

## Editorial

Liebe Leserinnen und Leser,

wahrscheinlich sind Sie mittlerweile von den Reisen des Sommers und den Konferenzen des Septembers zurückgekehrt; der Herbst ist plötzlich da, und die Winterzeit lässt die Tage noch kürzer erscheinen. Die meisten von uns sind wieder im Büro, schauen nur gelegentlich in den Regen oder Nebel da draußen, und können uns wieder den Artikeln zuwenden, die wir schreiben, und den Projekten, die wir planen, neu starten oder die uns bereits lange beschäftigen.

Viele spannende Projekte und Projektergebnisse zeigt auch diese neue Voice Message – schauen Sie gern, was im Kreis der Kolleginnen und Kollegen an Ergebnissen und Preisen zu vermelden ist.

Für Ihre eigenen Veröffentlichungen findet sich am Ende des Newsletters wieder die aktualisierte Liste der wichtigsten Konferenz-Deadlines – schon am 1.11. sind die Abstracts für die DAS-DAGA fällig, und für die Interspeech in Rotterdam (Deadline 12. Februar) können bereits Pläne gemacht werden.

In diesem Sinne hoffe ich, dass unsere Voice Message Ihnen einige neue Anregungen und vielleicht auch Inspiration für den Herbst bringt – einschließlich unserer neuen Kategorie ‚Open Science.‘ Ich wünsche Ihnen eine gute und spannende Zeit voller neuer Ideen und inspirierender Kooperationen!

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. Sebastian Möller (Technische Universität Berlin und DFKI) wurde 2024 zum ISCA Fellow ernannt, [for sustained contributions to the evaluation of speech transmission, processing, and speech technology](#). Ganz herzlichen Glückwunsch!
- Jean-Marie Lemerrier aus dem Arbeitsbereich Signal Processing der Universität Hamburg unter der Leitung von Prof. Dr.-Ing. Timo Gerkmann hat gemeinsam mit Eloi Moliner von der Aalto University in Finnland den Best Student Paper Award auf der IWAENC-Konferenz erhalten. Ausgezeichnet wurden

sie für das Paper „[Buddy: Single-Channel Blind](#)



[Unsupervised Dereverberation with Diffusion Models](#)“.

- Congratulations to Julian Wechsler from the



International Audio Laboratories Erlangen for receiving the Best Student Paper Award (2nd place) at the [2024 International Workshop on Acoustic Signal Enhancement \(IWAENC\)](#) in Aalborg for his paper [“Neural directional filtering: Far-field directivity control with a small microphone array”](#). The paper was co-authored by Srikanth Ray Chetupalli, Mhd Modar Halimeh, Oliver Thiergart, and Emanuël Habets.

- Die Deutsche Forschungsgemeinschaft fördert das Graduiertenkolleg [HEARAZ](#) (Hearable-zentrierte Assistenz: Vom Sensor zur Teilhabe), das gemeinsam von den Universitäten Oldenburg und Bremen beantragt wurde. Ziel ist es, Hörgeräte zu sogenannten Hearables weiterzuentwickeln, die ihre Trägerinnen und Träger als Gesundheitsassistenten im Alltag unterstützen. Diese Hearables sollen unter anderem vor Sturzgefahr warnen, kognitive Veränderungen erkennen und den Verlauf von Parkinson-Erkrankungen dokumentieren. Oldenburg

bringt seine langjährige Erfahrung in der Hörerätforschung sowie Expertise aus der Geriatrie und Neurologie ein. Bremen ergänzt das Projekt mit Know-how in der Biosignal- und Sprachverarbeitung.

- Vom 07.-09-10.2024 fand in Düsseldorf das Kick-Off Meeting für die zweite Phase des DFG-Schwerpunktprogramms AUDICTIVE, SPP 2236 statt. Im Fokus der zweiten Phase steht die Erweiterung der Erkenntnisse, Theorien und Modelle der auditiven Wahrnehmungs- und Kognitionsforschung auf umfassende und komplexe Alltagsszenarien in Verbindung mit interaktiven virtuellen- und Mixed-Reality-Technologien. In AUDICTIVE kooperieren Forschende in insgesamt acht gemeinsamen Projekten; das Schwerpunktprogramm wird koordiniert von Prof. Dr.-Ing. Janina Fels, RWTH Aachen.

