

Editorial

Liebe Leserinnen und Leser,

die letzten Monate haben viele große Veränderungen gebracht, in unserem Land und im globalen Kontext. Diese Änderungen rücken für mich die Notwendigkeit technologischer Souveränität und verantwortungsvoller Technologieentwicklung noch schärfer in den Fokus. In diesem Sinn sehe ich auch die Rolle der ITG und unserer Community weiter gestärkt, als Ort nicht nur der wissenschaftlichen und technologischen Weiterentwicklung, sondern auch als Quelle wissenschaftlich fundierter Informationen für anstehende politische und wirtschaftliche Entscheidungen, sowohl in unserem Land als auch hoffentlich zunehmend im europäischen Raum.

In diesem Sinn sind sowohl ein neues Positionspapier zu Large Language Models (LLMs) als auch unsere Konferenzaktivitäten zu nennen. Für den weiteren gemeinsamen Austausch bietet sich in diesem Jahr besonders die ITG Fachtagung Sprachkommunikation an, zu der Sie die herzliche Einladung ebenfalls in dieser Ausgabe finden - natürlich zusammen mit den Informationen zu den wichtigen und größeren Konferenzen unseres Arbeitsgebiets in diesem Jahr. Ich freue mich darauf, diese Möglichkeiten zu einem umfassenden Austausch über die wissenschaftlichen und gesellschaftlichen Aspekte unserer Arbeit zu nutzen.

Sehr motivierend hierzu sind die vielen Erfolge und Publikationen aus unserer Runde, von denen Sie einige hier finden, und die bei allen aktuellen Unwägbarkeiten unsere starken Möglichkeiten zu positiven Veränderungen in den Vordergrund rücken.

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. Walter Kellermann](#) vom Lehrstuhl für Multimediakommunikation und Signalverarbeitung (LMS) der FAU Erlangen-Nürnberg erhält den *IEEE SPS Leo L. Beranek Meritorious Service Award 2024* für seine außerordentlichen Leistungen in der [IEEE Signal Processing Society](#). Die Auszeichnung wird im Rahmen der Award Ceremony der

[IEEE International Conference on Acoustics, Speech and Signal Processing \(ICASSP\)](#) überreicht, die vom 6. bis 11. April 2025 in Hyderabad, Indien, stattfindet.

- Julius Richter, Simon Welker, Jean-Marie Lemerrier and Bunlong Lay vom Arbeitsbereich [Signal Processing \(SP\)](#) unter der Leitung von Timo Gerkmann wurden mit dem VDE ITG Preis 2024 ausgezeichnet. Prämiert wurde die Veröffentlichung: Julius Richter, Simon Welker, Jean-Marie Lemerrier,



Bunlong Lay, Timo Gerkmann, "[Speech Enhancement and Dereverberation with Diffusion-based Generative Models](#)", IEEE/ACM Trans. Audio, Speech, Language Proc., Vol. 31, pp. 2351 - 2364, 2023.

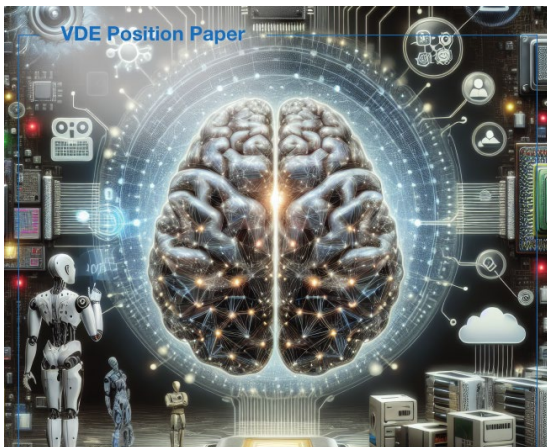
- Am 1. April 2024 hat [Prof. Sebastian J. Schlecht](#) als Nachfolger von Prof. Walter Kellermann die Professur für Signalverarbeitung am Lehrstuhl für Multimediakommunikation und Signalverarbeitung (LMS) der FAU Erlangen-Nürnberg angetreten. Prof. Schlecht promovierte 2017 unter der Leitung von Prof. Emanuël Habets an den International Audio Laboratories Erlangen und war von 2019 bis 2024 als [Professor of Practice](#) für Sound in Virtual Reality an der Aalto University in Helsinki, Finnland, tätig. Neben zentralen Themen der Audio- und akustischen Signalverarbeitung liegt ein besonderer Forschungsschwerpunkt auf Virtual and Augmented Reality, mit interdisziplinärer Anbindung an Psychoakustik, Raumakustik, maschinelles Lernen und kreative Anwendungen.



Quelle:
FAU/Georg
Pöhlein

Aus dem Fachausschuss AT 3

● ITG Positionspapier zu großen Sprachmodellen
Der Fachausschuss AT 3 hat das [Positionspapier](#) rechtzeitig zum ITG Hauptstadtforum am 26. November 2024 veröffentlicht. Es beschreibt die Technologien, Anforderungen an Daten und Hardware und relevante Anwendungen großer Sprachmodelle, und diskutiert deren Potentiale und Risiken. Die abschließenden Empfehlungen betonen die Notwendigkeit des weiteren Ausbaus der Infrastruktur für industrielle und akademische Forschung, der Verstärkung des Forschungstransfers und einer wissenschaftlich fundierten Diskussion zu Risiken und Regulierung, gemeinsam mit den Chancen, die sich hieraus auch für den europäischen Forschungsraum ergeben.



