VDE ITG **Voice Message** **FEB 2024**

Newsletter des ITG-Fachausschusses Sprachkommunikation (AT 3), 23. Ausgabe

Editorial

Liebe Leserinnen und Leser, mit der nunmehr 23. Ausgabe darf ich mich als Editor der Voice Message verabschieden! Als wir im Oktober 2016 starteten, dreimal im Jahr diesen kleinen Rundbrief aufzulegen, hatte ich nicht mit einer so treuen und aktiven Community gerechnet: Vielen Dank an Sie alle für die tollen Beiträge über die vielen Jahre. Mein Dank gilt auch Reinhold Haeb-Umbach, waren wir doch lange Jahre zusammen in der Leitung des Fachausschusses tätig. Last but not least: Die Fachausschussleitung haben ja mittlerweile Rainer Martin unterstützt von Dorothea Kolossa übernommen und ich freue mich, dass Dorothea von nun an die Editorenrolle der Voice Message übernimmt. Meine neuen Aufgaben sind nun die Mitarbeit im ITG-Vorstand sowie die Leitung des Fachbereiches Audiotechnik (AT). In diesem Zusammenhang einen herzlichen Dank an unseren langjährigen treuen FB-Sprecher Reinhard Lerch!

Ihr Tim Fingscheidt

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, Awards

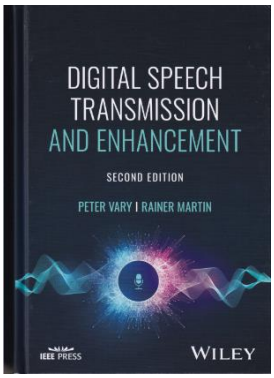
● Wie schon in der letzten Voice Message notiert, wurde im vergangenen Herbst in Aachen eine neue Fachausschussleitung gewählt. Die neuen Fachausschussleiter, Rainer Martin und Dorothea Kolossa, möchten an dieser Stelle einen kurzen Einblick in die laufende und zukünftige Arbeit geben. Nachdem sich die Voice Message als Plattform für den Austausch von Nachrichten innerhalb des Ausschusses in den letzten Jahren hervorragend etabliert hat, möchten wir in den kommenden Jahren verstärkt auch zur Außenwahrnehmung der ITG beitragen. Hierzu gehört die Positionierung der ITG im Hinblick auf technologisch wichtige Themen (hier haben wir mit *large language models* (LLMs) ein ganz heißes Eisen in der Schmiede!) aber auch die Unterstützung der Nachwuchsförderung und Nachwuchsrekrutierung. Auch hier bietet es sich an, über die Themen LLM und etwas allgemeiner, KI im Kontext von Alltagsanwendungen, die Sichtbarkeit der ITG für Schüler und Studenten zu stärken. Darüber hinaus bleibt in den kommenden Jahren natürlich auch die Fortführung und Organisation der Fachtagung "Sprachkommunikation" ein zentraler Punkt auf der Agenda des Fachausschusses. Wie bisher auch, freut sich das neue Leitungsteam über Anregungen und Mitwirkung aus dem Kreis der Mitglieder! Schließlich ist noch Folgendes zu berichten:

Prof. Dr.-Ing. Tim Fingscheidt (TU Braunschweig) ist zum VDE ITG Fellow ernannt worden. Er ist nach Peter Vary (2018) und Henning Puder (2021) nun das dritte Mitglied des Fachausschusses, dem diese besondere Ehre zuteil wird. Hiermit werden seine herausragenden Verdienste, sein Engagement und seine wissenschaftlichen Leistungen auf dem Gebiet der Sprachkommunikation zur robusten Sprachübertragung und Sprachverbesserung gewürdigt. Der Fachausschuss AT 3 gratuliert!

Rainer Martin und Dorothea Kolossa

Bücher, Dissertationen

● Peter Vary and Rainer Martin: Digital Speech Transmission and Enhancement, 2nd Edition, 2024 John Wiley & Sons, Ltd.



The second edition of "Digital Speech Transmission and Enhancement" has been thoroughly updated to encompass foundational principles and the latest advancements in the theory and practice of speech signal processing and its applications. While the first edition primarily focused on mobile communications, this revised edition gives increased attention to speech enhancement for hearing aids and human-machine interfaces, reflecting their growing importance over the past decade.

The book is divided into three main parts: fundamentals, speech coding, and speech enhancement. The fundamentals section covers models of speech production and hearing, spectral transformations, low-latency spectral analysis-synthesis, filter bank equalizers, stochastic signal processing, and estimation theory. The speech coding section explores quantization, differential waveform coding, and particularly the concepts of code excited linear prediction (CELP), presenting relevant speech codec standards such as the Adaptive Multi-Rate codec (AMR) and the Enhanced Voice Services codec (EVS) for cellular and IP communication. The speech enhancement section addresses topics including error concealment, bandwidth extension, near-end listening enhancement, noise and reverberation reduction in single and dual-channel scenarios, acoustic echo cancellation, and beamforming. Furthermore, recent machine learning techniques in speech signal processing are addressed.

