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
Research Assistant 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: top 3%**
- **Advisor:** Prof. *Ming Li* (Duke Kunshan University)

Research Interests

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

Publications & Preprints

Tingle Li, Yichen Liu, Andrew Owens, Hang Zhao. "Listen for Texture: Learning Visual Styles from Audio-Visual Associations". Submitted to *CVPR* 2022.

Chenzhuang Du, Jiaye Teng, **Tingle Li**, Yichen Liu, Yue Wang, Yang Yuan, Hang Zhao. "Modality Laziness: Everybody's Business is Nobody's Business". Submitted to *ICLR* 2022.

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

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

Tingle Li, Jiawei Chen, Haowen Hou, Ming Li. "Sams-Net: A Sliced Attention-based Neural Network for Music Source Separation". *In Proc. ISCSLP*, 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. Odyssey*, 2020.

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

Awards & Achievements

ISCA Travel Grant Award, Interspeech 2021 *Jun. 2021*
Best Undergraduate Dissertation (**top 1%**) *Jun. 2020*
1st for SID & **3rd** for SAD, Fearless Steps Challenge Phase-2, Interspeech 2020 *May. 2020*
Presidential Scholarship for Outstanding Students (**top 5%**) *Dec. 2017, 2018, 2019*
1st Prize, China Students Innovation and Entrepreneurship Competition (**top 8%**) *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*

Listen for Texture: Learning Visual Styles from Audio-Visual Associations

- Proposed a novel task of *audio-driven image stylization*, which aims to convert the texture of a visual scene conditioned on sound (demo video & paper).
- Presented a contrastive-based audio-visual GAN model that learns from unlabeled data to manipulate visual scene texture via audio.
- Curated a dataset that contains egocentric hiking videos from the internet.
- Evaluated our proposed method on the dataset to show that it can generate plausible images that match the content of the audio.

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

CVC: Contrastive Learning for Non-parallel Voice Conversion

- Explored a voice conversion model based on noise contrastive estimation loss, which enables one-way conversion in the non-parallel voice conversion setting (project page).
- Conducted experiments to show that our proposed method simplifies the training process, reduces training time, and improves speech intelligibility.

Neural Dubber: Dubbing for Videos According to Scripts

- Introduced a novel task called *automatic video dubbing*, which aims to synthesize speech temporally synchronized with the video given scripts (project page).
- Proposed an audio-visual text-to-speech model that can generate high-quality, lip-synced speech.
- Performed experiments to show that our model is on par with baselines, and even outperforms them with image-based speaker embedding, in terms of speech quality.

Modality Laziness: Everybody's Business is Nobody's Business

- Discussed an optimization problem in multi-modal fusion methods, where encoders from multi-modal training suffer from learning insufficient representations of each modality (paper).
- Proved a theorem that with multi-modal data as inputs, the fusion model is prone to saturate since it can see powerful features from different modalities and ignores the features that are hard to learn but still important.
- Proposed a dichotomy method to mitigate the issue on various multi-modal datasets.

Data Science Research Center — Duke Kunshan University
Advisor : Prof. *Ming Li* — *Jul. 2019 – Jul. 2020*

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

- Proposed a method that leverages the attention mechanism to incorporate the mixture spectrogram and target speaker embedding (project page).
- Performed experiments to show that it generalizes well to practical situations.
- Collaborated with Xiaomi Corporation to explore the possibility of landing in applications.

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

- Proposed a method in which the scope of attention is narrowed down to the intra-chunk features that are most likely to affect each other (project page).
- Conducted experiments to show that it performs better with fewer parameters than baselines.

Optimal Mapping Loss: A Faster Loss for End-to-End Speaker Diarization

- Developed a novel optimal mapping loss between the output and ground-truth speaker sequences through the Hungarian algorithm (paper).
- Conducted experiments to verify that it can reduce time complexity while retaining performance.