

**Editorial**

Liebe Leserinnen und Leser, Ehrungen und Berufungen im Bereich der Sprachakustik oder Sprachverarbeitung sind immer ein schöner Anlass für eine Neuausgabe der Voice Message. Was uns auch freut: Es ist immer mal wieder schwierig, die eingegangenen Beiträge auf zwei Seiten zu kürzen. Dann lassen wir es halt.

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**

● Smartphone, Smart Home, Smart Car – Intelligente Systeme umgeben uns. Verstehen wir sie? Der Bachelorstudiengang [Informationstechnik - Smarte Systeme](#) wird von der Otto-von-Guericke-Universität Magdeburg und der Hochschule Magdeburg-Stendal gemeinsam durchgeführt und von National Instruments (NI) technologisch unterstützt. Elektro- und Informationstechnik mit spezieller Ausrichtung auf Technologien der Smarten Systeme werden an NI-Systemen hands-on praxisnah ausgebildet, und das über die gesamte Studiendauer. Die Studierenden lernen so die Bearbeitung von Projekten, die auf industrielle Maßstäbe skalierbar sind. Diese Art der Ausbildung in Kooperation mit NI gibt es in Europa nur noch in Manchester (UK). Der Studiengang ist ab 2019 unbeschränkt offen mit Abitur und Fachhochschulreife. Beratung unter [bachelor.itsmsy@ovgu.de](mailto:bachelor.itsmsy@ovgu.de).



**Persönliches**

● Herr Dr.-Ing. Ingo Siegert ist im November 2018 zum [Juniorprofessor für Mobile Dialogsysteme](#) am Institut für Informations- und Kommunikationstechnik der Otto-von-Guericke-Universität Magdeburg berufen worden. Er befasst sich mit der Erforschung von Dialogstrategien mit Maschinen und mit mobilen Nutzern und bewegt sich dabei im Schnittpunkt der Forschungs-



gebiete Sprachsignalverarbeitung und Mensch-Maschine-Interaktion. Anwendungsgebiete sind Assistenzsysteme im häuslichen Bereich sowie im Automobilbereich.

● Am 5. November 2018 ist [Prof. Dr.-Ing. Peter Vary](#) von der Informationstechnischen Gesellschaft (ITG) im VDE zum ITG Fellow ernannt worden. Das Preisprüfergremium und der ITG-Vorstand haben mit dieser Entscheidung seine von der gesamten Fachwelt beachteten herausragenden Verdienste, sein Engagement und seine wissenschaftlichen Leistungen auf dem Gebiet des Mobilfunks gewürdigt. Die Urkunde wurde feierlich in der Berlin-Brandenburgischen Akademie der Wissenschaften im Rahmen des VDE Hauptstadtforums überreicht. Das Foto zeigt ihn gemeinsam mit Prof. Dr.-Ing. Hans D. Schotten, dem Vorsitzenden der Informationstechnischen Gesellschaft. Die Mitglieder in den Fachausschüssen gratulieren Prof. Peter Vary ganz herzlich zu dieser besonderen Auszeichnung.



**Projekte und Aktivitäten**

● Auf der **13. ITG-Fachtagung Sprachkommunikation** stellten Fachleute aus Wissenschaft und Industrie den über 150 Teilnehmern in 68 Beiträgen aktuelle Forschungsergebnisse und Entwicklungen auf dem Gebiet der Sprachsignalverarbeitung vor. Die Tagung wurde am Dept. für medizinische Physik und Akustik der Universität Oldenburg unter der Leitung von Prof. Simon Doclo veranstaltet. Höhepunkte der Tagung waren auch diesmal wieder die eingeladenen Übersichtsvorträge von Prof. Sharon Gannot, Prof. Shinji Watanabe und Dr. Thomas Brand. [Die Vortragsfolien sind online verfügbar.](#)



