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 on music and audio signal processing, including multitrack music generation [1–5], automatic instrumentation [6], automatic arrangement [1, 5], automatic harmonization [7], audio synthesis [8, 9], audio source separation [10] and symbolic music processing software [11, 12]. 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, and 3) *Self-supervised learning for audio and music*—learning sound separation and synthesis using unlabeled videos in the wild.

My research has been impactful in the field of music information research. My work on generating multi-instrument music using convolutional generative adversarial networks was the first deep learning model that tackles the challenge of multitrack music generation [1]. This work has inspired much follow-up research that reused our data processing pipeline, dataset, model and evaluation metrics. **Our proposed MuseGAN model led to a commercial implementation in the AWS DeepComposer, an AI-powered keyboard made and sold by Amazon** [13, 14]. In addition, my open-source software for symbolic music processing provides a backbone codebase for researchers to build upon and has been used by many researchers in their research.

I am determined to pursue a career in the academia and continue working on AI-powered music creation. My future research vision springs from two fundamental questions: 1) *How can AI help musicians or amateurs create music?* 2) *Can AI learn to create music in a way similar to how humans learn music?* With a growing momentum in both academia and industry on generative and creative AI, I am excited about working on multimodal learning for controllable music generation, interactive human-AI music co-creation, and post-production technology for music and audio.

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. 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 the field of signal processing.

References

[1] **Hao-Wen Dong***, Wen-Yi Hsiao*, Li-Chia Yang, and Yi-Hsuan Yang, "MuseGAN: Multi-Track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment," *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*, 2018. (* indicates equal contribution)

[2] **Hao-Wen Dong***, Wen-Yi Hsiao*, Li-Chia Yang, and Yi-Hsuan Yang, "MuseGAN: Demonstration of a Convolutional GAN Based Model for Generating Multi-track Piano-rolls," *Late-Breaking Demos of ISMIR*, 2017. (* indicates equal contribution)

[3] **Hao-Wen Dong** and Yi-Hsuan Yang, "Convolutional Generative Adversarial Networks with Binary Neurons for Polyphonic Music Generation," *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, 2018.

[4] **Hao-Wen Dong**, Ke Chen, Shlomo Dubnov, Julian McAuley, and Taylor Berg-Kirkpatrick, "Multitrack Music Transformer," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[5] Hao-Min Liu, **Hao-Wen Dong**, Wen-Yi Hsiao, and Yi-Hsuan Yang, "Lead sheet and Multi-track Piano-roll generation using MuseGAN," *GPU Technology Conference (GTC) Taiwan*, 2018.

[6] **Hao-Wen Dong**, Chris Donahue, Taylor Berg-Kirkpatrick, and Julian McAuley, "Towards Automatic Instrumentation by Learning to Separate Parts in Symbolic Multitrack Music," *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, 2021.

[7] Yin-Cheng Yeh, Wen-Yi Hsiao, Satoru Fukayama, Tetsuro Kitahara, Benjamin Genchel, Hao-Min Liu, **Hao-Wen Dong**, Yian Chen, Terence Leong, and Yi-Hsuan Yang, "Automatic Melody Harmonization with Triad Chords: A Comparative Study," *Journal of New Music Research (JNMR)*, 50(1):37–51, 2021.

[8] **Hao-Wen Dong**, Cong Zhou, Taylor Berg-Kirkpatrick, and Julian McAuley, "Deep Performer: Score-to-Audio Music Performance Synthesis," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

[9] **Hao-Wen Dong**, Gunnar A. Sigurdsson, Chenyang Tao, Jiun-Yu Kao, Yu-Hsiang Lin, Anjali Narayan-Chen, Arpit Gupta, Tagyoung Chung, Jing Huang, Nanyun Peng, and Wenbo Zhao, "CLIPSynth: Self-supervised Text-queried Sound Synthesis," *submitted to International Conference on Machine Learning (ICML)*, 2023.

[10] **Hao-Wen Dong**, Naoya Takahashi, Yuki Mitsufuji, Julian McAuley, and Taylor Berg-Kirkpatrick, "CLIPSep: Learning Text-queried Sound Separation with Noisy Unlabeled Videos," *Proceedings of International Conference on Learning Representations (ICLR)*, 2023.

[11] **Hao-Wen Dong**, Wen-Yi Hsiao, and Yi-Hsuan Yang, "Pypianoroll: Open Source Python Package for Handling Multitrack Pianoroll," *Late-Breaking Demos of ISMIR*, 2018.

[12] **Hao-Wen Dong**, Ke Chen, Julian McAuley, and Taylor Berg-Kirkpatrick, "MusPy: A Toolkit for Symbolic Music Generation," *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, 2020.

[13] https://www.amazon.com/dp/B07YGZ4V5B/.

[14] https://aws.amazon.com/blogs/aws/aws-deepcomposer-now-generally-available-with-new-features/.