

Xiangyue Zhang (章湘粤)

Phone: +86 18874623697 | Email: xiangyuezhang@whu.edu.cn | 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. I am also a third year **M.S. student** under the supervision of Prof. Zhigang Tu at **Wuhan University**. Currently, I am a research intern at **Galbot**, focusing on **Vision-Language-Action Modeling for Humanoid Robots**. My research interest lies in computer vision and graphics, with a current focus on **2D/3D Generation, VLA and Embodied AI. Expected graduation in June 2026.**

EDUCATION

The University of Tokyo

Ph.D. Student. Mechano-Informatics

Tokyo, Japan
Expected Oct. 2026– 2029

Wuhan University

M.S. Student. Computer Application

- **National Scholarship 2025. ¥20,000. Top 3%.**

Wuhan, Hubei
Sep. 2023– Now

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

StreamTalk: Key-Pose-Guided Diffusion for Real-Time Long-Term Co-Speech Gesture Generation

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

(under review)

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*)

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 AAI Conference on Artificial Intelligence (*AAAI 2026*)

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*)

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)

3D Animation
3D Motion Synthesis

3D Motion
Recognition

EXPERIENCE

Research Intern, Galbot

March. 2026 – now

- Advised by: Dr. Jinlu Zhang and Prof. Li Yi
- I participated in the research of Vision-Language-Action Modeling for Humanoid Robots and Embodied Human–Robot Interaction with Language-Grounded Physical Actions.

Research Intern, ByteDance

Dec. 2025 – March. 2026

- Advised by: Youjiang Xu
- During the internship, I participated in the research of large streaming motion generation models. At the same time, the development of a real-time streaming 3D digital human motion generation system.

Research Intern, Tongyi Lab, Alibaba Group

Jun. 2024 – Nov. 2025

- Advised by: Dr. Liefeng Bo, Prof. Steven Hoi and Dr. Jianfang Li
- During the internship, I participated in the code optimization and research of co-speech 3D motion generation. At the same time, investigated 2D work such as echomimic, hallo, liveportait, etc.