

# Jiatong Shi

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

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

## Education

---

- Carnegie Mellon University** *Pittsburgh, PA, U.S.A.*  
PHD IN LANGUAGE TECHNOLOGIES *2021.08 - present*  
• Advisor: Dr. Shinji Watanabe
- The Johns Hopkins University** *Baltimore, MD, U.S.A.*  
MS COMPUTER SCIENCE *2019.08 - 2021.05*  
• Advisor: Dr. Shinji Watanabe
- Renmin University of China** *Beijing, China*  
BS COMPUTER SCIENCE *2015.09 - 2019.06*  
• Second Major in Fintech  
• Undergrad research advisor: Dr. Qin Jin, Dr. Wei Xu, and Dr. Wei Du

## Research Experience

---

- Carnegie Mellon University - Language Technologies Institute** *Pittsburgh, PA, U.S.A.*  
ADVISOR: DR. SHINJI WATANABE *2021 - present*  
• Focus1: speech representation learning  
• Focus2: speech-to-speech translation  
• Focus3: low-resource, multilingual speech recognition  
• Focus4: music processing (e.g., singing voice synthesis)
- Frederick Jelinek Memorial Summer Workshop (JSALT, JHU Summer Workshop)** *Baltimore, MD, U.S.A.*  
ADVISOR: DR. ANN LEE, DR. SHINJI WATANABE, AND DR. HUNG-YI LEE *2022 Summer*  
• Focus: unsupervised ASR and its applications
- The Johns Hopkins University - Center for Language and Speech Processing (CLSP)** *Baltimore, MD, U.S.A.*  
ADVISOR: DR. SHINJI WATANABE *2019 - 2021*  
• Focus: low-resource, multilingual, robust speech recognition and applications in endangered language documentation
- Renmin University of China - Multi-level Multi-aspect Multimedia Analysis (AI.M3) Lab** *Beijing, China*  
ADVISOR: DR. QIN JIN *2018-2019*  
• Thesis: "Computer assisted language learning for young English learners"  
• Focus: computer assisted language learning (CALL) and music processing
- Renmin University of China - Lab for Information System Applications (LISA)** *Beijing China*  
CO-ADVISORS: DR. WEI XU AND DR. WEI DU *2017-2019*  
• Thesis: "Multimedia Analytics for Crowdfunding Prediction"  
• Focus: online data analytic and fin-tech applications

## Professional Experience

---

- Tencent America (Tencent AI Lab)** *Bellevue, WA, U.S.A.*  
SUPERVISOR: DR. CHUNLEI ZHANG AND DR. DONG YU *May. 2024 - Aug. 2024*  
• Project: Speech language models for speech processing.
- Meta AI** *Menlo Park, CA, U.S.A.*  
SUPERVISOR: ANNA SUN, DR. XUTAI MA, DR. HIROFUMI INAGUMA, AND DR. ILIA KULIKOV *May. 2023 - Dec. 2023*  
• Project: Self-supervised speech representation learning and its application to speech-to-speech translation.

### Tencent America (Tencent AI Lab)

SUPERVISOR: DR. CHUNLEI ZHANG AND DR. DONG YU

Bellevue, WA, U.S.A.

May. 2022 - Aug. 2022

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

### IBM Watson Research Center

SUPERVISOR: DR. GEORGE SAON

Yorktown Heights, NY, U.S.A.

