Thesis: Speech Enhancement Using Speech Synthesis Techniques

Abstract: Traditional speech enhancement systems reduce noise by modifying the noisy signal to make it more like clean signal, which suffers from two problems: under-suppression of noise and over-suppression of speech. These problems create distortions in enhanced speech and hurt the quality of the enhanced signal. We propose to utilize speech synthesis techniques for a higher quality speech enhancement system. Synthesizing clean speech based on the noisy signal could produce outputs that are both noise-free and high quality. We first show that we can replace the noisy speech with its clean resynthesis from a previously recorded clean speech dictionary from the same speaker (concatenative resynthesis). Next, we show that using a speech synthesizer (vocoder) we can create a ``clean'' resynthesis of the noisy speech for more than one speaker. We term this parametric resynthesis (PR). PR can generate better prosody from noisy speech than a TTS system which uses textual information only. Additionally, we can use the high-quality speech generation capability of neural vocoders for better quality speech enhancement. When trained on data from enough speakers, these vocoders can generate speech from unseen speakers, both male, and female, with similar quality as seen speakers in training. Finally, we show that using neural vocoders we can achieve better objective signal and overall quality than the state-of-the-art speech enhancement systems and better subjective quality than an oracle mask-based system.

Committee:

- Professor Michael Mande, Mentor,Brooklyn College
- Professor Kyle Gorman, The Graduate Center
- Professor Rivka Levitan, Brooklyn College

Outside Member:

- Ron Weiss, Google Brain