梅嘉豪

(+86) 1826-8117-399 · jiahaomei@sjtu.edu.cn · **上海交通大学 • 计算机科学硕士** · Google Scholar

个人总结

上海交通大学计算机科学与技术 28 届硕士，研究方向为语音、音乐、音效的通用音频理解与生成和大语言模型。发表/预印论文 9 篇（NeurIPS、NAACL、ICME 等，Google Scholar 引用 100+），拥有 10 万小时级别音频数据处理与模型训练经验，具备从音频 Tokenizer 设计到端到端多模态生成的全链路研究经验。曾在阿里通义实验室和小米语音团队从事 LLM 评估和统一音频生成框架与混合音频生成研究。

教育背景

上海交通大学, 计算机科学, *工学硕士* 2025.9 – 2028.3

跨语言媒体智能实验室 (X-LANCE Lab)，导师：吴梦玥教授、俞凯教授。研究面向精神健康的人工智能系统，包括音频脑电解码、情感计算与多模态生成等方向。

华东师范大学, 计算机科学, *工学学士* 2021.9 – 2025.3

语言认知与知识计算实验室 (ICALK Lab)，导师：董道国研究员、贺樑教授。研究基于可控音乐生成和视觉-听觉跨模态对齐

工作经历

小米集团, 小爱 PLUS · 语音生成团队, *算法工程师* 2026.1 – 2026.7

导师：Heinrich Dinkel。负责端到端混合音频（语音、音乐、音效）生成研究

阿里巴巴, 通义实验室 · 自然语言智能团队, *算法工程师* 2024.1 – 2024.8

导师：吴宇宁、严明。负责大模型创作能力评估基准构建与统一音频生成框架研究

学术论文

会议论文

1. Yuning Wu, **Jiahao Mei**, Ming Yan, et al. "WritingBench: A Comprehensive Benchmark for Generative Writing." *NeurIPS*, 2025.
   本文提出并开源了一个覆盖 6 大领域、100 子领域、共 1239 条 Query 的长文本创作综合 Benchmark，并设计了动态评估框架，达到了 83% 人类一致性，显著超越静态评估标准。以此框架筛选高质量训练数据，可以使 7B 模型写作能力接近闭源 SOTA。
2. 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.
   本文提出了一个开源的多模态多智能体故事视频生成框架 MM-StoryAgent，通过多阶段写作 pipeline 与全模态（图像、语音、音效）专家智能体的协同工作，实现了高质量沉浸式有声绘本视频的自动化生成。该项目在 ModelScope 获得 85K+ 次访问。
3. Jialing Zou*, **Jiahao Mei***, Xudong Nan, et al. "TEAdapter: Supply Vivid Guidance for Controllable Text-to-Music Generation." *IEEE ICME*, 2024.
   本文提出了一种用于 Diffusion 模型的轻量级插件 TEAdapter，通过提取 Teacher Music 中的和弦、旋律与乐器特征实现细粒度的可控音乐生成，并设计了基于结构功能（Intro/Chorus/Outro）的多 Adapter 协同与 Inpainting 机制，有效解决了长音频生成的结构连贯性问题。
4. Jialing Zou*, **Jiahao Mei***, Guangze Ye, et al. "EMID: An Emotional Aligned Dataset in Audio-Visual Modality." *ACM MM Workshop*, 2023.
   本文构建了高质量的音乐-图像跨模态匹配 EMID 数据集（包含超过 30k+ 数据对），创新性地将音乐与图像的情感一致性作为跨模态对齐的主要依据，以支持艺术治疗等领域的生成和检索任务。
5. Kaiyuan Liu, **Jiahao Mei**, Hengyu Zhang, et al. "Moyun: A Diffusion-Based Model for Style-Specific Chinese Calligraphy Generation." *ACM MM Workshop*, 2025.
   本文提出基于 Vision Mamba 和 TripleLabel 机制的中文书法生成模型，用于生成指定书法家、字符和风格的中文书法。本项目构建超过 190 万张中文书法图像的 Mobao 数据集，并在该数据集上进行了训练和测试。Moyun 模型在书法结构保真度和风格匹配方面达到了优秀水平。

预印本

6. **Jiahao Mei**, Heinrich Dinkel, Yadong Niu, et al. "Dasheng AudioGen: A Unified Model for Generating Coherent Audio Scenes from Text." *arXiv*, 2025.

本文提出 Dasheng AudioGen，一个面向通用音频场景生成的统一文本到音频框架，通过结构化多视角描述与语义-声学统一表征，实现了语音、音乐、音效及环境声的端到端协同生成，并在多个音频类别中性能逼近真实录音。

7. **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.
   本文提出 Latent Affective Representation Alignment 机制，实现了对音乐生成模型的连续细粒度的情感控制。该方法能够接受连续的效价-唤醒（valence-arousal）值作为输入，有效地解耦了情感属性与音乐内容，并在情感准确性和生成质量上显著优于基线方法。
8. Xuenan Xu*, **Jiahao Mei***, Zihao Zheng, et al. "UniFlow-Audio: Unified Flow Matching for Audio Generation from Omni-Modalities." *arXiv*, 2025.
   本文提出第一个完全开源的基于 Flow Matching 的统一音频生成框架，并创新提出 Dual-Fusion 机制统一建模了 Time-Align 和 Non-Time-Align 两大类音频生成任务。UniFlow-Audio 支持文本、音频和视频等全模态输入，在 TTS、TTA 等七项任务上展现出优秀性能。
9. Heinrich Dinkel, Xingwei Sun, Gang Li, **Jiahao Mei**, et al. "DashengTokenizer: One Layer is Enough for Unified Audio Understanding and Generation." *arXiv*, 2025.
   本文提出 DashengTokenizer，一种面向音频理解与生成的统一连续音频 tokenizer，通过将声学信息注入冻结语义特征，在语音、音乐与环境声理解任务上显著优于主流 codec/tokenizer 基线，并在 TTA、TTM 与语音增强生成任务中取得优于 VAE 基线的效果。

技术能力

- **语言**: 普通话，英语（CET-6 544），日语
- **编程**: Python, C/C++, Shell

奖项与荣誉

- **上海市优秀毕业生**, 2024
- **本科生国家奖学金** (CNY 10000), 华东师范大学 2/115, 2024
- **华鑫奖学金** (CNY 15000), 华东师范大学 3/115, 2024
- **中国国际大学生创新大赛上海市金奖**（项目负责人）, 2024
- **中国高校计算机大赛-网络挑战赛全国三等奖**（项目负责人）, 2024
- **全国大学生电子商务" 创新、创意及创业" 挑战赛上海赛区一等奖**（项目负责人）, 2024
- **上海市大学生计算机应用能力大赛三等奖**（项目负责人）, 2024
- **华东师范大学最具活力项目奖**（项目负责人）, 2024
- **哪吒奖学金** (CNY 10000), 华东师范大学 2/115, 2023
- **大学生创新创业训练计划项目国创优秀结题、优秀论文**（项目负责人）, 2023
- **" 汇创青春" 上海大学生文化创意作品展示季一等奖**, 2023