Large Language Models are Transformers in Artificial Intelligence, Industry, Education, and Society

by VDE ITG

VDE

Allen Autoren und Reviewern sei für die gute Zusammenarbeit herzlich gedankt!

● Die **ITG-Fachtagung Sprachkommunikation** findet vom 24.-26. September in Berlin die statt, mit Prof. Sebastian Möller als Chair, Prof. Timo Gerkmann als Co-Chair und Prof. Dorothea Kolossa als Technical Program Chair. Wie immer sind Beiträge aus dem gesamten, breiten Feld der Sprachkommunikation herzlich eingeladen [\[CfP\]](#). In diesem Jahr haben wir darüber hinaus als Fokusthema die Sprachkommunikation in der Zeit der Foundation Models in den Blick genommen.

Journalartikel

● R. Sinha, C. Rollwage, S. Doclo (2024), [Variants of LSTM cells for single-channel speaker-conditioned target speaker extraction](#).

Speaker-conditioned target speaker extraction aims to estimate the target speaker from a mixture of speakers utilizing auxiliary information about the target speaker. In this paper, we propose two variants of long short-term memory (LSTM) cells that are customized for speaker-conditioned target speaker extraction. The first variant customizes both the forget gate and input gate of the LSTM cell, while the second variant introduces a new gate, referred to as auxiliary-modulation gate. Both in unidirectional and bidirectional mode, experimental results on 2- or 3-speaker mixtures and noisy mixtures show that both proposed LSTM variants outperform standard LSTM cells for single-channel target speaker extraction, where the best performance is obtained using the auxiliary-gated LSTM cells.

● Jean-Marie Lemercier, Julius Richter, Simon Welker, Eloi Moliner, Vesa Välimäki, Timo Gerkmann (2024), ["Diffusion Models for Audio Restoration: A Review"](#). This IEEE Signal Processing Magazine article reviews diffusion models for speech enhancement and music restoration. Diffusion models elegantly combine data-driven learning with stochastic differential equations to perform gradual, step-by-step denoising, yielding natural and high sound quality.

● E. Seidel, G. Enzner, P. Mowlae, T. Fingscheidt (2024), [Neural Kalman Filters for Acoustic Echo Cancellation: Comparison of Deep Neural Network-Based Extensions](#).

Kalman filtering is a powerful approach to adaptive filtering. The frequency-domain adaptive Kalman filter (FDKF), based on the concept of the acoustic state space, provides a unifying solution to the adaptive filter update and the related stepsize control. It was conceived for the problem of acoustic echo cancellation and, as such, is frequently applied in hands-free systems. This article briefly recapitulates the linear FDKF and investigates how it can be further supported by deep neural networks (DNNs) in various ways, specifically to overcome the challenges and limitations related to the usually required estimation of process and observation noise covariances. While the mere FDKF comes with very low computational complexity, its neural Kalman filter variants may deliver faster (re)convergence, better echo cancellation, and even exceed the FDKF in its excellent double-talk near-end speech preservation both under linear and nonlinear loudspeaker conditions. To provide a synopsis of the state of the art, this article contributes a comparison of a range of DNN-based extensions of FDKF in the same training framework and using the same data.

- Florian Hilgemann, Egke Chatzimoustafa, Peter Jax (2024), [Data-Driven Uncertainty Modeling for Robust Feedback Active Noise Control in Headphones](#)

Active noise control (ANC) has become popular for reducing noise and enhancing user comfort in headphones. Although feedback control offers an effective way to implement ANC, it is restricted by the uncertainty of the controlled system (e.g., arising from differing wearing situations). Widely used unstructured models that capture these variations tend to overestimate the uncertainty and thus restrict ANC performance. As a remedy, this work explores uncertainty models derived from experimentally determined measurement data to improve ANC performance for over-ear and in-ear headphones. The controller optimization based on these models is described, and an ANC prototype is implemented to assess the performance benefit. Extensive measurements with human wearers confirm the robustness and indicate a performance improvement over conventional methods. The results allow to increase the active attenuation of ANC headphones by several decibels at no loss of stability.

- M. Bokaei, J. Jensen, S. Doclo, J. Østergaard (2024), [Low-Latency Deep Analog Speech Transmission Using Joint Source Channel Coding](#).

This paper proposes a configurable low-latency deep Joint Source-Channel Coding (JSCC) system for speech transmission. The system can be configured for varying signal-to-noise ratios, wireless channel conditions and bandwidths. A joint source-channel encoder based on deep neural networks is used to compress and transmit analog-coded information, while a configurable decoder reconstructs speech from noisy compressed signals. The system latency is adaptable, achieving a minimum latency of 2 ms with a lightweight architecture of 25 k parameters. The simulation results demonstrate that the proposed system outperforms conventional separate source-channel coding systems in terms of speech quality and intelligibility, particularly in low-latency and noisy channel conditions.

