

## Editorial

Liebe Leserinnen und Leser,

heute freue ich mich sehr, Ihnen die erste Berlin-Edition der Voice Message zu senden! Seit Tim Fingscheidt im Oktober 2016 diesen Rundbrief ins Leben gerufen hat, habe ich gern und mit Spannung die Voice Message gelesen, und mich sehr gefreut, dass ich so aus unserer Community die aktuellsten Entwicklungen erfahren konnte – selbst wenn Konferenzbesuche ausgefallen sind, und obwohl wir an den verschiedensten Baustellen arbeiten und uns teils nur selten im wirklichen Leben begegnen.

Jetzt darf ich diese Aufgabe in einer äußerst spannenden und produktiven Zeit großer Umbrüche übernehmen: Für unsere Community eröffnen die rasanten Entwicklungen der Large Language Models, der Foundation Models und der multimodalen Technologien immense neue Chancen. Dabei stehen wir vor der Herausforderung, wie wir in dieser sich schnell verändernden Forschungslandschaft unsere Zielvorstellung eines tiefen Verständnisses unserer Methoden verwirklichen können. Gleichzeitig können wir uns durch die neuen Potentiale und durch die in vieler Hinsicht schneller und leichter gewordenen Entwicklungsprozesse zu ganz neuen Möglichkeiten hin inspirieren lassen, und wir selbst können versuchen, unsererseits zu neuen Inspirationen beizutragen.

Ich hoffe, dass wir mit der Voice Message ein wenig dazu beitragen können, die hierzu wichtigen Verbindungen zwischen unseren Instituten und Fachgebieten aufrechtzuerhalten und zu vertiefen. Hierfür freue ich mich über Ihre Hinweise und Ideen, sei es zur Form des Newsletters, zu anderen Verbreitungswegen oder zu neuen Themen.

Mit herzlichen 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. Walter Kellermann, der seit 1999 eine Professur für Nachrichtentechnik am Lehrstuhl für Multimediakommunikation und Signalverarbeitung (LMS) der FAU Erlangen-Nürnberg innehatte, wurde am 25. März 2024 mit einer Festveranstaltung des Lehrstuhls in den Ruhestand verabschiedet.

Zu einem von Prof. André Kaup moderierten Kolloquium trugen Prof. Rudolf Rabenstein (LMS), Prof. Alan Oppenheim (MIT, USA), Dr. Gary Elko (mh acoustics, USA), Prof. Patrick Naylor (Imperial College London, UK) und Dr. Roland Maas (Amazon, Seattle, USA) sowie Kollegen und Freunde bei. Im Anschluss waren alle Gäste zu einer Abendveranstaltung eingeladen, die im Fraunhofer Institut für Integrierte Schaltungen stattfand.



- Am 14. Juni 2024 feierte das [Institut für Kommunikationssysteme](#) (IKS) der RWTH Aachen sein 60-jähriges Bestehen. Das Institut, ursprünglich gegründet als Institut für Nachrichtengeräte und Datenverarbeitung (IND), wurde im Jahr 1964 mit der Berufung von Prof. Dr.-Ing. Hans Jörg Tafel ins Leben gerufen.

1988 übernahm [Prof. Dr.-Ing. Peter Vary](#) die Leitung des Instituts und führte es durch eine bedeutende Entwicklungsphase der Sprachsignalverarbeitung.

2015 wurde [Prof. Dr.-Ing. Peter Jax](#) die Leitung



übertragen. Unter seiner Führung wurde das Institut in „Institut für Kommunikationssysteme“ umbenannt. Dieses Jubiläum feiert sechs Jahrzehnte wegweisender Arbeit und bedeutender Beiträge im Bereich der Kommunikationssysteme.

- Der [Group Technical Achievement Award](#) der European Association for Signal Processing (EURASIP) geht in diesem Jahr an Prof. Dr.-Ing. Rainer Martin (Ruhr-Universität Bochum) „for contributions to statistical signal processing of speech and audio in hearing devices and for mobile communications through extensive activities of his research group.“

- Die nächste Sitzung unseres [Fachausschusses „Sprachkommunikation“](#) wird im Zusammenhang mit dem [International Workshop on Acoustic Signal Enhancement](#) in Aalborg stattfinden, voraussichtlich am Mittwoch, 11. September 2024, in der Mittagspause der Tagung. Für Mitglieder, die nicht an der Tagung in Aalborg teilnehmen, wird eine online-Teilnahme per Zoom eingerichtet. Alle Mitglieder des Fachausschusses mögen sich bitte diesen Termin schon jetzt vormerken!

## Bücher, Dissertationen

- Johannes Fabry: [Least-Squares Methods for Individualization of Hearables with Active Noise Cancellation](#), Institut für Kommunikationssysteme, RWTH Aachen, (Betreuer: Prof. Peter Jax, Prof. Simon Doclo)

Excessive noise exposure poses significant public health risks, including hearing loss, stress, sleep disturbances, and cardiovascular diseases. Active Noise Cancellation (ANC) in headphones helps mitigate these risks by reducing ambient noise. However, current ANC headphones face limitations due to processing power and battery life, employing time-invariant or simple adaptive filters that offer suboptimal performance and limited adaptability. Dr.-Ing. Johannes Fabry developed methods for the calibration and individualization of ANC headphones under real-world conditions as well as efficient adaptive algorithms that deliver near-optimal performance with only minimal prior knowledge about the acoustic system.



