

Xiangyue Zhang (章湘粤)

Email: x-zhang@mi.t.u-tokyo.ac.jp | WeChat: zxy812343664

My homepage: <https://xiangyuezhong.com>



I'm an incoming Ph.D. student 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**.

EDUCATION

The University of Tokyo

Ph.D. Student. Mechano-Informatics

- **The University of Tokyo Fellowship, 2026–2029. Top 1%.**

Tokyo, Japan
Oct. 2026– 2029

Wuhan University

M.S. Student. Computer Application

- **National Scholarship 2025. ¥20,000. Top 3%.**
- **Wangzhizhuo Scholarship 2026. ¥10,000. Top 0.3%.**

Wuhan, Hubei
Sep. 2023– Jun. 2026

Central South University

Bachelor of Geomatics

- GPA: 89.95/100, Minor: Data Science and Big Data Technology

Changsha, Hunan
Sep. 2019– Jun. 2023

SELECTED PUBLICATIONS (First Author)

PersonaGesture: **Single-Reference** Co-Speech Gesture **Personalization** for Unseen Speakers

Xiangyue Zhang, Yiyi Cai, Kunhang Li, et. al.

arXiv preprint, 2026 [\[arxiv\]](#) [\[code\]](#) [\[website\]](#)

StreamTalk: **Streaming** Co-speech Gesture Generation With **Key-pose Anchoring**

Xiangyue Zhang, Jianfang Li, Jiaxu Zhang, et. al.

European Conference on Computer Vision (**ECCV 2026**) [\[arxiv coming soon\]](#) [\[code coming soon\]](#) [\[website\]](#)

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 | [\[arxiv\]](#) [\[code\]](#) [\[website\]](#)

1.2k+ 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, 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 ([ECCV 2026](#)).