VDE ITG **Voice Message** **OKT 2022**

Newsletter des *neuen* ITG-Fachausschusses Sprachkommunikation (AT 3)

Editorial

Liebe Leserinnen und Leser,

Ein ITG und ein EURASIP-Preis, ein Sieg bei der DCASE Challenge, eine tolle IWAENC-Tagung in Bamberg, Artikel und Dissertationen... *Fünf* Seiten Voice Message, da machen wir keine langen Worte vorab!

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

● Kristina Tesch und Prof. Dr.-Ing. Timo Gerkmann erhalten den VDE ITG Preis 2022, welcher jährlich von der ITG für eine herausragende wissenschaftliche Veröffentlichung verliehen wird. Ausgezeichnet wird die Publikation „Nonlinear Spatial Filtering in Multichannel Speech Enhancement“ erschienen in den IEEE/ACM Transactions on Audio, Speech, and Language Processing im April 2021 [Paper und Audiobeispiele]. Die Arbeit zeigt, basierend auf analytischen Schätzern, das Potenzial von nichtlinearen räumlich-spektralen Filtern gegenüber der klassischen Kombination eines linearen Beamformers mit nachgeschaltetem Postfilter und bietet eine theoretische Grundlage für weitere Forschung zu DNN-basierten räumlichen-spektralen nichtlinearen Filtern.

● Vom 5.-8. September 2022 fand der *International Workshop on Acoustic Signal Enhancement* (IWAENC 2022) in Bamberg statt. Nach zweimaliger pandemiebedingter Verschiebung konnte mit ca. 200 Teilnehmer*innen vor Ort der traditionelle Charakter dieser seit 1989 bestehenden Konferenz gewahrt werden, so dass die meisten der Wissenschaftler*innen aus 20 Ländern und vier Kontinenten erstmals nach mehreren Jahren wieder den unmittelbaren persönlichen Austausch genießen konnten. Wissenschaftliche Höhepunkte stellten drei exzellente ‚Keynote lectures‘ dar: Gary Elko (mh acoustics, USA) präsentierte ‚*The evolution of microphone array beamformers*‘ von den ersten Systemen aus der Zeit des ersten Weltkriegs bis hin zu

den aktuellsten ‚Wearables‘. Shoko Araki (NTT, Japan) analysierte und illustrierte in ‘*Pushing the limits of speech enhancement technology*’ die jüngsten von ‚Deep Learning‘ geprägten Ansätze zur Sprachverbesserung für Dialogsysteme und schließlich zeigte Toon van Waterschoot mit ‘*Spatial acquisition, digital archiving, and interactive auralization of church acoustics*’ anhand der komplexen Akustik von Kirchen die Möglichkeiten und Grenzen der digitalen Erfassung komplexer akustischer Szenarios auf. Neben fast 90 vielbesuchten Posterpräsentationen wurden sechs ausgewählte Konferenzbeiträge in einer zentralen Vortragssitzung vorgestellt und Tobias Cord-Landwehr (Universität Paderborn) wurde für seinen Beitrag ‚*Monaural Source Separation: From Anechoic to Reverberant Environments*‘ mit dem ‚Best Student Paper Award‘ ausgezeichnet. Als weitere Kernkomponente des Workshops fanden auch die neun hochprofessionellen Demonstrationen überwiegend industrieller Aussteller großes Interesse. Nicht unerwähnt sollte auch ein gemeinsam von einer DFG-Forschergruppe und einem EU-ITN unmittelbar vor dem IWAENC 2022 veranstalteter ‚Satellite Workshop‘ zum Thema ‚*Signal Processing and Machine Learning for Spatially Distributed Microphones*‘ bleiben, der mit 130 registrierten Teilnehmern unerwartet große Resonanz fand (siehe auch nachfolgender Bericht).

