Xiangyue Zhang (章湘粤)

Phone: +86 18874623697 | Email: x-zhang@mi.t.u-tokyo.ac.jp | WeChat: zxy812343664
Address: Faculty of Information Science, Wuhan University, Wuhan 430079
My homepage: https://xiangyuezhang.com

I'm an incoming Ph.D. student in Mechano-Informatics at **The University of Tokyo**, supervised by Tatsuya Harada. My research interest lies in computer vision and graphics, with a current focus on **2D/3D Motion Generation, Agent, and Embodied AI. Expected graduation in September 2029.**

EDUCATION

The University of Tokyo — Tokyo, Japan
Ph.D. Student. Mechano-Informatics — Expected Oct. 2026– 2029

Wuhan University — Wuhan, Hubei
M.S. Student. Computer Application — Sep. 2023– Jun. 2026
- **National Scholarship 2025. ¥20,000. Top 3%.**
- **Wangzhizhuo Scholarship 2026. ¥10,000. Top 0.3%.**

Central South University — Changsha, Hunan
Bachelor of Geomatics — Sep. 2019– Jun. 2023
- GPA: 89.95/100, Minor: Data Science and Big Data Technology

SELECTED PUBLICATIONS (First Author)

StreamTalk: Streaming Co-speech Gesture Generation With Key-pose Anchoring
Xiangyue Zhang, Jianfang Li , Jiaxu Zhang, et. al.
Manuscript, 2026

PersonaGesture: Single-Reference Co-Speech Gesture Personalization for Unseen Speakers
Xiangyue Zhang, Yiyi Cai, Kunhang Li, et. al.
arXiv preprint, 2026 [arxiv] [code] [website]

SemTalk: Holistic Co-speech Motion Generation with Frame-level Semantic Emphasis
Xiangyue Zhang, Jianfang Li, Jiaxu Zhang, et. al.
International Conference on Computer Vision (***ICCV 2025***) [arxiv] [code] [website]

GlobalDiff: Mitigating Error Accumulation in Co-Speech Motion Generation via Global Rotation Diffusion and Multi-Level Constraints
Xiangyue Zhang, Jianfang Li, Jiaxu Zhang, Jianqiang Ren.
The 40th Annual AAAI Conference on Artificial Intelligence (***AAAI 2026***) [arxiv] [code] [website]

EchoMask: Speech-Queried Attention-based Mask Modeling for Holistic Co-Speech Motion Generation
Xiangyue Zhang, Jianfang Li, Jiaxu Zhang, et. al.
The 33rd ACM International Conference on Multimedia (***ACM MM 2025***) [arxiv] [code] [website]

Robust 2D Skeleton Action Recognition via Decoupling and Distilling 3D Latent Features
Xiangyue Zhang, Yifan Jia, Jiaxu Zhang, Yijie Yang, Zhigang Tu.
IEEE Transactions on Circuits and Systems for Video Technology, 2025 (***T-CSVT*, *IF: 11.1***)

PROJECT

Auto Deep Researcher 24x7 | [arxiv] [code] [website] — **1.1k+ Stars**
- Built an autonomous 24/7 deep-learning experiment agent that handles code editing, training launch, zero-cost monitoring, result parsing, and iterative planning with a leader-worker architecture and persistent research memory.

Douyin Digital Human "Hua Aotian", ByteDance Intelligent Creation — Dec. 2025 – Mar. 2026
- Developed a single-IP diverse semantic motion generation pipeline with multi-IP SFT data, weighted pretraining constraints, LoRA adaptation, retrieval-based action tagging, and emotion/arm-bucket controls.
- Built an initial real-time streaming motion framework with diffusion-forcing generation and fine/coarse text annotation, supporting concurrent speech-to-motion and text-to-motion control.

EXPERIENCE

Research Intern, Qingyun Program, Tencent Visvise Jun. 2026 – Present

- Advised by: Dr. Ronghui Li
- Working on 2D/3D interactive animation, with a focus on controllable digital-human motion generation and interaction.

Research Intern, Intelligent Creation Team, ByteDance Dec. 2025 – Mar. 2026

- Advised by: Youjiang Xu
- Researched large-scale streaming motion generation for digital humans, building a real-time 3D motion synthesis system with speech/text-conditioned control and diffusion-forcing streaming inference.

Research Intern, Tongyi Lab, Alibaba Group Jun. 2024 – Nov. 2025

- Advised by: Dr. Liefeng Bo, Prof. Steven Hoi and Dr. Jianfang Li
- Developed semantic-rhythm disentanglement and speech-queried latent mask modeling for diverse, speech-guided co-speech motion generation, leading to SemTalk (ICCV 2025) and EchoMask (ACM MM 2025).
- Scaled long-horizon and streaming generation with global-rotation diffusion constraints and key-pose retrieval anchoring, leading to GlobalDiff (AAAI 2026) and StreamTalk.