May. 2021 - Aug. 2021

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

### Tencent America (Tencent AI Lab)

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

Bellevue, 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

54. Taiqi He, Kwanghee Choi, Lindia Tjuatja, **Jiatong Shi**, Nate Robinson, Graham Neubig, Shinji Watanabe, David Mortensen, and Lori Levin. 2024. "WAV2GLOSS: Generating Interlinear Glossed Text from Speech". (Accepted by ACL 2024)
53. Rongjie Huang, Chunlei Zhang, Yongqi Wang, Dongchao Yang, Jinchuan Tian, Luping Liu, Zhenhui Ye, Ziyue Jiang, Xuankai Chang, **Jiatong Shi**, Chao Weng, Zhou Zhao, and Dong Yu. 2024. "Revisiting Voice Large Language Models as Scalable Multi-Lingual and Multi-Task Learners". (Accepted by ACL 2024)
52. Dongchao Yang\*, Jinchuan Tian\*, Xu Tan, Rongjie Huang, Songxiang Liu, Xuankai Chang, **Jiatong Shi**, Sheng Zhao, Jiang Bian, Xixin Wu, Zhou Zhao, and Helen Meng. 2024. "UniAudio: An Audio Foundation Model Toward Universal Audio Generation". (Accepted by ICML 2024)
51. Shu-wen Yang, Heng-Jui Chang\*, Zili Huang\*, Andy T. Liu\*, Cheng-I Lai\*, Haibin Wu\*, **Jiatong Shi**, Xuankai Chang, Hsiang-Sheng Tsai, Wen-Chin Huang, Tzu-hsun Feng, Po-Han Chi, Yist Y. Lin, Yung-Sung Chuang, Tzu-Hsien Huang, Wei-Cheng Tseng, Kushal Lakhotia, Abdelrahman Mohamed, Shang-Wen Li, Shinji Watanabe, and Hung-yi Lee. 2024. "A Large-Scale Evaluation of Speech Foundation Models". (Accepted by TASLP)
50. **Jiatong Shi**, Hirofumi Inaguma, Xutai Ma, Ilia Kulikov, and Anna Sun. 2024. "Multi-resolution HuBERT: Multi-resolution Speech Self-Supervised Learning with Masked Unit Prediction". *Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*, SpotLight.
49. Rongjie Huang\*, Mingze Li\*, Dongchao Yang\*, **Jiatong Shi\***, Xuankai Chang, Zhenhui Ye, Yuning Wu<sup>+</sup>, Zhiqing Hong, Jiawei Huang, Jinglin Liu, Yi Ren, Zhou Zhao, Shinji Watanabe. 2024. "AudioGPT: Understanding and Generating Speech, Music, Sound, and Talking Head". *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 23802-23804.
48. Takashi Maekaku, **Jiatong Shi**, Xuankai Chang, Yuya Fujita, and Shinji Watanabe. 2024. "HuBERTopic: Enhancing semantic representation of HuBERT through self-supervision utilizing topic model". *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 11741-11745.
47. Chien-yu Huang, Ke-Han Lu, Shih-Heng Wang, Chi-Yuan Hsiao, Chun-Yi Kuan, Haibin Wu, Siddhant Arora, Kai-Wei Chang, **Jiatong Shi**, Yifan Peng, Roshan Sharma, Shinji Watanabe, Bhiksha Ramakrishnan, Shady Shehata, and Hung-yi Lee. 2024. "Dynamic-SUPERB: Towards A Dynamic, Collaborative, and Comprehensive Instruction-Tuning Benchmark for Speech". *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 12136-12140.
46. Xuankai Chang, Brian Yan, Kwanghee Choi, Jeeweon Jung, Yichen Lu, Soumi Maiti, Roshan Sharma, **Jiatong Shi**, Jinchuan Tian, Shinji Watanabe, Yuya Fujita, Takashi Maekaku, Pengcheng Guo, Yao-Fei Cheng, Pavel Denisov, Kohei Saijo, and

