

Jiatong Shi

6403 Gates and Hillman Centers, 4902 Forbes Ave, Pittsburgh, PA 15213

✉ jiatongs@andrew.cmu.edu | 🏠 shijt.site | 📷 ftshijt

Education

Carnegie Mellon University

PHD IN LANGUAGE TECHNOLOGIES

- Advisor: Dr. Shinji Watanabe

Pittsburgh, PA, U.S.A.

2021.08 - present

The Johns Hopkins University

MS COMPUTER SCIENCE

- Advisor: Dr. Shinji Watanabe

Baltimore, MD, U.S.A.

2019.08 - 2021.05

Renmin University of China

BS COMPUTER SCIENCE

- Second Major in Fintech
- Undergrad research advisor: Dr. Qin Jin, Dr. Wei Xu, and Dr. Wei Du

Beijing, China

2015.09 - 2019.06

Research Experience

Carnegie Mellon University - Language Technologies Institute

ADVISOR: DR. SHINJI WATANABE

- Focus1: speech-to-speech translation
- Focus2: low-resource, multilingual speech recognition
- Focus3: music processing (singing voice synthesis, automatic song writing)

Pittsburgh, PA, U.S.A.

2021 - present

Frederick Jelinek Memorial Summer Workshop (JSALT, JHU Summer Workshop)

ADVISOR: DR. ANN LEE, DR. SHINJI WATANABE, AND DR. HUNG-YI LEE

- Focus: unsupervised ASR and its applications

Baltimore, MD, U.S.A.

2022 Summer

The Johns Hopkins University - Center for Language and Speech Processing (CLSP)

ADVISOR: DR. SHINJI WATANABE

- Focus: low-resource, multilingual, robust speech recognition and applications in endangered language documentation

Baltimore, MD, U.S.A.

2019 - 2021

Renmin University of China - Multi-level Multi-aspect Multimedia Analysis (AI.M3) Lab

ADVISOR: DR. QIN JIN

- Thesis: "Computer assisted language learning for young English learners"
- Focus: computer assisted language learning (CALL) and music processing

Beijing, China

2018-2019

Renmin University of China - Lab for Information System Applications (LISA)

CO-ADVISORS: DR. WEI XU AND DR. WEI DU

- Thesis: "Multimedia Analytics for Crowdfunding Prediction"
- Focus: online data analytic and fin-tech applications

Beijing China

2017-2019

Professional Experience

Tencent AI Lab

SUPERVISOR: DR. CHUNLEI ZHANG AND DR. DONG YU

- Project: Towards streaming speaker diarization systems in noisy meeting scenarios.

Seattle, WA, U.S.A.

May. 2022 - Aug. 2022

IBM Watson Research Center

SUPERVISOR: DR. GEORGE SAON

- Project: Lattice generation, confusion network processing, and confidence measure estimation for RNN transducer

Yorktown Heights, NY, U.S.A.

May. 2021 - Aug. 2021

Tencent AI Lab

CO-SUPERVISORS: DR. CHUNLEI ZHANG, DR. CHAO WENG, AND DR. DONG YU

Seattle, WA, U.S.A.

May. 2020 - Dec. 2020

- Project 1: the joint-framework for target speaker extraction and speech recognition
- Project 2: end-to-end speaker diarization with generalized neural speaker clustering

Tencent Ethereal Audio Lab

SUPERVISOR: WEI (DENNIS) XIAO

Shenzhen, Guangdong, China

Jun. 2019 – Aug. 2019

- Project 1: psycho-acoustic loss design for speech super-resolution
- Project 2: real-time speech super-resolution system for Tencent Meeting (i.e., Voov Meeting)

NetEase Youdao - AI Application

SUPERVISOR: YIXUAN XIAO

Beijing, China

May. 2018 - Nov. 2018

- Project: a prototype system for computer assisted language learning

Publications

* equal contribution; + mentored students;

