



## Editorial

Liebe Leserinnen und Leser, auf der wieder einmal hervorragend gelungenen ITG Fachtagung Speech Communication (diesmal an der RWTH Aachen – mit gleich zwei Best Paper Awards für die Gastgeber!) tagte auch der Fachausschuss Sprachkommunikation: Lesen Sie zum neu gewählten Sprecherteam unter „Latest News“. Unsere Arbeitsfelder wurden geschärft, wie man der [Webseite](#) entnehmen kann: „Modellierung, Signalverarbeitung und maschinelles Lernen für“ insgesamt 13 Themengebiete. Außerdem: Zwei Artikel mit Code und vieles mehr ...

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

- In der Fachausschusssitzung vom 20.09.2023 an der RWTH Aachen wurde ein neues Leitungsteam des Fachausschusses Sprachkommunikation gewählt. Als neuer Leiter wurde Prof. Dr.-Ing. Rainer Martin von der Ruhr-Universität Bochum gewählt, im Team mit seiner Stellvertreterin Prof. Dr.-Ing. Dorothea Kolossa von der TU Berlin. Der Fachausschuss wünscht beiden gutes Gelingen in ihrer neuen Rolle und dankt den langjährigen bisherigen Leitern Prof. Dr. Reinhold Haeb-Umbach (Universität Paderborn) und Prof. Dr.-Ing. Tim Fingscheidt (TU Braunschweig) für ihre hervorragende und nachhaltige Arbeit.

- Vom 20. bis 22. September 2023 fand die ITG Conference on Speech Communication an der RWTH Aachen statt. Die Konferenz wurde von der

Informationstechnischen Gesellschaft in Zusammenarbeit mit dem Fachausschuss Sprachkommunikation AT3, dem technischen Programmkomitee und dem [Institut für Kommunikationssysteme](#) (IKS) unter der Leitung von Prof. Dr. Peter Jax organisiert.

Mit 120 Teilnehmer\*innen (siehe Gruppenfoto) stieß die Konferenz auf eine erfreulich hohe Resonanz. Über drei Tage hinweg wurde intensiv über Theorien, Algorithmen und Anwendungen in den Bereichen Sprach- und Audioverarbeitung diskutiert. Das bewährte Format aus abwechselnden Vorträgen und Poster-Präsentationen, das den Workshop-Charakter früherer Konferenzen beibehält, ermöglichte einen regen wissenschaftlichen Austausch. Insgesamt wurden 55 Beiträge präsentiert und diskutiert, darunter sechs ausgewählte Konferenzbeiträge, die in einer zentralen Vortragsitzung vorgestellt wurden, sowie eine strukturierte Sitzung zum Thema „Generative Methods for Speech Processing“.

Das Programm wurde schließlich durch drei exzellente eingeladene Übersichtsvorträge abgerundet. Dr. Hendrik Witt (TeamViewer) berichtete zum Thema „Leap into the Industrial Metaverse: Where Wearables and Sensors Meet Reality“, Prof. Dr. Nilesh Madhu (Ghent University) illustrierte auf anschauliche Weise „A Communication Love Story Through the Ages“ und Prof. Dr. Frank Kurth (Fraunhofer FKIE) referierte über „Information Extraction from Speech and Audio Sensor Data – Challenges for Applied Research in the Security Domain“.

Zwei punktgleiche Beiträge wurden mit dem „ITG - Best Paper Award“ ausgezeichnet. Till Hardenbicker (IKS, RWTH Aachen) erhielt den Best Paper Award für seinen Beitrag zum Thema „Fast Tracking of Time-Variant Systems Using Local Affine Subspaces“, und Florian Hilgemann (IKS, RWTH Aachen) wurde für sein Paper „Design of Low-order IIR Filters Based on Hankel Nuclear Norm Regularization for Achieving Acoustic Transparency“ ausgezeichnet.

Die nächste Tagung in dieser Reihe findet im September 2025 in Berlin statt.