- Hsiu-Hsuan Wang. 2024. "Exploring Speech Recognition, Translation, and Understanding with Discrete Speech Units: A Comparative Study". *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 11481-11485.
45. Yi-Hui Chou<sup>+</sup>, Kalvin Chang, Meng-Ju Wu, Winston Ou, Alice Wen-Hsin Bi, Carol Yang, Bryan Y. Chen, Rong-Wei Pai, Po-Yen Yeh, Jo-Peng Chiang, Lu-Tshiann Phoann, Winnie Chang, Chenxuan Cui, Noel Chen, and **Jiatong Shi**. 2023. "Evaluating Self-supervised Speech Models on a Taiwanese Hokkien Corpus". *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1-8.
  44. Wen-Chin Huang, Lester Phillip Violeta, Songxiang Liu, **Jiatong Shi**, and Tomoki Toda. 2023. "The Singing Voice Conversion Challenge 2023". *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1-8.
  43. **Jiatong Shi**, William Chen, Dan Berrebbi, Hsiu-Hsuan Wang, Wei-Ping Huang, En-Pei Hu, Ho-Lam Chuang, Xuankai Chang, Yuxun Tang<sup>+</sup>, Shang-Wen Li, Abdelrahman Mohamed, Hung-yi Lee, and Shinji Watanabe. 2023. "Findings of the 2023 ML-SUPERB Challenge: Pre-Training and Evaluation over More Languages and Beyond". *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1-8.
  42. Yifan Peng, Jinchuan Tian, Brian Yan, Dan Berrebbi, Xuankai Chang, Xinjian Li, **Jiatong Shi**, Siddhant Arora, William Chen, Roshan Sharma, Wangyou Zhang, Yui Sudo, Muhammad Shakeel, Jee-Weon Jung, Soumi Maiti, and Shinji Watanabe. 2023. "Reproducing Whisper-Style Pre-training Using an Open-Source Toolkit and Publicly Available Data". *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1-8.
  41. William Chen, **Jiatong Shi**, Brian Yan, Dan Berrebbi, Wangyou Zhang, Yifan Peng, Xuankai Chang, Soumi Maiti, and Shinji Watanabe. 2023. "Joint Prediction and Denoising for Large-Scale Multilingual Self-Supervised Learning". *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1-8.
  40. Yui Sudo, Shakeel Muhhammad, Brian Yan, **Jiatong Shi**, and Shinji Watanabe. 2023. "4D: Joint Modeling of CTC, Attention, Transducer, and Mask-predict Decoders". *Proceedings of Interspeech*, pp. 3312-3316.
  39. **Jiatong Shi**, Dan Berrebbi\*, William Chen\*, En-Pei Hu\*, Wei-Ping Huang\*, Ho Lam Chung\*, Xuankai Chang, Shang-Wen (Daniel) Li, Abdelrahman Mohamed, Hung-yi Lee, and Shinji Watanabe. 2023. "ML-SUPERB: Multilingual Speech Universal PERFORMANCE Benchmark". *Proceedings of Interspeech*, pp. 884-888.
  38. **Jiatong Shi**, Yun Tang, Hirofumi Inaguma, Hongyu Gong, Juan Pino, and Shinji Watanabe. 2023. "Exploration on HuBERT with Multiple Resolution" *Proceedings of Interspeech*, pp. 3287-3291.
  37. Sweta Agrawal, Antonios Anastasopoulos, Luisa Bentivogli, Ondřej Bojar, Claudia Borg, Marine Carpuat, Roldano Cattoni, Mauro Cettolo, Mingda Chen, William Chen, Khalid Choukri, Alexandra Chronopoulou, Anna Currey, Thierry Declercq, Qianqian Dong, Kevin Duh, Yannick Estève, Marcello Federico, Souhir Gahbiche, Barry Haddow, Benjamin Hsu, Phu Mon Htut, Hirofumi Inaguma, Dávid Javorský, John Judge, Yasumasa Kano, Tom Ko, Rishu Kumar, Pengwei Li, Xutai Ma, Prashant Mathur, Evgeny Matusov, Paul McNamee, John P. McCrae, Kenton Murray, Maria Nadejde, Satoshi Nakamura, Matteo Negri, Ha Nguyen, Jan Niehues, Xing Niu, Atul Kr. Ojha, John E. Ortega, Proyag Pal, Juan Pino, Lonneke van der Plas, Peter Polák, Elijah Rippeth, Elizabeth Salesky, **Jiatong Shi**, Matthias Sperber, Sebastian Stüker, Katsuhito Sudoh, Yun Tang, Brian Thompson, Kevin Tran, Marco Turchi, Alex Waibel, Mingxuan Wang, Shinji Watanabe, and Rodolfo Zevallos. 2023. "Findings of the IWSLT 2023 Evaluation Campaign", *Proceedings of the 20th International Conference on Spoken Language Translation (IWSLT)*, pp. 1-61
  36. Brian Yan\*, **Jiatong Shi**\*, Soumi Maiti, William Chen, Xinjian Li, Yifan Peng, Siddhant Arora, and Shinji Watanabe. 2023. "CMU's IWSLT 2023 Simultaneous Speech Translation System", *Proceedings of the 20th International Conference on Spoken Language Translation (IWSLT)*, pp. 235-240
  35. Biran Yan\*, **Jiatong Shi**\*, Yun Tang, Hirofumi Inaguma, Yifan Peng, Siddharth Dalmia, Peter Polák, Patrick Fernandes, Dan Berrebbi, Tomoki Hayashi, Xiaohui Zhang, Zhaoheng Ni, Moto Hira, Soumi Maiti, Juan Pino, Shinji Watanabe. 2023. "ESPnet-ST-v2: Multipurpose Spoken Language Translation Toolkit", *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 400-411
  34. Tao Qian<sup>+</sup>, Fan Lou, **Jiatong Shi**, Yuning Wu<sup>+</sup>, Shuai Guo<sup>+</sup>, Xiang Yin, and Qin Jin. 2023. "UniLG: A Unified Structure-aware Framework for Lyrics Generation", *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 983-1001.
  33. William Chen, Brian Yan, **Jiatong Shi**, Yifan Peng, Soumi Maiti, and Shinji Watanabe. 2023. "Improving Massively Multilingual ASR with Auxiliary CTC Objectives" *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.
  32. **Jiatong Shi**, Chan-Jan Hsu, Holam Chung, Dongji Gao, Paola Garcia, Shinji Watanabe, Ann Lee, and Hung-yi Lee. 2023. "Bridging Speech and Text Pre-trained Models with Unsupervised ASR" *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.

