Haohan Guo

Xi'an, China hhguo1995@gmail.com (+86)18829234898

Education

Master of Engineering, Supervisor: Lei Xie, Northwestern Polytechnical University, China. 2017 - 2020
Bachelor of Engineering, Northwestern Polytechnical University, China 2013 - 2017

Research Experience

Exploiting Syntactic Features in a Parsed Tree to Improve End-to-End TTS 2018.7-2018.11
Accepted by INTERSPEECH 2019. Authors: Haohan Guo, Frank K. Soong, Lei He, Lei Xie
In this study we investigated syntactic parsing derived features embedded in a parsed tree for improving end-to-end TTS synthesis performance. Experimental results show that syntactic features can indeed improve the quality of the synthesized speech, and word relation based features we proposed yield the best performance.

A New GAN-based End-to-End TTS Training Algorithm 2018.11-2019.4
Accepted by INTERSPEECH 2019. Authors: Haohan Guo, Frank K. Soong, Lei He, Lei Xie
To avoid exposure bias in AR model based end-to-end TTS, we propose a new GAN-based, end-to-end TTS training algorithm. Experimental results show that schedule sampling can improve the model generalization, but is harmful for speech quality. Our proposed algorithm improves both output quality and generalization of the model.

Conversational End-to-End TTS for Voice Agent 2019.5-2019.10
Accepted by SLT 2021. Authors: Haohan Guo, Shaofei Zhang, Frank K. Soong, Lei He, Lei Xie
We design a new recording scheme to create a clean corpus with spontaneous conversational speaking style, and propose a conversational end-to-end TTS model with a conversational context encoder to enhance its performance on the conversation.

Many-to-many Singing Voice Conversion using Non-parallel Data 2020.5-
We propose a MelGAN-based SVC to improve the quality and robustness. Experiments show its better capability than conventional approaches in singing quality and singer similarity.

Project Experience

Development of the offline HMM-based TTS system 2016.10-2017.12
Developed offline HMM-based English TTS system. Maintained English front-end modules. Investigated the fixed-point algorithm, and implemented it on the vocoder for real-time decoding (nearly 4 times).

Development of the NN-based TTS system 2016.10-2017.12
Developed NN-based Mandarin TTS system. Implemented the pipeline of Mandarin TTS system. Studied emotional TTS algorithm, e.g. speaker code, multi-head network and finetune based approaches.

Adaptive multi-speaker speech synthesis based on i-vector 2017.2-2017.4
Investigated the impact of the i-vector and data size on multi-speaker TTS.
Implemented a demo which can synthesize high-quality speech for a new speaker.

Optimization of auto-regressive neural vocoder 2019.11-2020.5
Implemented Mel-spectrogram based LPCNet, and optimized it to achieve faster inference speed.
Optimized single-gaussian WaveRNN to achieve both better sound quality and faster speed.

Work Experience

Speech Intern, chumenwenwen (Mobvoi), Beijing 2016.07–2016.10
Maintain the models and code of TTS front-end modules.

Research Intern, Microsoft (MSRA and STCA), Beijing 2018.04–2019.09
Study the end-to-end TTS model, and try to improve its quality and stability.

Research Intern, Tencent AI lab, Beijing 2020.05–2020.12
Study many-to-many singing conversion based on un-parallel singing corpus.

Researcher, Sogou Inc, Beijing 2020.12–
Develop commercial singing voice conversion system.

Awards

First Prize Scholarship, *Northwestern Polytechnical University* 2013 - 2017
Second Prize of ACM Programming Competition, *Shannxi Province* 2015 - 2016