VDE ITG **Voice Message** **FEB 2022**

Newsletter der ITG-Fachausschüsse Sprachakustik (AT 3) und Sprachverarbeitung (AT 4)

Editorial

Liebe Leserinnen und Leser,

in dieser Voice Message stehen ganz offensichtlich Erlanger Wissenschaftler*innen im Mittelpunkt: Sie gewinnen Preise, schreiben Unmengen an Journalartikeln und last but not least: Sie organisieren den IWAENC 2022 ganz bewusst wieder in Präsenz. Wir freuen uns alle auf ein persönliches Wiedersehen – allerdings nicht in Erlangen, sondern am Veranstaltungsort in Bamberg. Lesen Sie selbst.

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

● Nach zweimaliger Verschiebung aufgrund der anhaltenden Corona-Pandemie findet der International Workshop on Acoustic Signal Enhancement (IWAENC) vom 05. bis 08. September 2022 in Bamberg statt! Das dreitägige Programm umfasst Keynote-Talks, Poster- und Vortrags-sitzungen, aber auch Demo-Sessions. IWAENC 2022 soll in Präsenzform stattfinden, um dem einzigartigen Charakter dieses Workshops auch weiterhin gerecht zu werden, auf dem sich seit 1989 Forscherinnen und und Forscher aus aller Welt austauschen und vernetzen. Eine virtuelle Teilnahme wird ermöglicht, falls aufgrund von Reisebeschränkungen oder Reisewarnungen eine Teilnahme vor Ort nicht möglich ist. Weitere Informationen zur Einreichung von Beiträgen sind auf www.iwaenc2022.org zu finden.

● Unmittelbar vor Beginn des IWAENC 2022 findet am 05.09.2022 von 13:00h – 17:00h ein gemeinsamer Workshop der DFG-Forschungsgruppe *Acoustic Sensor Networks (ASN)* sowie des Marie Skłodowska-Curie Actions European Training Networks *Service-Oriented, Ubiquitous, Network-Driven Sound (SOUNDS)* statt. Er steht unter dem Titel „Signal Processing and Machine Learning for Spatially Distributed Microphones“. Neben einer Keynote von Tuomas Virtanen wird es interessante Vorträge, Poster, und vor allem Demonstrationen zu obigem Thema geben. Näheres wird in Kürze auf den Webseiten der IWAENC zu lesen sein, auch die kostenlose Anmeldung kann hier vorgenommen werden. Liebe Kolleginnen und Kollegen, verpasst also nicht diesen spannenden Workshop unmittelbar vor der Welcome Reception der IWAENC!

Persönliches

● Mohamed Elminshawi from the International Audio Laboratories Erlangen (Prof. Dr. Emanuël Habets) received the VDE Bayern Award 2021. In his Master Thesis he focused on target speaker extraction, which is the task of isolating the speech signal of a desired speaker from a mixture of audio sources, such as other interfering speakers or background noise, with the help of a pre-recorded speech reference signal from the same person. He showed that this approach can be extended to extract non-speech target sources as well, e.g., piano or guitar. He also proposed a new technique to exploit the temporal dynamics of the reference signal by modifying the architecture of the neural network.

Journalartikel

● T. Robotham, O.S. Rummukainen, M. Kurz, M. Eckert and E.A.P. Habets
Comparing direct and indirect methods of audio quality evaluation in virtual reality scenes of varying complexity
As many quality evaluation methods are used to assess uni-modal audio or video content without considering perceptual, cognitive, and interactive aspects present in virtual reality (VR) settings, little is known regarding the repercussions of the employed evaluation method, content, and subject behavior on the quality ratings in VR. The conducted study uses four subjective audio quality evaluation methods (viz. multiple-stimulus with and without reference for direct scaling, and rank-order elimination and pairwise comparison for indirect scaling) to investigate the contributing factors present in multi-modal 6-DoF VR on quality ratings of real-time audio rendering. Our results show all methods that do not employ a reference produce similar results. However, rank-order elimination proved to be the fastest method, required the least amount of repetitive motion, and yielded the highest discrimination between spatial conditions. Scene complexity was found to be a main effect within results, with behavioral and task load index results implying more complex scenes and interactive aspects of 6-DoF VR can impede quality judgments.

● D. Mirabilii and E.A.P. Habets
Wind speed and direction estimation based on the spatial coherence of closely spaced microphones
Wind-induced noise recorded with a compact microphone array can be exploited to infer the mean velocity of a free-field airflow. In this work, a model-based method to estimate the wind flow speed and direction is proposed that uses spectro-spatial correlations of closely spaced microphone signals. The accuracy of the proposed method is investigated

across a range of wind speed between 0.5 and 12 m/s and all directions, using observation lengths from 5 seconds to 1 hour.