31. **Jiatong Shi**, Yun Tang, Ann Lee, Hirofumi Inaguma, Changhan Wang, Juan Pino, and Shinji Watanabe. 2023. "Enhancing Speech-to-Speech Translation with Multiple TTS Targets" *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.
30. Dongji Gao\*, **Jiatong Shi**\*, Shun-Po Chuang, Leibny Paola Garcia, Hung-yi Lee, Shinji Watanabe, and Sanjeev Khudanpur. 2023. "EURO: ESPnet Unsupervised ASR Open-source Toolkit" *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.
29. Yuning Wu<sup>+</sup>, **Jiatong Shi**, Tao Qian<sup>+</sup>, and Qin Jin. 2023. "PHONEix: Acoustic Feature Processing Strategy for Enhanced Singing Pronunciation with Phoneme Distribution Predictor" *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.
28. Antonios Anastasopoulos, Loïc Barrault, Luisa Bentivogli, Marcelly Zanon Boito, Ondřej Bojar, Roldano Cattoni, Anna Currey, Georgiana Dinu, Kevin Duh, Maha Elbayad, Clara Emmanuel, Yannick Estève, Marcello Federico, Christian Federmann, Souhir Gahbiche, Hongyu Gong, Roman Grundkiewicz, Barry Haddow, Benjamin Hsu, Dávid Javorský, Věra Kloudová, Surafel Lakew, Xutai Ma, Prashant Mathur, Paul McNamee, Kenton Murray, Maria Nădejde, Satoshi Nakamura, Matteo Negri, Jan Niehues, Xing Niu, John Ortega, Juan Pino, Elizabeth Salesky, **Jiatong Shi**, Matthias Sperber, Sebastian Stüker, Katsuhito Sudoh, Marco Turchi, Yogesh Virkar, Alexander Waibel, Changhan Wang, and Shinji Watanabe. 2022. "Findings of the IWSLT 2022 Evaluation Campaign", *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT)*, pp. 298-307
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" *IEEE Spoken Language Technology Workshop (SLT)*, pp. 1128-1135.
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, and Hung-yi Lee. 2022. "SUPERB @ SLT 2022: Challenge on Generalization and Efficiency of Self-Supervised Speech Representation Learning" *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*, pp. 1096-1103.
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*, pp. 1656-1660.
24. **Jiatong Shi**, Shuai Guo<sup>+</sup>, Tao Qian<sup>+</sup>, 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*, pp.4277-4281.
23. Shuai Guo<sup>++</sup>, **Jiatong Shi**\*, Tao Qian<sup>+</sup>, Shinji Watanabe and Qin Jin. 2022. "SingAug: Data Augmentation for Singing Voice Synthesis with Cycle-consistent Training Strategy" *Proceedings of Interspeech*, pp.4272-4276.
22. Dan Berrebbi<sup>+</sup>, **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*, 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*, 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*, 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)*, 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, and 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 (ACL)*, pp. 8479-8492.
17. Tao Qian<sup>+</sup>, **Jiatong Shi**, Shuai Guo<sup>+</sup>, Peter Wu<sup>+</sup>, 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<sup>+</sup>, **Jiatong Shi**, Yifan Zhong, Shinji Watanabe, and Alan W Black. 2021. "Acoustic Cross-lingual Transfer using Language Similarity". *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1050-1057.
13. Peter Wu<sup>+</sup>, 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 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*, 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)*, 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 Yoloxóchitl 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 (EACL)*, 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*, 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*, 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)*, Association of Information Systems, pp. 1886- 1898.