PUBLISHED

27. Yen Meng, Hsuan-Jui Chen, **Jiatong Shi**, Shinji Watanabe, Paola Garcia, Hung-yi Lee, and Hao Tang. 2022. "On Compressing Sequences for Self-Supervised Speech Models" (accepted by SLT 2022)
26. Tzu-hsun Feng, Annie Dong, Ching-Feng Yeh, Shu-wen Yang, Tzu-Quan Lin, **Jiatong Shi**, Kai-Wei Chang, Zili Huang, Haibin Wu, Xuankai Chang, Shinji Watanabe, Abdel-rahman Hohamed, Shang-Wen Li, Hung-yi Lee. 2022. "SUPERB @ SLT 2022: Challenge on Generalization and Efficiency of Self-Supervised Speech Representation Learning" (accepted by SLT 2022)
25. **Jiatong Shi**, George Saon, David Haws, Shinji Watanabe and Brian Kingsbury. 2022. "VQ-T: RNN Transducers using Vector-Quantized Prediction Network States" *Proceedings of Interspeech 2022*, pp. 1656-1660.
24. **Jiatong Shi**, Shuai Guo⁺, Tao Qian⁺, Nan Huo, Tomoki Hayashi, Yuning Wu, Fangzheng Xu, Xuankai Chang, Huazhe Li, Peter Wu, Shinji Watanabe and Qin Jin. 2022. "Muskits: an End-to-end Music Processing Toolkit for Singing Voice Synthesis" *Proceedings of Interspeech 2022*, pp.4277-4281.
23. Shuai Guo⁺, **Jiatong Shi**^{*}, Tao Qian⁺, Shinji Watanabe and Qin Jin. 2022. "SingAug: Data Augmentation for Singing Voice Synthesis with Cycle-consistent Training Strategy" *Proceedings of Interspeech 2022*, pp.4272-4276.
22. Dan Berrebbi⁺, **Jiatong Shi**, Brian Yan, Osbel López-Francisco, Jonathan Amith and Shinji Watanabe. 2022. "Combining Spectral and Self-Supervised Features for Low Resource Speech Recognition and Translation" *Proceedings of Interspeech 2022*, pp.3533-3537.
21. Keqi Deng, Shinji Watanabe, **Jiatong Shi** and Siddhant Arora. 2022. "Blockwise Streaming Transformer for Spoken Language Understanding and Simultaneous Speech Translation" *Proceedings of Interspeech 2022*, pp.1746-1750.
20. Yui Sudo, Shakeel Muhammad, Kazuhiro Nakadai, **Jiatong Shi** and Shinji Watanabe. 2022. "Streaming Automatic Speech Recognition with Re-blocking Processing Based on Integrated Voice Activity Detection" *Proceedings of Interspeech 2022*, pp.4641-4645.
19. Brian Yan, Patrick Fernandes, Siddharth Dalmia, **Jiatong Shi**, Yifan Peng, Dan Berrebbi, Xinyi Wang, Graham Neubig, and Shinji Watanabe. 2022. "CMU's IWSLT 2022 Dialect Speech Translation System", *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022)*, pp. 298-307
18. Hsiang-Sheng Tsai, Heng-Jui Chang, Wen-Chin Huang, Zili Huang, Kushal Lakhotia, Shu-wen Yang, Shuyan Dong, Andy T. Liu, Cheng-I Lai, **Jiatong Shi**, Xuankai Chang, Phil Hall, Shang-Wen Li, Shinji Watanabe, Abdelrahman Mohamed, Hung-yi Lee. 2022. "SUPERB-SG: Enhanced Speech processing Universal PERFORMANCE Benchmark for Semantic and Generative Capabilities" *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL2022)*, pp. 8479-8492.
17. Tao Qian⁺, **Jiatong Shi**, Shuai Guo⁺, Peter Wu⁺, and Qin Jin. 2022. "Training strategies for automatic song writing: a perspective with a unified framework" *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4738-4742.
16. Chunlei Zhang, **Jiatong Shi**, Chao Weng, Meng Yu, and Dong Yu. 2022. "Towards End-to-end Speaker Diarization with Generalized Neural Speaker Clustering" *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8372-8376.