- Nima Nabizadeh: [Extracting and modelling sequential information from procedural task instructions](#), Elektronische Systeme der Medizintechnik, TU Berlin. (Betreuer: Prof. Dorothea Kolossa, Prof. Heiko Wersing)

Um Assistenzsysteme für komplexe Aufgaben zu gestalten, sollen maschinelle Modelle lernen, Wissen aus textuellen Instruktionen zu extrahieren und damit explizite Modelle für Verfahrensabläufe gewinnen. Die Dissertation von Nima Nabizadeh konzentriert sich genau auf diese Informationsextraktion aus Anweisungen, und fokussiert dabei Reparaturanleitungen und Kochrezepte als Wissensquellen. Dabei zeigt die Arbeit insbesondere den Wert einer Einbeziehung von hierarchischen und sequenziellen, auch langfristigen Abhängigkeiten, die jeweils zu wesentlichen Verbesserungen in der

Vorhersage des nächsten, auszuführenden Schritts im Reparatur- oder Kochprozess führen.

## Journalartikel

- Benjamin Stodt, Daniel Neudek, Stephan Getzmann, Edmund Wascher, Rainer Martin (2024): [Comparing auditory distance perception in real and virtual environments and the role of the loudness cue: A study based on event-related potential](#).

The perception of the distance to a sound source is relevant in many everyday situations, not only in real spaces, but also in virtual reality (VR) environments. In the present study, we compared the detection of changes of the distance to a sound source and its neurocognitive correlates in a real and a virtual reverberant environment, using an active auditory oddball paradigm and EEG measures. The main goal was to test whether the experiments in the virtual and real environments produced equivalent behavioral and EEG results. The participants had to indicate a target position via a joystick response (“near” or “far”). Sounds were emitted either by real loudspeakers in the real environment or rendered and played back for the corresponding positions via headphones in the virtual environment. Within both environments, loudness of the auditory stimuli was either unaltered or the loudness cue was manipulated. The EEG analysis focused on mismatch negativity (MMN), P3a, and P3b as correlates of deviance detection, attentional orientation, and context-updating/stimulus evaluation. Behavioral data showed that detection of the target positions was reduced within the virtual environment, especially when loudness was matched. Except for slight latency shifts in the virtual environment, EEG analysis indicated comparable patterns within both environments, independent of loudness settings. Thus, while the neurocognitive processing of changes in distance appears to be similar in virtual and real spaces, a proper representation of loudness appears to be crucial to achieve good task performance in virtual acoustic environments.

- E. Seidel and P. Mowlae and T. Fingscheidt, [Convergence and Performance Analysis of Classical, Hybrid, and Deep Acoustic Echo Control](#), Acoustic echo cancellation (AEC) and suppression (AES) are widely researched topics. However, only few papers about hybrid or deep acoustic echo control provide a solid comparative analysis of their methods as it was common with classical signal processing approaches. There can be distinct differences in the behavior of an AEC/AES model which cannot be fully represented by a single metric or test condition, especially when comparing classical signal processing and machine-learned approaches.

These characteristics include convergence behavior, reliability under varying speech levels or far-end signal types, and robustness to adverse conditions such as harsh nonlinearities, room impulse response changes, or delayed echo. A first contribution of this article is to present an extended set of test conditions and metrics that yields a proper characterization of an AEC/AES model and provides researchers with a useful toolbox to benchmark their systems. Second, we evaluate multiple AEC/AES models, representing classical, machine-learned, or hybrid paradigms, in various test conditions. We provide an analysis and new insights into their strengths and weaknesses and identify limitations of common metrics in some cases. Our entire toolbox of evaluation metrics and testing conditions is available on GitHub.

- [Christoph Boeddeker, Aswin Shanmugam Subramanian, Gordon Wichern, Reinhold Haeb-Umbach, Jonathan Le Roux, TS-SEP: Joint Diarization and Separation Conditioned on Estimated Speaker Embeddings](#)

Since diarization and source separation of meeting data are closely related tasks, we propose an approach to perform the two objectives jointly. It builds upon the target speaker voice activity detection (TS-VAD) diarization approach, which assumes that initial speaker embeddings are available. We replace the final combined speaker activity estimation network of TS-VAD with a network that produces speaker activity estimates at a time-frequency resolution. Those act as masks for source extraction, either via masking or via beamforming. The technique can be applied both for single-channel and multi-channel input and, in both cases, achieves a new state-of-the-art word error rate (WER) on the LibriCSS meeting data recognition task. We further compute speaker-aware and speaker-agnostic WERs to isolate the contribution of diarization errors to the overall WER performance.

## Tagungen

- [EUSIPCO](#), 26.-30.08.2024, Lyon, Frankreich  
[Keine Anmeldung von Beiträgen mehr]
- [Interspeech](#), 01.-05.09.2024, Kos, Griechenland  
[Keine Anmeldung von Beiträgen mehr]
- [IWAENC](#), 09.-12.09.2024, Aalborg, Dänemark  
[Keine Anmeldung von Beiträgen mehr]
- [SLT](#), 09.-12.12.2024, Macau, China  
[Keine Anmeldung von Beiträgen mehr]
- [DAS-DAGA](#), 17.-20.03.2025, Kopenhagen  
Paper-Deadline: Noch offen
- [ICASSP](#), 06.-11.04.2025, Hyderabad, India  
Paper-Deadline: 09.07.2024 [[Cfp](#)]
- 16. ITG-Fachtagung Sprachkommunikation,  
24.-26.09.2025, Berlin, Deutschland  
Paper-Deadline: Noch offen