Research Week 2020

using conditional adversarial networks for intelligibility improvement for dysarthric speech and laryngectomees

Tuan Dinh, Ph.D., Alexander Kain

dintu@ohsu.edu

OHSU

Keywords

conditional adversarial nets, speech intelligibility

Abstract

We explored voice conversion systems to improve speech intelligibility of 1) dysarthric speech and 2) laryngectomees.

In the first case, we explore the potential of conditional generative adversarial networks (cGANs) to learn the mapping from habitual speech to clear speech. We evaluated the performance of cGANs in three tasks: 1) speaker-dependent one-to-one mappings, 2) speaker-independent many-to-one mappings, and 3) speaker-independent many-to-many mappings. In the first task, cGANs outperformed a traditional deep learning (DNN) mapping in term of average keyword recall accuracy and the number of speakers with improved intelligibility. In the second task, we showed that without clear speech, we can significantly improve intelligibility of the habitual speech of one of three speakers. In the third task which is the most challenging one, we improved the keyword recall accuracy for two of three speakers.

In the second case, we aim to improve speech of laryngectomees in term of intelligibility and naturalness. We predict the voicing and voicing degree for laryngectomees from speech spectra using a deep neural network. We use a logarithmically falling synthetic F0 for statement phrases. Spectra are converted to synthetic target spectra using a cGAN.