Dissertationen

- Sebastian Nagel: [Interactive Reproduction of Binaurally Recorded Signals](#), Institut für Kommunikationssysteme, RWTH Aachen, (Betreuer: Prof. Peter Jax, Prof. Michael Vorländer)

When binaural headphone signals are adapted to head movements using head-tracking, listeners perceive sound sources as fixed in the environment, rather than attached to or inside the head. Existing approaches are typically based on descriptions of virtual scenes or spatial sound field sampling using microphone arrays. This dissertation pursues this goal based on previously binaurally rendered or binaurally

recorded signals. The proposed approach can be entirely implemented in headphones. Since no other data interfaces or formats are required apart from the usual stereo transmission, such headphones can offer the user increased immersion with any existing media device and platform.

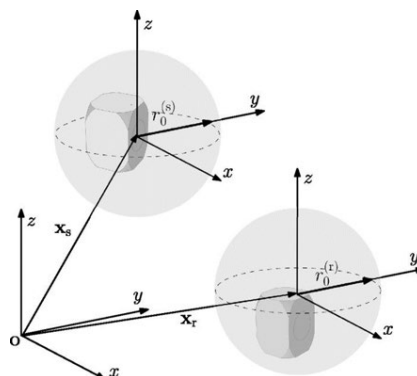


- Patrick Meyer: [Contributions to Automatic Interaction Analysis of Meetings](#), TU Braunschweig (Betreuer: T. Fingscheidt)

For automatic interaction analysis (AIA) of meetings based on acoustic data, high-quality audio recordings of each meeting participant are required. This is best achieved by multichannel audio recordings, where each meeting participant is equipped with a close-talk microphone and recorded in an individual target microphone channel. However, so-called crosstalk is a common problem with such audio recordings. Consequently, the speech signals of the participants are disturbed by each other, which strongly affects the performance of speech analysis methods for an AIA. To solve this issue, this thesis presents innovative contributions to the automatic processing and analysis of multichannel close-talk audio recordings of meetings suffering from crosstalk.

Open Science: Data and Algorithms

- **Intent on simulating more realistic room transfer functions?** Unlike existing tools like [gpuRIR](#) and [pyroomacoustics](#), [DEISM](#) allows you to model *sources and receivers with arbitrary directivities in non-shoebox rooms*. Whether you are working with loudspeakers, cellphones, or human heads, DEISM provides accurate simulations with *angle-dependent reflection coefficients*—a crucial but often-overlooked factor in room acoustics. Built on the latest research from Xu et al. ([JASA](#), [IWAENC](#)), DEISM pushes the boundaries of acoustic simulation. **Try it now and let us know what you think!**



Stellenangebote

- Zum Thema Voice Conversion sind derzeit eine Promotions- und eine Postdoktorandenstelle in Graz ([Signal Processing and Speech Communication Lab](#)) und Wien ([Speech and Hearing Science Lab](#)) ausgeschrieben.
- Im Fachgebiet [mtec](#) von Prof. Dorothea Kolossa ist eine [DFG-Projektstelle](#) (100% TVL-E13) in einem neuen Kooperationsprojekt mit [Prof. Boaz Rafaely](#) (Ben-Gurion Universität, Beer-Sheva) zu besetzen. Das Projekt ist zunächst für 3 Jahre bewilligt, und hat die mehrkanalige audiovisuelle Störfreiung von Sprachsignalen zum Ziel.

Tagungen

Der [ISCA/ITG-Workshop on Diversity in Large Speech and Language Models](#) hat erstmals am 20.2.2025 in Berlin stattgefunden. Er betrachtete Fragen der linguistischen und inhaltlichen Anpassung von LLMs an diverse Nutzergruppen – und von Nutzern an LLMs – u.a. mit einer Keynote von [Rachel Ostrand](#) (IBM Research) und einer Panel-Diskussion zu Fragen der inklusiven Technologiegestaltung und zu technologischer Souveränität im europäischen Raum.

Termine und Fristen

- [DAS-DAGA](#), 17.-20.03.2025, Kopenhagen, DK
[Keine Anmeldung von Beiträgen mehr]
- [ICASSP](#), 06.-11.04.2025, Hyderabad, India
[Keine Anmeldung von Beiträgen mehr]
- [Interspeech](#), 17.-21.8.2025, Rotterdam, Netherlands
[Keine Anmeldung von Beiträgen mehr]
- [EUSIPCO](#), 08.-12.09. 2025, Palermo, Italy
Paper-Deadline: 13.03.2025 [[CfP](#)]
- [MLSP](#), 31.8.-03.09. 2025, Istanbul, Turkey
Paper-Deadline: 22.04.2025 [[CfP](#)]
- [WASPAA](#), 12.-15.10.2025, Tahoe, CA, USA
Abstract-Deadline: 23.04. 2025 [[CfP](#)]
- [16. ITG-Fachtagung Sprachkommunikation](#)
24.-26.09.2025, Berlin, Deutschland
Paper-Deadline: 23.05. 2025 [[CfP](#)]