Research Statement

Hao-Wen Dong

My research aims to **empower music creation with machine learning**. I build intelligent systems that learn to compose, arrange and synthesize music. My goal is to **lower the barrier of entry for music composition and democratize music creation**.

With the revolutionary transformation brought by AI in many fields, the advancement of AI technology will also reshape the $20-billion-worth global music industry in the next decade. On one hand, we have witnessed major progress in automatic music composition, which has long been considered a grand challenge of AI. On the other hand, our expectations of *AI Music* today has expanded to cover the whole music creation process—from composition, arrangement, sound production, recording to mixing. With a growing momentum in both academia and industry, AI-powered music creation has been gaining attentions in the broader AI community, and **it is now an exciting time to pursue research in this emerging field of *AI Music***. From a musical perspective, technology has always been a driving factor of music evolution. For example, the study of acoustics and musical instrument making fostered the development of classical music; the invention of synthesizers and drum machines helped popularize electronic music. From a technical perspective, music possesses a unique complexity in that music follows rules and patterns while being creative and expressive at the same time. I envision the future development of AI Music to be a two-way process—**new technology creates new music; new music inspires new technology**.

Motivated by this belief, I study a wide range of topics centering AI-powered music creation, including multitrack music generation [1–5], automatic instrumentation [6], automatic arrangement [1, 5], automatic harmonization [7], music performance synthesis [8] and symbolic music processing software [9, 10]. My research can be roughly categorized into three main pillars:

1. **Multitrack music generation**—generating new music contents automatically
2. **Assistive music creation tools**—assisting humans in creating and performing music
3. **Infrastructure for music generation research**—providing the building blocks for developing music generation systems

Selected Projects

MuseGAN In [1, 2], I investigated generating five-track pop music excerpts using convolutional generative adversarial networks (GANs). To provide better controllability to the user, I developed a special mechanism that allows the user to control the characteristics for each generated track on top of the overall music style. Further, I compiled a dataset of more than twenty thousand pop songs to train the proposed model. The proposed MuseGAN model represents **the first deep neural network that can generate multitrack, polyphonic (i.e., multi-pitch) music from scratch** (see Fig. 1). It was featured as one of the backbone models implemented in the AWS DeepComposer, an AI-powered

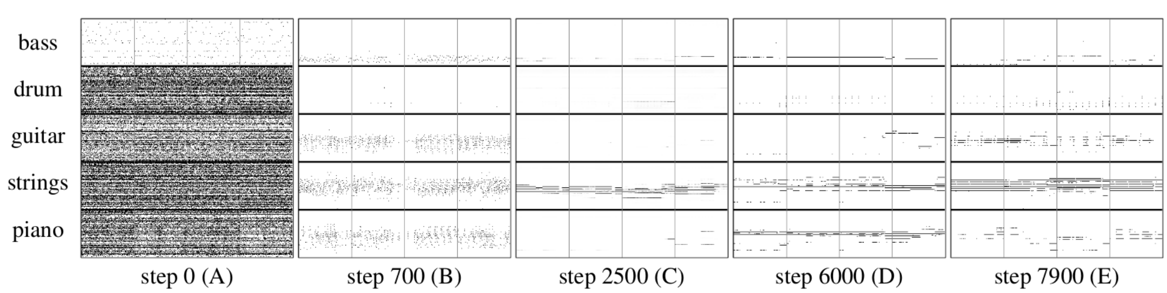

Figure 1: The MuseGAN system [1, 2] gradually learns the characteristics of each instrument, e.g., the monophonic (i.e., single-pitch) bass line, repetitive drum patterns and chord progressions played by the guitar, strings and piano (demo).

keyboard made and sold by Amazon.[1,2]

Arranger Instrumentation refers to the process where a musician arranges a solo piece for a certain ensemble such as a string quartet or a rock band. This can be challenging for amateur composers as it requires domain knowledge of each target instrument. I developed **the first deep learning model for automatic instrumentation** [6]. An automatic instrumentation system can suggest potential instrumentation for amateur composers, especially useful when arranging for an unfamiliar ensemble. By framing this problem as a sequential multi-class classification problem, I showed that the proposed model can produce convincing instrumentations for a solo piece (see Fig. 2). Further, I explored a real-time capable setting where the proposed model can be integrated into an intelligent keyboard. By performing automatic instrumentation on the keyboard inputs, it can empower a musician to play multiple instruments on a single keyboard at the same time.