- Das Quality and Usability Lab (QULab) unter der Leitung von Prof. Dr.-Ing. Sebastian Möller an der TU Berlin arbeitet seit Jahrzehnten eng mit der International Telecommunication Union (ITU) zusammen. Aktuell entwickelt die Quality Group unter Dr. Robert Spang ein Modell zur Sprachqualitätsvorhersage, das für die Standardisierung durch die ITU in Betracht gezogen wird. Das Projekt (P.SAMD) zielt darauf ab, ein single-ended, nicht-intrusives Modell zu schaffen, das die Qualität empfangener Sprachsignale vorhersagen kann, wie sie von menschlichen Zuhörern in Telekommunikationsszenarien wie Mobiltelefongesprächen oder Videokonferenzen wahrgenommen wird. Anders als bisherige Standards wie POLQA nutzt P.SAMD ausschließlich das empfangene Signal zur Bewertung und benötigt kein Referenzsignal.

Die Doktorandin Wafaa Wardah entwickelt das Modell weiter und verwendet den Audio Spectrogram Transformer (AST), der von Forschern am MIT, USA, entwickelt und vortrainiert wurde, als Merkmalsextraktor. Das P.SAMD-Modell erzeugt daraus mithilfe eines Regressionsmodells Qualitätsbewertungen in verschiedenen Dimensionen, z.B. allgemeine Sprachqualität, Rauschen, Färbung und Lautstärke. Das Modell ermöglicht Telekommunikationsanbietern, die Sprachqualität allein anhand des empfangenen Signals zu überwachen, ohne Hörtests im Labor zu benötigen. Die internationale Zusammenarbeit mit anderen Parteien der [ITU Studygroup 12](#), wie Orange Innovation, Head Acoustics und Rohde & Schwarz, war hierfür von unschätzbarem Wert. Das Projekt soll auf der ITU SG12-Sitzung im Januar 2025 abgeschlossen werden.

## Journalartikel

- M. Tammen, S. Doclo (2023), [Parameter Estimation Procedures for Deep Multi-Frame MVDR Filtering for Single-Microphone Speech Enhancement](#).

Aiming at exploiting temporal correlations across consecutive time frames in the STFT domain, multi-frame algorithms for single-microphone speech enhancement have been proposed. Recently, it was shown that integrating the fully differentiable MFMVDR filter into an end-to-end supervised learning framework allows for a high estimation accuracy of the required parameters, i.e., the speech inter-frame correlation vector and the interference covariance matrix. In this paper, we investigate different covariance matrix structures, namely Hermitian positive-definite, Hermitian positive-definite Toeplitz, and rank-1. The main differences between the considered structures lie in the number of parameters that need to be estimated as well as the required linear algebra operations. Experimental results on the deep noise suppression challenge dataset show that the estimation procedure using the Hermitian positive-definite matrix structure yields the best performance, closely followed by the rank-1 matrix structure at a much lower complexity. Furthermore, imposing the MFMVDR filter structure instead of directly estimating the multi-frame filter coefficients consistently improves the speech enhancement performance.

- Christian Gruber, Gerald Enzner, Rainer Martin (2024), [Physically Architected Recurrent Neural Networks for Nonlinear Dynamical Loudspeaker Modeling](#)

The nonlinear behavior of loudspeakers is of great interest in a number of audio processing algorithms, as it may have a detrimental effect on their performance. These algorithms may be further enhanced when an accurate model of the loudspeaker's input-output behavior is available. A variety of approaches has been investigated in the past to solve this task via nonlinear adaptive system identification. Their modeling capabilities are often limited due to a mismatch with electroacoustic principles of real loudspeakers. This paper therefore presents a novel approach using recurrent neural networks (RNN) specifically designed to match the dynamical loudspeaker's physical model behavior. By means of the physical model and its corresponding state-space representation, we derive three conceptually different RNN architectures, which are initially trained on synthetic audio data in order to gain insights into the required training procedure and limitations. Further training and evaluation of the preferred architecture on real loudspeaker recordings shows consistent improvements of the mean-squared modeling error compared to a linear system model and to nonlinear baseline algorithms.

