Chang ZENG (曾 暢 ソウ チョウ)

Speech Signal Processing/Deep Learning

Tokyo, Japan | (81)07085400924 | zengchang.elec@gmail.com

| github.com/zengchang233 | scholar.google.com/citations?user=gfGyn49j-MkC&hl=en

I am a Ph.D. candidate with 5 years of speech signal processing/sequence-to-sequence (S2S)/deep learning experiences. I have explored speaker recognition in universities and speech recognition in the industry. My research/work interest includes speech/speaker recognition and generative speech AI such as TTS and voice conversion. I would like to dedicate myself to **speech signal processing/speech AI/Machine Learning/Deep Learning** field for my long-term future work.

Education

National Institute of Informatics (SOKENDAI) **2020.10 - present**

Doctor of Informatics

Sponsored by the Japanese government MEXT scholarship, supervised by Prof. Yamagishi.

The University of Tokyo (東京大学) **2017.10 - 2020.03**

Master of Electrical Engineer and Information System

Supervised by Prof. Minematsu.

Tianjin University **2012.09 - 2016.07**

Bachelor of Engineering

Publications

Peer-Reviewed

1. **Zeng, C.**, Wang, X., Cooper, E., Miao, X., & Yamagishi, J. (2022). Attention back-end for automatic speaker verification with multiple enrollment utterances. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6717-6721). Project page. *(Top Conference in Speech and Audio Processing)*
2. **Zeng, C.**, Zhang, L., Liu, M., & Yamagishi, J. (2022). Spoofing-Aware Attention based ASV Back-end with Multiple Enrollment Utterances and a Sampling Strategy for the SASV Challenge 2022. In *Proc. Interspeech 2022, 2883–2887. (Top Conference in Speech and Audio Processing)*
3. **Zeng, C.**, Wang, X., Miao, X., Cooper, E., & Yamagishi, J. (2023). Improving Generalization Ability of Countermeasures for New Mismatch Scenario by Combining Multiple Advanced Regularization Terms. In *Proc. Interspeech 2023, 1998-2002. (Top Conference in Speech and Audio Processing)*
4. Wang, C., **Zeng, C (co-first author)**., & He, X. (2023). Xiaoicesing 2: A High-Fidelity Singing Voice Synthesizer Based on Generative Adversarial Network. In *Proc. Interspeech 2023, 5401-5405.* Project page. *(Top Conference in Speech and Audio Processing)*
5. Zhu, W., Wang, Z., Lin, J., **Zeng, C.**, & Yu, T. (2023) SSI-Net: A MULTI-STAGE SPEECH SIGNAL IMPROVEMENT SYSTEM FOR ICASSP 2023 SSI CHALLENGE. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-2) *(Top Conference in Speech and Audio Processing)*
6. Liu, M., Wang, L., Lee, K. A., Zhang, H., **Zeng, C.**, & Dang, J. (2021). DeepLip: A Benchmark for Deep Learning-Based Audio-Visual Lip Biometrics. In *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 122-129). Project Page.
7. Liu, M., Lee, K. A., Wang, L., Zhang, H., **Zeng, C.**, & Dang, J. (2023) Cross-Modal Audio-Visual Co-learning for Text-independent Speaker Verification. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). Project page. (Top Conference in Speech and Audio Processing)
8. Liu, X., Liu, M., Zhang, L., Zhang, L., **Zeng, C.**, Li, K., Li, N., Lee, K., Wang, L., & Dang, J. (2022). Deep Spectro-temporal Artifacts for Detecting Synthesized Speech. In *ACMMM workshop. 69–75.*
9. Li, K., Li, S., Lu, X., Masato, A., Liu, M., Zhang, L., **Zeng, C.**, Wang, L., Dang, J., & Unoki, M. (2022). Data Augmentation Using McAdams-Coefficient-Based Speaker Anonymization for Fake Audio Detection. In *Proc. Interspeech 2022, 664-668. (Top Conference in Speech and Audio Processing)*

