal auch eine Special Issue in unserem Themenbereich.

Ganz unten finden Sie wie immer die Liste der wichtigsten Konferenzen in unserem Themenbereich. Diesmal habe ich auch, aus aktuellem Anlass, die IJCAI-ECAI mit aufgenommen, die im kommenden Jahr an der Universität Bremen von Tanja Schultz mit ausgerichtet wird – vielleicht ein guter Grund für den einen oder die andere von uns, unsere Arbeiten in diesen Kontext zu stellen, oder jedenfalls die IJCAI-ECAI zu besuchen und dort miteinander ins Gespräch zu kommen?

Und sonst sind ja nun der Herbst und das neue Semester ernstlich gestartet – ich hoffe, Sie können die etwas ungemütlichere Jahreszeit gut nutzen, für spannende Aufgaben und Projekte.

Mit freundlichen Grüßen
Dorothea Kolossa

Sie wünschen ein Abo oder haben einen Beitrag? Sehr gerne! Bitte melden Sie sich einfach per Email *unter Hinweis darauf, ob Sie nur Abonnent, oder Abonnent und auch möglicher Autor sein möchten! Wir weisen aus datenschutzrechtlichen Gründen darauf hin, dass Sie unter gleicher Emailadresse jederzeit Auskunft über Ihre gespeicherten Daten erfragen können, sowie die Löschung Ihrer Kontaktdaten erwirken können.*

Latest News, Awards

● Prof. Nilesh Madhu ist zum 01.09.2025 auf die W3-Professur „Signal Processing for Medical Applications“ der Fakultät für Elektrotechnik an der Helmut-Schmidt-Universität Hamburg berufen worden. Seit 2018 leitet er zudem an der Universiteit Gent ASPIRE@IDLab, eine Forschungsgruppe, die sich mit der KI-gestützten Verarbeitung von (bio-)Signalen beschäftigt – mit Anwendungen in der Kommunikation, Gesundheitsvorsorge usw. Die neue Professur verfolgt ein breites Forschungs-spektrum zur Entwicklung und Unterstützung intelligenter, interpretierbarer Systeme im Gesundheitswesen.

Der thematische Schwerpunkt liegt an der Schnittstelle von Signalverarbeitung, maschinellem Lernen und biomedizinischer Technik und konzentriert sich auf die folgenden drei Kernbereiche: (a) Algorithmen zur Audio- und Sprachverbesserung in Kommunikations- und Assistenztechnologien; (b) Multimodale Signalanalyse und -modellierung für diagnostische und therapeutische Anwendungen, einschließlich Informationsfusion für Brain-Computer-Interfaces (BCI); (c) Analytik und Repräsentationslernen als methodische Grundlage für interpretierbare KI, die eine transparente und anwendungsspezifische Anpassung von Rechenmodellen ermöglicht.

● Andreas Brendel (Fraunhofer IIS, Erlangen) erhielt für seine Doktorarbeit *From Blind to Semi-Blind Acoustic Source Separation based on Independent Component Analysis*, für die er von 2016 bis 2022 unter der Betreuung von Prof. Walter Kellermann am Lehrstuhl für Multimediakommunikation und Signalverarbeitung (FAU Erlangen-Nürnberg) forschte, den *Best PhD Award* der European Association for Signal Processing (EURASIP).

Quelle: A. Brendel

Die Verleihung des Preises fand im Rahmen der European Signal Processing Conference (EUSIPCO 2025) in Palermo (Italien) statt.

● Renzheng Shi und Prof. Tim Fingscheidt von der TU Brauschweig haben einen Best Paper Award auf der ITG Fachtagung Speech Communication verliehen bekommen. Ausgezeichnet wurde das Paper mit dem Titel “A Very-Low Delay High-Performance Speech Vocoder Based on the Encodec Speech Decoder”, das einen auf einem neuronalen Audio-Codec beruhenden Vocoder vorschlägt, der aus einer

spektralen Mel-Repräsentation ein Zeitsignal rekonstruiert. Im Gegensatz zu den meisten verfügbaren Vocodern beruht der hier entwickelte Ansatz nicht auf Sequenz-zu-Sequenz-Verarbeitung, sondern arbeitet niedriglatent mit einem rahmenweisen Vorschub. Dabei erzielt der vorgeschlagene Vocoder auf allen untersuchten Metriken eine bessere Performanz als die Referenzverfahren.

