

# Curriculum Vitae of Chao Zhang

Machine Intelligence Laboratory  
Cambridge University Engineering Department

Email: cz277@cam.ac.uk  
Web: <http://mi.eng.cam.ac.uk/~cz277>

## EDUCATION

- PhD in Information Engineering** **2012.10 – 2017.1**
- Department of Engineering, Cambridge University, UK
  - Supervisor: Prof. Phil Woodland (FREng.)
- BEng & MSc in Computer Science and Technology** **2005.8 – 2012.7**
- Tsinghua University, Beijing, China

## RESEARCH INTERESTS

Speech processing, text generation, recommender system, deep learning, machine learning, large scale optimisation, artificial intelligence and neuroscience.

## WORK EXPERIENCE

- Senior Research Scientist at Google AI** **2021.3 – present**
- Visiting Fellow at Cambridge University** **2021.3 – present**
- Research Associate in Speech and Language Processing** **2017.9 – 2021.2**
- Machine Intelligence Laboratory, University of Cambridge
- Consultant in Speech and Language Processing** **2018.5 – 2021.2**
- JD.com AI Research
    - Working on ASR, TTS, speaker recognition, text generation, and multimodal intelligence.
    - Ranking 3rd in DCASE 2019 and 3rd in FFSVC 2020.
- Research Student at University of Cambridge** **2012 – 2017**
- HTK V3.5 development IEEE ASRU 2015 MGB challenge, 1st in all tasks.
    - Created the ANN modules in HTK, PyHTK, and rebuilt all Cambridge ASR pipelines.
  - Machine Intelligence Laboratory, University of Cambridge.
    - IEEE ASRU 2015 MGB challenge, 1st in all tasks.
    - DARPA BOLT Program 2014 evaluation, 1st in telephony speech recognition task.
    - iARPA Babel Program 2013 evaluation, 1st in Pashto, Tagalog & Turkish.
- Research Intern at Apple Siri** **2017**
- CICC stock index prediction system development** **2010 – 2011**
- Asset Management Department, China International Capital Corporation
    - Created a CRF based index prediction system (25M RBM investment and 22% return rate).

## BOOK & JOURNAL PUBLICATIONS

1. Q. Li, **C. Zhang**, P.C. Woodland, "Combining frame-synchronous and label-synchronous systems for speech recognition," in *Submission*, 2021.
2. C. Wingfield, **C. Zhang**, B. Devereux, E. Fonteneau, A. Thwaites, X. Liu, P. Woodland, W. Marslen-Wilson, L. Su, "Closing the circle between artificial and brain neural networks for speech recognition," in *Submission*, 2021. (**co-first author**)

3. A. Haider, **C. Zhang**, F.L. Kreyssig, P.C. Woodland, "A distributed optimisation framework to combine curvature information with natural gradient for discriminative sequence training," *Neural Networks (to appear)*, 2021. **(co-first author)**
4. G. Sun, **C. Zhang**, P.C. Woodland, "Combination of deep speaker embeddings for diarisation," *Neural Networks*, Vol. 141, pp. 372-384, 2021. **(co-first author)**
5. **C. Zhang**, Z. Yang, X. He, L. Deng, "Multimodal intelligence: Representation learning, information fusion, and applications," *IEEE JSTSP*, Vol. 14(4), pp. 3-19, 2020.
6. X. He, L. Deng, R. Rose, M. Huang, I. Trancoso, **C. Zhang**, "Introduction to the special issue on deep learning for multi-modal intelligence across speech, language, vision, and heterogeneous signals," *IEEE JSTSP*, Vol. 14(4), pp. 1-2, 2020. **(editorial paper, co-lead guest editor)**
7. C. Wingfield, L. Su, X. Liu, **C. Zhang**, P.C. Woodland, A. Thwaites, E. Fonteneau, W.D. Marslen-Wilson, "Relating dynamic brain states to dynamic machine states: Human and machine solutions to the speech recognition problem," *PLOS Computational Biology*, Vol. 13(9), pp. 1-25, 2017.
8. S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, A. Ragni, V. Valtchev, P. Woodland, **C. Zhang**, *The HTK Book (for HTK version 3.5)*, Cambridge University Engineering Dept., 2015.
9. **C. Zhang**, Y. Liu, Y.-Q. Xia, X. Wang, C.-H. Lee, "Reliable accent specific unit generation with discriminative dynamic Gaussian mixture selection for multi-accent Chinese speech recognition," *IEEE Trans. on ASLP*, Vol. 21(10), pp. 2073-2084, 2013.
10. **C. Zhang**, Y. Liu, T.F. Zheng, "Acoustic model reconstruction for multi-accent Chinese speech recognition," *Journal of Tsinghua Univ. (Sci. & Tech.)*, Vol. 51, pp. 1161-1166, 2011. **(best paper)**