Under Review

1. **Zeng, C.**, Miao, X., Wang, X., Cooper, E., & Yamagishi, J. (2022). Joint Speaker Encoder and Neural Back-end Model for Fully End-to-End Automatic Speaker Verification with Multiple Enrollment Utterances. *Submitted to Computer Speech & Language.*
2. Wang, C., **Zeng, C (co-first author)**., Chen, J., & He, X. (2023). HiFi-WaveGAN: Generative Adversarial Network with Auxiliary Spectrogram-Phase Loss for High-Fidelity Singing Voice Generation. *Submitted to ASRU 2023.* Project page.
3. Wang X., **Zeng, C (co-first author)**., Wang, C. (2023). CrossSinger: A Cross-Lingual Multi-Singer High-Fidelity Singing Voice Synthesizer Trained on Monolingual Singers. *Submitted to ASRU 2023.* Project page.
4. Tang, H., Liu, Z., **Zeng, C.**, & Li, X. (2023). Beyond Universal Transformer: block reusing with adaptor in Transformer for automatic speech recognition. *Submitted to ASRU 2023.*

Work Experiences

Xiaobing.ai **2022.07 - 2023.07**

Intern Avatar Researcher

Responsibilities

- I am focusing on building a generative model for high-fidelity (48kHz) singing voice generation tasks collaborating with other researchers and engineers.

Achievements

- Developed a GAN-based acoustic model called XiaoiceSing2 to generate a mel-spectrogram with high accuracy. The paper has been accepted by Interspeech 2023;
- Developed a GAN-based vocoder model called HiFi-WaveGAN to reconstruct the waveform from the mel-spectrogram predicted by XiaoiceSing2 and submitted the paper to ASRU 2023;
- Improved XiaoiceSing2 to CrossSinger, which is a cross-lingual multi-singer singing voice synthesizer. The paper has been submitted to ASRU 2023;
- Improving these works into a multi-singer singing voice synthesis system, which can realize zero-shot singing voice synthesis;
- Exploring to realize a general vocoder to cover singing/speech/accompanying scenarios.

ALIBABA **2020.04 - 2020.09**

Speech Recognition Researcher

Responsibilities

- In Alibaba, I belonged to a department that is responsible for TAOBAO living. To prevent live broadcasters from violating laws and regulations, I developed two systems with my colleagues.

Achievements

- Developed a large-scale speaker recognition system that aims at identifying the broadcaster in each living room as the exact person;
- Developed a spoken word detection system aims at catching some illegal words spoken by broadcasters;
- Explored SSL for speech representation and developed an end-to-end speech recognition system with knowledge distillation based on ESPNet.

ALIBABA **2019.07 - 2019.09**

Intern Speech Processing Researcher

Achievements

- Developed a neural network-based end-to-end speaker recognition system to replace the traditional I-Vector/PLDA system.

Skills

Tools and Languages	Python, CPP, Shell, Git
Speech Toolkits	PyTorch, SpeechBrain, WeNet, WeSpeaker, Kaldi, Espnet
Communication	Chinese (native), English (business level), Japanese (N2-125)

Activities

Competitions

- 4th/77 place for VoxCeleb Speaker Recognition Challenge 2019.
- 2nd/110 place for Zhijiang Cup Speech Recognition for Conversational Scenario 2021.
- 4th/42 place for Audio Deep Synthesis Detection Challenge 2022 track1 (Low-quality Fake Audio Detection, LF).
- 5th/27 place for Audio Deep Synthesis Detection Challenge 2022 track2 (Partially Fake Audio Detection, PF).

Academic Activities

- Research assistant at Yamagishi Lab, National Institute of Informatics
- ICASSP 2022 Oral Presentation.
- Interspeech 2022 Oral Presentation.
- Interspeech 2023 Poster Presentation.
- Reviewer of IEEE Open Journal of Signal Processing.

Open Source

- WeSpeaker contributor
- ASV-Subtools contributor
- Attention backend