- Tim Fingscheidt vom Institut für Nachrichtentechnik der TU Braunschweig wurde seitens der Deutschen Forschungsgemeinschaft (DFG) ein Vorhaben zur Untersuchung der Informationsfusion in der automatischen Spracherkennung nach dem Turbo-Prinzip zuerkannt. Das Ziel des Vorhabens ist es, das Potential des iterativen Turbo-Fusionsverfahrens in der automatischen Spracherkennung weiter zu entfalten und offene Fragestellungen zu klären. So soll unter anderem durch quantitative Analysen verstanden werden, wie die Performanzgewinne der Turbo-Fusion entstehen und in welchen Anwendungen (z.B. akustische Sensornetzwerke und andere dezentralisierte Systeme) eine iterative Informationsfusion weitere Gewinne bringt. Ultimatives Ziel wäre die hyperparameterfreie Fusion von nichtbinären asynchronen Informationsquellen.

- Interested in doing research related to multi-channel wind noise reduction? We made the code available to [generate multichannel wind noise](#) samples with different wind properties (e.g., speed and direction). More information can be found in the accompanying paper "Multi-channel wind noise simulation based on the Corcos model" by Daniele Mirabilli and Emanuël Habets presented at IWAENC 2018.

## Datenbasen

- Der [Voice Assistant Conversation Corpus \(VACC\)](#) wurde an der Otto-von-Guericke Universität Magdeburg in Zusammenarbeit des [Lehrstuhls Kognitive Systeme](#) und der [Universitätsklinik für Psychosomatische Medizin und Psychotherapie](#) erstellt. Er ermöglicht Studien zur Mensch-Maschine-Interaktion unter definierten Randbedingungen: 1) Art des Dialoges (formal, informell), 2) Anzahl der Interaktionspartner (Mensch-Maschine vs. Mensch-Mensch-Maschine). Die Interaktion erfolgt hierbei mit dem kommerziellen Sprachassistenzsystem Amazon ALEXA. 27 deutsche Sprecher sollten zuerst Übungstermine bestimmen, wobei der Kalender des Probanden nur über ALEXA erfragt werden konnte. Anschließend sollten mithilfe von ALEXA Quizfragen gelöst werden. Beide Aufgaben werden einmal alleine und einmal zusammen mit einem Interaktionspartner durchgeführt. Neben 17h Audioaufnahmen enthält der Datensatz verschiedene Kontextinformationen (Sprecher-Turns, Off-Talk, Lachen, etc.), Transkriptionen sowie neben soziodemografischen Informationen auch Selbsteinschätzungen über die Interaktionen und Änderungen im Sprechverhalten während der Interaktion.

## Dissertationen

- Hendrik Meutzner: *Supporting Human-Machine Interaction by Robust Automatic Speech Recognition*, Ruhr-Universität Bochum (Betreuerin: D. Kolossa)



An der Schnittstelle zwischen Mensch und Maschine sind Nutzbarkeit und Zuverlässigkeit zentrale Kriterien. Die Arbeit von Herrn Meutzner untersucht, wie der Trade-Off zwischen Nutzbarkeit und Sicherheit bei akustischen CAPTCHAS besser gelingen und wie automatische Spracherkennung, u.a. durch [hybride audiovisuelle Modelle](#), robuster werden kann.

- Michael Stadtschnitzer: [Robust Speech Recognition for German and Dialectal Broadcast Programmes](#), Rheinische Friedrich-Wilhelms-Universität Bonn (Betreuer: C. Schmidt)

This thesis focuses on the development and the optimization of the German broadcast LVCSR system, which is part of the Fraunhofer IAIS audio mining system, over the course of several years, dealing with robustness-related problems that arise for German broadcast media and also dealing with the requirements for the employment of the ASR system in a productive audio mining system for the industrial use including stability, decoding time and memory consumption. We approach the following three problems: the continuous development and optimization of the German broadcast LVCSR system over a long period, rapidly finding the optimal ASR decoder parameters automatically and dealing with German dialects in the German broadcast LVCSR system. To guarantee superb performance over long periods of time, we regularly re-train the system using the latest algorithms and system architectures that became available by the research community, and evaluate the performance of the algorithms on German broadcast speech.



