

# (郭浩瀚) Haohan GUO

Ph.D. Student

📍 Hong Kong SAR, China ✉ [hguo@se.cuhk.edu.hk](mailto:hguo@se.cuhk.edu.hk) 🏠 [hhguo.github.io](https://github.com/hhguo)

## Education

**Ph.D.**, The Chinese University of Hong Kong, Hong Kong SAR, China. 2021 -  
**M.S.**, Northwestern Polytechnical University, Xi'an Shaanxi, China 2017 - 2020  
**B.S.**, Northwestern Polytechnical University, Xi'an Shaanxi, China 2013 - 2017

## Research Interests

Speech Synthesis, (Singing) Voice Conversion, Speech & Language Processing, Deep Learning

## Selective Research Projects

**Large Language Model (LLM) based TTS System** 2023.06-2023.11  
This work proposes scaling TTS into a large system with over 1B parameters and 100 hours of training data. The paper "*BASE TTS: Lessons from building a billion-parameter Text-to-Speech model on 100K hours of data*" was published at Arxiv.

**Compact Speech Representations for High-Quality TTS** 2021.10-2023.05  
This work aims to learn compact speech representations as acoustic features. Compact features are easier to predict, and have higher generalization. Our current works focus on the VQ-VAE based discrete representations, and we propose Multi-Stage Multi-Codebook VQ-VAE to learn better compact representations for TTS. The paper "*A Multi-Stage Multi-Codebook VQ-VAE Approach to High-Performance Neural TTS*" was accepted by INTERSPEECH 2022. The journal "*MSMC-TTS: Multi-Stage Multi-Codebook VQ-VAE Based Neural TTS*" was published at TASLP.

**GAN-based High-Quality Singing Voice Conversion** 2020.05-2021.07  
This work aims to utilize GAN based waveform generation to enhance PPG-based singing voice conversion. The MelGAN-based SVC system is first proposed to convert PPGs to audios directly for better conversion quality. Then, the harmonic signals based method is proposed to obtain better singing quality. The paper "*Improving Adversarial Waveform Generation based Singing Voice Conversion with Harmonic Signals*" was accepted by ICASSP 2022.

**Conversational TTS** 2019.05-2019.10  
It is a preliminary study to end-to-end TTS in the conversational scenario, including the design of a conversational TTS corpus, and the implementation of a high-quality conversational end-to-end TTS system. This work was done during the internship at MSRA. The paper "*Conversational End-to-End TTS for Voice Agent*" was accepted by SLT 2021.

**An Investigation on "Exposure Bias" in Auto-Regressive TTS** 2018.11-2019.04  
We investigated the "exposure bias" problem of auto-regressive mechanism, and proposed a GAN-based training algorithm to improve both output quality and generalization of TTS. This work was done during the internship at MSRA. The paper "*A New GAN-based End-to-End TTS Training Algorithm*" was accepted by INTERSPEECH 2019.

**Exploiting Syntactic Features for End-to-End TTS** 2018.07-2018.11  
This work aims to utilize syntactic features to improve the naturalness of the generated speech. Experimental results show that the proposed "word relation based features" in a parsed tree enhances TTS significantly. This work was done during the internship at MSRA. The paper "*Exploiting Syntactic Features in a Parsed Tree to Improve End-to-End TTS*" was accepted by INTERSPEECH 2019.

## Work Experience

**Amazon**, Cambridge, UK, Applied Scientist Intern 2023.06–2023.11  
**Xiaohongshu Inc**, Beijing, Research Intern 2022.05–2022.08  
**Sogou Inc**, Beijing, Full-Time Researcher 2020.12–2021.08  
**Tencent AI Lab**, Beijing, Research intern 2020.05–2020.12  
**Microsoft (MSRA and STCA)**, Beijing, Research intern 2018.04–2019.09  
**Chumenwenwen (Mobvoi)**, Beijing, Intern 2016.07–2016.10

## Publications

Mateusz Łajszczak, Guillermo Cámara, Yang Li, Fatih Beyhan, Arent van Korlaar, Fan Yang, Arnaud Joly, Álvaro Martín-Cortinas, Ammar Abbas, Adam Michalski, Alexis Moinet, Sri Karlapati, Ewa Muszyńska, **Haohan Guo**, Bartosz Putrycz, Soledad López Gambino, Kayeon Yoo, Elena Sokolova, Thomas Drugman. *BASE TTS: Lessons from building a billion-parameter Text-to-Speech model on 100K hours of data*. Arxiv, 2024.

Jiawen Kang, Lingwei Meng, Mingyu Cui, **Haohan Guo**, Xixin Wu, Xunying Liu, Helen Meng. *Cross-Speaker Encoding Network for Multi-Talker Speech Recognition*. ICASSP, 2024.

**Haohan Guo**, Fenglong Xie, Jiawen Kang, Yujia Xiao, Xixin Wu and Helen Meng. *QS-TTS: Towards Semi-Supervised Text-to-Speech Synthesis via Vector-Quantized Self-Supervised Speech Representation Learning*. Arxiv, 2023.

**Haohan Guo**, Fenglong Xie, Frank K. Soong, Xixin Wu, Helen Meng. *MSMC-TTS: Multi-Stage Multi-Codebook VQ-VAE Based Neural TTS*. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023.

**Haohan Guo**, Fenglong Xie, Frank K. Soong, Xixin Wu, Helen Meng. *A Multi-Stage Multi-Codebook VQ-VAE Approach to High-Performance Neural TTS*. INTERSPEECH, 2022.

**Haohan Guo**, Hui Lu, Xixin Wu, Helen Meng. *A Multi-Scale Time-Frequency Spectrogram Discriminator for GAN-based Non-Autoregressive TTS*. INTERSPEECH, 2022.

**Haohan Guo**, Zhiping Zhou, Fanbo Meng, Kai Liu. *Improving Adversarial Waveform Generation based Singing Voice Conversion with Harmonic Signals*. ICASSP, 2022.

**Haohan Guo**, Shaofei Zhang, Frank K. Soong, Lei He, Lei Xie. *Conversational End-to-End TTS for Voice Agents*. SLT, 2020.

**Haohan Guo**, Frank K. Soong, Lei He, Lei Xie. *A New GAN-based End-to-End TTS Training Algorithm*. INTERSPEECH, 2019.

**Haohan Guo**, Frank K. Soong, Lei He, Lei Xie. *Exploiting Syntactic Features in a Parsed Tree to Improve End-to-End TTS*. INTERSPEECH, 2019

## Awards

**First Prize Scholarship**, Northwestern Polytechnical University

2013 - 2017

**Silver Award of ACM Programming Competition**, Shannxi Province

2015 - 2016