Jiahao Mei

(+86) 1826-8117-399 · jiahaomei@sjtu.edu.cn
Shanghai Jiao Tong University · M.Eng. in Computer Science
Homepage · Google Scholar

Education

Shanghai Jiao Tong University, Computer Science, *M.Eng.* Shanghai | 2025.9 – 2028.3

X-LANCE Lab. Advisors: Prof. Mengyue Wu and Prof. Kai Yu. Research: audio understanding & generation, EEG representation learning, and affective computing.

East China Normal University, Computer Science, *B.Eng.* Shanghai | 2021.9 – 2025.3

ICALK Lab. Advisors: Prof. Daoguo Dong and Prof. Liang He. Research on controllable music generation and visual-audio cross-modal alignment.

Work Experience

Xiaomi, XiaoAI PLUS · Speech Generation Team, *Algorithm Engineer* Beijing | 2026.1 – 2026.7

Mentor: Heinrich Dinkel

- End-to-end mixed audio (speech, music, sound effects) generation system pretraining and optimization [Dasheng AudioGen]

Alibaba, Tongyi Lab · NLP Team, *Algorithm Engineer* Hangzhou | 2024.1 – 2024.8

Mentors: Yuning Wu, Ming Yan

- Multi-agent storybook generation system [MM-StoryAgent], deployed as a product
- LLM creative writing evaluation benchmark [WritingBench]
- Unified audio generation framework research [UniFlow-Audio]

Publications

Unified Audio Understanding & Generation

1. **Jiahao Mei**, Heinrich Dinkel, Yadong Niu, et al. "Dasheng AudioGen: A Unified Model for Generating Coherent Audio Scenes from Text." *arXiv*, 2026.
   Proposed Dasheng AudioGen, a unified text-to-audio framework for general audio scene generation. Through structured multi-perspective descriptions and semantic-acoustic unified representations, it achieves end-to-end co-generation of speech, music, sound effects, and ambient sounds, with performance approaching real recordings across multiple audio categories.
2. Xuenan Xu*, **Jiahao Mei***, Zihao Zheng, et al. "UniFlow-Audio: Unified Flow Matching for Audio Generation from Omni-Modalities." *arXiv*, 2025.
   Proposed the first fully open-source unified audio generation framework based on Flow Matching. Introduced the Dual-Fusion mechanism to unify Time-Aligned and Non-Time-Aligned audio generation tasks. Supports text, audio, and video inputs, achieving strong performance across seven tasks including TTS and TTA.
3. Heinrich Dinkel, Xingwei Sun, Gang Li, **Jiahao Mei**, et al. "DashengTokenizer: One Layer is Enough for Unified Audio Understanding and Generation." *arXiv*, 2026.
   Proposed DashengTokenizer, a unified continuous audio tokenizer for audio understanding and generation. By injecting acoustic information into frozen semantic features, it significantly outperforms mainstream codec/tokenizer baselines on speech, music, and environmental sound understanding tasks, and achieves superior results over VAE baselines in TTA, TTM, and speech enhancement generation tasks.

Controllable Music Generation

4. Jialing Zou*, **Jiahao Mei***, Xudong Nan, et al. "TEAdapter: Supply Vivid Guidance for Controllable Text-to-Music Generation." *IEEE ICME*, 2024.
   Proposed TEAdapter, a lightweight plugin for diffusion models that enables fine-grained controllable music generation by extracting chord, melody, and instrument features from teacher music. Designed multi-adapter collaboration with structural functions (Intro/Chorus/Outro) and inpainting mechanisms to address structural coherence in long audio generation.

5. **Jiahao Mei**, Xuenan Xu, Zeyu Xie, et al. "LARA-Gen: Enabling Continuous Emotion Control for Music Generation Models via Latent Affective Representation Alignment." *arXiv*, 2025.
   Proposed the Latent Affective Representation Alignment mechanism for continuous fine-grained emotion control in music generation. Accepts continuous valence-arousal values as input, effectively decoupling emotional attributes from musical content, significantly outperforming baselines in emotion accuracy and generation quality.

Multimodal Generation

6. Xuenan Xu, **Jiahao Mei**, Chenliang Li, et al. "MM-StoryAgent: Immersive Narrated Storybook Video Generation with a Multi-Agent Paradigm across Text, Image and Audio." *NAACL*, 2025.
   Proposed an open-source multi-modal multi-agent story video generation framework that achieves high-quality immersive narrated storybook video through multi-stage writing pipeline and full-modality (image, speech, sound effects) expert agents. The project received 85K+ visits on ModelScope.
7. Kaiyuan Liu, **Jiahao Mei**, Hengyu Zhang, et al. "Moyun: A Diffusion-Based Model for Style-Specific Chinese Calligraphy Generation." *ACM MM Workshop*, 2025.
   Proposed a Chinese calligraphy generation model based on Vision Mamba and TripleLabel mechanism. Built the Mobao dataset with over 1.9 million calligraphy images. Moyun achieves excellent performance in structural fidelity and style matching.

Benchmark & Dataset

8. Yuning Wu, **Jiahao Mei**, Ming Yan, et al. "WritingBench: A Comprehensive Benchmark for Generative Writing." *NeurIPS*, 2025.
   Proposed an open-source comprehensive benchmark covering 6 domains, 100 sub-domains, and 1,239 queries for long-form writing evaluation. Designed a dynamic evaluation framework achieving 83% human agreement, significantly outperforming static metrics. Training data filtered by this framework enables 7B models to approach closed-source SOTA writing quality.
9. Jialing Zou*, **Jiahao Mei***, Guangze Ye, et al. "EMID: An Emotional Aligned Dataset in Audio-Visual Modality." *ACM MM Workshop*, 2023.
   Constructed a high-quality music-image cross-modal matching dataset EMID (30k+ pairs), innovatively using emotional consistency between music and images as the primary basis for cross-modal alignment, supporting generation and retrieval tasks in art therapy and related domains.

Skills

- **Languages**: Mandarin (native), English (CET-6 544), Japanese
- **Programming**: Python, C/C++, Shell

Awards & Honors

- Outstanding Graduate of Shanghai, 2024
- National Scholarship for Undergraduates (CNY 10,000), ECNU, Rank 2/115, 2024
- Huaxin Scholarship (CNY 15,000), ECNU, Rank 3/115, 2024
- China International College Students' Innovation Competition, Shanghai Gold Award (Team Leader), 2024
- China Collegiate Computing Contest - Cyber Challenge, National 3rd Prize (Team Leader), 2024
- National E-commerce "Innovation & Entrepreneurship" Competition, Shanghai 1st Prize (Team Leader), 2024
- Shanghai Collegiate Computer Application Competition, 3rd Prize (Team Leader), 2024
- ECNU Most Vibrant Project Award (Team Leader), 2024
- Nezha Scholarship (CNY 10,000), ECNU, Rank 2/115, 2023
- National Innovation Training Program, Outstanding Completion & Paper (Team Leader), 2023
- "Huichuang Youth" Shanghai Cultural Creativity Exhibition, 1st Prize, 2023