

Yinghao Ma (马英浩)

Email | +44 (0)7741575063 | [nicolaus625.github.io](https://github.com/nicolaus625) | [Linkedin](#) | [Google Scholar](#)

EDUCATION

Queen Mary University of London (QMUL) 09/2022 – 01/2027 (expected)

PhD: AI & Music, School of Electronic Engineering and Computer Science. Supervised by Dr Emmanouil Benetos

- Research interests: **Large language models (LLMs) for music understanding & generation**
- Awards and Honours: Google PhD Fellowship 2025 (\$60k USD research fund)
- Leadership: Co-founder of the [Multimodal Art Projection](#) community ([Hugging-face](#))

Carnegie Mellon University (CMU) 09/2020 – 08/2022

MS: Music & Technology, School of Music. Supervised by Prof. Richard M. Stern

- Overall GPA: 4.03/4.00 (Ranked 1st in the cohort)
- Awards and Honours: Fellowship for master students that covers 50% of tuition fee
U.S. National Music Honor Society member. (Theta Xi, Pi Kappa Lambda reward).
- Music Background: Recorded Chinese musical version of Beethoven's serenade for 250th anniversary of his birth during COVID to cheer others up. Released on CMU DL course web

Peking University (PKU) 09/2016 – 07/2020

BS: Mathematics & Applied Mathematics (Probability Theory), School of Mathematical Sciences

- Awards and Honours: Outstanding graduates of the School of Mathematical Science at PKU
Preliminary excellence prize for S.-T.Yau College Student Mathematics Contests
- Music Background: One of the student conductors in the orchestra of the Chinese Music Institute at PKU
Highest amateur certification in Chinese Flute, China Conservatory of Music

(CO)-FIRST AUTHOR PUBLICATION

- **Ma, Y.**, Xia, H., Gao, H., ... & Benetos, E. (2026). "CMI-RewardBench: Evaluating Music Reward Models with Compositional Multimodal Instruction." International Conference on Machine Learning (ICML), 2026.
- Tang, X.* , Lei, X.* , Zhu, C.* , Chen, S., Yuan, R., Li, Y., ... & **Ma, Y.*** (2025). "AutoMV: An Automatic Multi-Agent System for Music Video Generation." arXiv preprint arXiv:2512.12196. [Under double blind review.]
- Ma, Z.* , **Ma, Y.***, Zhu, Y.* , , et al. "MMAR: A Challenging Benchmark for Deep Reasoning in Speech, Audio, Music, and Their Mix." Advances in Neural Information Processing Systems (NeurIPS), 2025.
- **Ma, Y.**, Li, S., Yu, J. et al. "CMI-Bench: A Comprehensive Benchmark for Evaluating Music Instruction Following." International Society for Music Information Retrieval (ISMIR), 2025.
- Li, Y.* , **Ma, Y.***, Zhang, G.* , et al. "OmniBench: Towards The Future of Universal Omni-Language Models." Advances in Neural Information Processing Systems (NeurIPS), 2025.
- **Ma, Y.**, Øland, A., Ragni, A., et al. "Foundation models for music: A survey." [Huggingface Daily Paper top3.]
- Qu, X.* , Bai, Y.* , **Ma, Y.***, et al. "MuPT: A Generative Symbolic Music Pretrained Transformer." International Conference on Learning Representations (ICLR) 2025.
- Deng, Z.* , **Ma, Y.***, Liu, Y. et al. "MusLingo: Bridging Music and Text with Pre-trained Language Models for Music Captioning and Query Response." North American Chapter of the Association for Computational Linguistics 2024.
- Li, D.* , **Ma, Y.***, et al. "Mertech: Instrument Playing Technique Detection Using Self-supervised Pretrained Model with Multi-task Finetuning." International Conference on Acoustics, Speech & Signal Processing (ICASSP) 2024.
- Li, Y.* , Yuan, R.* ; Zhang, G.* , **Ma, Y.***; et al. "MERT: Acoustic Music Understanding Model with Large-Scale Self-supervised Training." International Conference on Learning Representations (ICLR) 2024.
- Yuan, R.* , **Ma, Y.***, Li, Y.* , et al. "MARBLE: Music Audio Representation Benchmark for universal Evaluation." Advances in Neural Information Processing Systems (NeurIPS), 2023.
- **Ma, Y.**, Yuan, R., Li, Y., et al. "On the Effectiveness of Speech Self-Supervised Learning for Music." International Society for Music Information Retrieval (ISMIR), 2023.

RESEARCH EXPERIENCE

Evaluating Music Reward Models (RMs) w/ Compositional Multimodal Instruction 10/2025 – 01/2026

Supervised by Dr Emmanouil Benetos, Centre for Digital Music, Queen Mary University of London

- Built CMI-RewardBench to evaluate RMs on text-, lyrics-, & audio-conditioned music generation.
- Curated preference datasets, including 110k pseudo-labeled samples and 4k expert annotations.
- Developed RMs with strong human-correlation performance and supported reranking via top-k filtering.