- Andreas Brendel erhält den [VDE Bayern Award 2023](#) für seine Dissertation „From Blind to Semi-Blind Acoustic Source Separation based on Independent Component Analysis“, die am [Lehrstuhl für Multimediakommunikation und Signalverarbeitung](#) der FAU Erlangen-Nürnberg unter der Betreuung von Prof. Walter Kellermann 2022 abgeschlossen wurde.
- Am 29.09.2023 fand das 75. (!) Treffen der ITG-Fachgruppe „Signalverarbeitung und maschinelles Lernen“ am Arbeitsbereich [Signal Processing \(SP\)](#) von Timo Gerkmann mit ca. 35 Teilnehmenden an der Universität Hamburg statt. Neben sieben Vorträgen aus dem Arbeitsbereich SP bot das [Programm](#) auch auch fünf spannende externe Beiträge aus der Fachgruppe.

#### Journalartikel

- J.-M. Lemerrier, J. Richter, S. Welker, T. Gerkmann, "[StoRM: A Diffusion-based Stochastic Regeneration Model for Speech Enhancement and Dereverberation](#)" [[arxiv](#)][[code](#)][[audio](#)]  
Diffusion models are a promising novel approach to generative speech restoration which receives a great amount of attention world-wide. We have shown that they may outperform their predictive counterparts for non-additive corruption types or when they are evaluated on mismatched conditions. However, diffusion models suffer from a high computational burden and may generate wrong phonemes in low SNRs. By first applying a predictive approach and then a generative approach, we show that these negative effects can be reduced quite effectively while also mending the speech distortions introduced by predictive approaches.

- Z. Xu, Z. Zhao, and T. Fingscheidt  
[Coded Speech Quality Measurement by a Non-Intrusive PESQ-DNN](#) [[code](#)]  
Wideband codecs such as AMR-WB or EVS are widely used in (mobile) speech communication. Evaluation of coded speech quality is often performed subjectively by an absolute category rating (ACR) listening test. However, the ACR test is impractical for online monitoring of speech communication networks. Perceptual evaluation of speech quality (PESQ) is one of the widely used metrics instrumentally predicting the results of an ACR test. However, the PESQ algorithm requires an original reference signal, which is usually unavailable in network monitoring, thus limiting its applicability. NISQA is a new non-intrusive neural-network-based speech quality measure, focusing on super-wideband speech signals. In this work, however, we aim at predicting the well-known PESQ metric using a non-

intrusive PESQ-DNN model. We illustrate the potential of this model by predicting the PESQ scores of wideband-coded speech obtained from AMR-WB or EVS codecs operating at different bitrates in noisy, tandeming, and error-prone transmission conditions. We compare our methods with the state-of-the-art network topologies of QualityNet, WaweNet, and DNSMOS — all applied to PESQ prediction — by measuring the mean absolute error (MAE) and the linear correlation coefficient (LCC). The proposed PESQ-DNN offers the best total MAE and LCC of 0.11 and 0.92, respectively, in conditions without frame loss, and still is best when including frame loss. Note that our model could be similarly used to non-intrusively predict POLQA or other (intrusive) metrics.

#### Tagungen (nach Paper Deadline sortiert)

- [DAGA](#), 18.-21.03.2024, Hannover  
[Anmeldung von Beiträgen](#) bis 01.11.2023
- [ASRU](#), 16.-20.12.2023, Taipei, Taiwan  
[keine Einreichungen mehr]
- [Interspeech](#), 01.-05.09.2024, Jerusalem, Israel  
Paper Deadline: noch offen
- [EUSIPCO](#), 26.-30.08.2024, Lyon, Frankreich  
Paper Deadline: 03.03.2024 [[CfP](#)]
- [ICASSP](#), 14.-19.04.2024, Seoul, Korea  
[keine Einreichungen mehr]
- [IWAENC](#), 09.-12.09.2024, Aalborg, Dänemark  
Paper Deadline: 01.05.2024 [[CfP](#)]

#### 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 in Hörgeräten. [Weitere Infos hier.](#)