Neben der technischen Exzellenz ist der IWAENC auch für sein traditionell besonders attraktives Beiprogramm bekannt. Nach pandemiebedingtem Ausfall des IWAENC 2020 wollten die Organisatoren Versäumtes nachholen und hohe Erwartungen erfüllen: Nach der von professionellen Jazzmusikern untermalten ‚Welcome reception‘ wurden Führungen durch Bamberg als Weltkulturerbe und als Zentrum des Bierbrauens von herrlich warmen Spätsommerabenden begünstigt. Den eindrucksvollen Höhepunkt bildete das Bankett im Schloss Weißenstein, mit Barockmusik begleitet von Künstler*innen der Nürnberger Hochschule für Musik.

● Am 05.09.2022 fand in Bamberg der Workshop "Signal Processing and Machine Learning for Spatially Distributed Microphones" statt. Er wurde gemeinsam organisiert von der DFG Forschungsgruppe FOR 2457 "Akustische Sensornetze" (ASN) und von dem Marie Sklodowska-Curie europäischen Trainingsnetzwerk "Service-Oriented, Ubiquitous, Network-Driven Sound" (SOUNDS). Mit 130 Teilnehmern aus aller Welt stieß der Workshop auf außerordentlich hohes Interesse. Durch diese gemeinsame Veranstaltung konnte der Staffelstab der Forschung von der sich dem Laufzeitende nähernden Forschungsgruppe an das erst kürzlich gestartete Trainingsnetzwerk übergeben werden.

Mitarbeiter der DFG Forschungsgruppe ASN

Persönliches

● Bei der DCASE 2022 Challenge Task 4 "Sound event detection in domestic environments" hat Janek Ebbers vom Fachgebiet Nachrichtentechnik der Universität Paderborn unter rund 100 eingereichten Systemen den ersten Platz belegt (siehe https://dcase.community/challenge2022/task-sound-event-detection-in-domestic-environments-results)! Die Aufgabe lautete, einen Klassifikator zu entwickeln, der typische Geräusche, die in einem Haushalt auftreten (z.B. Spülmaschine, Telefon, Staubsauger, etc.), erkennen kann. Eine besondere Herausforderung bestand darin, dass der überwiegende Teil der Trainingsdaten nicht annotiert war, d.h. keine Klassenlabels hatte, und der annotierte Teil der Trainingsdaten darüber hinaus nur schwach annotiert war. Damit ist gemeint, dass zwar die Ereignisklasse angegeben ist, nicht jedoch, wann das akustische Ereignis innerhalb der Aufnahme auftritt. Gleichwohl sollte das zu entwickelnde System neben der Ereignisklasse auch den Zeitstempel ausgeben, wann das Ereignis auftrat, wobei natürlich mehrere akustische Ereignisse gleichzeitig auftreten können.

● Ms. Maja Taseska, currently a Senior Researcher and Research Manager at Microsoft, received the 2022 EURASIP Best PhD Award for her Ph.D. thesis entitled "Informed Spatial Filters for Speech Enhancement: Noise and Interference Reduction, Blind Source Separation, and Acoustic Source Tracking". She conducted her research at the International Audio Laboratories Erlangen, a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer IIS under the supervision of Prof. Emanuël Habets.

● Seit dem 01.10.2022 leitet Prof. Dr.-Ing. Dorothea Kolossa das Fachgebiet „Elektronische Systeme der Medizintechnik" (MTEC) an der TU Berlin. Sie sucht dafür ab sofort auch zwei wissenschaftliche Mitarbeiter*innen für Lehre & Forschung in Sprach- und Audiotechnologien und multimodalen Lern–verfahren für die Medizintechnik. Begeisterung für die Forschung, hervorragende mathematische Fähigkeiten, solide Programmierkenntnisse und Elektronikaffinität sind dabei besonders wichtig – Prof. Kolossa freut sich über Bewerbungen!