#### **CONFERENCE PUBLICATIONS**

1. G. Sun, **C. Zhang**, P.C. Woodland, "Tree-constrained pointer generator for end-to-end contextual speech recognition", in *Submission*, 2021.
2. X. Zheng, **C. Zhang**, P.C. Woodland, "Adapting GPT, GPT-2, and BERT language models for speech recognition", in *Submission*, 2021.
3. Z. Yuan, K. Ren, **C. Zhang**, H. Ao, F. Liu, W. Hu, X. Miao, "Personalised DNN recommendation systems based on user-dependent gating mechanism", in *Submission*, 2021.
4. D. Jiang, **C. Zhang**, P.C. Woodland, "Variable frame rate acoustic models using minimum error reinforcement learning", in *Proc. Interspeech*, 2021.
5. G. Xu, W. Song, Z. Zhang **C. Zhang**, X. He, B. Zhou, "Improving prosody with cross-utterance BERT embeddings for TTS", in *Proc. ICASSP*, 2021.
6. W. Wu, **C. Zhang**, P.C. Woodland, "Emotion recognition by fusing time synchronous and time asynchronous representations", in *Proc. ICASSP*, 2021.
7. G. Sun, **C. Zhang**, P.C. Woodland, "Transformer language models with LSTM-based cross-utterance information representation", in *Proc. ICASSP*, 2021.
8. G. Sun, D. Liu, **C. Zhang**, P.C. Woodland, "Content-aware speaker embeddings for speaker diarisation", in *Proc. ICASSP*, 2021.
9. W. Xue, G. Quan, **C. Zhang**, G. Ding, X. He, B. Zhou, "Neural Kalman filtering for speech enhancement", in *Proc. ICASSP*, 2021.
10. W. Song, X. Yuan, Z. Zhang **C. Zhang**, Y. Wu, X. He, B. Zhou, "DIAN: Duration informed autoregressive network for voice cloning", in *Proc. ICASSP*, 2021.
11. Q. Li, F. Kreyssig, **C. Zhang**, P.C. Woodland, "Discriminative neural clustering for speaker diarisation", in *Proc. SLT*, 2021. **(best paper)**
12. W. Xue, Y. Tong, **C. Zhang**, G. Ding, X. He, B. Zhou, "Sound event localization and detection based on multiple DOA beamforming and multi-task learning", in *Proc. Interspeech*, 2020.
13. W. Song, G. Xu, Z. Zhang, **C. Zhang**, X. He, B. Zhou, "Efficient WaveGlow: An improved WaveGlow vocoder with enhanced speed", in *Proc. Interspeech*, 2020.

