

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

RESEARCH

Machine Intelligence Lab, University of Cambridge October 2023 - Now

Research Associate Cambridge, UK

- Contextual knowledge integration in large language models (LLMs).
- Teaching commitment: Lecturer of Machine Learning and Machine Intelligence (MLMI) 14 - Advanced Speech Recognition. MPhil project advisor for MLMI 2024.

AI-Lab-Speech, ByteDance-Tsinghua Univerity Joint Project July 2023 - September 2023

Research Intern Beijing, China

- Manager: Dr Wei Li, Prof. Chao Zhang
- Multi-modal large language model (LLM) for the cognition-driven understanding of speech, audio and video together, such as Tiktok short videos.

Machine Intelligence Lab, University of Cambridge October 2019 - October 2023

PhD Student Cambridge, UK

- Supervisor: Prof. Phil Woodland
- Neural-symbolic contextual knowledge integration using tree-based pointer generator (TCPGen)
- Audio-visual contextual knowledge extraction and integration for meeting and presentation transcription and understanding systems
- Contextual knowledge integration with structured knowledge bases in spoken language understanding (SLU) and spoken dialogue systems driven by the speech-end

Google Brain, Google June 2019 - September 2019

Summer Research Intern in Dr Yonghui Wu's Group Mountain View, US

- Supervisor: Dr Yu Zhang
- Fine-grained variational auto-encoder (VAE) for prosody modelling in end-to-end text-to-speech synthesis

Machine Intelligence Lab, University of Cambridge October 2018 - June 2019

M.Eng. Research Project Cambridge, UK

- Supervisor: Prof. Phil Woodland
- Cross-utterance language model which extract contextual information from surrounding utterances

- Supervisor: Prof. Phil Woodland
- Speaker diarisation using combination of speaker embeddings from deep neural networks

PUBLICATIONS

- ◇ **G. Sun**, W. Yu, C. Tang, X. Chen, T. Tan, W. Li, L. Lu, Z. Ma & C. Zhang “Fine-grained Audio-Visual Joint Representations for Multimodal Large Language Models”, submitted to *ICLR*, 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”, submitted to *ICLR*, 2024.
- ◇ N. Lashkarashvili, W. Wu, **G. Sun**, P. C. Woodland “Parameter Efficient Finetuning for Speech Emotion Recognition and Domain Adaptation”, to appear in *ICASSP*, 2024.
- ◇ C. Tang, W. Yu, **G. Sun**, X. Chen, T. Tan, W. Li, L. Lu, Z. Ma & C. Zhang “Extending Large Language Models for Speech and Audio Captioning”, to appear in *ICASSP*, 2024.
- ◇ W. Yu, C. Tang, **G. Sun**, X. Chen, T. Tan, W. Li, L. Lu, Z. Ma & C. Zhang “Connecting Speech Encoder and Large Language Model for ASR”, to appear in *ICASSP*, 2024.
- ◇ Q. Zhao, **G. Sun**, C. Zhang, M. Xu & T. F. Zheng “Enhancing Quantised End-to-End ASR Models via Personalisation”, to appear in *ICASSP*, 2024.
- ◇ Y. Li, C. Yu, **G. Sun**, W. Zu, Z. Tian, Y. Wen, W. Pan, C. Zhang, J. Wang, Y. Yang & F. Sun, “Cross-Utterance Conditioned VAE for Speech Generation”, submitted to *IEEE Transactions on Audio, Speech, and Language Processing*, 2023.
- ◇ **G. Sun**, C. Zhang & P. C. Woodland “Knowledge-Aware Audio-Grounded Generative Slot Filling for Limited Annotated Data”, submitted to *Computer Speech & Language*, 2023.
- ◇ **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”, submitted to *IEEE Transactions on Audio, Speech, and Language Processing*, 2023.
- ◇ **G. Sun**, C. Zhang & P. C. Woodland “End-to-end Spoken Language Understanding with Tree-constrained Pointer Generator”, in *Proc. ICASSP*, 2023.
- ◇ Evonne P. C. Lee, **G. Sun**, C. Zhang & P. C. Woodland “Spectral Clustering-aware Learning of Embeddings for Speaker Diarisation”, in *Proc. ICASSP*, 2023.
- ◇ **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**, D. Liu, C. Zhang & P. C. Woodland “Content-Aware Speaker Embeddings for Speaker Diarisation”, 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.

WORK EXPERIENCE

ARM Ltd. Jun. 2017 - Sept. 2017, Hardware Engineering Intern, Sheffield and Cambridge, UK.

UENI Ltd. Jun. 2016 - Sept. 2016, Software Engineering Intern, London, UK.

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

- Undergraduate teaching: first-year and third-year supervision
- Reviewer: 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