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 generative **AI for music and audio**, including multitrack music generation [1–5], automatic instrumentation [6], automatic arrangement [1, 5], automatic harmonization [7], music performance synthesis [8], audio source separation [9], text-to-audio synthesis [10, 11] and symbolic music processing software [12, 13]. 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. **Multimodal learning for audio and music**—learning sound separation and synthesis using unlabeled videos in the wild systems

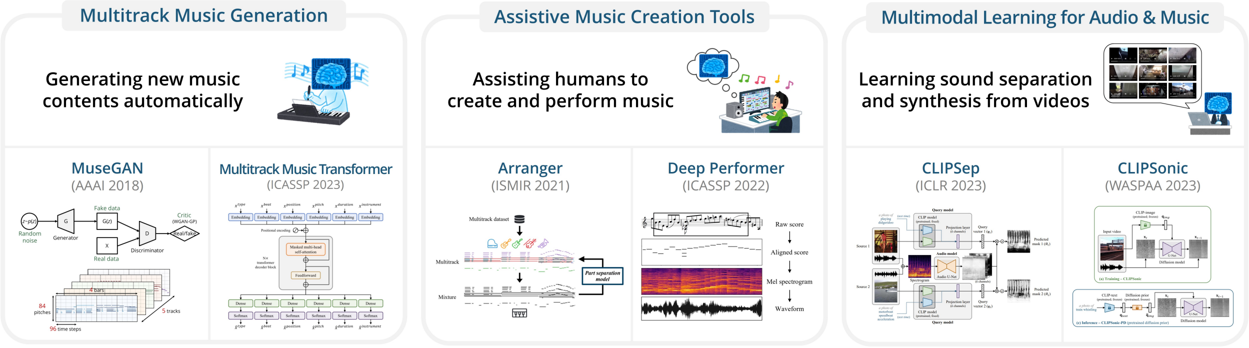

Figure 1: An overview of my research

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 [14, 15]. 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.

Future Directions

I am determined to pursue a career in the academia and continue working on generative AI for music and audio. 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?* This section outlines some future research directions that I am excited to pursue. To support my research group, I will actively apply for NSF funding and seek industrial collaborations with music tech companies, including Adobe, Dolby, Sony and Yamaha, which I have worked with in the past.

Learning music through listening. Existing data-driven approaches for music generation usually rely on *reading* large collections of music scores. Unlike machines, however, humans learn music mostly through listening and practicing music rather than reading scores over and over again. I want to build intelligent systems that can learn to compose music in a more human-like way. In particular, can a machine learn symbolic music composition through listening to a large collection of musical audio data? Can we improve a music generation model by equipping it with the knowledge of how different musical instruments sound in real world? Some recent work [16, 17] has shown preliminary results towards this direction, and I believe this will be the next frontier of automatic music composition.

Interactive human-AI music co-creation. While recent deep learning-based music generation system can create short, plausible music excerpts, they offer limited interactivity and controllability [18, 19]. My research on automatic arrangement [1, 5–7] has touched on this topic. In future work, I want to extend my research and explore real time capable music generation systems for improvisations and live performances.

Multimodal learning for music and audio generation. Lately, self-supervised contrastive learning has revolutionized the field of multimodal learning [20]. This lays the foundations for creative applications to film, video and audiobook generation. In particular, I want to explore generating background music, foley sounds and sound effects for videos and stories. I will seek collaborations with other faculty members in computer vision and natural language processing to pursue research along this direction.

Post-production technology for music and audio. There is a growing research momentum in intelligent music production [21]. While my past work primarily focuses on the pre-production and production stages of music creation, in the future, I want to explore AI-powered post-production technology for music and audio, including sound editing, spatial audio processing and auto-mixing. I will seek industrial collaborations with music tech companies in this direction.

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. 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 data science.

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,” *AAAI Conference on Artificial Intelligence* ***(AAAI)***, 2018.

[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,” *ISMIR Late-Breaking Demos*, 2017.

[3] Hao-Wen Dong and Yi-Hsuan Yang, “Convolutional Generative Adversarial Networks with Binary Neurons for Polyphonic Music Generation,” *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,” *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,” *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,” *IEEE International Conference on Acoustics, Speech and Signal Processing* ***(ICASSP)***, 2022.

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

[10] 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: Learning Text-to-audio Synthesis from Videos using CLIP and Diffusion Models,” *CVPR Workshop on Sight and Sound*, 2023.

[11] Hao-Wen Dong, Xiaoyu Liu, Jordi Pons, Gautam Bhattacharya, Santiago Pascual, Joan Serrà, Taylor Berg-Kirkpatrick, and Julian McAuley, “CLIPSonic: Text-to-Audio Synthesis with Unlabeled Videos and Pretrained Language-Vision Models,” *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* ***(WASPAA)***, 2023.

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

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

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

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

[16] Prafulla Dhariwal, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever, “Jukebox: A Generative Model for Music,” *arXiv preprint arXiv:2005.00341*, 2020.

[17] Rodrigo Castellon, Chris Donahue, and Percy Liang, “Codified audio language modeling learns useful representations for music information retrieval,” *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, 2021.

[18] Jean-Pierre Briot, Gaëtan Hadjeres, and François-David Pachet, “Deep Learning Techniques for Music Generation–A Survey,” *arXiv preprint arXiv:1709.01620*, 2017.

[19] Cheng-Zhi Anna Huang, Hendrik Vincent Koops, Ed NewtonRex, Monica Dinculescu, and Carrie J. Cai, “AI Song Contest: Human-AI Co-Creation in Songwriting,” *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, 2020.

[20] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever, “Learning Transferable Visual Models From Natural Language Supervision,” *Proceedings of International Conference on Machine Learning*, 2021.

[21] Brecht De Man, Ryan Stables, and Joshua D. Reiss, *Intelligent Music Production*, Routledge, 2019.