SHITONG XU

Personal Website: https://xu-shitong.github.io ⋄ **GitHub:** https://github.com/xu-shitong
Email: shitong.xu@cs.ox.ac.uk ⋄ **Phone:** (+86) 13701470078, (+44) 7873923492

Fourth-year DPhil student in Computer Science at the University of Oxford, developing models for understanding acoustic semantic and spatial information. Research spans **target speaker extraction, spatial audio simulation and rendering, and sound source localization**, with work published at leading conferences including **NeurIPS** and **Interspeech**.

Research interests: Signal Processing, Multimodal Machine Learning, Physics Informed Machine Learning.

EDUCATION

University of Oxford, PhD in Computer Science 2023 - Expected 2028

- Researching under the supervision of Prof. Niki Trigoni and Prof. Andrew Markham.
- Focusing on audio signal processing, including source localization, extraction, spatial audio generation.
- Performed teaching assistance for Physics Informed Machine Learning (PIML) course.

Imperial College London, MEng in Computer Science 2019 - 2023

- Graduated with First Class Honours (76.03%).

SELECTED RESEARCH EXPERIENCE

- **Shitong Xu**, Chang Xu, Zilong Wang. *"ClinALM: Clinically Knowledge Informed Audio-Language Modeling for Pathological Speech."* In submission
- Yuhang He*, **Shitong Xu***, Jia-Xing Zhong, Sangyun Shin, Niki Trigoni, Andrew Markham. *"SPEAR: Receiver-to-Receiver Acoustic Neural Warping Field."*, In submission
- Yiyuan Yang, **Shitong Xu**, Niki Trigoni, Andrew Markham. *"Geometry-Informed Distributed Acoustic Scene Understanding."* In **Interspeech 2026**.
- **Shitong Xu**, Yiyuan Yang, Niki Trigoni, Andrew Markham. *"Target Speaker Extraction through Comparing Noisy Positive and Negative Audio Enrollments."* In **NeurIPS 2025**.
- Yiyuan Yang, **Shitong Xu**, Niki Trigoni, Andrew Markham. *"Efficient and Microphone-Fault-Tolerant 3D Sound Source Localization."* In **Interspeech 2025**.

INTERN EXPERIENCE

Microsoft Research Asia, Machine Learning Group Mar 2026 - July 2026
Machine Learning Intern Beijing, China

- Developed a **pathological speech classification** framework based on a **Large Audio Language Model (LALM)**, jointly trained with pathological **speech recognition** as an auxiliary supervision task.
- Designed a **clinical knowledge-informed encoder** to explicitly encode structured clinical prior knowledge linking diseases to pathological speech attributes, producing embeddings used to condition the LALM backbone.
- Conducted experiments across four pathological speech datasets in four languages; Outperformed baseline by **3.2 and 4.6 macro-F1** on AVFAD and the Saarbruecken Voice Database, respectively.
- Internship research resulted in ClinALM (in submission).

Cisco, Platform Team Apr 2022 - Sep 2022
Software Engineer Intern London, UK

- Engineered and deployed an on-premises **Function-as-a-Service (FaaS)** platform on enterprise **Kubernetes** clusters, providing internal engineering teams with a streamlined environment for deploying serverless workloads.