

Editorial

Liebe Leserinnen und Leser,

hier ein sommerlicher Newsletter aus unserem Fachausschuss. Es gibt viel Neues zu Rängen aus internationalen Wettbewerben, besonders viele Artikel aus der hamburgischen Wissenschaft, wissenschaftlicher Nachwuchs in Aachen und – na klar – die ITG Fachtagung Sprachkommunikation im September in Aachen, zu der man sich ab sofort registrieren kann!

Ihr Tim Fingscheidt & Reinhold Häb-Umbach

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



- Matthias Kreuzer, Doktorand am [Lehrstuhl für Multimediakommunikation und Signalverarbeitung](#) der Friedrich-Alexander-Universität Erlangen-Nürnberg, wurde als einer der drei besten von 16 Teilnehmern bei der [IEEE Data Challenge on Industrial Systems' Health Monitoring using Vibration Signal Analysis](#) im Rahmen der [IEEE Conference on Prognostics and Health Management \(ICPHM\)](#), die vom 5. - 7. Juni 2023 in Montreal (Kanada) stattfand, ausgezeichnet.

- Der Arbeitsbereich [Signal Processing \(SP\)](#) der Universität Hamburg, unter Leitung von Timo Gerkmann, hat einen Platz in den Top 5 der ICASSP 2023 „Speech Signal Improvement“ Grand Challenge erreicht. Hierfür wurde erstmals eine diffusionsbasierte Sprachverbesserung mit einer algorithmischen Latenz ≤ 20 ms realisiert. [Der Beitrag wurde auf der ICASSP 2023 vorgestellt.](#) Audiobeispiele befinden sich [hier](#).

Bücher, Dissertationen

- Matthias Schrammen: [Front-End Signal Processing for Far-Field Speech Communication](#), RWTH Aachen University (Betreuer: Prof. Dr. Peter Jax, Prof. Dr. Nilesh Madhu)

Geräte für die Sprachkommunikation im Freisprechmodus ermöglichen eine sehr natürliche Art der menschlichen Kommunikation. In Smart-Home-Umgebungen oder Konferenzräumen sind oft auch mehrere Kommunikationsgeräte vorhanden.



Die im Rahmen der Dissertation von Matthias Schrammen entwickelten Algorithmen zur Front-End-Signalverarbeitung bieten Lösungen zur Schätzung aller Mikrofonpositionen, die unbekannt sind oder sich während des Betriebs ändern, sowie die gleichzeitige Verwendung von Beamforming und akustischer Echokompensation. Darüber hinaus sind die entwickelten Lösungen in der Lage, nichtlineare Echopfade zu verarbeiten und weisen eine geringe Rechenkomplexität auf, um auch für batteriebetriebene Geräte attraktiv zu sein.

Journalartikel

- A. Brendel, T. Haubner, and W. Kellermann [A Unifying View on Blind Source Separation of Convolutional Mixtures Based on Independent Component Analysis](#)

Convolutional Blind Source Separation (BSS) is a central problem in audio signal processing and methods based on Independent Component Analysis (ICA) are especially important in this field. Most of the currently used algorithms belong to one of the following three families: Frequency Domain ICA (FD-ICA), Independent Vector Analysis (IVA), and TRiple-N Independent component analysis for CONvolutional mixtures (TRINICON). While the relation between ICA, FD-ICA and IVA becomes apparent due to their construction, the relation to TRINICON is not well established yet. This paper fills this gap by providing an in-depth treatment of the common building blocks of these algorithms and their differences, and thus provides a common framework for all considered algorithms.

- J. Richter, S. Welker, J.-M. Lemerrier, B. Lay, T. Gerkmann, "[Speech Enhancement and Dereverberation with Diffusion-based Generative Models](#)", *IEEE/ACM Trans. Audio, Speech, Language Proc.*, 2023. [[doi](#)] [[arxiv](#)] [[code](#)] [[audio](#)]

In this article, we propose a diffusion-based generative model for speech enhancement based on stochastic differential equations (SDEs). In particular,

we propose to include a drift term into the SDE to achieve the required task adaptation. This allows for conditional generation with warm initialization which stabilizes and accelerates the convergence of the algorithm. The model has been shown to be applicable to a vast range of distortions by performing denoising, dereverberation, or bandwidth extension. The code is available online and already heavily used in the community.

- N. R. Prabhu, N. Lehmann-Willenbrock, T. Gerkmann, "[End-to-End Label Uncertainty Modeling in Speech Emotion Recognition using Bayesian Neural Networks and Label Distribution Learning](#)", *IEEE Transactions on Affective Computing*, 2023. [[doi](#)] [[arxiv](#)] [[code](#)]

In this paper we address the fact that in emotion recognition the labels themselves are subjective such that one ground truth arguably does not exist. We therefore propose to account for the distribution of labels obtained from multiple annotators using a Bayesian Neural Network.

- J.-M. Lemerrier, J. Thiemann, R. Koning, T. Gerkmann, "[A Neural Network-Supported Two-Stage Algorithm for Lightweight Dereverberation on Hearing Devices](#)", *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 18, 2023. [[doi](#)]

Here a lightweight combination of traditional signal processing and machine learning methods for speech dereverberation in hearing devices is proposed.

- H. Fang, D. Becker, S. Wermter, T. Gerkmann, "[Integrating Uncertainty into Neural Network-based Speech Enhancement](#)", *IEEE/ACM Trans. Audio, Speech, Language Proc.*, vol. 31, pp. 1587-1600, 2023. [[doi](#)]

In this article, we discuss how data and model uncertainties can be captured in neural-network-based speech enhancement.

Tagungen (nach Paper Deadline sortiert)

[ASRU](#), 16.-20.12.2023, Taipei, Taiwan,
Paper Deadline: 03.07.2023 [[CfP](#)]
[Interspeech](#), 20.-24.08.2023, Dublin, Irland
[keine Einreichungen mehr]

[ITG Conference on Speech Communication 2023](#)
20.-22.09.2023, Aachen
Registrierungsplattform ist nun geöffnet!

[EUSIPCO](#), 04.09.-08.09.2023, Helsinki, Finnland
[keine Einreichungen mehr]
[WASPAA 2023](#), 22.-25.10.2023, New Paltz, NY, USA,
[keine Einreichungen mehr]

Stellenanzeigen

- An der Professur für Sprachtechnologie und Kognitive Systeme der TU Dresden (Prof. Birkholz) gibt es mehrere zu besetzende Stellen für Wissenschaftler/Doktoranden (TVL-13, 100%) in den Bereichen Sprachtechnologie, Sensorik, und Phonetik. [Weitere Infos gibt es hier.](#)