#### ARXIV PREPRINT

7. Tejes Srivastava<sup>+</sup>, **Jiatong Shi**, William Chen, and Shinji Watanabe. "EFFUSE: Efficient Self-Supervised Feature Fusion for E2E ASR in Multilingual and Low Resource Scenarios".

6. Jee-weon Jung\*, Wangyou Zhang\*, **Jiatong Shi\***, Zakaria Aldeneh, Takuya Higuchi, Barry-John Theobald, Ahmed Hussien Abdelaziz, and Shinji Watanabe. "ESPnet-SPK: full pipeline speaker embedding toolkit with reproducible recipes, self-supervised front-ends, and off-the-shelf models"
5. **Jiatong Shi\***, Yueqian Lin\*, Xinyi Bai, Keyi Zhang, Yuning Wu, Yuxun Tang, Yifeng Yu, Qin Jin, and Shinji Watanabe. "Singing Voice Data Scaling-up: An Introduction to ACE-Opencpop and ACE-KiSing"
4. Yuning Wu<sup>+</sup>, Yifeng Yu\*, **Jiatong Shi**, Tao Qian<sup>+</sup>, Qin Jin. "A Systematic Exploration of Joint-training for Singing Voice Synthesis"
3. Yu-Kuan Fu, Liang-Hsuan Tseng, **Jiatong Shi**, Hung-yi Lee, and Shinji Watanabe. "Improving Speaker-preserving Cascaded Unsupervised Speech Translation with Denoising Back-translation"
2. Yushi Hu, Chunlei Zhang, **Jiatong Shi**, Jiachen Lian, Mari Ostendorf, Dong Yu. "ProsodyBERT: Self-Supervised Prosody Representation for Style-Controllable TTS"
1. Tomoki Hayashi, Ryuichi Yamamoto, Takenori Yoshimura, Peter Wu<sup>+</sup>, **Jiatong Shi**, Takaaki Saeki, Yooncheol Ju, Yusuke Yasuda, Shinnosuke Takamichi, Shinji Watanabe. "ESPnet2-TTS: Extending the Edge of TTS Research"