15. **Jiatong Shi**, Chunlei Zhang, Chao Weng, Shinji Watanabe, Meng Yu, and Dong Yu. 2022. "An Investigation of Neural Uncertainty Estimation for Target Speaker Extraction Equipped RNN Transducer". *Computer Speech and Language (CSL)*, 73: 10327.
14. Peter Wu⁺, **Jiatong Shi**, Yifan Zhong, Shinji Watanabe, and Alan W Black. 2021. "Acoustic Cross-lingual Transfer using Language Similarity". *Proceedings of 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1050-1057.
13. Peter Wu⁺, Paul Pu Liang, **Jiatong Shi**, Ruslan Salakhutdinov, and Shinji Watanabe, Louis-Philippe Morency. 2021. "Understanding the Tradeoffs in Client-side Privacy for Downstream Speech Tasks". *Proceedings of 2021 APSIPA Annual Summit and Conference*, pp. 841-848.
12. Shu-wen Yang, Po-Han Chi*, Yung-Sung Chuang*, Cheng-I Jeff Lai*, Kushal Lakhotia*, Yist Y Lin*, Andy T Liu*, **Jiatong Shi***, Xuankai Chang, Guan-Ting Lin, Tzu-Hsien Huang, Wei-Cheng Tseng, Ko-tik Lee, Da-Rong Liu, Zili Huang, Shuyan Dong, Shang-Wen Li, Shinji Watanabe, Abdelrahman Mohamed, and Hung-yi Lee. 2021. "SUPERB: Speech processing Universal PERformance Benchmark". *Proceedings of Interspeech 2021*, International Speech Communication Association, pp. 1194-1198.
11. Hirofumi Inaguma, Brian Yan, Siddharth Dalmia, Pengcheng Guo, **Jiatong Shi**, Kevin Duh, and Shinji Watanabe. 2021. "ESPnet-ST IWSLT 2021 Offline Speech Translation System" *Proceedings of the 18th International Conference on Spoken Language Translation (IWSLT 2021)*, pp.100-109.
10. **Jiatong Shi**, Jonathan D. Amith, Xuankai Chang, Siddharth Dalmia, Brian Yan, and Shinji Watanabe. 2021. "Highland Puebla Nahuatl-Spanish Speech Translation Corpus for Endangered Language Documentation". *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pp.53-63.
9. Jonathan D. Amith, **Jiatong Shi**, and Rey Castillo García. 2021. "End-to-End Automatic Speech Recognition: Its Impact on the Workflow for Documenting Yoloxóchtitl Mixtec". *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pp. 64-80.
8. **Jiatong Shi**, Chunlei Zhang, Chao Weng, Shinji Watanabe, Meng Yu, and Dong Yu. 2021. "Improving RNN Transducer with Target Speaker Extraction and Neural Uncertainty Estimation". *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6908-6912.
7. **Jiatong Shi***, Shuai Guo**+, Nan Huo, Yuekai Zhang, and Qin Jin. 2021. "Sequence-to-sequence Singing Voice Synthesis with Perceptual Entropy Loss". *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 76-80.
6. Pengcheng Guo, Florian Boyer, Xuankai Chang, Tomoki Hayashi, Yosuke Higuchi, Hirofumi Inaguma, Naoyuki Kamo, Chenda Li, Daniel Garcia-Romero, **Jiatong Shi**, Jing Shi, Shinji Watanabe Kun Wei, Wangyou Zhang, and Yuekai Zhang. 2021. "Recent Developments on ESPNet Toolkit Boosted by Conformer". *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.5874-5878.
5. **Jiatong Shi**, Jonathan D. Amith, Rey Castillo García, Esteban Guadalupe Sierra, Kevin Duh, and Shinji Watanabe. 2021. "Leveraging End-to-End ASR for Endangered Language Documentation: An Empirical Study on Yolóxochitl Mixtec". *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL2021)*, pp.1134-1145.
4. **Jiatong Shi**, Kunlin Yang, Wei Xu, and Mingming Wang. 2021. "Leveraging deep learning with audio analytics to predict the success of crowdfunding projects." *Journal of Supercomputing*, 77(7): 7833-7853.
3. **Jiatong Shi**, Nan Huo, and Qin Jin. 2020. "Context-aware Goodness of Pronunciation for Computer Assisted Pronunciation Training". *Proceedings of Interspeech 2020*, pp. 3057-3061.
2. Wenxin Hou, Yue Dong, Bairong Zhuang, Longfei Yang, **Jiatong Shi**, and Takahiro Shinozaki. 2020. "Large-Scale End-to-End Multilingual Speech Recognition and Language Identification with Multi-Task Learning". *Proceedings of Interspeech 2020*, pp. 1047-1051.
1. **Jiatong Shi**, Wei Du, and Wei Xu. 2018. "Identifying Impact Factors of Question Quality in Online Health Q&A Communities: an Empirical Analysis on MedHelp." *Proceedings of the 22nd Pacific Asia Conference on Information Systems (PACIS 2018)*, Association of Information Systems, pp. 1886- 1898.

IN REVIEW

Jiatong Shi, Chunlei Zhang, and Dong Yu. "Streaming Speaker Diarization with Improved Generalized Neural Speaker Clustering"

Jiatong Shi, Chan-Jan Hsu, Holam Chung, Dongji Gao, Paola Garcia, Shinji Watanabe, Ann Lee, and Hung-yi Lee. "Bridging Speech and Text Pre-trained Models with Unsupervised ASR"

Jiatong Shi, Yun Tang, Ann Lee, Hirofumi Inaguma, Changhan Wang, Juan Pino, and Shinji Watanabe. "Enhancing Speech-to-Speech Translation with Multiple TTS Targets"

Dongji Gao*, **Jiatong Shi***, Shun-Po Chuang, Leibny Paola Garcia, Hung-yi Lee, Shinji Watanabe, and Sanjeev Khudanpur. "EURO: ESPnet Unsupervised ASR Open-source Toolkit"

