

Editorial

Liebe Leserinnen und Leser,

die Voice Message wird volljährig: Dies ist die 18. Ausgabe. Und da gibt es gleich eine wichtige Neuigkeit: Wir sind nun in der ITG der neue Fachausschuss „Sprachkommunikation“, siehe auch der Leitartikel unter „Latest News“. Ansonsten gibt es Preise, Preise, Preise... Sehen wir uns in Bamberg?

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

● „Es wächst zusammen, was zusammen gehört.“ Hat Willy Brandt mit diesem wegweisenden Satz aus zwei Zeitungsinterviews rund um den Mauerfall 1989 womöglich auch unsere beiden Fachausschüsse „Sprachakustik“ (AT3) und „Sprachverarbeitung“ (AT4) gemeint? Spaß beiseite. Auf Anregung von Prof. Rainer Martin (Ruhr-Universität Bochum) diskutierten die Fachausschüsse bereits im letzten Jahr diese Möglichkeit und die Sinnhaftigkeit der Zusammenlegung beider Fachausschüsse. Und regten einen entsprechenden ITG-Vorstandsbeschluss an, und zwar unter dem gemeinsamen Namen „Sprachkommunikation“. Der war schnell gefunden, veranstalten wir doch seit langer Zeit gemeinsam und mit gutem Erfolg die gleichnamige ITG Fachtagung. In der Vorstandssitzung am 24.11.2021 dann wurde es offiziell: Die beiden vormaligen Fachausschüsse AT3 und AT4 werden nun unter dem Namen „Sprachkommunikation“ zusammengelegt. Bei der Gelegenheit wurden die gemeinsamen Arbeitsfelder aktualisiert: Schauen Sie doch einmal auf der [Webseite des Fachbereichs Audiotechnik unter AT3](#). Ach ja, so viel Analogie zur deutschen Wiedervereinigung gab es dann doch: Das neue Ganze übernahm das Kürzel des alten Einen: AT3.

● Ein Team von Wissenschaftlern der Universität Oldenburg hat in der „[Clarity Enhancement Challenge](#)“, einem internationalen Wettbewerb zur Verbesserung der Sprachverständlichkeit im Störgeräusch für Hörgeräte, den ersten Preis erzielt. Der Schlüssel zum Erfolg lag im Einsatz eines binauralen Systems, das Signale an beiden Ohren verarbeitet. Störgeräusch und Nachhall wurden in einem zweistufigen System unter Einsatz von sowohl traditioneller mehrkanaliger Signalverarbeitung als

auch tiefen neuronalen Netzen reduziert. In Hörversuchen mit hörbeeinträchtigten Probanden erzielte die Lösung der Oldenburger Forscher die besten Ergebnisse. [[Paper](#), [Video](#)]

Persönliches

● Auf dem Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE



2021) wurde Stepan Shishkin mit dem Best Student Paper Award für seinen Beitrag „[Active Learning for Sound Event Classification using Monte-Carlo Dropout and PANN Embeddings](#)“ ausgezeichnet. In seinem Beitrag wird eine aktive Lernmethode zur Klassifizierung von Schallereignissen vorgeschlagen, um den menschlichen Arbeitsaufwand bei der Datenannotation zu reduzieren. Seit 2020 arbeitet Stepan Shishkin am Fraunhofer Institutsteil Hör-, Sprach- und Audiotechnologie in Oldenburg und promoviert zugleich an der Universität Oldenburg (Betreuer: S. Doclo).

● Jahn Heymann, ehemaliger Doktorand am Fachgebiet Nachrichtentechnik der Universität Paderborn, erhält den [VDE Promotionspreis 2022 des VDE Nordrhein-Westfalen](#). Mit diesem Preis wird jährlich eine herausragende Dissertation aus dem Gebiet der Elektro- und Informationstechnik an nordrhein-westfälischen Hochschulen ausgezeichnet. Jahn erhält den Preis für seine Arbeit mit dem Titel [Robust multi-channel speech recognition with neural network supported statistical beamforming](#). Die Verleihung erfolgt im Rahmen einer Festveranstaltung am Nachmittag des 17.11.2022 im HNI in Paderborn.

● Auf der Jahrestagung für Akustik im März 2022 hat Wiebke Middelberg (Universität Oldenburg, Betreuer: S. Doclo) für ihre Masterarbeit den Studienpreis der Deutschen Gesellschaft für Akustik erhalten. In ihrer Masterarbeit mit dem Titel „GSC-based noise and interferer reduction for binaural hearing aids exploiting external microphones“ erforschte Frau Middelberg die Möglichkeiten zur Verbesserung der Sprachverständlichkeit in Hörgeräten, indem sie externe Mikrophone zur Reduzierung von Störgeräuschen und konkurrierenden Sprechern mitnutzte.



Projekte und Aktivitäten

- Der Sonderforschungsbereich „[Hörakustik: Perzeptive Prinzipien, Algorithmen und Anwendungen](#)“ (Sprecher: Volker Hohmann) an der Universität Oldenburg erhält Förderung für weitere vier Jahre. Der SFB arbeitet an Hörgeräten und Hörassistenzsystemen, die sich mit Hilfe künstlicher Intelligenz an unterschiedliche Umgebungen anpassen und sich dabei immer besser auf die individuellen Nutzerinnen und Nutzer einstellen. Der SFB verbindet verschiedene Disziplinen, insbesondere Akustik, Psychoakustik, Audiologie, Ingenieurwissenschaften und physikalische Modellierung. An dem Großprojekt sind neben der Universität Oldenburg mit der Jade Hochschule, dem Fraunhofer Institut für Digitale Medientechnologie IDMT, dem Hörzentrum Oldenburg, der RWTH Aachen und der TU München weitere führende Einrichtungen aus der Hörforschung beteiligt.

