

GUANGZHI (BRIAN) SUN

Department of Engineering, Cambridge, CB2 1PZ ◊ gs534@cam.ac.uk ◊ Personal Website

EDUCATION

University of Cambridge October 2019 - June 2023
PhD in Engineering Cambridge, UK

- Topic: Contextual knowledge integration in neural-based end-to-end conversational AI systems

University of Cambridge October 2015 - June 2019
M.Eng. and B.A in Information Engineering Cambridge, UK

- Result: Distinction for M.Eng. Class 1 with distinction (top 2%) in the third year and 1% in the first two years among all engineering students in the same year for B.A.

University of Hong Kong September 2014 - June 2015
HKU-Cambridge Joint Scheme Hong Kong, China

- Result: Dean's Honours List

EMPLOYMENT

Trinity College, University of Cambridge October 2024 - September 2028
Junior Research Fellow Cambridge, UK

- Reliability in audio and audio-visual large language models including hallucination and AI safety.

Machine Intelligence Lab, University of Cambridge October 2023 - September 2024
Research Associate Cambridge, UK

- Reliable multimodal Large Language Models including hallucination, controllability and AI safety.
- Lecturing Machine Learning and Machine Intelligence (MLMI) course: Advanced Speech Recognition.

AI-Lab-Speech, ByteDance-Tsinghua Univerity Joint Project July 2023 - September 2023
Research Intern (Manager: Dr Wei Li, Prof. Chao Zhang) Beijing, China

- Speech-enhanced audio-visual large language models (LLM) for video understanding.

Google Brain, Google June 2019 - September 2019
Summer Research Intern (Supervisor: Dr Yu Zhang) Mountain View, US

- Fine-grained variational auto-encoder (VAE) for prosody modelling in end-to-end text-to-speech synthesis

SELECTED PUBLICATIONS

- ◊ **G. Sun**, W. Yu, C. Tang, X. Chen, T. Tan, W. Li, L. Lu, Z. Ma, Y. Wang, & C. Zhang “video-SALMONN: Speech-Enhanced Audio-Visual Large Language Models”, *ICML*, 2024.
- ◊ **G. Sun**, X. Zhan & J. Such “Building Better AI Agents: A Provocation on the Utilisation of Persona in LLM-based Conversational Agents”, *CUI*, 2024. **Best Provocation Prize.**
- ◊ Z. Chen, H. Liu, W. Yu, **G. Sun**, H. Liu, J. Wu, C. Zhang, Y. Wang & Y. Wang “M³AV: A Multimodal, Multigenre, and Multipurpose Audio-Visual Academic Lecture Dataset”, *ACL*, 2024.
- ◊ **G. Sun**, S. Feng, D. Jiang, C. Zhang, M. Gašić & P. C. Woodland “Speech-based Slot Filling using Large Language Models”, *ACL findings*, 2024.

- ◇ C. Tang, W. Yu, **G. Sun**, X. Chen, T. Tan, W. Li, L. Lu, Z. Ma & C. Zhang “SALMONN: Towards Generic Hearing Abilities for Large Language Models”, *ICLR*, 2024.
- ◇ **G. Sun**, X. Zheng, C. Zhang & P. C. Woodland “Can Contextual Biasing Remain Effective with Whisper and GPT-2?”, *Proc. Interspeech*, 2023.
- ◇ **G. Sun**, C. Zhang & P. C. Woodland “Graph Neural Networks for Contextual ASR with the Tree-Constrained Pointer Generator”, *IEEE Transactions on Audio, Speech, and Language Processing*, 2024.
- ◇ **G. Sun**, C. Zhang & P. C. Woodland “Minimising Biasing Word Errors for Contextual ASR with the Tree-Constrained Pointer Generator”, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 31, pp. 345–354, 2022.
- ◇ **G. Sun**, C. Zhang & P. C. Woodland “Tree-constrained Pointer Generator with Graph Neural Network Encodings for Contextual Speech Recognition”, *Proc. Interspeech*, 2022. **Best Student Paper Award**.
- ◇ Y. Li, C. Yu, **G. Sun**, H. Jiang, F. Sun, W. Zu, Y. Wing, Y. Yang & J. Wang “Cross-Utterance Conditioned VAE for Non-Autoregressive Text-to-speech”, in *Proc. ACL*, 2022.
- ◇ **G. Sun**, C. Zhang & P. C. Woodland “Tree-constrained Pointer Generator for End-to-end Contextual Speech Recognition”, *Proc. ASRU*, 2021.
- ◇ **G. Sun**, C. Zhang & P. C. Woodland “Transformer Language Models with LSTM-based Cross-utterance Information Representation”, in *Proc. ICASSP*, Toronto, 2021.
- ◇ **G. Sun**, C. Zhang & P. C. Woodland “Combination of Deep Speaker Embeddings for Diarisation”, *Neural Networks*, Volume 141, pp. 372–384, 2021.
- ◇ **G. Sun**, Y. Zhang, R. J. Weiss, Y. Cao, H. Zen, Y. Wu “Fully-hierarchical Fine-grained Prosody Modeling for Interpretable Speech Synthesis”, in *Proc. ICASSP*, Barcelona, 2020.
- ◇ **G. Sun**, Y. Zhang, R. J. Weiss, Y. Cao, H. Zen, A. Rosenberg, B. Ramabhadran, Y. Wu “Generating Diverse and Natural Text-to-speech Samples Using a Quantized Fine-grained VAE and Autoregressive Prosody Prior”, in *Proc. ICASSP*, Barcelona, 2020.
- ◇ **G. Sun**, C. Zhang & P. C. Woodland, “Speaker Diarisation Using 2D Self-Attentive Combination of Embeddings”, in *Proc. ICASSP*, Brighton, 2019.

AWARDS

- Best Student Paper Award: Interspeech 2022
- Cambridge Commonwealth, European and International Trust: Research scholarship, 2019
- Trinity College Research Scholar: PhD research scholarship by Trinity College, 2019–2023
- Trinity College Examination Prize: Prize for getting first class results in each year, 2016–2018
- Stephen Bush Prize: Prize for academic excellence in the second year, 2017
- BP Prize: University prize for coming top in the first-year Tripos exam, 2016

ACTIVITIES

- Academic Activities: Technical program committee and session chair for UK and Ireland Speech 2024.
- Teaching: Supervision (tutorial) for 1st and 3rd year undergraduate students.
- Project Supervision: MPhil project supervisor 2023-2024. M.Eng. project co-supervisor 2024-2025.
- Reviewer: NeurIPS 2024, EMNLP 2023, ACL 2023, ACM Multimedia 2022, Interspeech 2022, ICASSP 2022, 2023, APSIPA Transactions on Signal and Information Processing, IEEE Journal of Selected Topics in Signal Processing
- Social activities: Music director of Cambridge University Chinese Orchestra Society; Committee member of Trinity Oriental Society; Team leader of HKU Fujia seed voluntary teaching program