AutoMV: A Musician-First Multi-Agent System for Music Video Generation 09/2025 – 12/2025

Collaborate with Dr Jiaheng Liu (刘佳恒), Nanjing University

- Investigated a system coordinating agents for music informatics, storyboard planning and visual generation.
- Proposed evaluation protocols combining music industry domain knowledge and LLM-based judging.
- Developed a musician-first, cheap and fast solution on whole-song generation with music-visual alignment.

A Challenging Benchmark for Deep Reasoning in Speech, Audio, Music, & Their Mix 02/2025 – 05/2025
Collaborate with Dr Xie Chen (陈谐), Shanghai Jiaotong University

- Constructed a 1,000-sample benchmark for deep reasoning over multiple categories real-world recordings.
- Designed hierarchical task taxonomies and chain-of-thought for perception, semantic, & cultural reasoning.
- Evaluated diverse audio-LLMs & reasoning models, showing MMAR's difficulty and systems' weaknesses.

Controlled Genre-Specific Acoustic Music Generation 12/2024 – 04/2025
Collaborate with Prof. Chenghua Lin (林成华), University of Manchester

OmniBench: A Reasoning Benchmark of LLMs with Omni-modality Input 06/2024 – 09/2024
Collaborate with Dr Wenhao Huang (黄文灏), 01.ai

Symbolic Music Scaling (SMS) Law for a Symbolic Music GPT 01/2024 – 03/2024
Collaborate with Dr Jie Fu (付杰), Hong Kong University of Science and Technology

- Developed and integrated the SMS Law into Symbolic Music GPT, optimising model scalability and efficiency, which led to enhanced performance with constrained computational resources.
- Analysed overfitted loss curve to predict the early stop points during training.

Bridging Music & Text with Pre-trained Models for Music Captioning and QA 07/2023 – 11/2023
Supervised by Dr Emmanouil Benetos, Centre for Digital Music, Queen Mary University of London

Instrument Playing Technique (IPT) Detection on World Music 06/2023 – 09/2023
Supervised by Dr Emmanouil Benetos, Centre for Digital Music, Queen Mary University of London

MARBLE: Music Audio Representation Benchmark for universal Evaluation 01/2023 – 06/2023
Supervised by Dr Emmanouil Benetos, Centre for Digital Music, Queen Mary University of London

Acoustic Music Understanding Model with Large-Scale Self-supervised Training 08/2022 – 05/2023
Collaborate with Prof. Chenghua Lin (林成华), University of Manchester

- Built self-supervised learning systems, acquiring 50k+ downloading of checkpoints on Huggingface.
- Replaced the pseudo-tag from MFCCs to Chroma music features for harmonic information.
- Utilised deep features like Encodec instead of k-means for scaling up models to 1 B parameters.

Learnable Front End for Music, Speech and Audio (Master thesis) 09/2021 – 07/2022
Master Thesis, Supervised by Prof. Richard Stern, Carnegie Mellon University

Chinese Flute Playing Technique Classification Based on FCNNs (undergraduate thesis) 02/2020 – 05/2020
Undergraduate Thesis, Supervised by Prof. Xiaou Chen (陈晓鸥), Peking University

Correspondence between Speech Melody and Pitch Contour in Sichuan Folk Songs 07/2019 – 09/2019
Research Assistant, Supervised by Prof. Zhiyao Duan (段智尧), University of Rochester

Automatic Musical Instrument Recognition and Timbre Recognition 02/2019 – 07/2019
Research Assistant, Supervised by Prof. Xiaou Chen (陈晓鸥), Peking University

INTERNSHIP EXPERIENCE

Exploring Flow Matching for Real-time Controlable Music Generation 06/2025 – 09/2025
Research intern, Microsoft Research. (Redmond, WA, USA), supervised by Dr. Hannes Gamper

Multimodal Understanding for Acoustic Music 08/2024 – 10/2024
Research intern, Research and Development, Yamaha Corporation (Hamamatsu, Japan), by Dr. Akira Maezawa

Cover Song Detection & Evaluation of Automatic Speech Recognition 05/2021 – 08/2021
Algorithm Engineer, Tencent Holdings Limited. (Beijing, China)

Tempo, Beat and Downbeat Detection in Chinese Pop Songs 06/2020 – 08/2020
Algorithm Engineer, Beijing Deepmusic Technology Co. (Beijing, China)

SELECTED ACADEMIC ACHIEVEMENT

Advanced Digital Signal Processing (DSP)

- 1st rank of the class. Proofread lecture notes' errata, listed in acknowledgements.

TA: ML4SP (21 Autumn), DSP (22 Spring), Applied Statistics (23 Autumn), DL4Audio&Music (24 Spring)

LEADERSHIP & ACTIVITIES

- Core organiser of ISMIR 2025 satellite event on [Large Language Models for Music and Audio](#).
- A organiser of ISMIR 2025&2026 [MIREX challenge](#), and InterSpeech 2026 [Audio Reasoning Challenge](#).
- An editor of the Transaction of International Society of Music Information Retrieval, LLM4Music special issue.