Multimodal Singing Voice Transcription: A Research Topic to Connect the Dots

Dr Ye Wang
School of Computing
National University of Singapore

Date: March 6, 2023 (Monday)
Time: 12:45 – 14:00
Venue: Room 104, Runme Shaw Building, HKU /
Online via Zoom (Hybrid mode)

Registration:

Chair: Dr Xiao Hu

Abstract:
Singing is presumed to be the oldest form of music making and can be found in human cultures around the world. In addition to a source of entertainment, singing has also many other benefits to humanity ranging from education to healthcare settings. Consequently, computational methods for singing analysis/synthesis have emerged as an active research topic in several research communities such as the music information retrieval (MIR), multimedia, and artificial intelligence/machine learning (AI/ML). In this talk, I will discuss our recent work on automatic singing voice transcription, in particular, automatic lyric transcription (ALT) which is a nascent field of study attracting increasing interest from both the speech and music information retrieval communities, given its significant application potential. ALT with audio data alone is a notoriously difficult task due to instrumental accompaniment and musical constraints resulting in degradation of both the phonetic cues and the intelligibility of sung lyrics. To tackle this challenge, we propose the Multimodal Automatic Lyric Transcription (MM-ALT) system, together with a new dataset, N20EM, which consists of audio recordings, videos of lip movements, and inertial measurement unit (IMU) data of an earbud worn by the performing singer. We first adapt the wav2vec 2.0 framework from automatic speech recognition (ASR) to the ALT task. We then propose a video-based ALT method and an IMU-based voice activity detection (VAD) method. In addition, we put forward the Residual Cross Attention (RCA) mechanism to fuse data from the three modalities (i.e., audio, video, and IMU). Experiments show the effectiveness of our proposed MM-ALT system, especially in terms of noise robustness. I will conclude this talk by touching some of our current and ongoing research that help connect the dots, as well as how such a research journey helped shape a new educational model.

About the speaker:
Ye Wang is an Associate Professor in the Computer Science Department at the National University of Singapore (NUS), and is also affiliated with the NUS Graduate School’s Integrative Sciences and Engineering Programme (ISEP), the Institute of Data Science (IDS), as well as the Institute for Applied Learning Sciences & Educational Technology (ALSET). He is an associate editor (AE) of Journal of New Music Research, and has just finished his term as AE of IEEE Transactions on Multimedia, and Distinguished Lecturer of Asia Pacific Signal and Information Processing Association (APSIPA). He established and directed the NUS Sound and Music Computing (SMC) Lab (https://smcnus.comp.nus.edu.sg/). His team received the ACM MM 2004 Best Student Paper Award, IEEE ISM 2012 Best Paper Award, ISMIR2014 Best Paper Award, APSIPA ASC 2017 Best Student Paper Award, and ACM MM 2022 Top Paper Award. Before joining NUS he was a member of the technical staff at Nokia Research Center in Tampere, Finland for 9 years. His research at NUS is centered on sound and music computing for human health and potential (SMC4HHP). His most recent projects involve the design and evaluation of systems to support 1) therapeutic gait training using Rhythmic Auditory Stimulation (RAS), 2) second language learning.

~ All are welcome ~
For enquiries, please contact the Office of Research, Faculty of Education at hkchow@hku.hk