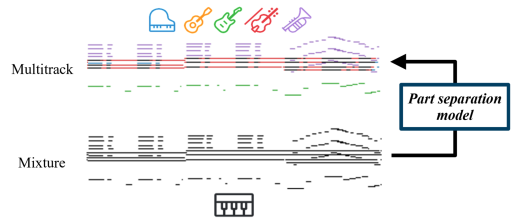

Figure 2: The proposed automatic instrumentation model [6] automatically assigns instruments to notes in a solo music piece, turning a piano piece into a pop band version (demo).

Deep Performer Synthesizers play a critical role and are intensively used in modern music production. However, existing synthesizers either requires an input with expressive timing or allows only monophonic inputs. I developed **the first deep learning based polyphonic synthesizer that can synthesize a score into a natural, expressive performance** [8]. In light of the similarities between text-to-speech (TTS) and score-to-audio synthesis, I leveraged the recent advances from TTS synthesis and proposed a transformer encoder-decoder model for polyphonic music synthesis. The proposed Deep Performer model can synthesize violin and piano music with clear polyphony and harmonic structures as well as smooth transitions between notes (see Fig. 3).

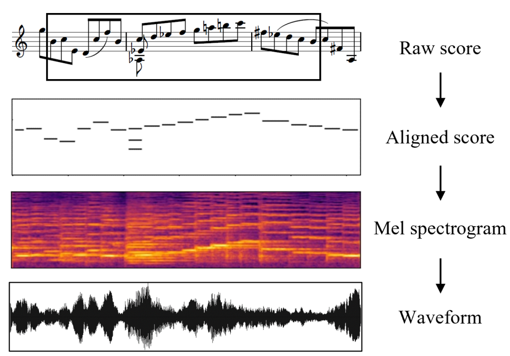

Figure 3: The Deep Performer system [8] synthesizes a musical score into a natural, expressive performance (demo).

Broader Impacts

I envision my research to be integrated into the music creation workflow for professional musicians and music amateurs. Through providing new tools and interfaces to make music, my research could lower the barrier for music composition and empower novices to create their own music. Moreover, it could provide content creators (e.g., TikTokers, YouTubers and Twitch streamers) with royalty-free materials to avoid unintended copyright infringement. My research could also find applications in music education and therapy, where creating personalized courses can be costly. Finally, we could gain insights into the future of human-AI music co-creation though the interactions between human and automatic music composition systems. I envision this to foster the discussions in human-AI relationships in other fields.

References

[1] H.-W. Dong et al., *AAAI*, 2018.
[2] H.-W. Dong et al., *ISMIR LBD*, 2017.
[3] H.-W. Dong et al., *ISMIR*, 2018.
[4] H.-W. Dong et al., *ICASSP*, 2023.
[5] H.-M. Liu et al., *GTC Taiwan*, 2018.
[6] H.-W. Dong et al., *ISMIR*, 2021.
[7] Y.-C. Yeh et al., *JNMR*, 50(1):37–51, 2021.
[8] H.-W. Dong et al., *ICASSP*, 2022.
[9] H.-W. Dong et al., *ISMIR LBD*, 2018.
[10] H.-W. Dong et al., *ISMIR*, 2020.

[1] https://www.amazon.com/dp/B07YGZ4V5B/
[2] https://aws.amazon.com/blogs/aws/aws-deepcomposer-now-generally-available-with-new-features/