Journalartikel

- S. Liebich and P. Vary
[Occlusion Effect Cancellation in Headphones and Hearing Devices – The Sister of Active Noise Cancellation](#)

The perception of one’s own voice influences the acceptance of hearing devices, such as headphones, headsets or hearing aids. When these devices fully or partially occlude the ear canal, the wearer’s own voice sounds boomy or like talking in a barrel. This is called occlusion effect. Occluding the ear canal results in an amplification of body-conducted sounds, mainly at low frequencies, and an attenuation of air-conducted sounds, predominantly at high frequencies, compared to the open ear. The contribution provides a comprehensive treatment of Occlusion Effect Cancellation (OEC) and its relation to Active Noise Cancellation (ANC) using digital signal processing. A novel effective filter structure is presented which offers some degree of adaptability and adjustability. Furthermore, digitally opening and closing the ear is evaluated by listening tests and objective measurements.

- Z. Xu, M. Strake, and T. Fingscheidt
[Deep Noise Suppression Maximizing Non-Differentiable PESQ Mediated by a Non-Intrusive PESQNet](#)

Speech enhancement employing deep neural networks (DNNs) for denoising is called deep noise suppression (DNS). DNS trained with mean squared error (MSE) losses cannot guarantee good perceptual quality. Perceptual evaluation of speech quality (PESQ) is a widely used metric for evaluating speech quality. However, the original PESQ algorithm is non-differentiable, therefore, cannot directly be used as

optimization criterion for gradient-based learning. In this work, we propose an end-to-end non-intrusive PESQNet DNN to estimate the PESQ scores of the enhanced speech signal. Thus, by providing a reference-free perceptual loss, it serves as a mediator towards the DNS training, allowing to maximize the PESQ score of the enhanced speech signal. Note that following this methodology, *a DNS can be optimized towards any metric given by any black-box software*. As further novelty, we propose to train the DNS and the PESQNet alternately to keep the PESQNet up-to-date and perform well specifically for the DNS under training. Our proposed method outperforms the baselines, comprising the same DNS trained with the MSE-based loss, the same DNS trained with an approximated differentiable PESQ loss, and the Interspeech 2021 DNS Challenge baseline.

- H. Schepker, F. Denk, B. Kollmeier, and S. Doclo
[Robust single- and multi-loudspeaker least-squares-based equalization for hearing devices](#)

To improve the sound quality of hearing devices, equalization filters can be used to achieve acoustic transparency, i.e., listening with the device in the ear is perceptually similar to the open ear. In this article, we propose a unified procedure to design single- and multi-loudspeaker equalization filters for hearing devices aimed at achieving acoustic transparency. To account for non-minimum phase components, we utilize a group delay compensation. To reduce comb-filtering artifacts, we propose a frequency-dependent regularization. Experimental results using measured acoustic transfer functions from a multi-loudspeaker earpiece show that the proposed equalization filter design procedure results in robust acoustic transparency and reduces the impact of comb-filtering artifacts. A comparison between single- and multi-loudspeaker equalization shows that for both cases a robust equalization performance can be achieved.

- A. Bohlender, L. Van Severen, J. Sterckx, N. Madhu
[DOA-guided source separation with direction-based initialization and time annotations using complex angular central Gaussian mixture models.](#)

Spatial clustering and time-frequency masking can be used to separate a mixture of multiple speakers and noise into the underlying signal components. Models such as the complex angular central Gaussian mixture model (cACGMM) allow a determination of the necessary separation parameters based on the given signal mixture. Thereby, no misfit between training and testing conditions arises – in contrast to approaches requiring labeled, training datasets. However, these models suffer from sensitivity to initialization and frequency permutation. These problems can be solved by incorporating additional a priori knowledge. Here we consider three techniques to overcome these limitations using direction of

arrival (DOA) estimates. First, we propose an initialization with simple DOA-based masks. Secondly, we derive speaker specific time annotations from the same masks to constrain the cACGMM. Thirdly, we employ an approach where the mixture components are specific to each DOA instead of each speaker. The results demonstrate that particularly the DOA-based initialization is effective to overcome both of the described limitations, and the performance is on par with using oracle information. Lastly, we also show that the proposed DOA-guided source separation works quite robustly in the presence of adverse conditions and realistic DOA estimation errors.

Tagungen (nach Paper Deadline sortiert)

- [SLT](#), 09.-12.01.2023, Doha, Qatar,
Paper Deadline: 15.07.2022 [[CfP](#)]
- [EUSIPCO](#), 29.08.-02.09.2022, Belgrad, Serbien
[keine Einreichungen mehr]
- [IWAENC 2022](#), 05.-08.09.2022, Bamberg,
Paper Deadline: 22.04.2022
[keine Einreichungen mehr, Anmeldung ab 01.07.2022]
- [Interspeech](#), 18.-22.09.2022, Incheon, Korea
[keine Einreichungen mehr]
- [ICASSP 2023](#), 04.-09.06.2023, Rhodos, Griechenland
Paper Deadline: 19.10.2022 [[CfP](#)]
- [DAGA 2023](#), 06.-09.03.2023, Hamburg,
Deadline für Abstracts: 01.11.2022 [[CfP](#)]
- [Interspeech 2023](#), 20.-24.08.2023, Dublin, Irland
Paper Deadline: Noch unbekannt
- [ITG Conference on Speech Communication 2023](#)
20.-22.09.2023 in Aachen
Paper Deadline: Noch unbekannt [[CfP](#)]

Stellenanzeigen

- Several full-time PhD positions in audio, speech and signal processing are available at the Internet and Data Sciences Lab, University of Gent, Belgium. Please contact [Prof. Dr.-Ing. Nilesh Madhu](#) for further details.