

Editorial

Liebe Leserinnen und Leser, keine Voice Message ist wie die andere. In der Sprachtechnologie gibt es Preise und öffentlichkeitswirksame Exponate zu melden (MS WISSENSCHAFT), diesmal vor allem jedoch Dissertationen und Journalartikel *en masse*. Ach ja, nicht zuletzt durch die Konjunktur im maschinellen Lernen auch einige Stellenangebote. Wie schön, zu diesen Zeiten in der Sprachverarbeitung tätig zu sein...

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



Künstliche Intelligenz im Fokus: OvGU auf der MS WISSENSCHAFT mit an Bord (Quelle: Ilja Hendel/Wid)

- Das IIKT (Institut für Informations- und Kommunikationstechnik) der Uni Magdeburg stellt zum Thema Sprachverarbeitung mithilfe von Amazons ALEXA ein Exponat auf der [MS WISSENSCHAFT](#) aus. Auf der Forschungsebene soll mit einem aktuell durchgeführten, ähnlich gearteten Projekt untersucht werden, wie Dialoge mit ALEXA natürlicher werden können. Mittels des ausgestellten Exponats soll auf einfache, interaktive Weise gezeigt werden, dass aktuelle KI-Systeme eben nicht immer zur Zufriedenheit der Nutzer funktionieren. Ziel des Forschungsprojekts ist es, dass sich technische Systeme auf den Nutzer einstellen – und nicht umgekehrt. Die Idee der MS WISSENSCHAFT ist ein schwimmendes *Science Center* mit Exponaten zum Ausprobieren und Mitmachen. Das Thema der Ausstellung orientiert sich am jeweiligen Wissenschaftsjahr und fokussiert 2019 die Künstliche Intelligenz. Sie tourt von Mai bis September durch Deutschland.

Persönliches

- Stefan Liebich, M.Sc., Johannes Fabry, M.Sc., Prof. Peter Jax und Prof. Peter Vary wurden mit dem Innovation Award 2018 der RWTH Aachen University für das Projekt „[Enhanced Audio Reality \(EAR\): Digitally opening the ear for headsets and hearing aids](#)“ ausgezeichnet. Der patentierte Ansatz löst das Problem des Okklusionseffektes durch die aktive Abstrahlung von Kompensationssignalen mit dem integrierten Lautsprecher. Eine korrekt balancierte Verarbeitung der beiden Mikrofonsignale resultiert in einer „digitalen Ohröffnung“ und einer wesentlich natürlicheren Wahrnehmung sowohl der eigenen Stimme als auch der Umgebung.



- Maja Taseka, who obtained her Ph.D. degree in Nov. 2017 from the University of Erlangen-Nürnberg, received the 2018 Doctoral Thesis Award from the Alumni Technische Fakultät Erlangen and was nominated for the ARD/ZDF Förderpreis »Frauen+Medientechnologie« 2019.

Projekte und Aktivitäten

- Im Rahmen des BMWi-Ideenwettbewerbs „Künstliche Intelligenz“ wurde das Projekt SPEAKER – Aufbau einer führenden Sprachassistentenplattform „Made in Germany“ ausgewählt. In der ersten Konzeptionsphase erstellen das Fraunhofer IAIS (Projektleitung Joachim Köhler) und das Fraunhofer IIS ein Planungskonzept. Ziel des Vorhabens ist die Schaffung einer führenden Sprachassistentenplattform „Made in Germany“, welche vor allem Anwendungen in der Industrie und in Business-to-Business-Produkten findet. Dieses System soll sowohl in ausgewählten Komponenten technologisch weltführend sein, als auch international Sichtbarkeit erlangen. Die Sprachassistentenplattform ist zentraler Bestandteil eines umfassenden Ökosystems für branchenübergreifende Sprachanwendungen und verbindet die Exzellenz der Forschung und Industrie unter besonderer Berücksichtigung des Mittelstandes.
- In der [Fakultät Elektrotechnik der Ostfalia HAW](#) wird ein Projekt zum Thema „Cochlea-Implantate“ durchgeführt. Dabei soll die Sprachverständlichkeit für CI-Träger verbessert werden. Die subjektive und objektive Validierung / Bewertung der Algorithmen erfolgt automatisiert mit Hilfe eines Kunstkopfmesssystems. Reale Umgebungsbedingungen

werden über ein Auralisationssystem, aufgebaut in einem reflexionsarmen Freifeldraum, simuliert.

Dissertationen

- Christian Hümmer: [*A Bayesian Network Approach to Selected Problems in Speech Signal Processing*](#)

Friedrich-Alexander-Universität Erlangen-Nürnberg (Betreuer: W. Kellermann)

Two distinct speech signal processing tasks are addressed from a unifying Bayesian network perspective. First, the problem of single-channel nonlinear acoustic echo cancellation (NAEC) is considered and in this context, a common heuristic choice of an optimum stepsize for the NLMS algorithm is derived as a special case of the Kalman filter. Furthermore, the properties of the elitist particle filter based on evolutionary strategies (EPFES) are analyzed when applied to identifying the nonlinear acoustic echo path model. As a second application, the task of environmentally-robust automatic speech recognition (ASR) is addressed.