## IN REVIEW

- Xuankai Chang, **Jiatong Shi**, Jinchuan Tian, Yuning Wu, Yuxun Tang, Yihan Wu, Shinji Watanabe, Yossi Adi, Xie Chen, and Qin Jin. "The Interspeech 2024 Challenge on Speech Processing Using Discrete Units".
- Yongyi Zang, **Jiatong Shi**, You Zhang, Ryuichi Yamamoto, Jionghao Han, Yuxun Tang, Shengyuan Xu, Wenxiao Zhao, Jing Guo, Tomoki Toda, Zhiyao Duan. "CtrSVDD: A Benchmark Dataset and Baseline Analysis for Controlled Singing Voice Deepfake Detection".
- Jiatong Shi**, Shih-Heng Wang<sup>+</sup>, William Chen, Martijn Bartelds, Vanya Bannihatti Kumar, Jinchuan Tian, Xuankai Chang, Dan Jurafsky, Karen Livescu, Hung-yi Lee, and Shinji Watanabe. "ML-SUPERB 2.0: Benchmarking Multilingual Speech Models Across Modeling Constraints, Languages, and Datasets".
- Jiatong Shi**, Xutai Ma, Hirofumi Inaguma, Anna Sun, and Shinji Watanabe. "MMM: Multi-Layer Multi-Residual Multi-Stream Discrete Speech Representation from Self-supervised Learning Model".
- Yuxun Tang<sup>+</sup>, Yuning Wu<sup>+</sup>, **Jiatong Shi**, and Qin Jin. "SingOMD: Singing Oriented Multi-resolution Discrete Representation Construction from Speech Models"
- Yuning Wu<sup>+</sup>, Chunlei Zhang, **Jiatong Shi**, Yuxun Tang<sup>+</sup>, and Qin Jin "TokSing: Singing Voice Synthesis based on Discrete Tokens"
- Shih-Heng Wang<sup>+</sup>, **Jiatong Shi**, Chien-yu Huang, Shinji Watanabe, and Hung-yi Lee. "Fusion of Self-Supervised Learning Discrete Representations for Automatic Speech Recognition".
- Yifeng Yu<sup>+</sup>, **Jiatong Shi**, Shinji Watanabe, and Yuning Wu. "VISinger2+: End-to-End Singing Voice Synthesis Augmented by Self-Supervised Learning Representation".
- Yifan Peng, Jinchuan Tian, William Chen, Siddhant Arora, Brian Yan, Yui Sudo, Muhammad Shakeel, Kwanghee Choi, **Jiatong Shi**, Xuankai Chang, Jee-weon Jung, and Shinji Watanabe. "OWSM v3.1: Better and Faster Open Whisper-Style Speech Models based on E-Branchformer".
- William Chen, Wangyou Zhang, Yifan Peng, Xinjian Li, **Jiatong Shi**, Jinchuan Tian, Xuankai Chang, Soumi Maiti, Karen Livescu, and Shinji Watanabe. "Towards Robust Speech Representation Learning for 4000+ Languages".
- Kristin Qi<sup>+</sup>, **Jiatong Shi**, Caroline Summerour, John Batsis, and Xiaohui Liang. "Exploiting Longitudinal Speech Data via Voice Assistant Systems for Early Detection of Cognitive Decline".
- Shuhua Li<sup>+</sup>, Qirong Mao, and **Jiatong Shi**. "PL-TTS: A Generalizable Prompt-based Diffusion TTS Augmented by Large Language Model".
- Kalvin Chang, Yi-Hui Chou<sup>+</sup>, **Jiatong Shi**, Hsuan-Ming Chen, Nicole Holliday, Odette Scharenborg, and David R. Mortensen. "Self-supervised Speech Representations Still Struggle with African American Vernacular English".
- Shin-Heng Wang<sup>+</sup>, Zih-Ching Chen,\*, **Jiatong Shi\***, Ming-To Chuang\*, Guan-Ting Lin, Kuan-Po Huang, David Harwath, Shang-Wen Li, and Hung-yi Lee. "How to Learn a New Language? An Efficient Solution for Self-Supervised Learning Models Unseen Languages Adaption in Low-Resource Scenario".

## Awards, Fellowships, Grants, & Notable Achievement

---

- 2023 **Best Paper Finalist at 2023 ASRU**, IEEE ASRU Committee  
**Best Paper Finalist at 2022 SLT**, IEEE SLT Committee  
**Best Project at 2022 SLT Hackathon (as a mentor)**, IEEE SLT Committee
- 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

---

- Mar. 2024. *Multi-resolution HuBERT*. Invited Talk, Algorithm Talk at Alibaba-NTU (Nanyang Technological University)
- Mar. 2024. *Speech Representation Learning: From Continuous to Discrete*. Invited Talk, Huawei
- Mar. 2024. *Speech-to-speech Translation: Task Overview*. Lecture, Speech Processing, CMU.
- Dec. 2023. *ML-SUPERB and ML-SUPERB Challenge*. Invited Talk, SPARKS Workshop at ASRU 2023, National Taiwan University
- Aug. 2023. *Language Resources and Technology: Inclusion, Diversity, and Sovereignty*. Panel Discussion, SIGUL Workshop, Interspeech@Dublin
- Jun. 2023. *ESPnet-Muskits and Its Integration to LLM*. Invited Talk, Music & Audio Processing Workshop, HKUST
- Apr. 2023. *Speech-to-speech Translation*. Lecture, Speech Processing, CMU.
- Feb. 2023. *ESPnet Tutorial 2: New Tasks and Models*. Lecture, Speech Processing, CMU.
- Nov. 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*. 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

