Huadai Liu

liuhuadai.github.io | github.com/liuhuadai | Google Scholar

Education

Zhejiang University *Sep 2021 – Mar 2024*
M.S. in Artificial Intelligence *Ranked Top 0.1% of program*
Shandong University *Sep 2016 – Jun 2020*
B.S. in Electronic Information

Selected Systems Contributions

- **Low-latency speech synthesis.** Core contributor to CosyVoice 2, a multilingual streaming TTS system with 14,500+ GitHub stars; designed a Rectified Flow integration yielding a **5–10× inference speedup** in production.
- **Large-scale audio learning.** Architected a distributed training framework supporting **100K+ hours** of multi-domain speech data, spanning data pipelines, large-scale training, evaluation, and streaming inference.
- **Multimodal audio generation.** Led ThinkSound, a video-to-audio system with 1,400+ GitHub stars; introduced multimodal-LLM chain-of-thought planning and the AudioCoT dataset.

Selected Research

AudioCALM: Continuous Autoregressive LM for Audio Generation *NeurIPS 2026 (under review)*

- Proposed Continuous Autoregressive Language Modeling (CALM): a flow-matching head replaces the LM softmax to predict rectified-flow velocities over continuous audio latents, with block-causal AR-Flow attention enabling streaming, arbitrary-length speech, sound, and music generation in one LLM backbone.
- Introduced A-MoME (Asymmetric Mixture-of-Modality-Experts) to address asymmetric local–global text–audio mismatch in joint training, matching modality-specific SOTA while surpassing prior unified baselines across all three domains.

STAR-VAE: Structured Topology-Aware Regularization for Audio *ICML 2026*

- Formalized the Rate–Distortion–Regularity Trilemma in audio VAEs; proposed channel-indexed KL regularization aligning latent capacity with audio's spectral hierarchy.
- Combined STAR with a hybrid CNN–Mamba architecture and LLM-based Flow Matching, achieving SOTA reconstruction and generation on AudioCaps and Song-Describer.

PrismAudio: Decomposed CoT and Multi-dimensional Rewards for V2A *ICLR 2026*

- First V2A framework integrating reinforcement learning with decomposed CoT planning across semantic, temporal, aesthetic, and spatial dimensions; proposed Fast-GRPO with hybrid ODE–SDE sampling to reduce training overhead.
- Released the AudioCanvas benchmark and achieved SOTA across all four perceptual dimensions on VGGSound and AudioCanvas.

ThinkSound: Chain-of-Thought Reasoning for Audio Generation *NeurIPS 2025*

- First framework to integrate chain-of-thought reasoning from multimodal LLMs into video-to-audio generation; built the large-scale AudioCoT dataset bridging visual, textual, and acoustic reasoning.

FlashAudio: Rectified Flows for Ultra-Fast Text-to-Audio Generation *ACL 2025 Oral (SAC Highlight)*

- First work to apply Rectified Flows to TTA, enabling single-step inference at **400× real time** on one RTX 4090 while surpassing strong diffusion baselines requiring hundreds of steps.

Publications

Selected first or co-first-author publications. Full list on Google Scholar.

1. **H. Liu**, K. Luo, W. Xue, et al. "AudioCALM: Continuous Autoregressive Language Modeling for Universal Audio Generation." **NeurIPS 2026** (under review).
2. **H. Liu**, K. Luo, W. Xue, et al. "STAR-VAE: Structured Topology-Aware Regularization for Audio Reconstruction and Generation." **ICML 2026**.
3. **H. Liu**, K. Luo, W. Xue, et al. "PrismAudio: Decomposed Chain-of-Thought and Multi-dimensional Rewards for V2A Generation." **ICLR 2026**.
4. **H. Liu**, K. Luo, J. Wang, et al. "ThinkSound: Chain-of-Thought Reasoning in Multimodal LLMs for Audio Generation and Editing." **NeurIPS 2025**.
5. **H. Liu**, T. Luo, K. Luo, et al. "OmniAudio: Generating Spatial Audio from 360-Degree Video." **ICML 2025**.
6. **H. Liu**, J. Wang, R. Huang, et al. "FlashAudio: Rectified Flows for Fast and High-Fidelity Text-to-Audio Generation." **ACL 2025** (Oral, SAC Highlight).
7. **H. Liu**, J. Wang, X. Li, et al. "MEDIC: Zero-shot Music Editing with Disentangled Inversion Control." **ACM MM 2025** (Oral).
8. **H. Liu**, R. Huang, Y. Liu, et al. "AudioLCM: Efficient and High-Quality Text-to-Audio Generation with Minimal Inference Steps." **ACM MM 2024**.
9. **H. Liu**, R. Huang, J. He, et al. "Wav2SQL: Direct Generalizable Speech-To-SQL Parsing." **ACL 2024**.
10. **H. Liu**, R. Huang, X. Lin, et al. "TranSpeech: Speech-to-Speech Translation With Bilateral Perturbation." **ICLR 2023**.
11. **H. Liu**, R. Huang, X. Lin, et al. "ViT-TTS: Visual Text-to-Speech with Scalable Diffusion Transformer." **EMNLP 2023**.
12. **H. Liu**, R. Huang, X. Lin, et al. "AV-TranSpeech: Audio-Visual Robust Speech-to-Speech Translation." **ACL 2023**.
13. **H. Liu**, Y. Ren, R. Huang, et al. "ProDiff: Progressive Fast Diffusion Model for High-Quality Text-to-Speech." **ACM MM 2022**.
14. **H. Liu**, et al. "AntCritic: Argument Mining for Free-Form and Visually-Rich Financial Comments." **LREC-COLING 2024**.

Open Source

- **CosyVoice** (core contributor, 14,500+ stars) — multilingual streaming TTS. github.com/FunAudioLLM/CosyVoice
- **ThinkSound & PrismAudio** (project lead, 1,400+ stars) — video-to-audio generation and reasoning. github.com/FunAudioLLM/ThinkSound
- **AudioLCM** (project lead, 1,100+ stars) — efficient text-to-audio generation. github.com/Text-to-Audio/AudioLCM

Honors & Awards

- **ACL 2025 SAC Highlight Award** — Senior Area Chair top recommendation
- **National Scholarship** (top 1% nationally) 2023
- **Outstanding Graduate of Zhejiang Province** 2024
- **Outstanding Graduate of Zhejiang University** 2024

Academic Service

- **Area Chair**, ACL Rolling Review 2026
- **Outstanding Reviewer Award**, ECCV 2026
- **Conference Reviewer:** NeurIPS, ICML, ICLR, ACL, EMNLP, CVPR, ECCV, AAAI, ACM Multimedia
- **Journal Reviewer:** IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP)

Technical Skills

Research Areas: Audio and speech generation, multimodal learning and reasoning, diffusion and flow matching, reinforcement learning for generative models

Engineering: Large-scale distributed training (FSDP, DeepSpeed), PyTorch, scalable inference systems, production deployment