Skysong Innovations
Advancing the Arizona State University Knowledge Enterprise

1475 N. Scottsdale Road, Suite 200
Scottsdale, AZ 85287-3538
Phone: 480 884 1996 Fax: 480 884 1984

Case ID:M21-217L

Published: 5/3/2022

Inventors

Jianwei Zhang

Visar Berisha

Suren Jayasuriya

Contact

Jovan Heusser
jovan.heusser@skysonginnovations.com

Model to Restore Degraded Speech

Removing interferences and improving quality of degraded speech via speech enhancement (SE) is important in many applications such as telecommunications, speech recognition, voice over IP, hearing aids, and more. One major class of SE techniques includes machine learning, which is effective at enhancing speech quality, however, they require complex network structures with a large number of parameters.

Vocoders can generate high quality speech waveforms based on an input conditioner (e.g. a mel-spectrum). One such vocoder, DiffWave is diffusion model-based and provides state-of-the-art synthesized speech quality, a relatively short waveform generation time and a small number of parameters. However, DiffWave has been primarily used for generative modeling tasks such as unsupervised speech generation, where the data distribution of audio is learned by the model.

Researchers at Arizona State University have created a neural network architecture based on a modification of the DiffWave model, that restores the original speech signal. Mel-spectrum, the input conditioner, is replaced with a deep CNN upsampler, which is trained to alter the degraded speech mel-spectrum to match that of the original speech. While the model is trained using the original speech waveform, it is conditioned on the degraded speech mel-spectrum, resulting in improved speech quality.

This neural network architecture replaces the mel-spectrum up sampler in DiffWave with a deep CNN resulting in improved speech quality, less training time and overall better performance.

-

Potential Applications

- Speech enhancement of degraded speech
  - Mobile phones
  - VoIP

- Teleconferencing
- Hearing aids
- Speech recognition

Benefits and Advantages

- Improved quality of speech degraded by LPC-10 compression, AMR-NB compression and signal clipping
- Compared to the original DiffWave architecture, this scheme achieves better performance on several objective perceptual metrics and in subjective comparisons
- Requires less training time and less parameters to synthesize high-quality audio
- Can revert the deterministic transformation
- Tested on 128 randomly selected samples of degraded speech

For more information about this opportunity, please see

Zhang et al - arXiv - 2021

For more information about the inventor(s) and their research, please see

Dr. Berisha's departmental webpage

Dr. Jayasuriya's departmental webpage