Rui-Chen Zheng | 郑瑞晨

Phone: (+86)181-8926-6050 | **Email:** zhengruichen@mail.ustc.edu.cn

Education

M.S. Student in Signal and Information Processing *2021.09 - Now*
University of Science and Technology of China *Hefei, Anhui, China*
- Supervised by Prof. Zhen-Hua Ling
- GPA: 3.96/4.3

Bachelor of Electronic Information Engineering *2017.09 – 2021.06*
University of Science and Technology of China *Hefei, Anhui, China*
- Thesis: Method and Practice on Text-to-speech Without Text
- GPA: 3.89/4.3
- Minor in Business Administration

Research Experience

Speech Reconstruction from Silent Lip and Tongue Articulation by Pseudo Target Generation and Domain Adversarial Training
- Supervised by Prof. Zhen-Hua Ling
- We study the task of speech reconstruction from ultrasound tongue images and optical lip videos recorded in a silent speaking mode, where people only activate their intra-oral and extra-oral articulators without producing sound. We propose to employ a method built on pseudo target generation and domain adversarial training with an iterative training strategy to improve the intelligibility and naturalness of the speech recovered from silent tongue and lip articulation. To be specific, pseudo targets were first generated for silent articulations in order to enable the same supervised training paradigm as when vocalized articulations were used as input. Besides, by adding a domain discriminator behind the encoder, the model is able to learn common features existed in silent and vocalized articulation. Iterative training strategy was conducted to obtain better pseudo targets. Experiments show that our proposed method significantly improves the intelligibility and naturalness of the reconstructed speech in both silent and vocalized speaking mode compared to the baseline.
- Demo-Page: https://github.com/ZhengRachel/ImprovedTaLNet-demo
- Accepted by ICASSP2023.

Incorporating Ultrasound Tongue Images for Audio-Visual Speech Enhancement Through Knowledge Distillation.
- Supervised by Prof. Zhen-Hua Ling
- Audio-visual speech enhancement (AV-SE) aims to enhance the degarded speech with extra visual information assistance, of which lip videos are the most commonly used, and has been shown to be more effective than audio-only speech enhancement. This paper proposes incorporating ultrasound tongue images to improve the performance of AV-SE systems further. To address the challenge of acquiring ultrasound tongue images during inference, a knowledge distillation method is proposed during the training stage, enabling an audio-lip speech enhancement student model to learn from a pre-trained audio-lip-tongue speech enhancement teacher model. Experimental results demonstrate significant improvements in the quality and intelligibility of the speech enhanced by the proposed method compared to the traditional audio-lip speech enhancement baselines. We also evaluate the enhanced speech's phoneme error rate (PER). The proposed method can reduce the PER of several tongue-related phonemes when applying automatic speech recognition (ASR) to enhanced speech.
- Submitted to INTERSPEECH2023.

Selected Honors

- Honor Rank for Top 5% Graduates of USTC. *2021.06*
- Huawei Scholarship 2020.12
- USTC Outstanding Student Scholarship, Gold Award *2019.12 & 2018.12*
- Top-Notch Program Funding *2019.12 & 2018.12*

Skills

Language – Chinese/English
- TOEFL iBT: 105
- CET-6 607

Programming Language
- Python/Matlab/C

Other Expericence

Admitted to UCLA CSST Summer Internship Program *2020.07*