Xuenan Xu, Ph.D.

✉ wsntxxn@gmail.com   in xu-xuenan   🌐 http://wsntxxn.github.io/

Education

2019 – present	**Ph.D., Shanghai Jiao Tong University** in Computer Science and Technology.
2023.10 – 2024.4	**Visiting Ph.D., University of Surrey** in Centre for Vision, Speech and Signal Processing (CVSSP).
2017.7 – 2017.8	**Summer Exchange Student, University of California, Berkeley**.
2015 – 2019	**B.Sc., Shanghai Jiao Tong University** in Material Science and Engineering.

Projects

2021.6 – 2021.8	**AI Speech Co.** Algorithm Developer. Trained a lightweight door activity detector model for deployment using pre-training and data augmentation.

Awards and Achievements

2022.7	Ranking **1st** in DCASE Task 6, subtask 1 & 2
2021.10	**Zhaozhumulan Scholarship** of Shanghai Jiao Tong University.
2021.7	Ranking **2nd** in DCASE Task 6
2020.10	**Huawei Scholarship** of Shanghai Jiao Tong University.
2020.7	Ranking **4th** in DCASE Task 6
2017.10	**First Prize** in Chinese Mathematics Competitions (Shanghai Division)
2016.10	**National Scholarship** of Shanghai Jiao Tong University.

Research Publications

Journal Articles

1. **X. Xu**, Z. Xie, M. Wu, and K. Yu, "Beyond the status quo: A contemporary survey of advances and challenges in audio captioning," *IEEE/ACM Trans. Audio, Speech Lang. Process.*, 2023.

2. H. Dinkel, S. Wang, **X. Xu**, M. Wu, and K. Yu, "Voice activity detection in the wild: A data-driven approach using teacher-student training," *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 29, pp. 1542–1555, 2021.

Conference Proceedings

1. **X. Xu**, X. Xu, Z. Xie, P. Zhang, M. Wu, and K. Yu, "A detailed audio-text data simulation pipeline using single-event sounds," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2024, pp. 1091–1095.

2. G. Li, **X. Xu**, L. Dai, M. Wu, and K. Yu, "Diverse and vivid sound generation from text descriptions," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, IEEE, 2023, pp. 1–5.

3. Z. Xie, **X. Xu**, M. Wu, and K. Yu, "Enhance temporal relations in audio captioning with sound event detection," in *Proc. ISCA Annu. Conf. Int. Speech Commun. Assoc.*, 2023, pp. 4179–4183.

4   **X. Xu**, Z. Zhang, Z. Zhou, *et al.*, "Blat: Bootstrapping language-audio pre-training based on audioset tag-guided synthetic data," in *Proc. ACM Int. Conf. Multimedia*, New York, NY, USA: ACM, 2023, pp. 2756–2764.

5   G. Li, **X. Xu**, H. Dinkel, M. Wu, and K. Yu, "Category-adapted sound event enhancement with weakly labeled data," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, IEEE, 2022, pp. 851–855.

6   G. Li, **X. Xu**, M. Wu, and K. Yu, "Navigating audio-visual event detection across mismatched modalities," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, IEEE, 2022, pp. 1975–1979.

7   S. Lou, **X. Xu**, M. Wu, and K. Yu, "Audio-text retrieval in context," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, IEEE, 2022, pp. 4793–4797.

8   **X. Xu**, M. Wu, and K. Yu, "Diversity-controllable and accurate audio captioning based on neural condition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 971–975.

9   Z. Zhou, Z. Zhang, **X. Xu**, Z. Xie, M. Wu, and K. Q. Zhu, "Can audio captions be evaluated with image caption metrics?" In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 981–985.

10   **X. Xu**, H. Dinkel, M. Wu, Z. Xie, and K. Yu, "Investigating local and global information for automated audio captioning with transfer learning," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 905–909.

11   **X. Xu**, H. Dinkel, M. Wu, and K. Yu, "A lightweight framework for online voice activity detection in the wild.," in *Proc. ISCA Annu. Conf. Int. Speech Commun. Assoc.*, 2021, pp. 371–375.

12   **X. Xu**, H. Dinkel, M. Wu, and K. Yu, "Audio caption in a car setting with a sentence-level loss," in *Proc. Int. Symp. Chin. Spoken Lang. Process.*, 2021, pp. 1–5.

13   **X. Xu**, H. Dinkel, M. Wu, and K. Yu, "Text-to-audio grounding: Building correspondence between captions and sound events," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 606–610.

14   **X. Xu**, H. Dinkel, M. Wu, and K. Yu, "A crnn-gru based reinforcement learning approach to audio captioning," in *Proc. Detection Classification Acoust. Scenes Events*, 2020, pp. 225–229.

Skills

Languages	▌	Proficiency in English (TOFEL 104), Mandarin Chinese.
Coding	▌	Python, shell, c++, java, LaTeX, …
Misc.	▌	Football, Saxphone.