Tingle Li

Research Assistant, Tsinghua University

(86)166-7516-4723
tingle.li@outlook.com
https://github.com/Tinglok
https://tinglok.netlify.com

Education

Tsinghua University Beijing, China
Post-bachelor at Institute for Interdisciplinary Information Sciences *Jul. 2020 - Present*
- Research-oriented program during COVID-19, also affiliated to Shanghai Qi Zhi Institute.
- **Advisor:** Prof. *Hang Zhao*

Tiangong University Tianjin, China
B.Eng. in Computer Science and Technology *Sep. 2016 - Jun. 2020*
- Pilot class of Artificial Intelligence (an elite program for top 10% students)
- **GPA:** 3.74/4.0 **Ranking:** 3/124
- **Advisor:** Prof. *Ming Li* (Duke University)
- **Selected Courses:** Machine Learning (97), Mathematical Foundation of Artificial Intelligence (98), Data Mining (91), Computer Vision (91), Numerical Analysis (94), Applied Statistics (91), Linear Algebra (93), Probability and Statistics (93), Advanced Mathematics (94), Data Structure (94), Algorithm Design and Analysis (93), High-level Programming Language (95)

Research Interests

Speech Processing: Source Separation, Voice Conversion, Music Information Retrieval
Multi-modal Learning: Audio-visual Learning, Self-supervised Learning

Publications

Chenzhuang Du*, **Tingle Li***, Yichen Liu*, Zixin Wen, Tianyu Hua, Yue Wang, Hang Zhao. "Improving Multi-Modal Learning with Uni-Modal Teachers". Submitted to NeurIPS 2021.

Chenxu Hu, Qiao Tian, **Tingle Li**, Yuxuan Wang, Hang Zhao. "Neural Dubber: Dubbing for Silent Videos According to Scripts". Submitted to NeurIPS 2021.

Tingle Li*, Yichen Liu*, Chenxu Hu*, Hang Zhao. "CVC: Contrastive Learning for Non-parallel Voice Conversion". *In Proc. INTERSPEECH*, Brno, Czechia, August 2021.

Tingle Li, Jiawei Chen, Haowen Hou, Ming Li. "Sams-Net: A Sliced Attention-based Neural Network for Music Source Separation". *In Proc. ISCSLP*, Hong Kong, China, January 2021. (**Oral Presentation**)

Qingjian Lin*, **Tingle Li***, Lin Yang, Junjie Wang, Ming Li. "Optimal Mapping Loss: A Faster Loss for End-to-End Speaker Diarization". *In Proc. Speaker Odyssey*, Tokyo, Japan, November 2020.

Tingle Li, Qingjian Lin, Yuanyuan Bao, Ming Li. "Atss-Net: Target Speaker Separation via Attention-based Neural Network". *In Proc. INTERSPEECH*, Shanghai, China, October 2020.

Qingjian Lin, **Tingle Li**, Ming Li. "The DKU Speech Activity Detection and Speaker Identification Systems for Fearless Steps Challenge Phase-02". *In Proc. INTERSPEECH*, Shanghai, China, October 2020.

Awards & Achievements

Best Undergraduate Dissertation (**top 1%**) *Jun. 2020*
Presidential Scholarship for Outstanding Students (**top 5%**) *Dec. 2017, 2018, 2019*
1st for SID and **3rd** for SAD among 50 teams, Fearless Steps Challenge Phase-02 *May. 2020*
1st Prize, China Students Innovation and Entrepreneurship Competition *May. 2019*
3rd Prize, Lan Qiao Cup National Selection Competition *May. 2017*

Research Experience

Electrical Engineering and Computer Science — University of Michigan
Advisor : Prof. *Andrew Owens* — *Apr. 2021 - Present*

Material Converter: Converting Materials of Visual Objects with Audio

- Given a source video and target material audio as input, this task aims to convert the source video to a new video that corresponds to the target material audio.
- Explored an audio-visual GAN model containing two distinct training objectives for converting texture and preserving structure, which is able to generate consistent videos given distinguishable material audio.

Institute for Interdisciplinary Information Sciences (IIIS) — Tsinghua University
Advisor : Prof. *Hang Zhao* — *Jul. 2020 - Present*

CVC: Contrastive Learning for Non-parallel Voice Conversion

- Given a speech signal from source speakers, this task aims to convert it to the timbre of the target speaker while preserving the speech content.
- Explored a voice conversion model based on noise contrastive estimation (infoNCE) loss, which enables one-way conversion in the non-parallel voice conversion setting, while comparatively improving speech intelligibility and effectively reducing training time.

Neural Dubber: Dubbing for Silent Videos According to Scripts

- Introduced a novel task called silent video dubbing. Given the video script and the silent video as input, this task aims to synthesize speech that is temporally synchronized with the video.
- Proposed Neural Dubber, a multi-modal text-to-speech (TTS) model, which can generate lip-synced mel-spectrograms in parallel.
- Experimental results show that Neural Dubber is on par with FastSpeech 2, and even outperforms FastSpeech 2 with image-based speaker embedding, in terms of speech quality.

Speech and Multimodal Intelligent Information Processing Lab — Duke Kunshan University
Advisor : Prof. *Ming Li* — *Jul. 2019 – Jul. 2020*

Atss-Net: Target Speaker Separation via Attention-based Neural Network

- Given a referenced utterance of the target speaker, and a mixed utterance containing the target speaker, this task aims at filtering the target speaker's voice from the mixed utterance.
- Proposed a target speaker separation model based on attention neural network, which leveraged the attention mechanism to fuse the mixed spectrogram and the target speaker embedding.
- Collaborated with Xiaomi Corporation to land in application, where demos are available here.

Sams-Net: A Sliced Attention-based Neural Network for Music Source Separation

- Given a musical utterance, our goal is to recover the individual stems from the music (i.e. vocals, drums, bass and others).
- Proposed a new attention mechanism called Sliced Attention, where the scope of attention is narrowed down to the intra-chunk features that are most likely to affect each other.
- Our model has achieved the state-of-the-art performance, although it contained fewer parameters compared with baselines, where demos are available here.