QIBING BAI

+86 132-7888-2787 ◇ 12032871@mail.sustech.edu.cn ◇ https://p1ping.github.io

EDUCATION

The Chinese University of Hong Kong - Shenzhen *08/2023 – Present*
Ph.D. in Computer Science *Shenzhen, China*
Supervisor: Haizhou Li
Research Interests: Speech Synthesis, Speech Representation Learning, Speech Translation

Southern University of Science and Technology *09/2020 – 06/2023*
Master of Engineering *Shenzhen, China*
Supervisor: Yu Zhang, Jimmy Liu, Tom Ko

Central South University *09/2016 – 06/2020*
Bachelor's in Electronic Engineering *Changsha, China*
Overall GPA: 90/100 Ranking: 6/62

RESEARCH EXPERIENCE

Augmentation-Aware Speech Representation Learning *06/2022 – 10/2022*

Data augmentation techniques, such as utterance mixing, can enhance the capacity of SSL models on some speech tasks but may compromise their generalization on tasks requiring augmentation-related information. In this project, we proposed an augmentation-aware learning framework that aims to balance this trade-off by guiding the model to learn augmentation-related information. We designed an augmentation-aware objective that predicts the strength of augmenting signals. Based on HuBERT, our method showed improved performance on several downstream tasks.

Lightweight and Configurable Speech Representation Learning *12/2021 – 03/2022*

A speech pre-training framework is proposed to meet various computational constraints. We used HuBERT as the teacher model and distilled its hidden representations to train a student model. A once-for-all method is devised to generate sub-networks of different sizes from the student model. The student model surpassed the HuBERT teacher in performance. A sub-network with 27M parameters achieved better results than DistilHuBERT.

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

We aimed to improve the rendering of rising intonation in declarative questions (e.g. You are going to school?) for Cantonese TTS. we constructed a dataset with declarative questions and used BERT to provide syntactic and semantic information for Tacotron2. Several implicit and explicit modeling approaches are explored and evaluated subjectively and objectively. All modeling approaches surpassed the baseline. The explicit method was more effective and flexible than the implicit ones.

WORK EXPERIENCE

Direct Speech-to-Speech Translation (Internship at ByteDance) *02/2022 – 03/2023*

- Objective: To bridge the performance gap between end-to-end speech-to-speech translation (S2ST) models and cascade models in terms of translation quality, speech intelligibility, and timbre preservation.
- Outcome: We synthesized target audios for GigaST (a pseudo En-Zh ST dataset) and trained Transformer-based models with various strategies. The best end-to-end model achieved a BLEU score of 36.9, only slightly lower than a cascade model (ST + TTS). The model was deployed for internal testing on Lark. We also experimented with back-translation to train the model with realistic speaker-parallel data.

PUBLICATIONS

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

Rui Wang*, **Qibing Bai***, Junyi Ao, Long Zhou, Zhixiang Xiong, Zhihua Wei, Yu Zhang, Tom Ko, and Haizhou Li. *LightHuBERT: Lightweight and Configurable Speech Representation Learning with Once-for-All Hidden-Unit BERT*, Proc. 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*, Proc. Interspeech, 2022

Fangfang Zhou, **Bing Bai**, Yitao Wu, Minghui Chen, Zengsheng Zhong, Rongchen Zhu, Yi Chen, Ying Zhao, *FuzzyRadar: Visual Analysis for Understanding Fuzzy Clusters*, Journal of Visualization, 2019

HONORS & REWARDS

Excellent Teaching Assistant, Southern University of Science and Technology *2022*
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

Programming & Toolkits Python, C/C++, PyTorch, fairseq, ESPnet
Languages Mandarin (native), English (C1, TOEFL 103), French (≈A1)

SOCIETY MEMBERSHIPS

Student Member, International Speech Communication Association (ISCA) *2022 – present*

EXTRACURRICULAR ACTIVITIES

Class President, Central South University *2019 – 2020*
Volunteer for Language Day, Southern University of Science and Technology *2021*