● Die neunte Ausgabe der CHiME Challenge ist gestartet. Zwei Tasks werden angeboten: Task 1, „Multi-modal context-aware recognition“, widmet sich einem echten Cocktailparty-Szenario: Bis zu vier parallel verlaufende Gespräche sollen aus den audio-visuellen Aufzeichnungen identifiziert und verschriftet werden. In Task 2, „Enhancing Conversations to address hearing impairment“, sollen Techniken entwickelt werden, um Hörgeschädigten die Unterhaltung in lauten Umgebungen zu erleichtern. Die Challenge wurde zwar schon offiziell eröffnet, aber es ist noch genügend Zeit, um bis zum 07.02.2026 Ergebnisse einzureichen. Der CHiME-Workshop findet am 04.05.2026 als Satellitenworkshop zur ICASSP 2026 statt.

Journalartikel und Special Issues

● F. Hilgemann, P. Jax (2025), Optimum Fixed-Pole Adaptive Filtering for Active Noise Control in Headphones. The implementation of least mean square (LMS)-type algorithms for active noise control (ANC) in headphones entails a trade-off between performance and computational complexity. In both aspects, adaptive infinite impulse response (IIR) filters with pre-determined pole locations, commonly referred to as fixed-pole adaptive filters (FPAFs), offer advantages over widely used finite impulse response (FIR) filters. The optimal selection of the poles requires oracle knowledge but remains ambiguous under practical conditions. As a remedy, a novel loss based on the steady-state error is proposed. The associated pole location optimization, which yields stable low-order FPAFs, is described. Using real-world measurement data, the performance and the degree of robustness against transfer path variations are investigated. The results indicate an improvement of several decibels over conventional algorithms with similar implementation costs.

● The Computer Speech & Language Special Issue Security and Privacy in Speech Communication is now out! It features cutting-edge research on privacy, fairness, and trust in speech technologies—from voice anonymization and spoofing detection to user trust and ethical design. The issue was edited by Ingo Siegert, Jennifer Williams, and Sneha Das, and invited contributions that address responsible, transparent, and human-centered speech AI.

● T. von Neumann, C. Boeddeker, M. Delcroix, R. Haeb-Umbach (2025), Word Error Rate Definitions and Algorithms for Long-Form Multi-Talker Speech Recognition The predominant metric for evaluating speech recognizers, the Word Error Rate (WER), has been extended in different ways to handle transcripts produced by long-form multi-talker speech recognizers. These systems process long transcripts containing multiple speakers and complex speaking patterns, so that the classical WER cannot be applied. There are speaker-attributed approaches that count speaker confusion errors, such as the concatenated minimum-permutation WER, and speaker-agnostic approaches, which aim to ignore speaker confusion errors. As these WERs evaluate different aspects and error types (e.g., temporal misalignment), we present a unified description of the existing WERs and highlight when to use which metric. To further analyze how many errors are caused by speaker confusion, we propose a diarization-invariant WER, which ignores speaker attribution errors. Since some WER definitions have high computational complexity, we also introduce a greedy algorithm to approximate these with high precision (<0.1% deviation in our experiments) and with polynomial instead of exponential complexity.

Stellenangebote

● Die neue Professur „Signal Processing for Medical Applications“ ist derzeit aktiv auf der Suche nach motivierten Wissenschaftler*innen, um das Team zu stärken und die etablierte Forschung im Bereich der Audio- und Sprachverbesserung weiter voran-zutreiben. Anfragen und Initiativbewerbungen können gern direkt an mailto:madhun@hsu.hamburg gesendet werden.

Tagungen

Hier finden Sie die Tagungen des kommenden Jahres im Überblick, mit vielen besonders spannenden und sehenswerten Konferenzorten:

ASRU	06. - 10.12.2025, Honolulu, HI, USA [Keine Anmeldung von Beiträgen mehr]
DAGA	23.-26.03.2026, Dresden, Deutschland Einreichungen bis 1.11.25
ICASSP	04. - 08.05.2026, Barcelona, Spanien [Keine Anmeldung von Beiträgen mehr]
IJCAI-ECAI	15. - 21.08.2026, Bremen, Deutschland Einreichungen bis ca. Ende Januar
EUSIPCO	31.08. - 04.09.2026, Brügge, Belgien Einreichungen bis 06.02.26
IWAENC	07.09. - 10.09.2026, Cremona, Italien Einreichungen bis 24.04.26
Interspeech	27.09. - 01.10.2026, Sydney, Australien Einreichungen bis 25.02.26