Xintong Wang

🏠 Haidian District, Beijing, 100080, China

📫 walston874848612@163.com ☎ (+86) 185-1377-5531 🐙 oshindow

📚 Education

Beijing Forestry University	**Beijing, China**
Bachelor, Mathematics and Applied Mathematics	2018–2022

💼 Experience

Xiaoice Research — **Beijing, China**
Machine Learning Engineer — Jul 2022–Present

Singing Voice Synthesis
- Developed a multi-language **acoustic model** via Generative Adversarial Network, which is a single model for three languages (Mandarin/English/Japanese), transforming a monolingual singer into a multilingual singer.
- Introduced energy parameter into acoustic model and develop an energy parameter prediction model.
- Developed and robusted a iSTFT-based **vocoder** via Generative Adversarial Network, which is predicting phase using STFT spectrogram, generating 48k high-fidelity audio high consistency with the ground truth.
- Investigated diffusion-based acoustic models, such as DiffSinger.

Speech Recognition
- Deploy a CPU-based gRPC speech recognition service using the **Wenet framework**, with the Wenet pre-trained model for Chinese trained on 10,000 hours.
- Develop a phoneme correction module based on carefully annotated data to correct the position of voiced consonants after automatic alignment by MFA. Further develop **phoneme alignment** models based on HMM-GMM and CTC, using Kaldi, k2, Funasr toolkits.

Xiaoice Waveland Team — **Beijing, China**
Intern — May 2021–Jul 2022

Language Model
- Build, prune and merge **N-gram Language Models** from large scale corpus of several domains.
- Compile Language Model, Pronunciation Lexicon and Acoustic Model into an HCLG.fst, and decode with Kaldi to obtain a better CER on benchmarks

Auto popular songs medley
- Given a **MIDI** library and a specific chords progression, develop a program that automatically selects segments from the library that match the given chords progression. Create a new musical score using these selected segments, and use XiaoiceSing1.8 to synthesize a mashup song.

🔧 Projects

Streaming Mandarin Speech Recognition — Jun 2019–Jun 2021
- Train HMM-GMM based SR models on Kaldi and Transducer-based Streaming SR models on ESPnet
- Extract Transducer module from ESPnet

✋ Peer-Reviewed Publications

Wang,Xintong, and Chuangang Zhao, "A 2D Convolutional Gating Mechanism for Mandarin Streaming Speech Recognition." Information-an International Interdisciplinary Journal, vol. 12, no. 4, 2021, p. 165.

🏆 Awards

COMAP Mathematical Contest in Mathematical Contest in Modeling ; Honorable Mention (26%)	2020
Outstanding Student Scholarship	2019 Fall - 2020 Fall
Academic Excellence Scholarship	2019 Fall - 2020 Fall