QIBING BAI

(+86)132-7888-2787 ⋄ 12032871@mail.sustech.edu.cn

EDUCATION

Southern University of Science and Technology (SUSTech)   *09/2020 – 06/2023*
Master student, Department of Computer Science and Engineering
Supervisor: Yu Zhang, Tom Ko
Research Interests: Speech Synthesis, Speech Pre-training

Central South University (CSU)   *09/2016 – 06/2020*
BEng, Electronic Engineering
GPA: 89.6/100   Ranking: 6/62

RESEARCH EXPERIENCE

Model Compression for Speech Pre-training   *12/2021 – 3/2022*

Motivation: Current compression methods for pre-trained speech models either greatly harm the model performance, or cannot meet various computational constraints.

Method: We choose HuBERT as the teacher model, and take its hidden representations to train a student. We propose a two-stage training strategy: 1) pre-training distillation to adapt the model, and 2) once-for-all training to obtain sub-networks of various sizes. The resulted model of stage.1 outperforms the HuBERT teacher. A sub-network with 27M parameters outperforms DistilHuBERT.

Modeling of Rising Intonation in Cantonese TTS   *06/2021 – 10/2021*

Motivation: Rising intonation is key to distinguishing declarative questions (e.g. You are going to school?) and statements in Cantonese. Limited by insufficient training data and weak capacity, neural TTS models usually fail to render such sentences rising intonation.

Method: We add declarative questions to the dataset, and utilize BERT to provide necessary syntactic and semantic information for Tacotron2. Several implicit and explicit modeling approaches are investigated. Subjective and objective evaluations show the effectiveness of the explicit modeling method.

WORK EXPERIENCE

Direct Speech-to-Speech Translation (Internship at ByteDance)   *03/2022 – 09/2022*

Motivation: There is still a performance gap between end-to-end S2ST models and cascade systems. We are interested in improving translation quality and speech intelligibility for E2E S2ST.

Result: We synthesize target audios for GigaST (a pseudo En-Zh ST dataset). We train Transformer-based models with various strategies. The best E2E model achieves a BLEU score of 36.9, only 0.4 behind a cascade model (ST + TTS). The model is deployed for internal experience on Lark.

PUBLICATIONS

Qibing Bai, Tom Ko, Yu Zhang, *A Study of Modeling Rising Intonation in Cantonese Neural Speech Synthesis*, Interspeech 2022

Rui Wang*, **Qibing Bai***, et al. *LightHuBERT: Lightweight and Configurable Speech Representation Learning with Once-for-All Hidden-Unit BERT*, Interspeech 2022

Qianqian Dong, Fengpeng Yue, Tom Ko, Mingxuan Wang, **Qibing Bai** and Yu Zhang, *Leveraging Pseudo-labeled Data to Improve Direct Speech-to-Speech Translation*, Interspeech 2022

Fangfang Zhou, **Bing Bai**, et al. *FuzzyRadar: Visual Analysis for Understanding Fuzzy Clusters*, Journal of Visualization, 2019

HONORS & REWARDS

Graduation with Distinction, Central South University	*2020*
The HollySys Scholarship, Central South University	*2019*
The First Prize Scholarship, Central South University	*2019*
Outstanding Student, Central South University	*2018*

SKILLS & OTHERS

Toolkits & Programming	fairseq, ESPnet, PyTorch, Python, C/C++
Languages	Mandarin (native), English (TOFEL: 103), French (beginner)