Friedrich-Alexander-Universität
Department of Electrical Engineering

EEI-Colloquium

LipVoicer: Generating Speech from Silent Videos Guided by Lip-Reading

Prof. Sharon Gannot
Bar-Ilan University, Israel

Chair: Prof. Dr.-Ing. Walter Kellermann

Tuesday, 26.3.2024, 12:00 a.m.
Live: Lecture room 05.025 (Cauerstr. 7, 5th floor, 91058 Erlangen)

Abstract: Lip-to-speech involves generating a natural-sounding speech synchronized with a soundless video of a person talking. Despite recent advances, current methods still cannot produce high-quality speech with high levels of intelligibility for challenging and realistic datasets. This talk presents LipVoicer, a novel method that generates high-quality speech, even for in-the-wild and rich datasets, by incorporating the text modality. Given a silent video, we first predict the spoken text using a pre-trained lip-reading network. We then condition a diffusion model on the video and use the extracted text through a classifier-guidance mechanism where a pre-trained automatic speech recognition (ASR) serves as the classifier. We demonstrate the effectiveness of LipVoicer through human evaluation, which shows that it produces more natural and synchronized speech signals than competing methods (demo page: https://lipvoicer.github.io). The presented LipVoicer is a joint work of Yochai Yemini, Aviv Shamsian, Lior Bracha, Sharon Gannot, and Ethan Fetaya.

Short Bio: Sharon Gannot obtained his Ph.D. in electrical engineering from Tel-Aviv University, Israel, in 2000. He is a full professor in the Faculty of Engineering at Bar-Ilan University, Israel. He serves as a Senior Area Chair for IEEE Transactions on Audio, Speech, and Language Processing, member of the Senior Editorial Board of IEEE Signal Processing Magazine, and Chair of the IEEE Signal Processing Society Data Science Initiative. Previously, he served as Chair of the IEEE Audio and Acoustic Signal Processing Technical Committee in 2017–2018. He has also held other roles, such as General Co-chair of the 2010 International Workshop on Acoustic Signal Enhancement and the 2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. He will serve as General Co-chair of Interspeech 2024.
Sharon Gannot received the 2022 European Association for Signal Processing Group Technical Achievement Award. His research interests include statistical signal processing and machine learning, with applications to speech enhancement, noise reduction, speaker separation and diarization, dereverberation, and speaker localization and tracking utilizing single-microphone and microphone arrays, as well as audio-visual processing. Sharon Gannot is a Fellow of the IEEE.

Lecture series by the Department of Electrical Engineering, Electronics and Information Technology