HUAN ZHANG

huan.zhang@qmul.ac.uk ◇ Linkedin ◇ Github ◇ Google Scholar

RESEARCH INTEREST AND EXPERTISE

My research expertise centers around machine learning and deep learning for multimodal understanding and generation with a focus on music. I have hands-on experience building efficient and controllable generative models, spanning audio, music, and text, with a strong focus on building scalable, prompt-driven systems. In terms of techniques, I am experienced in using representation learning, contrastive training, transformer-based and diffusion-based architectures for both discriminative and generative tasks.

EDUCATION

Queen Mary University of London, London, UK Sept 2021 - Aug 2025 (Expected)
Ph.D student in Artificial Intelligence and Music, Center for Digital Music

Carnegie Mellon University, Pittsburgh, PA Sept 2017 - May 2021
Bachelor of Science in Music and Technology, minor in Computer Science, GPA: 3.7/4.0

PROFESSIONAL EXPERIENCE

Teaching Fellow - Undergraduate and Graduate Thesis Project Supervision Sept 2023 – Aug 2025
Queen Mary University of London
· Led supervision on deep learning topics including computer vision, audio generation, and data analysis, mentoring 20 students each year through hands-on machine learning research projects.

Research and Development Intern Oct 2024 - Jan 2025
Yamaha Corporation, Hamamatsu, Japan
· Developed a controllable diffusion-based generative model that integrates symbolic music and textual inputs for music performance generation.

Research and Development Intern June 2024 – Sept 2024
Sony CSL, Tokyo, Japan
· Fine-tuned large-scale audio-language models for question answering in the music education domain.
· Performed statistical modelling and data analysis on rating-based human annotation datasets.

Research and Development Intern Dec 2020 – May 2021
Tencent Music Entertainment, Shenzhen, China
· Implemented deep metric learning approaches (e.g., triplet loss) for singing quality assessment tasks.
· Built an audio classification system using multi-label tagging for short-form user-generated content, applying convolutional and transformer-based architectures.
· Conducted competitive landscape analysis of AI-driven educational tools in Chinese market.

Research and Development Intern June 2020 – Oct 2020
Kuaishou Technology, Beijing, China
· Built a transformer-based sequence-to-sequence model for conditional generation tasks.
· Designed end-to-end training pipeline leveraging public symbolic music datasets, integrating evaluation metrics for sequence alignment and output consistency.

RESEARCH HIGHLIGHTS

Query-Based Feedback Generation with LLMs for Expressive Performance Assessment
· Proposed **LLaQo**, a query-driven music performance coach that leverages AudioMAE and Vicuna-7B to provide formative feedback across 12 performance dimensions (e.g., pitch, timing, articulation).
· Constructed an instruction-tuned dataset from public sources and a newly collected corpus (*NeuroPiano*) for performance rating and difficulty analysis.
· Demonstrated SOTA results in subjective and objective benchmarks against leading audio-language models (LTU, Mu-LLaMA) in music performance QA tasks.

LLM-Based Agents for Audio Creation and Editing
· Leveraged LLMs, such as LLAMA, to synthesize and edit audio content based on user instructions and available recordings.
· Developed a controllable framework for audio generation by coordinating various generative models, such as text-to-speech and text-to-music models.
· Evaluated the proposed system ability on audio drama where models should manipulate audio content without explicit user commands.

Controllable Expressive Audio Rendering from Text and Score
· Introduced **RenderBox**, a controllable diffusion-based framework that renders expressive performance audio from symbolic score and natural language prompts.

- Employed curriculum learning to gradually teach the model controllability over parameters such as speed, articulation, and performer style.
- Utilized the VAE-learnt audio latent representation from stable-audio-open, and also experimented with conditioning mechanism such as controlnet.

Benchmarking Audio Representations for Piano Expertise and Technique Understanding

- Proposed a benchmarking protocol for audio encoders on piano performance understanding tasks: *expertise ranking*, *difficulty estimation*, and *technique detection*.
- Released the Pianism-Labelling Dataset (PLD) and conducted cross-model comparisons (AudioMAE, MERT, Jukebox, DAC), highlighting the transferability of representations.

HONORS/AWARDS

ISMIR Diversity and Inclusion Grant (2023, 2024)	July 2023 / 2024
SPS ICASSP Travel Grant 2025	Jan 2025
UKRI PhD studentships in Artificial Intelligence and Music (AIM) receiver	Sept 2021
CMU Summer Internship Experience Fund (SIEF) Receiver	April 2019
Ranked top 1000 in Putnam Mathematical Competition	Dec 2017
Carnegie Mellon University Dean's List	2017-2019

INVOLVEMENT AND LEADERSHIP

Peer reviewing

- Reviewed 40+ conference and journal papers of ISMIR, ICASSP, IJCNN, TISMIR, IEEE/ACM Transactions on Audio Speech and Language Processing, EURASIP Journal on Audio, Speech, and Music Processing

WiMIR Mentor 2022-present

- Hold one-to-one meeting with undergraduate student signed up to Women in Music Information Retrieva (WiMIR) initiative, providing them with insights in research scene, application and academia life.

