PhD studentship (Full-time)

Institution	Xi'an Jiaotong-Liverpool University, China
School	School of Advanced Technology
Supervisors	*Please list all the names in the supervisory team. It should be consistent with the information on your approved PGRS proposal.* Principal supervisor: Professor/Dr. Yin Cao (XJTLU) Co-supervisor: Professor/Dr……..(XJTLU) Co-supervisor: Professor/Dr. Meng Fang (UoL)
Application Deadline	Open until the position is filled
Funding Availability	Funded PhD project (world-wide students)
Project Title	Spatial Audio Language Model for Understanding and Editing
Contact	Please email yin.cao@xjtlu.edu.cn (XJTLU principal supervisor's email address) with a subject line of the PhD project title. The principal supervisor's profile is linked here: https://scholar.xjtlu.edu.cn/en/persons/YinCao https://scholar.google.com/citations?user=J9edRm4AAAAJ&hl=en

Requirements:

A Master's degree with Merit and a Bachelor's degree with first-class or upper second-class honors are required for PhD admissions. Exceptional candidates holding only a Bachelor's degree may be considered on an individual basis in certain disciplines.

Evidence of good spoken and written English is essential. The candidate should have an IELTS (or equivalent) score of 6.5 or above, if the first language is not English. This position is open to all qualified candidates irrespective of nationality.

Degree:

The student will be awarded a PhD degree from the University of Liverpool (UK) upon successful completion of the program.

Funding:

The PhD studentship is available for three years subject to satisfactory progress by the student.

The award includes:

- **Full tuition fees** for three years (currently equivalent to **RMB 99,000 per annum**)

- **Conference funding** up to **RMB 16,500** for attending international conferences (e.g., ICASSP, Interspeech, DCASE)

- **A monthly living allowance** will be provided to support daily expenses throughout the study period

The scholarship holders are expected to conduct the majority of their research at XJTLU in Suzhou, China. However, they may apply for a short-term research visit to the University of Liverpool if the project requires it.

Project Description:

This PhD project focuses on the development of a Spatial Audio Language Model (SALM) — a novel framework that enables machines to understand and edit spatial audio using natural language.

Current audio-language models often ignore spatial cues like direction of arrival (DOA), limiting their effectiveness in immersive or interactive settings such as augmented reality (AR), robotics, and intelligent assistants. SALM addresses this limitation by bridging spatial audio and language through contrastive learning and structured multi-modal embeddings.

1. Key Goals of the Project

- Enable natural language-driven spatial audio editing — e.g., "move the barking dog sound to the left."
- Develop dual-branch audio encoders to disentangle and align semantic and spatial components.
- Create large-scale, synthetic spatial audio-caption datasets using spatial simulation and generative language models.
- Support zero-shot and few-shot sound understanding and retrieval in 3D environments.

2. What You'll Work On

The PhD student will:

- Design new machine learning architectures for spatial and semantic disentanglement.
- Implement text-driven spatial audio manipulation tools using learned embeddings.
- Contribute to the construction of publicly available datasets and benchmarks.
- Explore how models can generalize across diverse acoustic environments and language variations.

3. Research Methods

- Use First-Order Ambisonics (FOA) and room simulation to spatialize existing audio datasets.
- Train encoders with spatial and semantic contrastive losses, aligned with natural

language embeddings (e.g., RoBERTa).

- Develop and evaluate text-guided editing mechanisms, including directional shift and sound relocation.
- Conduct generalization experiments using robustness tests and embedding interpolation techniques.

4. Innovation and Impact

This project proposes a paradigm shift in how machines interact with sound — moving beyond classification to interpretation, manipulation, and reasoning. It bridges key domains:

- Sound event localization & detection (SELD)
- Contrastive audio-language modeling
- Natural language-guided scene editing

5. Training and Development

The student will gain strong expertise in:

- Spatial acoustics and machine listening
- Deep learning (e.g., CNNs, Transformers)
- Multi-modal AI and natural language processing

You will be encouraged to participate in top-tier conferences such as ICASSP, Interspeech, and DCASE, with funding support available. This is an exciting opportunity to publish, collaborate, and build a career at the frontier of AI and audio.

Applications span immersive media production, human-computer interaction, assistive technologies, and AR/VR spatial interfaces.

For more information about doctoral scholarship and PhD programme at Xi'an Jiaotong-Liverpool University (XJTLU), please visit

https://www.xjtlu.edu.cn/en/admissions/global/entry-requirements/

https://www.xjtlu.edu.cn/en/admissions/global/fees-and-scholarship

How to Apply:

Interested applicants are advised to email yincao@xjtlu.edu.cn (XJTLU principal supervisor's email address) the following documents for initial review and assessment (please put the project title in the subject line).

- CV
- Two formal reference letters
- Personal statement outlining your interest in the position
- Certificates of English language qualifications (IELTS or equivalent)
- Full academic transcripts in both Chinese and English (for international students, only the English version is required)

- Verified certificates of education qualifications in both Chinese and English (for international students, only the English version is required)
- PDF copy of Master Degree dissertation (or an equivalent writing sample) and examiners reports available