Beitrag ID: **31** Beitragskennung: **119** Typ: **Poster**

Diffusion Models for Audio-Visual Speech Enhancement

This poster showcases a selection of our work on diffusion models for speech enhancement. While diffusion models have proven successful in natural image generation, we adopt them for speech enhancement by introducing a task-adopted diffusion process in the complex short-time Fourier domain. Our results show competitive performance compared to strong predictive methods, while generalization is better when evaluated in a mismatched training scenario. However, for very challenging input, the model tends to produce speech-like sounds without semantic meaning. To address this problem, we condition the diffusion model on visual input with the speaker's lips, resulting in improved speech quality and intelligibility. This improvement is reflected in a reduced word error rate of a downstream automatic speech recognition model.

Find me @ my poster

Keywords

Diffusion models
Speech Enhancement
Audio-Visual
Generative Models

TentID

Autor: RICHTER, Julius

Co-Autor: Herr GERKMANN, Timo