● Kristina Tesch: Non-linear Spatial Filtering for Multi-channel Speech Enhancement and Separation, Universität Hamburg (Betreuer: Prof. Dr.-Ing. Timo Gerkmann)

Wenn Sprachsignale mit mehreren Mikrofonen aufgezeichnet werden, dann können für die Reduktion von Hintergrundgeräuschen und Störsprechern nicht nur zeitlich-spektrale Informationen herangezogen werden, sondern auch räumliche Informationen. Während traditionelle Ansätze hierfür ein lineares Filter mit einem separaten zeitlich-spektralen Postfilter kombinieren, prozessieren moderne tiefe neuronale Netze (DNNs) räumliche und zeitlich-spektrale Informationen gemeinsam mit einem nicht-linearen Verarbeitungsmodell.
Die Dissertation von Kristina Tesch untersucht die Vorteile eines solchen gemeinsamen nicht-linearen Verarbeitungsmodells aus einer statistischen Perspektive und demonstriert einen deutlichen Leistungsgewinn für nicht-gaußverteilte Störsignale. Um diesen Vorteil in praktischen Szenarien zu realisieren, werden DNN-basierte nicht-lineare räumliche Filter entwickelt und mit einem Steuerungsmechanismus kombiniert, der es erlaubt, die Extraktionsrichtung flexibel zu bestimmen. Die Forschung kulminiert in der Entwicklung einer Echtzeitdemonstration eines solchen nicht-linearen DNN-basierten Filters für die Sprecherextraktion (Video).

● Maximilian Kentgens: Signal Processing Concepts for User Movement in Scene-Based Spatial Audio, Institut für Kommunikationssysteme, RWTH Aachen, (Betreuer: Prof. Peter Jax, Prof. Boaz Rafaely)

Die Dissertation zielt auf künftige immersive Kommunikationssysteme ab, bei denen sich ein Benutzer virtuell in allen sechs rotatorischen und translatorischen Freiheitsgeraden ("6DoF") durch eine entfernte akustische Szene bewegen kann. Dahinter hinter steht die Vision einer Face-to-Face-Telefonie, bei der sich die Beteiligten perzeptiv am selben Ort befinden. Konkret wurden von Dr.-Ing. Maximilian Kentgens verschiedene adaptive und nicht-adaptive Signalverarbeitungskonzepte erarbeitet und evaluiert, um aus einem an einer einzigen Stelle im Raum aufgenommenen Higher-Order-Ambisonics-Signal plausible räumliche Tonsignale für andere Raumpositionen zu extrapolieren.

Journalartikel

● G. Enzner and S. Voit, Hybrid-Frequency-Resolution Adaptive Kalman Filter for Online Identification of Long Acoustic Responses with Low Input-Output Latency (HyKF)
Das Kalman Filter im Frequenzbereich [Enzner, Vary, 2006] (engl. *frequency-domain adaptive Kalman filter*, FDKF) ist ein modellbasiertes Verfahren für die adaptive akustische Echokompensation. Es verbindet die Funktion eines adaptiven Filters mit der einer optimalen Schrittweitensteuerung und wurde beliebt für seine Robustheit im Gegensprechfall. Neuerdings wurde das Verfahren oftmals verknüpft mit der datengetriebenen Optimierung für die akustische Echounterdrückung. Mit dieser Motivation stellt der oben genannte Artikel eine neue und akkuratere Herleitung sowie die resultierenden Modifikationen des Algorithmus vor. Der neue „HyKF“ bedient zwei praktisch wichtige Anforderungen: eine bedeutend schnellere Konvergenz zur adaptiven Lösung sowie die Unterstützung sehr langer Filter für die akustische Systemidentifikation. Diese Verbesserungen gelingen durch die Verwendung unterschiedlicher FFT Längen für das Eingangs- und Ausgangssignal des adaptiven Filters sowie durch interne Umrechnungen nach Bedarf der Zustandsvariablen im Kalman Filter.

● A. Chinaev, N. Knaepper, and G. Enzner
Online Distributed Waveform-Synchronization for Acoustic Sensor Networks with Dynamic Topology [code]
Acoustic sensing by multiple devices connected in a wireless acoustic sensor network (WASN) creates new opportunities for multichannel signal processing. However, the autonomy of agents in such a network still necessitates the alignment of sensor signals to a common sampling rate. It has been demonstrated that waveform-based estimation of sampling rate offset (SRO) between any node pair can be retrieved from asynchronous signals already exchanged in the network, but connected online operation for network-wide distributed sampling-time synchronization still presents an open research task. This is especially true if the WASN experiences topology changes due to failure or appearance of nodes or connections. In this work, we rely on an online waveform-based closed-loop SRO estimation and compensation unit for nodes pairs. For WASNs hierarchically organized as a directed minimum spanning tree (MST), it is then shown how local synchronization propagates network-wide from the root node to the leaves. Moreover, we propose a

network protocol for sustaining an existing network-wide synchronization in case of local topology changes. In doing so, the dynamic WASN maintains the MST topology after reorganization to support continued operation with minimum node distances. Experimental evaluation in a simulated apartment with several rooms proves the ability of our methods to reach and sustain accurate SRO estimation and compensation in dynamic WASNs.

