THE CHINESE UNIVERSITY OF HONG KONG
Department of Electronic Engineering
Seminar

Bridging Simulation, Understanding, and Generation in AI-Driven Audio and Music

Professor Qiuqiang Kong

Assistant Professor
Department of Electronic Engineering

Date: January 12, 2026 (Monday)
Time: 2:30 p.m. – 3:30 p.m.
Venue: Room 222, 2/F, Ho Sin Hang Engineering Building, CUHK

Abstract

Audio is a fundamental modality of human perception, conveying rich information about events, environments. This talk presents a unified research framework for music and audio simulation, understanding, and generation. **On the simulation side,** we investigate acoustic simulation techniques, including both wave-based and geometric acoustics, to generate physically grounded audio data under diverse acoustic conditions. The simulated audio is then leveraged to build robust music understanding systems that generalize beyond curated datasets. **On the understanding side**, we study universal sound and music source separation, aiming to separate arbitrary sounds in unconstrained environments, and propose state-of-the-art Transformer-based models for this task. We further explore music transcription and audio understanding with large language models, leading to a system that achieved 1st place in the DCASE challenge, demonstrating strong reasoning and generalization capabilities. **On the generation side,** we focus on speech and music synthesis, restoration, and spatial audio rendering from low-resolution mono recordings. These generation tasks close the loop between perception, understanding, and synthesis, enabling applications such as immersive audio reproduction and intelligent audio editing. Together, this research establishes a coherent pipeline that connects acoustic simulation, audio understanding, and generative modeling, contributing to both the theory and practice of audio AI.

Biography

Qiuqiang Kong is currently an assistant professor at the EE department of the Chinese University of Hong Kong. He received his Ph.D. degree from the University of Surrey, Guildford, UK, in 2019. Following his Ph.D., he joined ByteDance as a research scientist. His research topic includes the classification, detection, separation, and generation of general sounds and music. He was known for developing pretrained audio neural networks (PANNs) for audio tagging and was awarded the IEEE SPS Young Author Best Paper in 2023. He won the detection and classification of acoustic scenes and events (DCASE) challenge in 2017 and 2025. He was known for transcribing the largest piano MIDI dataset GiantMIDI-Piano in the world. He has co-authored over 80 papers in journals and conferences, including IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP), ICASSP, INTERSPEECH, IJCAI, DCASE, EUSIPCO, LVA-ICA. He was a frequent reviewer for world well known journals and conferences, including TASLP, TMM, SPL, TKDD, JASM, EURASIP, Neurocomputing, Neural Networks, ISMIR, CSMT. UK. He is currently an associate editor for IEEE/ACM TASLP.

***** ALL ARE WELCOME *****

For enquiries, please contact Ms. Natalie Tsang, Department of Electronic Engineering, on 3943-8277.