## Journalartikel

- W. Zhang, C. Hofmann, M. Bürger, T. D. Abhayapala and W. Kellermann: [Spatial Noise-Field Control With Online Secondary Path Modeling: A Wave-Domain Approach](#)  
This paper proposes a wave-domain adaptation algorithm for multichannel active noise control (ANC) with online secondary path modelling to cancel tonal

noise over an extended region of a two-dimensional plane in a reverberant room. The design is based on exploiting the diagonal-dominance property of the secondary path in the wave domain. The proposed wave-domain secondary path model is applicable to both concentric and nonconcentric circular loudspeakers and microphone array placement, and is also robust against array positioning errors. Through simulation-based verification in comparison with existing methods, the proposed algorithm demonstrates more efficient adaptation with low-level auxiliary noise.

- Z. Zhao, H. Liu and T. Fingscheidt:

#### [Convolutional Neural Networks to Enhance Coded Speech](#)

In this work we propose two postprocessing approaches applying convolutional neural networks (CNNs) either in the time domain or the cepstral domain to enhance speech that has been transmitted via a speech coder and decoder, without any modification of the codecs. The proposed approaches show significant quality improvements for several narrowband and wideband speech codecs in various conditions, including clean, tandeming, error-prone transmission and noisy conditions. For the narrowband G.711 codec, the quality could be improved to be even significantly better than uncoded, a surprising result, proving the power of CNN speech enhancement. The source code for the cepstral domain approach to enhance G.711-coded speech is made available in the [GitHub project](#).

- F.-R. Stöter, S. Chakrabarty, B. Edler and E.A.P. Habets: [CountNet: Estimating the number of concurrent speakers using supervised learning](#)

Estimating the maximum number of concurrent speakers from a single-channel mixture is an essential first step to address various audio-based tasks such as blind source separation, speaker diarization, and audio surveillance. In the paper, we first study how recent advances in deep learning may be exploited for the task of speaker count estimation. In particular, we show that convolutional recurrent neural networks outperform recurrent networks used in a previous study. Through comprehensive evaluation, we then compare the best-performing method to several baseline methods. Finally, the proposed method is compared to human performance, and insight into the strategy used by the proposed method is provided.

### Tagungen (nach Paper Deadline sortiert)

- [EUSIPCO](#) 02.-06.09.2019 in A Coruna, Spanien  
Paper Deadline: 18.02.2019 [CfP]
- [ESSV \(Elektronische Sprachsignalverarbeitung\)](#) 06.-08.03. 2019 in Dresden, [keine Einreichungen mehr]
- [DAGA](#) 18.-21.03.2019 in Rostock  
[keine Einreichungen mehr]
- [Interspeech](#) 15.-19.09.2019 in Graz, Österreich  
Paper Deadline: 29.03.2019 [CfP]
- [WASPAA](#) 20.-23.10.2019 in Mohonk Mountain House, New Paltz, NY, USA  
Paper Deadline 18.04.2019 [CfP]
- [ICASSP](#) 12.-17.05.2019 in Brighton, UK  
[keine Einreichungen mehr]
- [ASRU](#) 14.-18.12.2019 in Sentosa, Singapur  
Paper Deadline: 01.07.2019 [CfP]
- [ICASSP](#) 04.-09.05.2020 in Barcelona, Spain  
Paper Deadline: still unknown []

### Stellenangebote

- Die Forschungsgruppe „Sprach- und Audio-signalverarbeitung“ von Prof. Walter Kellermann am [Lehrstuhl für Multimediakommunikation und Signalverarbeitung](#) der Friedrich-Alexander Universität Erlangen-Nürnberg sucht zum nächstmöglichen Zeitpunkt eine(n) wissenschaftliche(n) Mitarbeiter(in) sowie eine(n) promovierte(n) Wissenschaftler(in) („Postdoc“). Bewerbungen mit den üblichen Unterlagen (Anschreiben, Lebenslauf, Zeugnisse etc.) sind bitte an [Walter Kellermann](#) zu schicken.