● Z. Xu, S. Elshamy, Z. Zhao, T. Fingscheidt
Components loss for neural networks in mask-based speech enhancement
Estimating time-frequency domain masks for single-channel speech enhancement using deep learning methods is a vital field of research. In this article, we propose a novel *components loss* (CL) for the training of neural networks for mask-based speech enhancement. During the training process, the proposed CL offers separate control over preservation of the speech component quality, suppression of the noise component, and preservation of a naturally sounding residual noise component. The new CL is compared to several baseline losses, comprising the conventional mean squared error (MSE) loss w.r.t. speech spectral amplitudes or w.r.t. an ideal-ratio mask, auditory-related loss functions, such as the perceptual evaluation of speech quality (PESQ) loss and the perceptual weighting filter loss, and also the recently proposed SNR loss with two masks. Detailed analysis suggests that the proposed CL obtains a better or at least a more balanced performance across all employed instrumental quality metrics, including SNR improvement, speech component quality, enhanced total speech quality, and particularly also delivers a natural sounding residual noise component. For unseen noise types, we excel even perceptually motivated losses by an about 0.2 points higher PESQ score. The recently proposed so-called SNR loss with two masks not only requires a network with more parameters due to the two decoder heads, but also falls behind on PESQ and POLQA and particularly w.r.t. residual noise quality. Note that the proposed CL shows significantly more 1st ranks among the evaluation metrics than any other baseline. It is easy to implement, and code is provided at https://github.com/ifnspaml/Components-Loss.

● A. Ilic Mezza, E.A.P. Habets, M. Mueller and A. Sarti
Unsupervised domain adaptation via principal subspace projection for acoustic scene classification
Existing acoustic scene classification (ASC) systems often fail to generalize across different recording devices. In this work, we present an unsupervised domain adaptation method for ASC based on data standardization and feature projection. Using the TUT Urban Acoustic Scenes 2018 Mobile Development dataset, we show that the proposed method can provide an absolute increment of over 10% compared to state-of-the-art unsupervised adaptation methods. Furthermore, the proposed method consistently outperforms a recent ASC model that ranked first in Task 1-A of the 2021 DCASE Challenge when evaluated on various unseen devices from the TAU Urban Acoustic Scenes 2020 Mobile Development dataset.

● A. Prinn, A. Walther and E.A.P. Habets
Estimation of locally reacting surface impedance at modal frequencies using an eigenvalue approximation technique
The accuracy of computational acoustic models is often limited by a lack of reliable information concerning the frequency-dependent impedance of surface materials. This lack of information stems from the unavailability of reliable measurement methods for low frequencies. In this work, an approach is proposed, using eigenvalue analysis, for estimating the locally reacting, frequency-dependent impedance of a sound-absorbing sample. It is shown, using finite element simulations of an impedance tube and a small reverberation room, that the proposed method can provide reasonable estimates of the surface impedance of a sample placed on a boundary surface.

Tagungen (nach Paper Deadline sortiert)

DAGA 2022, 21.-24.03.2022, Stuttgart,
[keine Einreichungen mehr]
Interspeech, 18.-22.09.2022, Incheon, Korea
Paper Deadline: 21.03.2022 [CfP]
IWAENC 2022, 05.-08.09.2022, Bamberg,
Paper Deadline: 22.04.2022 [CfP]
<Der IWAENC findet in Präsenz statt!>
ICASSP 2022, 22.-27.05.2022, Singapur
[keine Einreichungen mehr]
SLT, 09.-12.01.2023, Doha, Qatar,
Paper Deadline: Juli 2022 [CfP]
EUSIPCO, 29.-02.09.2022, Belgrad, Serbien
[keine Einreichungen mehr]
DAGA 2023, 06.-09.03.2023, Hamburg,
Paper Deadline: [noch offen]
ICASSP 2023, 04.-09.06.2023, Kos, Griechenland
Paper Deadline: [noch offen]
ITG Conference on Speech Communication 2023
20.-22.09.2023 in Aachen

Stellenanzeigen

● Am Institut für Kommunikationsakustik in der Fakultät für Elektrotechnik und Informationstechnik der Ruhr-Universität Bochum ist die Stelle eines **Akademischen Rats** / einer **Akademischen Rätin (A13)** mit der Option einer Verbeamtung auf Lebenszeit zu besetzen. Eine sehr gute Promotion sowie einschlägige Erfahrungen im Bereich der Sprach- und Audiosignalverarbeitung und des maschinellen Lernens werden vorausgesetzt. [Kontakt]

● Das Institut für Nachrichtentechnik der TU Braunschweig sucht eine(n) **wissenschaftliche(n) Mitarbeiter*in** (TV-L E13) für ein Forschungsprojekt im Themenfeld Speech Enhancement. [Kontakt]