

吴恒涛

(+86) 197-2067-1057 · eternity_w@sjtu.edu.cn · GitHub @EternalWavee · Homepage

教育背景

上海交通大学, 计算机科学与技术 (永强班), 工学学士

2024.9 - 2028.6 (预计)

GPA: 4.03/4.30; 核心学分: 92.63/100; 排名: 6/110

- 核心课程: 问题求解与实践 (100); 算法与复杂性 (99); 概率统计 (98); 计算机组成 (97); 操作系统 (97)
- 研究兴趣: 全双工语音交互, 智能语音技术, 自然语言处理
- 英语成绩: CET4: 625; CET6: 565

科研经历



X-Lance Lab | 语音交互组, 科研实习生, 导师: 陈谐教授

2025.08 - 至今

- 系统学习深度学习、语音识别、语音合成及全双工交互等相关理论与技术方法。
- 参与语音交互相关算法设计与系统开发, 重点探索流式语音处理、用户打断及全双工语音交互等关键技术。



Microsoft | Jinyu Li's Speech Science Team, 科研实习生, 导师: 卢恒

2026.10 - 至今

- 围绕全双工 Voice Agent 开展研究, 重点关注记忆机制与强化学习在实时语音交互中的应用。

科研论文

- XTurnix: Large-Scale Self-Supervised Turn Control through Two-State Binary Decisions**

Zhanxun Liu, Yifan Duan, **Hengtao Wu**, Chen Yang, Qinyuan Cheng, Kun Wang, Xingyu Zeng, Xipeng Qiu, Chaochao Lu, Xie Chen. [arXiv](#) | [GitHub](#)

- Audio-Oscar: A Multi-Agent System for Complex Audio Scene Generation, Orchestration, and Refinement**

Yifan Duan *, Qixiang Xu *, **Hengtao Wu ***, Zhanxun Liu, Wenhao Guan, Junxi Liu, Ziyang Ma, Kelu Xu, Xie Chen. [arXiv](#) | [GitHub](#)

项目经历

Audio-Oscar: A Multi-Agent System for Complex Audio Scene Generation, Orchestration, and

Refinement 共同第一作者 [Project Page](#)

2026.1 - 2026.5

- 项目背景:** 面向复杂音频场景生成任务, 提出多智能体协作框架, 将自然语言剧本转化为包含语音、音效、背景音乐、歌曲与混音结果的完整音频场景。
- 核心工作:** 设计了端到端生成流水线, 构建角色建模、语音规划、时间轴编排、音频生成、编辑规划与最终混音等模块, 实现从文本输入到混合音频输出的自动化生成。
- 技术贡献与成果:** 采用结构化 JSON 连接多智能体状态, 协调各类生成组件, 提升复杂音频事件的可控性、一致性与音频生成表现力。

X-Live: Enabling Continuous Real-Time User Interaction During Asynchronous AI Agent Task

Execution 项目负责人 [Project Page](#)

2026.7 - 至今

- 项目背景:** 面向现有 AI Agent 执行复杂任务时交互中断、用户需被动等待的问题, 设计实时交互式 Agent 前端, 使后台 Agent 持续执行任务的同时, 前台模型能够与用户保持自然、低延迟的实时对话。
- 核心工作:** 基于 Qwen3-30B-A3B-Instruct 构建实时交互模型, 通过 *delegate*、*update*、*inspect*、*cancel* 等工具实现后台任务委派、需求动态更新、状态查询与任务取消, 并支持用户在任务执行过程中持续补充要求、询问进度及进行开放式交流。
- 技术贡献与成果:** 构建面向异步 Agent 交互的训练数据与 SFT 流程, 重点优化任务执行期间的持续对话、动态意图理解与主动 backchannel 能力; 进一步探索全双工语音交互, 使模型能够在用户与后台 Agent 之间实现自然、连续且可打断的实时协调。

XTurnix: Large-Scale Self-Supervised Turn Control through Two-State Binary Decisions

第三作者 [Project Page](#)

2026.7 - 2026.10

- 项目背景:** 面向实时语音对话系统中的轮次控制问题, 将模型在监听与说话状态下的行为统一建模为两组状态条件二分类决策, 使单一模型能够基于完整多轮对话历史判断继续监听、开始响应、继续说话或停止说话。
- 核心工作:** 参与自监督数据构建与模型训练流程, 从带时间戳的双人对话转写中恢复交错的 Token 时间线, 并通过状态机自动生成因果动作标签; 结合合成多轮样本缓解四类状态动作的数据不均衡问题, 训练基于 0.6B 参数语言模型的轻量级文本轮次控制器。

- **技术贡献与成果:** 第一个包含历史的语义 Turn-Detection 模型。构建包含 550 万条因果动作样本的训练集，并在四个公开基准及自建平衡多轮基准上完成统一评测；模型在自建基准上达到 89.06% 准确率，较最强第三方基线提升 20.31 个百分点。

TerraScope: An Edge-Deployable Multimodal System for Remote-Sensing Image Understanding

核心开发者 · 揭榜挂帅挑战杯 Project Page 2026.6 - 2026.9

- **项目背景:** 面向受限算力与离线部署场景，基于 MiniCPM-V 4.6 构建遥感多模态大模型，将图像描述、场景分类、目标计数、目标定位、双时相变化分析和语义分割六类任务统一为自然语言交互范式。
- **核心工作:** 负责主模型完整研发链路，包括多源遥感数据整理、统一指令数据构建、模型微调及系统评测；围绕六类遥感任务完成统一生成式建模，并推动模型接入自然语言交互与可视化应用系统。
- **技术贡献与成果:** 构建“多模态主模型、专项能力增强与视觉专家协同”的分层架构，在保留通用遥感理解能力的同时强化超小目标定位与像素级分割；打通数据构建、模型训练、任务路由到结果可视化的端到端链路，形成支持多任务统一交互、模块化扩展及离线部署的完整遥感智能分析方案。

竞赛情况

- MCM 2026 Meritorious Winner , 2026.05
- 中国青年科技创新“揭榜挂帅”擂台赛

技术能力

- **编程与系统:** 熟练使用 Python、C/C++ 进行算法开发与系统实现；熟悉 PyTorch 深度学习框架，以及 Linux、Shell、Git 等工程工具。
- **大模型与语音智能:** 熟悉大语言模型训练与应用流程，包括指令微调（SFT）、Agentic RL、推理优化与 Agent 系统构建；具备语音交互、语音识别、音频生成及多模态模型相关研究与开发经验。
- **Agent 与实时交互系统:** 熟悉 Tool Calling、ReAct、Multi-Agent 协作等 Agent 架构；具备实时语音 Agent、异步任务执行、状态管理与交互策略设计经验。

荣誉奖项

- 信弘天和永强班奖学金（Top 1%）(2025-2026)
- 上海交通大学本科生 B 等优秀奖学金（Top 1%）(2025-2026)
- 上海交通大学本科生 B 等优秀奖学金（Top 5%）(2024-2025)