● T. Kabzinski and P. Jax
A Flexible Framework for Expectation Maximization-Based MIMO System Identification for Time-Variant Linear Acoustic Systems
Quasi-continuous system identification of time-variant linear acoustic systems can be applied in various audio signal processing applications when numerous acoustic transfer functions must be measured. In this article, the underlying multiple-input-multiple-output (MIMO) system identification problem is treated in a state-space model as a joint estimation problem for states, representing impulse responses, and state-space model parameters using the expectation maximization (EM) algorithm. Limitations of prior work are addressed by imposing different model structures, especially for dependencies within a (transformed) state vector. This results in block diagonal matrix structures, for which M-step update rules are derived. Making assumptions about this model structure and choosing a block size for a given application define the computational complexity. In examples, improvements of up to 10 dB in relative system distance were found.

● K. Tesch, T. Gerkmann, "Multi-channel Speech Separation Using Spatially Selective Deep Non-linear Filters" [arxiv] [audio] [video]
To enhance the spatial processing in a multi-channel source separation scenario, in this work, we propose a deep neural network (DNN) based spatially selective filter (SSF) that can be spatially steered to extract the speaker of interest by initializing a recurrent neural network layer with the target direction. We compare the proposed SSF with a common end-to-end direct separation (DS) approach trained using utterance-wise permutation invariant training (PIT), which only implicitly learns to perform spatial filtering. We show that the SSF has a clear advantage over a DS approach with the same underlying network architecture when there are more than two speakers in the mixture, which can be attributed to a better use of the spatial information. Furthermore, we find that the SSF generalizes much better to additional noise sources that were not seen during training and to scenarios with speakers positioned at a similar angle.

● S. Thaleiser and G. Enzner, "Binaural-Projection Multichannel Wiener Filter (BP-MWF) for Cue-Preserving Binaural Speech Enhancement" [audio]
Methoden der binauralen Sprachsignalverbesserung werden oft in räumliche Optimalfilter mit Nebenbedingung zum Erhalt der Richtungsinformation und in sog. „common-gain filter“ (zur gleichen spektralen Filterung links und rechts) unterschieden. Dieser Artikel definiert für die räumliche Optimalfilterung eine komplexwertige strikte Nebenbedingung für den Erhalt frequenzbezogener räumlicher Hörmerkmale, d.h. interauraler Pegel- und Zeitdifferenzen (ILD und ITD), und weist in diesem Falle eine neue Lösung in der Klasse der „common-gain filter“ nach. Diese kann zugleich auch als mehrkanaliges Wienerfilter (MWF) mit binauraler Projektion (BP) in Richtung der verrauschten Sprache interpretiert werden, woraus der Name „binaural-projection MWF“ (BP-MWF) hervorgeht. Mit neuartiger metrischer und optischer Auswertung wird gezeigt, dass die räumlichen Hörmerkmale durch die BP-MWF Bearbeitung wie vorgesehen erhalten bleiben. Geräuschfilterung und Sprachqualität gemessen in segmentellem SNR, PESQ und STOI werden gegenüber Vergleichsverfahren verbessert und ein formeller Hörversuch bestätigt diese instrumentelle Bewertung.

Tagungen (nach Paper Deadline sortiert)

Interspeech, 01.-05.09.2024, Kos, Griechenland
Paper Deadline: 02.03.2024 [CfP]
EUSIPCO, 26.-30.08.2024, Lyon, Frankreich
Paper Deadline: 03.03.2024 [CfP]
DAGA, 18.-21.03.2024, Hannover
[Keine Anmeldung von Beiträgen mehr]
ICASSP, 14.-19.04.2024, Seoul, Korea
[keine Einreichungen mehr]
IWAENC, 09.-12.09.2024, Aalborg, Dänemark
Paper Deadline: 01.05.2024 [CfP]
ICASSP, 06.04.-11.04.2025, Hyderabad, Indien
Paper Deadline: Noch offen
SLT, 09.-12.12.2024, Macau, China
Paper Deadline: Noch offen

Stellenanzeigen

● Am Institut für Nachrichtentechnik der TU Braunschweig wird eine wiss. Mitarbeiterin/ein wiss. Mitarbeiter (TVL-13, 100%) gesucht im Bereich Speech Enhancement mittels hocheffizienter neuronaler Netze. Weitere Infos hier.