- María Luis Valero: *Acoustic Echo Reduction for Multiple Loudspeakers and Microphones: Complexity Reduction and Convergence Enhancement*, International Audio Laboratories Erlangen (Betreuer: E.A.P. Habets)

The focus of the thesis lies on the complexity reduction and convergence enhancement of acoustic echo cancellation (AEC) for acoustic setups that comprise several loudspeakers or microphones. First, and irrespective of the acoustic setup, methods to reduce the computational cost and to enhance the convergence rate of partitioned-block frequency-domain adaptive filters are presented. Secondly, linear-periodically time-varying coherence reduction methods for the convergence enhancement of multi-channel AEC are theoretically analyzed and optimized. Regarding AEC for setups that comprise several microphones, an algorithm is presented, which is capable of reducing the computational cost of multi-microphone AEC by considering the relation across microphones of acoustic echoes instead of the acoustic propagation paths. Finally, mechanisms to cope with the residual echoes that remain after multi-channel and multi-microphone AEC are also provided.

- Markus Niermann: [*Digital Enhancement of Speech Perception in Noisy Environments*](#), RWTH Aachen University (Betreuer: P. Vary, zweiter Gutachter: R. Martin).

The perception of loudspeaker speech from public address systems or mobile phones is often affected by local background noise. For improving speech intelligibility and for reducing listening effort, "Near-



End Listening Enhancement" (NELE) processing can be applied. The loudspeaker signal is adaptively modified by taking the instantaneous background noise into account. The main novel contributions of this thesis are frequency domain NELE weighting rules without need for complex optimization and NELE time domain filter structures with low latency and complexity. Furthermore, NELE solutions for public address systems with massive crosstalk and combined NELE operation with far-end noise reduction in mobile telephony are proposed.

- Johannes Stahl: [*Contributions to Single-Channel Speech Enhancement with a Focus on the Spectral Phase*](#), Technische Universität Graz (Betreuer: P. Mowlae, G. Kubin)

Single-channel speech enhancement refers to the reduction of noise signal components in a single-channel signal composed of both speech and noise. Spectral speech enhancement methods are among the most popular approaches to solving this problem. Since the short-time spectral amplitude has been identified as a highly perceptually relevant quantity, most conventional approaches rely on processing the amplitude spectrum only, ignoring any information that may be contained in the spectral phase. As a consequence, the noisy short-time spectral phase is neither enhanced for the purpose of signal reconstruction nor is it used for refining short-time spectral amplitude estimates. This thesis investigates the use of the spectral phase and its structure in algorithms for single-channel speech enhancement.

Journalartikel

- A. Brendel and W. Kellermann: [*Distributed Source Localization in Acoustic Sensor Networks Using the Coherent-to-Diffuse Power Ratio*](#)

Acoustic source localization in an acoustic sensor network (ASN) based on estimates of the coherent-to-diffuse power ratio (CDR) is investigated. It is shown that the distance between an acoustic source and an observing microphone pair is reflected by the CDR and can be estimated by Gaussian process regression. Subsequently, the obtained distance estimates are used by an ASN for estimating the position of the acoustic source. The computational complexity of the algorithm is low and the amount of data to be shared between the sensor nodes can be adjusted.

- M. Luis Valero and E.A.P. Habets:
[Low-complexity multi-microphone acoustic echo control in the short-time Fourier transform domain](#)

This paper proposes a low-complexity architecture for multi-microphone echo control. The proposed architecture uses a low-complexity multi-microphone acoustic echo canceller followed by a beamformer for joint noise and residual echo reduction. To implement the latter, it is necessary to retrieve the relation of residual echoes across microphones for which two estimation methods are presented.

- H. Schepker, L. T. T. Tran, S. Nordholm, and S. Doclo:
[Null-steering beamformer based feedback cancellation for multi-microphone hearing aids with incoming signal preservation](#)

This paper proposes several methods to design a fixed beamformer aiming at canceling the acoustic feedback contribution for an earpiece with multiple microphones. Experimental results with a custom earpiece show that the proposed null-steering beamformer using an RTF-based constraint enables to increase the added stable gain by about 40 dB while preserving the quality of the incoming signal.

- W. Nogueira, J. Abel, and T. Fingscheidt:
[Artificial speech bandwidth extension improves telephone speech intelligibility and quality in cochlear implant users](#)

Hearing impaired persons, and particularly hearing-aid and cochlear implant (CI) users, often have difficulties communicating over the telephone. This work proposes to artificially extend the standard bandwidth of telephone speech to improve its intelligibility for CI users. Nine CI users were tested with and without the proposed artificial bandwidth extension algorithm. The experimental evaluation shows a significant improvement in speech intelligibility and speech quality over the phone using the artificial bandwidth extension. These promising results support the potential of ABE, which could be incorporated into a commercial speech processor or a smartphone-based pre-processor that streams the telephone speech to the CI.

