SEMINAR ANNOUNCEMENT

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
College of Design & Engineering
Website: https://cde.nus.edu.sg/ece/

Area: Signal Analysis & Machine Intelligence

Host: Prof Li Haizhou

Jointly organized with
IEEE Systems, Man and Cybernetics Singapore Chapter
Chinese and Oriental Languages Information Processing Society
Teochew Doctorate Society, Singapore

TOPIC	:	**Deep Generative Models for Text-to-Speech Synthesis**
SPEAKER	:	**Xu Tan** **Microsoft Research Asia**
DATE	:	**Oct 13 (Thu), 2022**
TIME	:	**11:00AM to 12:00PM in Singapore/Beijing Time**
WEBINAR	:	**Join Zoom Meeting** **https://nus-sg.zoom.us/j/81384926288?pwd=SFhLSC9ncG9ZZS9NWENyRmlJM3ZiZz09** **Meeting ID: 813 8492 6288** **Passcode: 072973**

ABSTRACT

Text-to-speech (TTS) synthesis is a typical data generation task, where the mapping between text and speech is one-to-many. Thus, the speech data can be represented as a conditional distribution given text, which can be well captured by generative models. As the development of deep learning, deep generative models (e.g., Autoregressive Models, GANs, VAEs, Flow Models, Diffusion Models) have been applied on TTS and achieved good progress. In this talk, I review some typical deep generative models, and introduce our recent research that leverages deep generative models to achieve high-quality speech synthesis: 1) NaturalSpeech, which achieves human-level quality in TTS using VAE, Flow and GAN; 2) PriorGrad/InferGrad/BinauralGrad, which designs better diffusion models for speech synthesis. At last, I will point out some future research directions on deep generative models for speech synthesis.

BIOGRAPHY

Xu Tan is a Principal Researcher and Research Manager at Microsoft Research Asia. His research interests cover deep learning, natural language processing, speech, and AI music. He has achieved human-level quality in machine translation and speech synthesis, and won several champions on WMT machine translation competition and Blizzard speech synthesis challenge. He has designed pre-trained language model MASS, text-to-speech system FastSpeech/NaturalSpeech, and AI music project Muzic, which receive broad attention in the community, and has transferred many research works to the important products (e.g., Azure, Bing) in Microsoft. He serves as the action editor or area chair of some AI journals/conferences (e.g., TMLR, NeurIPS, AAAI), and is an executive member of the committee on Speech, Dialogue and Auditory Processing and a member of the standing committee on Computational Art in China Computer Federation (CCF).