

Editorial

Liebe Leserinnen und Leser,

es gibt tolle Nachrichten aus der Speech Community. So freuen wir uns über die Ernennung von Prof. Walter Kellermann (FAU Erlangen-Nürnberg) zum EURASIP Fellow, und über Prof. Dorothea Kolossa zur Wiederwahl in den ITG-Vorstand. Ein kurzes Statement von ihr findet sich unter „Persönliches“. Des Weiteren gibt es DCASE Challenge Awards und viele Journalartikel dieses Mal. Ist Home Office besonders produktiv?

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

• In Anerkennung herausragender Leistungen auf dem Gebiet der digitalen Signalverarbeitung ernennt die [European Association for Signal Processing \(EURASIP\)](#) jedes Jahr bis zu vier Wissenschaftlerinnen und Wissenschaftler zum "EURASIP Fellow". Dies ist die höchste Auszeichnung, die die EURASIP vergibt. Einer der neuen EURASIP Fellows des Jahres 2021 ist Prof. Walter Kellermann vom [Lehrstuhl für Multimediakommunikation und Signalverarbeitung \(LMS\)](#) an der [FAU Erlangen-Nürnberg](#), der diese Auszeichnung für seine Arbeiten zu Theorie und Anwendung der akustischen Signalverarbeitung erhält. Er ist damit der fünfte Wissenschaftler in Deutschland, der diese Auszeichnung seit ihrer erstmaligen Verleihung im Jahr 2007 erhält. Die Preisverleihung findet im August 2021 in Dublin (Irland) im Rahmen der [EUSIPCO 2021](#) statt.

Persönliches

• Die rapiden Entwicklungen der Informations- und Kommunikations-Technologien – gerade auch des maschinellen Lernens – führen zu beeindruckend beschleunigten Innovationszyklen. Eine Aufgabe, die daher zunehmend wichtiger wird, ist die schnelle Kommunikation wissenschaftlicher Erkenntnisse und technologischer Entwicklungen zwischen Unternehmen und Universitäten, zur Politikberatung, und über Positionspapiere und (soziale) Medien in die Gesellschaft hinein.

Das betrifft sowohl die Chancen und Möglichkeiten der neuen Methoden als auch den sicheren Umgang

mit Internettechnologien und Big-Data Analytics, wo Sprach- und Audiosignale oft eine zentrale Rolle einnehmen. Hier möchte ich den Dialog zwischen der wissenschaftlich-technischen Welt und der breiteren Gesellschaft fördern, besonders auch mit dem Ziel, junge Menschen für die Möglichkeiten eines Ingenieursstudiums zu begeistern.

Wenden Sie sich gern direkt an mich, mit Ideen, Fragen, Wünschen und Anliegen an den ITG-Vorstand! [Prof. Dr.-Ing. Dorothea Kolossa](#)

• Janek Ebberts vom Fachgebiet Nachrichtentechnik der Universität Paderborn (Prof. Häb-Umbach) hat bei der DCASE Challenge "Sound Event Detecton and Separation in Domestic Environment" unter 72 Teilnehmern den 4. Platz erreicht [\[Paper\]](#). Außerdem wurde der Beitrag mit dem "Reproducible System Award" ausgezeichnet.

Projekte und Aktivitäten

• Spin-Off vom Institut für Kommunikationssysteme der RWTH Aachen: Am 04. Februar 2021 haben das Team um Stefan Liebich, Johannes Fabry, Raphael Brandis und Elfed Howells die [Elevear GmbH](#) gegründet. Die Grundlagen für die Occlear-Technologie stammen aus ihrer Zeit am Institut für Kommunikationssysteme der RWTH Aachen. Hierbei werden optimierte Filter mit ultra-geringer Latenz eingesetzt, um eine aktive Unterdrückung des Okklusionseffektes in Hörgeräten, Hearables und Kopfhörern zu realisieren.



Das Projekt wurde 2019 bereits mit dem RWTH Innovation Award ausgezeichnet und war ein Pilotprojekt des RWTH Innovation Sprint 2019/2020 zur Prototypenentwicklung und Vorgründungsförderung. Seit Anfang Dezember 2020 erhält Elevear zudem das EXIST Gründerstipendium und wurde Anfang Januar 2021 als eines von zwölf herausragenden Gründerteams in das RWTH Incubation Program aufgenommen.

