SEMINAR ANNOUNCEMENT

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
COLLEGE OF DESIGN AND ENGINEERING
Website: https://cde.nus.edu.sg/ece

Area: Signal Analysis & Machine Intelligence

Host: Associate Professor Thomas Yeo Boon Thye

TOPIC	:	**Accented Text-to-Speech Synthesis with Limited Data**
SPEAKER	:	**Mr. Zhou Xuehao** **Graduate Student, ECE Dept, NUS**
DATE	:	**Friday, 23 June 2023**
TIME	:	**12.30PM to 1.30PM**
VENUE	:	**Join Zoom Meeting:** **https://nus-sg.zoom.us/j/84900292463?pwd=Q295YklGZ3IvMVE5K052RTNLSGkwdz09** **Meeting ID: 849 0029 2463** **Passcode: 370266**

ABSTRACT

This paper presents an accented text-to-speech (TTS) synthesis framework with limited training data. We study two aspects concerning accent rendering: phonetic (phoneme difference) and prosodic (pitch pattern and phoneme duration) variations. The proposed accented TTS framework consists of two models: an accented front-end for grapheme-to-phoneme (G2P) conversion and an accented acoustic model with integrated pitch and duration predictors for phoneme-to-Mel-spectrogram prediction. The accented front-end directly models the phonetic variation, while the accented acoustic model explicitly controls the prosodic variation. Specifically, both models are first pretrained on a large amount of data, then only the accent-related layers are fine-tuned on a limited amount of data for the target accent. In the experiments, speech data of three English accents, i.e., General American English, Irish English, and British English Received Pronunciation, are used for pre-training. The pretrained models are then fine-tuned with Scottish and General Australian English accents, respectively. Both objective and subjective evaluation results show that the accented TTS frontend fine-tuned with a small accented phonetic lexicon (5k words) effectively handles the phonetic variation of accents, while the accented TTS acoustic model fine-tuned with a limited amount of accented speech data (approximately 3 minutes) effectively improves the prosodic rendering including pitch and duration. The overall accent modeling contributes to improved speech quality and accent similarity.

BIOGRAPHY

Zhou Xuehao received his B.Eng in Automation from University of Electronic Science and Technology of China, and M.Sc in Electrical Engineering from National University of Singapore in 2017 and 2018 respectively. He is now a Ph.D student under the supervisions of Prof. YEO, Boon Thye Thomas and Prof. Li Haizhou at the Department of Electrical and Computer Engineering, National University of Singapore. His research interests include accented text-to-speech and cross-lingual text-to-speech.