---

Spring 2024	<b>Speech Processing (CS492/692)</b> , Teaching Assistant	CMU
Spring 2023	<b>Speech Processing (CS492/692)</b> , Teaching Assistant	CMU
Fall 2022	<b>Speech Recognition and Understanding (CS751)</b> , Teaching Assistant	CMU
Spring 2022	<b>Natural Language Processing (CS411/611/711)</b> , Teaching Assistant	CMU
Spring 2021	<b>Principles of Programming Language (CS425/625)</b> , Course Assistant	JHU
Fall 2020	<b>Introduction to Human Language Technologies (CS465/665)</b> , Course Assistant	JHU
Spring 2020	<b>Machine Learning (CS475/675)</b> , Course Assistant	JHU

## Students Mentored

---

- 2024- **Xincan Feng**, PhD at NAIST
- 2023- **Jionghao Han**, MS ECE at CMU
- 2023- **Shin-Heng Wang**, MCS at NTU
- 2022- **Yifeng Yu**, MSMT at Georgia Tech
- 2022- **Yuxun Tang**, MCS at RUC
- 2022- **Yuning Wu**, MCS at RUC
- 2022- **Tejes Srivastava**, MCS at UChicago (was Undergrad at UC Davis, REUSE Internship)
- 2022-2023 **Yi-Hui (Sophia) Chou**, MISC at CMU
- 2023 **Yueqian Lin**, Undergraduate at DKU
- 2023 **Xinyi Bai**, Google
- 2021-2023 **Tao Qian**, Shanghai High School (was MCS at RUC)
- 2022-2023 **Tanisha Saxena**, Undergraduate at CMU (AI mentor program)
- 2021-2022 **Dan Berrebbi**, Apple (was 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)

## Professional Activities

---

### PEER REVIEWER

- Journals:
  - IEEE Signal Processing Letters
  - Journal of Selected Topics in Signal Processing
  - IEEE Transactions on Audio, Speech, and Language Processing
  - Computer Speech and Language
  - Language Resources and Evaluation
  - APSIPA Transactions on Signal and Information Processing
  - Transactions on Multimedia Computing Communications and Applications
- Conferences:
  - Interspeech
  - ACL Rolling Review (ARR)
  - ACM Multimedia
  - IEEE SLT
  - AAAI
  - ICASSP
  - ASRU

### OPEN-SOURCE CONTRIBUTION



- Toolkit:
  - ESPnet (<https://github.com/espnet/espnet>)
  - S3PRL (<https://github.com/s3prl/s3prl>)
  - Muskits (Merged to ESPnet, deprecated url at <https://github.com/SJTMusicTeam/Muskits>)
  - ParallelWaveGAN (<https://github.com/kan-bayashi/ParallelWaveGAN>)
  - AudioGPT (<https://github.com/AIGC-Audio/AudioGPT>)
  - Fairseq (<https://github.com/facebookresearch/fairseq>)
- Dataset:
  - Highland Puebla Nahuatl (<http://www.openslr.org/92/>)
  - Yoloxóchitl Mixtec (<http://www.openslr.org/89/>)
  - Totonac (<http://www.openslr.org/107/>)
  - KiSing (<http://shijt.site/index.php/2021/05/16/>)
  - ML-SUPERB (<https://arxiv.org/abs/2305.10615>)

## ORGANIZING SERVICES

- Benchmark and Challenges:
  - Speech Processing Using Discrete Speech Unit Challenge at Interspeech 2024
  - ML-SUPERB (Multilingual SUPERB) at ASRU 2023
  - SVCC (Singing Voice Conversion Challenge) at ASRU 2023
  - IWSLT since 2022(Simultaneous Track)
  - SUPERB
- Workshops:
  - SPARKS Workshop