- B. Cauchi, K. Siedenburg, J. F. Santos, T. H. Falk, S. Doclo, and S. Goetze:
[Non-intrusive speech quality prediction using modulation energies and LSTM-network](#)

This paper proposes a novel non-intrusive measure to predict the quality of processed speech by combining modulation energy features and a recurrent neural network with long short-term memory cells. For various speech enhancement algorithms, the proposed measure yields a higher correlation with perceptual speech quality than benchmark intrusive and non-intrusive measures.

- S. Chakrabarty and E.A.P. Habets:
[Time-Frequency Masking Based Online Multi-Channel Speech Enhancement With Convolutional Recurrent Neural Networks](#)

The paper proposes to use a convolutional recurrent neural network to estimate ideal binary and ratio masks using a sequence of multi-channel inputs. The estimated masks are used in two different ways for speech enhancement: i) to estimate real-valued gains which are applied to a single microphone, and ii) to estimate the power spectral density matrices of the desired and undesired signals for beamforming.

- S. Chakrabarty and E.A.P. Habets:
[Multi-Speaker DOA Estimation Using Deep Convolutional Networks Trained With Noise Signals](#)

This paper proposes a classification-based direction-of-arrival (DOA) estimator for multiple speakers. The classification is performed on a time-frame level using a convolutional neural network. The network can be trained using simulated impulse responses and noise signals only, and was evaluated also with multiple speakers recorded in a classroom.

Tagungen (nach Paper Deadline sortiert)

- [ASRU](#) 14.-18.12.2019 in Sentosa, Singapur
Paper Deadline: 01.07.2019 [CfP]
- [EUSIPCO](#) 02.-06.09.2019 in A Coruna, Spanien
[keine Einreichungen mehr]
- [Interspeech](#) 15.-19.09.2019 in Graz, Österreich
[keine Einreichungen mehr]
Die Paper Deadline ist zwar schon vorbei, aber die [Interspeech 2019](#) in Graz kann mit einigen Highlights aufwarten: Zuwachs an Papereinreichungen von über 30% im Vergleich zum Vorjahr, [Survey Presentations, kostenlose Kinderbetreuung für Konferenzteilnehmer*innen, Weiterbildung für Lehrer*innen der Sekundarstufe](#) und ein [Hackathon](#) – Anmeldungen sind noch möglich, so wie auch zu diversen [Satellite Workshops](#) und [Student Events](#).
- [WASPAA](#) 20.-23.10.2019 in Mohonk Mountain House, New Paltz, NY, USA
[keine Einreichungen mehr]
- [ICASSP](#) 04.-09.05.2020 in Barcelona, Spain
Paper Deadline: 21.10.2019 [CfP]
- [DAGA](#) 16.-19.03.2020 in Hannover,
Deadline für Beitragsmeldung: 01.11.2019
- [EUSIPCO](#), 24.-28.08.2020, Amsterdam, Niederlande,
Paper Deadline: 21.02.2020 [CfP]
- [Interspeech](#), 02.-06.09.2020, Shanghai, China,
Paper Deadline noch unbekannt
- [ITG Conference on Speech Communication](#)
9.-11.09.2020 in Aachen, Paper Deadline noch unbekannt

Stellenangebote

- Am Institut für Akustik und Sprachkommunikation der TU Dresden ist ab 1. Juli eine 50% Stelle TVL-13 im Bereich stille Spracherkennung auf Basis von Radarsensoren zu vergeben. Bei Interesse bitte melden bei Prof. [Peter Birkholz](#).
- Die Abteilung NetMedia des Fraunhofer-Institutes für Intelligente Analyse- und Informationssysteme IAIS sucht zum nächstmöglichen Zeitpunkt eine(n) wissenschaftliche(n) Mitarbeiter(in) im Bereich der Spracherkennung. Bewerbungen mit den üblichen Unterlagen (Anschreiben, Lebenslauf, Zeugnisse etc.) sind bitte an [Joachim Köhler](#) zu schicken.
- In der Abteilung Signalverarbeitung des Departments für medizinische Physik und Akustik an der Universität Oldenburg werden 2 Postdocs in den Bereichen „Machine learning-based speech enhancement algorithms“ und „Individualized signal processing for hearing devices“ gesucht. [\[Link\]](#)
- An der [Technischen Universität Graz](#) erfolgt die Neueinrichtung einer unbefristeten Universitätsprofessur für Akustik. Der Schwerpunkt in der Forschung soll auf der Technischen Akustik liegen, das federführende Engagement im interuniversitären Studium [Elektrotechnik-Toningenieur](#) erfordert auch eine musikalisch-künstlerische Ausbildung. Interessent*innen und Vorschläge von potentiellen Kandidat*innen bitte bei [Gernot Kubin](#) melden.
- Das [Signal Processing and Speech Communication Laboratory](#) der TU Graz sucht eine*n Universitätsassistent*in zur Promotion an der Schnittstelle zwischen phonetischer Forschung und maschinellem Lernen in der Spracherkennung. Die engagierte Mitarbeit in der Lehre zur Signalverarbeitung wird erwartet. Interessent*innen melden sich bitte bei [Gernot Kubin](#) und [Barbara Schuppler](#).