Projekte und Aktivitäten

● Das Fraunhofer IAIS stellt im Rahmen der KI-NRW Initiative einen Sprachassistenten für die Lernplattform »Open Roberta®«! bereit. Mit dieser Lernplattform lassen sich Sprachdialoge über eine No-Code Programmieroberfläche mit einem visuellen Editor entwickeln. Mehr Informationen dazu sind unter https://showroom.ki.nrw/roberta-speaker/ zu finden. Der KI-NRW Showcase wurde im Rahmen des vom Bundesministerium für Wirtschaft und Klimaschutz (BMWK) geförderten SPEAKER-Projekts entwickelt.

● Forschende des Quality and Usability Lab der TU Berlin organisierten eine geteilte Aufgabe im Rahmen der GermEval 2022. Der Wettbewerb war als Textregressionsaufgabe konzipiert, bei der die Teilnehmenden Modelle zur Vorhersage der Komplexität eines deutschen Textes entwickelten. Von den 24 Teilnehmenden, die sich für den Wettbewerb angemeldet hatten, reichten zehn Teams ihre Ergebnisse ein, wobei sieben Teams die von den Organisatoren vorgegebene Baseline übertrafen. Weitere Einzelheiten zu dem Wettbewerb sind auf der Webseite zur Challenge zu finden.

● Demnächst startet am Fachgebiet Quality & Usability der TU Berlin das DFG-Project „Evaluierung von Konversationsqualität durch Crowdsourcing". In diesem Projekt sollen die Methoden der subjektiven Messung von Konversationsqualität für die Nutzung in Crowdsourcing angepasst werden. Dazu sollen jeweils zwei Crowdworker durch eine Voice-over-IP-Verbindung miteinander verschiedene Konversationsszenarien durchspielen und anschließend die wahrgenommene Qualität des Systems bewerten. Somit sollen die klassisch für das Labor entwickelten Konversationstests für die Nutzung in diversen Umgebungen über das Internet adaptiert werden.

Bücher, Dissertationen

● Ingo Siegert, Stefan Hillmann, Benjamin Weiss, Jessica M. Szczuka, and Alexey Karpov (eds.)
Towards Omnipresent and Smart Speech Assistants
The functionality of digital voice assistant systems is constantly growing. The attractiveness of such devices is based on their ease of use as they allow to conduct online searches and orders as well as smart home services by simply calling up the device.

However, the implications of voice-based interaction are not always clear to the user, especially since today's voice assistants are sometimes only better remote controls. In future, however, they should not only process simple commands, but also enable a natural and smooth interaction and be omnipresent. In addition to an improved speech recognition, this will require enhanced speech understanding and more intelligent dialog guidance.
While state-of-the art systems are mainly conceptualized for young adults and middle-aged people, future systems should adapt to the user in order to meet the needs of different (vulnerable) user groups ranging from young children to the elderly. This will be accompanied by efforts to make systems more understandable and users more sophisticated. Consequently, legal aspects resulting from the spread of voice assistants and the stricter data protection regulations are important.
The research topic organized by the editors covers 11 articles from 34 different authors from various research fields, including linguistics, psychology, usability/user experience studies as well as the technical perspective. One apparent focus of this research topic was on analyzing and assessing user experience. Both, different user groups and situations are taken into account. However, we hope to see the aforementioned perspective on more sophisticated dialogs represented in the near future.

● Andreas Brendel
From Blind to Semi-Blind Acoustic Source Separation based on Independent Component Analysis, LMS, Friedrich-Alexander Universität Erlangen-Nürnberg (Betreuer: W. Kellermann)
Typical acoustic scenes consist of multiple superimposed sources, where some of them represent desired signals, but often many of them are undesired sources, e.g., interferers or noise. In this thesis, the problem of acoustic source separation and extraction is treated by Convolutive Blind Source Separation (CBSS) approaches based on independent component analysis. Here, a special focus lies on the analysis of well-known CBSS methods, the analysis and derivation of fast converging update schemes and the development of a generic semi-blind source separation framework.