[\[Elevear auf LinkedIn\]](#)

- Am Quality and Usability Lab der Technischen Universität Berlin ist das Softwarecampus-Projekt „Qualitätsmessung durch Simulation von Konversationen“ gestartet. Das Projekt baut auf einem Simulationsansatz auf, der bisher mit Hilfe von simulierten Telefonnetzwerken die Interaktion in Konversationen modellierte. Dieser Ansatz soll nun auf echte Netzwerke hin weiterentwickelt werden. So wird die Simulation menschlicher Interaktionen und ein natürlicher Sprecherwechsel über ein Mobilfunknetzwerk simuliert, um so Rückschlüsse auf die Beeinträchtigungen auf echte Konversationsituationen ziehen zu können. Mit Hilfe der gewonnenen Erkenntnisse soll schließlich ein Schätzer der Konversationsqualität erstellt werden.

Bücher, Dissertationen

- Christian Hofmann: [Efficient System Identification using Parallel Connections and Cascades](#), Friedrich-Alexander Universität Erlangen-Nürnberg (Betreuer: W. Kellermann)

This PhD thesis introduces efficient methods for Acoustic Echo Cancellation (AEC) for hands-free spoken-language human-machine interfaces and communication devices. After a unifying view on state-of-the-art methods for linear multichannel and nonlinear single-channel system identification, the concepts of significance-aware filtering and source-specific system identification are introduced and verified. These concepts exploit parallel connections and cascades to decompose the system identification problem into smaller sub-problems, which are much easier to solve than the original problem. This allows AEC at reduced computational effort without sacrificing accuracy, as well as good trade-offs between AEC accuracy and even further increased efficiency. The thesis is completed by an outlook on the extension of the introduced concepts to nonlinear multichannel systems.

- Stefan Uhrig: [Human Processing of Transmitted Speech Varying in Perceived Quality](#), Quality and Usability Lab der Technischen Universität Berlin (Betreuer: Prof. Dr.-Ing. Sebastian Möller)

In der Arbeit von Stefan Uhrig geht es um das Thema der menschlichen Informationsverarbeitung bei der Sprachqualitätsmessung. In der Dissertation wird ein psychophysiologisches Modell zur Messung des Einflusses von variabler Sprachübertragungsqualität auf innere, perzeptuell-kognitive Prozesse in passiven Hörern vorgestellt und anhand dreier experimenteller Studien validiert. Auf neuronaler Analyseebene zeigen sich hierbei komplexe Interaktionseffekte zwischen der wahr-genommenen Sprachqualität, dem unmittelbaren Reizkontext sowie dem semantischen Sprachinhalt.

Journalartikel

- A. Brendel, T. Haubner, and W. Kellermann
[A Unified Probabilistic View on Spatially Informed Source Separation and Extraction Based on Independent Vector Analysis](#)

Signal separation and extraction are important tasks for devices recording audio signals in real environments which, aside from the desired sources, often contain several interfering sources such as background noise or concurrent speakers. Blind Source Separation (BSS) provides a powerful approach to address such problems. However, BSS algorithms typically treat all sources equally and do not resolve uncertainty regarding the ordering of the separated signals at the output of the algorithm. This paper addresses this problem by incorporating prior knowledge into the adaptation of the demixing filters in a probabilistic framework. The proposed framework allows to incorporate prior knowledge about the demixing filters in a generic way and unifies several known and newly proposed algorithms using a probabilistic view. For all algorithmic variants, we provide efficient update rules. The performance of a large variety of representative algorithmic variants, including very recent algorithms, is compared in this paper.

- J. Fabry, S. Kühl, and P. Jax
[On the Steady State Performance of the Kalman Filter Applied to Acoustical Systems](#)

In acoustical applications, a first order Markov model is often used to describe the time-variant nature of transversal filters. The Kalman filter is the optimal unbiased estimator for such a Markov model. It inherently calculates the uncertainty of the current state estimate by its state error covariance matrix. In contrast to the broadband Kalman filter, the covariance matrix of the exact Kalman filter depends on properties of the input signal. We propose a solution for the steady state covariance matrix based on the discrete-time algebraic Riccati equation, which depends on the process parameters of the Markov model as well as properties of the input signal. We show how this result can be used to predict the steady state performance of the Kalman filter for system identification.

- M. Strake, B. Defraene, K. Fluyt, W. Tirry, and T. Fingscheidt
[Speech Enhancement by LSTM-Based Noise Suppression followed by CNN-Based Speech Restoration](#)

Deep learning-based approaches have notably improved the performance of speech enhancement algorithms under challenging conditions, but still introduce speech distortions if strong noise suppression shall be achieved. In this article, we propose to address this problem by using a two-stage

approach, first performing noise suppression and subsequently restoring natural sounding speech, using specifically chosen neural network topologies and loss functions for each task.