Music Involvement

- Classical Piano Performance: In this performance video I played Bach's Well Tempered Clavier in b minor, book 1, and the Jeux D'eau by Maurice Ravel. Performance Video (Youtube) (Bilibili)
- With full musicianship training from solfege, counterpoint to orchestration, I was also able provide music theory tutorials for my colleges and peers.

SKILLS

Concepts: Machine Listening, Deep Learning, Music Information Retrieval, Audio Signal Processing, Natural Language Processing, Computer Vision, Machine Learning (regression, clustering, hypothesis testing, GBMs, neural networks)
Coding: Python, JavaScript, HTML, C++; PyTorch, Numpy/SciPy, Amazon AWS, Docker, SQL
Audio technologies: Pro Tools, Audacity, Logic Pro X, Adobe Audition, MuseScore, Bela

COURSEWORKS

QMECS7006P - Music Informatics	Spring 2022
QMECS7013P - Deep Learning for Audio and Music	Spring 2022
QMECS707 - Digital Signal Processing	Fall 2021
QMECS741P - Music Perception	Fall 2021
CMU10605 - Machine Learning with Large Datasets	Fall 2021
CMU15780 - Graduate Artificial Intelligence	Spring 2021
CMU10701 - Introduction to Machine Learning	Spring 2020
CMU15210 - Parallel and Sequential Data Structures and Algorithms	Fall 2019
CMU15213 - Introduction to Computer Systems	Fall 2019
CMU15323 - Computer Music Systems and Information Processing	Spring 2019
CMU11411 - Natural Language Processing	Fall 2019

PUBLICATION LIST

H. Zhang, J. Liang, H. Phan, W. Wang, E. Benetos. (2025) "From Aesthetics to Human Preferences: Comparative Perspectives of Evaluating Text-to-Music Systems", in proceedings of *IEEE International Workshop on Machine Learning for Signal Processing (MLSP) 2025*.

H. Zhang, A. Maezawa, S. Dixon. "RenderBox: Expressive Performance Rendering with Text Control", under review.

H. Zhang, V. K.M. Cheung, H. Nishioka, S. Dixon, S. Furuya. "LLaQo: Towards a query-based coach in expressive performance assessment", in proceedings of *International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2025*.

H. Zhang, S. Chowdhurry, J. Liang, S. Dixon, G. Widmer. "DExter: Learning and Controlling Performance Expression with Diffusion Models", *Applied Sciences. 2024; 14(15):6543*, https://doi.org/10.3390/app14156543.

H. Zhang, J. Liang, S. Dixon. "From Audio Encoders to Piano Judges: Benchmarking Performance Understanding for Solo Piano", in proceeding of *25th International Society for Music Information Retrieval Conference (ISMIR 2024)*.

J. Liang, **H. Zhang**, H. Liu, Y. Cao, Q. Kong, X. Liu, W. Wang, M. Plumbley, H. Phan, E. Benetos . "WavCraft: Audio Editing and Generation with Large Language Models", accepted by *ICLR 2024 Workshop LLMAgents*.

A. Morsi, **H. Zhang**, A. Maezawa, S. Dixon, X. Serra. "Simulating Piano Performance Mistakes for Music Learning", in proceedings of *Sound and Music Computing Conference (SMC 2024)* (Best Paper Award).

H. Zhang, E. Karystinaios, S. Dixon, G. Widmer, C. E. Cancino-Chacón. "Symbolic Music Representations for Classification Tasks: A Systematic Evaluation", in proceeding of *24th International Society for Music Information Retrieval Conference (ISMIR 2023)*.

H. Zhang, S. Dixon. "Disentangling the Horowitz Factor: Learning Content and Style from Expressive Piano Performance", in proceeding of *International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2023*.

H. Zhang, J. Tang, S. Rafee, G. Wiggins, G. Fazekas, S. Dixon. "ATEPP: A Dataset of Automatically Transcribed Expressive Piano Performance", in proceeding of *23rd International Society for Music Information Retrieval Conference (ISMIR 2022)*.

H. Zhang, Y. Jiang, Y. Jiang, P. Hu. "Learn By Referencing: Towards Deep Metric Learning for Singing Assessment", in proceeding of *22nd International Society for Music Information Retrieval Conference (ISMIR 2021)*