● Stefan Kühl
Adaptive Algorithms for the Identification of Time-Variant Acoustic Systems, Institut für Kommunikationssysteme der RWTH Aachen (Betreuer: P. Jax)
This thesis considers the different aspects of single and multi-channel system identification of time-variant acoustic systems for diverse scenarios. It compares measurement procedures from the field of acoustics and tracking algorithms from communication applications in a joint framework and contributes a novel proof of their mathematical equivalence for periodic excitation. Furthermore, it provides new insights into the relationship of different system identification algorithms like the NLMS algorithm, the RLS algorithm, and the Kalman filter. It proposes novel concepts and algorithms for specific applications, e.g., by considering available a priori information.

Journalartikel

● M. Karbasi, S. Zeiler, and D. Kolossa
Microscopic and Blind Prediction of Speech Intelligibility: Theory and Practice
Being able to estimate speech intelligibility without the need for listening tests would confer great benefits to a wide range of speech processing applications. Here, in addition to studying the theoretical limitations of the performance of SI prediction, a fully blind speech intelligibility predictor is introduced, based on internal discriminance measures of automatic speech recognition (ASR). It is shown that this novel, blind estimator can predict intelligibility as well as—and often even with better accuracy than—the well-studied non-blind ASR-based approaches, and that its results are in good agreement with its theoretically derived performance potential.

● K. Tesch, T. Gerkmann
Insights into Deep Non-linear Filters for Improved Multi-channel Speech Enhancement
In a traditional setting, linear spatial filtering (beamforming) and single-channel post-filtering are performed separately. While in our award winning prior work we showed by means of theoretical analyses that a non-linear joint spatial-spectral filter

has the potential to outperform the classic sequential beamforming + postfiltering, in this work we show that this also holds in practice if the non-linear joint spatial-spectral filter is realized by means of a neural network. By carefully analyzing the role of spatial, temporal and spectral information, we are able to achieve state-of-the-art results with a small neural network composed of only three layers [audio examples].

● T. Peer, T. Gerkmann
Phase-Aware Deep Speech Enhancement: It's All About The Frame Length
Algorithmic latency in speech processing is dominated by the frame length used for Fourier analysis, which in turn limits the achievable performance of magnitude-centric approaches. As previous studies suggest the importance of phase grows with decreasing frame length, this work presents a systematical study on the contribution of phase and magnitude in modern Deep Neural Network (DNN)-based speech enhancement at different frame lengths. Results indicate that DNNs can successfully estimate phase when using short frames, with similar or better overall performance compared to using longer frames. Thus, interestingly, modern phase-aware DNNs allow for low-latency speech enhancement at high quality.

● A. Herzog and E.A.P. Habets
Distance estimation in the spherical harmonic domain using the spherical wave model
Estimating the distance of a sound source using a compact spherical microphone array is a challenging task due to the small dimensions of the array. In this work, different spherical harmonic domain distance estimation methods are investigated. Two existing methods are discussed, and a new distance estimation method is proposed. The proposed estimator is related to the ratio between the real and imaginary components of the intensity vector for first-order spherical harmonic domain signals. In addition, a frequency averaging method is proposed and investigated for broadband distance estimation. The existing and proposed methods are evaluated using simulated and measured data.

● B. Brüggemeier and P. Lalone
Perceptions and reactions to conversational privacy initiated by a conversational user interface
In 2019, media reports raised awareness about privacy and security violations in Conversational User Interfaces (CUI) like Alexa, Siri, and Google. The European General Data Protection Regulation (GDPR) and several other laws across the globe give users the right to control the processing of their data, for example, by requesting the deletion of their data. Furthermore, GDPR and other laws advise for seamless communication of user rights, which, currently, is poorly implemented in CUI. We used a bespoke data collection interface to generate speaking chatbots and made them available as tasks on the crowd-sourcing platform Mechanical Turk. With those chatbots, we simulated how privacy can be communicated in a dialogue between user and machine. We find that conversational privacy can affect user perceptions of privacy and security positively. Moreover, user choices suggest that users are interested in obtaining information on their privacy and security in dialogue form. We discuss the implications and limitations of this research.