We propose to use a long short-term memory (LSTM) network for noise suppression in the first stage, which is trained with a complex masked spectrum approximation loss providing high noise attenuation. Potential degradations of the speech component introduced by the first stage can subsequently be handled by the second stage, which performs speech restoration via spectral mapping with a convolutional encoder-decoder network. The proposed method improves speech quality (PESQ) over single-stage methods by about 0.1 points for unseen highly non-stationary noise types including interfering speech. Furthermore, it is able to increase intelligibility in low-SNR conditions.

- M. Niermann and P. Vary

[Listening Enhancement in Noisy Environments: Solutions in Time and Frequency Domain](#)

The objective of listening enhancement is to improve the speech intelligibility and the listening effort of speech, which is presented via a telephone or a public address system. This is also called Near-End Listening Enhancement (NELE). The loudspeaker signal is dynamically modified, taking the acoustical background noise at the near-end into account. Two strategies and two alternative implementation approaches are proposed.

The frequency domain approach does not require iterative optimization. The time domain approach is based on linear prediction techniques and even simpler. The proposed solutions outperform state-of-the-art in terms of computational complexity, memory requirement, continuous processor load, and latency.

- A. Schmidt, A. Brendel, T. Haubner and W. Kellermann

[Motor Data-Regularized Nonnegative Matrix Factorization for Ego-Noise Suppression](#)

Ego-noise, i.e., the noise a robot causes by its own motions, significantly corrupts the microphone signal and severely impairs the robot's capability to interact seamlessly with its environment. Therefore, suitable ego-noise suppression techniques are required. In this paper, we use a dictionary-based approach for ego-noise suppression. The estimation of the weights of the dictionary entries is very sensitive against other signals beside ego-noise contained in the mixture, e.g., speech. For increased robustness, we therefore propose to incorporate knowledge about the physical state of the robot, represented in terms of motor data, to the estimation of the weights. This is achieved by introducing a motor data-based regularization term to the estimation problem which promotes similar weights for similar physical states.

We analyze the proposed method and evaluate its ego-noise suppression performance for a large variety of different movements and demonstrate the superiority of the proposed method compared to an approach without using motor data.

- P. Meyer, S. Elshamy and T. Fingscheidt

[Multichannel Speaker Interference Reduction Using Frequency Domain Adaptive Filtering](#)

Crosstalk is a common problem in multichannel close-talk audio recordings, which occurs, when a target signal does not only couple into its dedicated microphone, but also into all other microphone channels. In this work, we extend a multichannel Wiener filter approach that considers all individual microphone channels to eliminate crosstalk in the target microphone channel. Therefore, an adaptive filter method in the frequency domain is used to obtain a well-performing estimate of the crosstalk (noise) component for the Wiener filter. As a result, the speech-to-crosstalk ratio is improved by up to 2.7dB at constant or even better speech component quality.

Tagungen (nach Paper Deadline sortiert)

[EUSIPCO](#), 23.-27.08.2021, Dublin, Irland,

Paper Deadline extended to 22.02.2021 [[CfP](#)]

[ESSV 2021](#), **03.03.-05.03.2021**, organisiert durch TU Berlin. Die Tagung findet virtuell statt.

[keine Einreichungen mehr]

[DAGA 2021](#), 15.-18.08.2021, Wien, Österreich,

Paper Deadline: 15.03.2021 [[CfP](#)]

[Interspeech](#), 30.08.-03.09.2021, Brno, Tschechien, (hybrid)

Paper Deadline: 26.03.2021 [[CfP](#)]

[ITG Conference on Speech Communication 2021](#)

29.09.-01.10.2021 in Kiel,

Paper Deadline: 23.05.2021 [[CfP](#)]

[ASRU](#), 13.-17.12.2021, Cartagena, Kolumbien,

Paper Deadline: 25.06.2021 (double-blind!)

[HSCMA](#) June 2021 **schedule cancelled, stay tuned.**

[ICASSP 2021](#), 06.-11.06.2021 in Toronto, Kanada

[keine Einreichungen mehr]

[ICASSP 2022](#), 23.-27.05.2022, Singapur

Paper Deadline: tbd.

[IWAENC](#), **05.-08.09.2022**, Bamberg,

Erneute Verschiebung aufgrund der anhaltenden Corona-Pandemie.

Stellenanzeigen

- Das Institut für Nachrichtentechnik der TU Braunschweig sucht zum 01.06.2021 eine/n Mitarbeiter/in (TV-L E13) für die Erforschung und Entwicklung einer mehrkanaligen Freisprechlösung für Leitstände in kritischen Infrastrukturen. Die Stelle ist zunächst auf 2 Jahre befristet. [[Kontakt](#)]