

Tutorial 3

Designing Controllable Synthesis System for Musical Signals

Hyeong-Seok Choi, Yusong Wu

Abstract

Advances in deep learning and signal processing research have made it possible to generate signals that at times can be difficult to distinguish from real samples. Despite the realistic output the models can produce, however, the controllability of the models is still constrained because of the black-box-like nature of many models.

In this tutorial, we aim to introduce considerations researchers can take into account for a better end-user experience. We would like to focus in particular on how to design deep generative models with intuitive control of music audio signals, specifically vocal and instrumental performance. To this end, we will first present a broad review of up-to-date generative models for singing voices and musical instrument performance. Then, we will share our own research results and insights regarding both the implicit and explicit controllability of the deep learning models. In the section on presenting controllable models for instrumental performance synthesis, we will include a walk-through of the building, training, and control of the DDSP and MIDI-DDSP models via Jupyter (Colab) Notebook with Python and Tensorflow.

The target audience for this tutorial is researchers who are interested in deep generative models for monophonic signals, especially for singing voice and musical instruments. We expect the audience to have a basic understanding of machine learning concepts for audio signal processing.

Biographies of Presenters

Hyeong-Seok Choi received his PhD from Seoul National University, South Korea, in 2022, with a thesis titled, “A Controllable Generation of Signals from Self-Supervised Representations” under the supervision of Prof. Kyogu Lee. His recent research interest is mainly in representation learning and controllable synthesis of speech and singing voices. He co-founded the audio technology startup company, Supertone, where he has been working as the lead of their research team. He contributed to the winning of the CES 2022 Innovation Awards Honoree: Software & Mobile Apps by proposing a real-time voice conversion technology.

Yusong Wu is a final-year research master at the University of Montreal and Mila in Montreal, Canada. He is co-advised by Prof. Aaron Courville and Prof. Cheng-Zhi Anna Huang and will become a Ph.D. student under the same advisors shortly. His research focuses on making better generative models for music creativity. His recent work “MIDI-DDSP: Detailed Control of Musical Performance via Hierarchical Modeling”, collaborating with Google Magenta, was accepted by ICLR 2022 for oral presentation.