● B. Popp, P. Lalone, and A. Leschanowsky
Chatbot Language–crowdsource perceptions and reactions to dialogue systems to inform dialogue design decisions
Conversational User Interfaces (CUI) are widely used, with about 1.8 billion users worldwide in 2020. For designing and building CUI, dialogue designers have to decide on how the CUI communicates with users and what dialogue strategies to pursue (e.g., reactive vs. proactive). Dialogue strategies can be evaluated in user tests by comparing user perceptions and reactions to different dialogue strategies. Simulating CUI and running them online, for example, on crowdsourcing websites, is an attractive avenue to collecting user perceptions and reactions, as they can be gathered time- and cost-effectively. However, developing and deploying a CUI on a crowd-sourcing platform can be laborious and requires technical proficiency from researchers. We present Chatbot Language (CBL) as a framework to quickly develop and deploy CUI on crowd-sourcing platforms without requiring a technical background. CBL is a library with specialized CUI functionality, which is based on the high-level language JavaScript.

● J. Skowronek, A. Raake, G. Berndtsson, O.S. Rummukainen, P. Usai, S.N.B. Gunkel, M. Johanson, E.A.P. Habets, L. Malfait, D. Lindero and A. Toet
Quality of experience in telemeetings and videoconferencing: A comprehensive survey
Telemeetings such as audiovisual conferences or virtual meetings play an increasingly important role in our professional and private lives. For that reason, system developers and service providers strive for an optimal experience for the user while at the same time optimizing technical and financial resources. This leads to the discipline of Quality of Experience (QoE), an active field originating from the telecommunication and multimedia engineering domains that strives for understanding, measuring, and designing the quality experience with multimedia technology. This paper provides the reader with an entry point to the large and still growing field of QoE of telemeetings by taking a holistic perspective, considering both technical and non-technical aspects, and by focusing on current and near-future services.

Tagungen (nach Paper Deadline sortiert)

DAGA 2023, 06.-09.03.2023, Hamburg,
Deadline für Abstracts: 01.11.2022 [CfP]
ESSV 2023, 01.03.-03.03.2023, München,
Einreichungsfrist Kurzfassung: 04.12.2022
SLT, 09.-12.01.2023, Doha, Qatar,
[keine Einreichungen mehr]
EUSIPCO, 04.09.-08.09.2023, Helsinki, Finland
Paper Deadline: 20.02.2023 [CfP]
Interspeech, 20.-24.08.2023, Dublin, Irland
Paper Deadline: 01.03.2023 [CfP]
WASPAA 2023, date open, New Paltz, NY, USA,
Paper Deadline: noch unbekannt
ICASSP 2023, 04.-09.06.2023, Rhodos, Griechenland
[keine Einreichungen mehr]
ITG Conference on Speech Communication 2023
20.-22.09.2023 in Aachen
Paper Deadline: 26.05.2023 [CfP]

Stellenanzeigen

● Das Quality and Usability Lab der TU Berlin sucht Mitarbeiter/innen für 3 Stellen (TVL-13, zunächst 2–3 Jahre Befristung) mit verschiedenen Forschungsschwerpunkten, die zum 01.10. oder 01.11.2022 zu besetzen sind: Forschung und Entwicklung zu Chatbot-basierter Unterstützung im Studium, Qualitätsmessung und -vorhersage bei konversationeller Sprache sowie multimodale Sprachdialogsysteme zur der Interaktion mit Rehapatienten. [Kontakt]

● Die Abteilung NetMedia des Fraunhofer-Instituts 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.

● Das Institut für Nachrichtentechnik der TU Braunschweig sucht laufend wiss. Mitarbeiter*innen im Bereich der automatischen Spracherkennung und des Speech Enhancement. Bewerbungen mit den üblichen Unterlagen (Anschreiben, Lebenslauf, Zeugnisse etc.) sind bitte an Tim Fingscheidt zu schicken.