Yuning Wu⁺, **Jiatong Shi**, Tao Qian, and Qin Jin. "PHONEix: Acoustic Feature Processing Strategy for Enhanced Singing Pronunciation with Phoneme Distribution Predictor"

Jessica Huynh*, **Jiatong Shi***, Xuankai Chang*, Jeremiah Milbauer*, and Shinji Watanabe. "Achieving Transparently Synthetic Speech with a Speaker Cocktail "

Yu-Kuan Fu, Liang-Hsuan Tseng, **Jiatong Shi**, Hung-yi Lee, and Shinji Watanabe. "Towards Unsupervised Speech-to-Speech Translation"

William Chen, Brian Yan, **Jiatong Shi**, Yifan Peng, Soumi Maiti, and Shinji Watanabe. "Improving Massively Multilingual ASR with Auxiliary CTC Objectives"

Yushi Hu, Chunlei Zhang, **Jiatong Shi**, Jiachen Lian, Mari Ostendorf, Dong Yu. "ProsodyBERT: Self-Supervised Prosody Representation for Style-Controllable TTS"

Yui Sudo, Shakeel Muhammad, Brian Yan, **Jiatong Shi**, and Shinji Watanabe. "4D: Joint Modeling of CTC, Attention, Transducer, and Mask-predict Decoders"

Awards, Fellowships, Grants, & Notable Achievement

- 2022 **CMU Presidential Fellowship**, Carnegie Mellon University
1st Place, IWSLT 2022 (the shared task in dialectal speech translation track)
7th Place (Finalist), AI Song Contest 2022
- 2021 **Research Fellowship**, Language Technologies Institute, Carnegie Mellon University
- 2019 **Outstanding Graduates**, Renmin University of China
Undergraduate Research Grants, Renmin University of China
Undergraduate Grants, in National Training Programs of Innovation and Entrepreneurship for Undergraduates
- 2018 **Award of Academic Excellence**, School of Information, Renmin University of China
Special Award, in National University Data-driven Innovation & Research Contest, Peking University
- 2017 **Award of Academic Excellence**, School of Information, Renmin University of China
- 2016 **Award of Academic Excellence**, School of Information, Renmin University of China

Presentations

Fall 2022. *Progress in Singing Voice Synthesis with Muskits*. Invited Talk, AIR lab, University of Rochester.

Sep. 2022. *End-to-End Unsupervised ASR and Its Application*. Invited Talk, Sphinx Lunch at CMU.

Sep. 2022. *ESPnet Tutorial 2: New Tasks and Models*. Guest Lecture, Speech Recognition and Understanding, CMU.

Jun. 2022. *ESPnet: an End-to-end Speech Processing Toolkit*. Invited Talk, 2022 JSALT Summer Workshop at JHU.

Apr. 2022. *Speech Technologies with Neural Networks*. Guest Lecture, Natural Language Processing, CMU.

Teaching Experience

Fall 2022	Speech Recognition and Understanding , Teaching Assistant	CMU
Spring 2022	Natural Language Processing (CS411/611/711) , Teaching Assistant	CMU
Spring 2021	Principles of Programming Language (CS425/625) , Course Assistant	JHU
Fall 2020	Introduction to Human Language Technologies (CS465/665) , Course Assistant	JHU
Spring 2020	Machine Learning (CS475/675) , Course Assistant	JHU

Students Mentored

- 2022- **Yuxun Tang**, Incoming MCS at RUC
- 2022- **Tanisha Saxena**, Undergraduate at CMU (AI mentor program)
- 2022- **Tejes Srivastava**, Undergraduate at UC Davis (REUSE Internship)
- 2022- **Yuning Wu**, MCS at RUC
- 2021- **Tao Qian**, MCS at RUC
- 2021-2022 **Dan Berrebbi**, MLT at CMU
- 2019-2022 **Shuai Guo**, Netease Thunder Light Studio (was MCS at RUC)
- 2021 **Peter Wu**, PhD student at UC Berkeley (was MSML at CMU)

Outreach & Professional Development

PEER REVIEWER

Journal: IEEE Signal Processing Letters, Journal of Selected Topics in Signal Processing, Computer Speech and Language, Language Resources and Evaluation, APSIPA Transactions on Signal and Information Processing, Transactions on Multimedia Computing Communications and Applications

Conference: Interspeech21, ARR since 2021, ACM Multimedia 2022, 2022 IEEE SLT, AAAI 2023, ICASSP 2023

OPEN-SOURCE CONTRIBUTION

Toolkit: ESPnet, S3PRL, Muskits, ParallelWaveGAN

Dataset: Highland Puebla Nahuatl, Yoloxóchitl Mixtec, Totonac, KiSing

ORGANIZING SERVICES

Benchmark and Challenges: SUPERB, IWSLT2022(Simultaneous Track)