SEMINAR ANNOUNCEMENT

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
Faculty of Engineering
Website: https://www.eng.nus.edu.sg/ece/

Area: Signal Analysis & Machine Intelligence

Host: Prof Li Haizhou

Co-organized by

Chinese and Oriental Languages Information Processing Society, Singapore
IEEE Singapore Systems, Man and Cybernetics Chapter

TOPIC	**:**	**Found Data and Speaker Modeling for Text-to-Speech Synthesis**
SPEAKER	**:**	**Dr. Erica Cooper,** **Postdoctoral Researcher,** **Yamagishi Lab, National Institute of Informatics,** **Japan**
DATE	**:**	**6 November 2019, Wednesday**
TIME	**:**	**3pm to 4pm**
VENUE	**:**	**E3-06-01, Engineering Block E3, Faculty of Engineering, NUS**

ABSTRACT

Recent advances in text-to-speech synthesis have resulted in synthetic speech with naturalness approaching that of actual human speech, so speech synthesis may appear to be a "solved problem." However, these types of systems typically require dozens of hours of high quality recorded data from a single speaker to work well. Can we synthesize speech in the voice of a different speaker, or a new style or language, without going through the lengthy and expensive process of collecting large amounts of new, high-quality data?

In the first part of my talk, I will describe experiments using "found data," or, data that was created or collected for a purpose other than text-to-speech synthesis. By focusing on acoustic and prosodic features that are characteristic of high-quality TTS data and using approaches such as data selection and adaptation, we are able to improve naturalness and intelligibility of synthesized voices trained on found data such as radio broadcast news and audiobooks. In the second part of my talk, I will describe a transfer learning based approach to multi-speaker synthesis and speaker adaptation. Using speaker embeddings extracted from a separately-trained speaker identification system, text-to-speech models can speak in the "voice" of a new individual using only a few seconds of that person's speech.

BIOGRAPHY

Dr. Erica Cooper received her BSc and MEng degrees from the Massachusetts Institute of Technology, and her PhD from Columbia University. Since February of this year she is now a postdoctoral researcher at the National Institute of Informatics in Tokyo. Her thesis work focused on discovering ways to create intelligible and natural-sounding text-to-speech voices from low-resource and found data. She has also previously worked on spoken keyword search, pronunciation modeling, natural language question answering, and deception detection. Her current research is in speaker modeling for end-to-end text-to-speech synthesis.