- R. Varzandeh, S. Doclo, V. Hohmann (2024), [Speech-aware Binaural DOA Estimation Utilizing Periodicity and Spatial Features in Convolutional Neural Networks](#).

In recent years, several supervised learning-based approaches have been proposed for estimating the direction of arrival (DOA) of a single talker in noisy and reverberant environments. In the absence of a voice activity detection (VAD), the estimated DOA may be erroneous due to speech pauses or noise dominance. In this paper, we consider a speech-aware DOA estimation system for binaural hearing aids, which does not require a separate VAD. This system utilizes a combination of spatial features with an auditory-inspired periodicity feature called periodicity degree (PD) as inputs of a convolutional neural network. Evaluations in reverberant environments with different background noises for both static and dynamic single-talker scenarios demonstrate that incorporating the PD feature provides an advantage for binaural DOA estimation in terms of accuracy and angular error.

- M. Ohlenbusch, C. Rollwage, S. Doclo (2024), [Modeling of speech-dependent own voice transfer characteristics for hearables with an in-ear microphone](#).

Many hearables contain an in-ear microphone, which may be used to capture the own voice of its user. However, due to occlusion, the in-ear microphone mostly records body-conducted speech, typically suffering from band-limitation effects and amplification at low frequencies. In this paper, we propose a speech-dependent model of the own voice transfer characteristics based on phoneme recognition, assuming a linear time-invariant relative transfer function for each phoneme. Experimental results based on recordings with a prototype hearable show that the proposed speech-dependent model enables a more accurate simulation of in-ear signals than a speech-independent model, especially under utterance and talker mismatch. Simulation results show that talker-averaged models generalize better to different talkers than individual models.

### Open Science: Data and Algorithms

- Do you need a specific noise field for developing, training or testing? You might have used the [INF-Generator](#) or [ANF-Generator](#). Recently, a new version of the ANF-Generator has been released with improvements based on the [JASA publication](#) by Daniele Mirabilii et al. from 2021. Now, a [Python](#) version is also available. By the way, did you know there also exists an official Python version of the [RIR-Generator](#) that you can install simply using `pip install rir-generator`?

### Stellenangebote

- Im Arbeitsbereich Signal Processing (SP) von Prof. Timo Gerkmann an der Universität Hamburg sind derzeit sechs Stellen auf Promotions- oder Postdoktorandenebene [ausgeschrieben](#).
- Im Fachgebiet mtec von Prof. Dorothea Kolossa ist eine Promotionsstelle (5 Jahre, mit Lehre im Bereich Spracherkennung) zum [Nutzerzentrierten Close-Captioning für Augmented-Reality-Anwendungen](#) zu besetzen.

### Tagungen

- [SLT](#), 02.-05.12.2024, Macau, China  
[Keine Anmeldung von Beiträgen mehr]
- [ISCA/ITG-Workshop on Diversity in Large Speech and Language Models](#)  
20.2.2025, Berlin, Germany [CfP]
- [DAS-DAGA](#), 17.-20.03.2025, Kopenhagen  
Paper-Deadline: 01.11.2024
- [ICASSP](#), 06.-11.04.2025, Hyderabad, India  
[Keine Anmeldung von Beiträgen mehr]
- [Interspeech](#), 17.-21.8.2025, Rotterdam, Netherlands  
Paper-Deadline: 12.02.2025 [CfP]
- [EUSIPCO](#), 08.-12.09. 2025, Palermo, Italy  
Paper-Deadline: 01.03.2025 [CfP]
- [16. ITG-Fachtagung Sprachkommunikation](#)  
24.-26.09.2025, Berlin, Deutschland  
Paper-Deadline: 15.05. 2025 [CfP]