14. Y. Tong, W. Xue, S. Huang, L. Fan, **C. Zhang**, G. Ding, X. He, "The JD AI speaker verification system for the FFSVC 2020 challenge", in *Proc. Interspeech*, 2020.
15. Y. Fathullah, **C. Zhang**, P.C. Woodland, "Improved large-margin softmax loss for speaker diarisation", in *Proc. ICASSP*, 2020.
16. Q. Li, **C. Zhang**, P.C. Woodland, "Integrating source-channel and attention-based sequence-to-sequence models for speech recognition", in *Proc. ASRU*, 2019. **(best paper)**
17. P. van Platen, **C. Zhang**, P.C. Woodland, "Multi-span acoustic modelling using raw waveform signals", in *Proc. Interspeech*, 2019.
18. W. Xue, Y. Tong, G. Ding, **C. Zhang**, T. Ma, X. He, B. Zhou, "Direct-path signal cross-correlation estimation for sound source localization in reverberation", in *Proc. Interspeech*, 2019.
19. **C. Zhang**, F.L. Kreyssig, Q. Liu, P.C. Woodland, "PyHTK: Python library and ASR pipelines for HTK", in *Proc. ICASSP*, 2019.
20. G. Sun, **C. Zhang**, P.C. Woodland, "Speaker diarisation using 2D self-attentive combination of embeddings", in *Proc. ICASSP*, 2019.
21. **C. Zhang**, P.C. Woodland, "Semi-tied units for efficient gating in LSTM and highway networks", in *Proc. Interspeech*, 2018.
22. Y. Wang, **C. Zhang**, M.J.F. Gales, P.C. Woodland, "Speaker adaptation and adaptive training for jointly optimised tandem systems", in *Proc. Interspeech*, 2018.
23. **C. Zhang**, P.C. Woodland, "High order recurrent neural networks for acoustic modelling", in *Proc. ICASSP*, 2018.
24. F.L. Kreyssig, **C. Zhang**, P.C. Woodland, "Improved TDNNs using deep kernels and frequency dependent grid-RNNs", in *Proc. ICASSP*, 2018.
25. **C. Zhang**, P.C. Woodland, "Joint optimisation of tandem systems using Gaussian mixture density neural network discriminative sequence training", in *Proc. ICASSP*, 2017.
26. F.L. Kreyssig, **C. Zhang**, P.C. Woodland, "Modular construction of complex deep learning architectures in HTK", in *UK Speech*, 2017
27. Q. Li, **C. Zhang**, F.L. Kreyssig, P.C. Woodland, "Experimental studies on teacher-student training of deep neural network acoustic models", in *UK Speech*, 2017
28. C. Wingfield, L. Su, B. Devereux, X. Liu, **C. Zhang**, P.C. Woodland, E. Fonteneau, A. Thwaites, W. Marslen-Wilson, "Multi-level representations in speech processing in brain and machine: evidence from EMEG and RSA", in *Cambridge Language Science Symposium*, 2016. **(best paper candidate)**
29. C. Wingfield, L. Su, X. Liu, **C. Zhang**, P. Woodland, A. Thwaites, E. Fonteneau, W. Marslen-Wilson, "Investigating human speech recognition: reverse-engineering the machine solution with EMEG and RSA", in *Proc. OHBM*, 2016.
30. P. Lanchantin, M.J.F. Gales, P. Karanasou, X. Liu, Y. Qian, L. Wang, P.C. Woodland, **C. Zhang**, "Selection of multi-genre broadcast data for the training of automatic speech recognition systems", in *Proc. Interspeech*, 2016.
31. **C. Zhang**, P.C. Woodland, "DNN speaker adaptation using parameterised sigmoid and ReLU hidden activation functions", in *Proc. ICASSP*, 2016.
32. J. Yang, **C. Zhang**, A. Ragni, M.J.F. Gales, P.C. Woodland, "System combination with log-linear models", in *Proc. ICASSP*, 2016.
33. L. Wang, **C. Zhang**, P.C. Woodland, M.J.F. Gales, P. Karanasou, P. Lauchantin, "Improved DNN-based segmentation for multi-genre broadcast audio", in *Proc. ICASSP*, 2016.
34. P.C. Woodland, X. Liu, Y. Qian, **C. Zhang**, M.J.F. Gales, P. Karanasou, P. Lanchantin, L. Wang, "Cambridge University transcription systems for the multi-genre broadcast challenge", in *Proc. ASRU*, 2015.

35. P. Lanchantin, M.J.F. Gales, P. Karanasou, X. Liu, Y. Qian, L. Wang, P.C. Woodland, **C. Zhang**, "The development of the Cambridge University alignment system for the multi-genre broadcast challenge", in *Proc. ASRU*, 2015.
36. P. Karanasou, M.J.F. Gales, P. Lanchantin, X. Liu, Y. Qian, L. Wang, P.C. Woodland, **C. Zhang**, "Speaker diarisation and longitudinal linking in multi-genre broadcast data", in *Proc. ASRU*, 2015.
37. R.C. van Dalen, J. Young, H. Wang, A. Ragni, **C. Zhang**, M.J.F. Gales, "Structured discriminative models using deep neural-network features", in *Proc. ASRU*, 2015. **(best paper candidate)**
38. X. Liu, F. Flego, L. Wang, **C. Zhang**, M.J.F. Gales, P.C. Woodland, "The Cambridge University 2014 BOLT conversational telephone Mandarin Chinese LVCSR system for speech translation", in *Proc. Interspeech*, 2015.
39. **C. Zhang**, P.C. Woodland, "Parameterised sigmoid and ReLU hidden activation functions for DNN acoustic modelling", in *Proc. Interspeech*, 2015.
40. **C. Zhang**, P.C. Woodland, "A general artificial neural network extension for HTK", in *Proc. Interspeech*, 2015.
41. H. Wang, A. Ragni, M.J.F. Gales, K.M. Knill, P.C. Woodland, **C. Zhang**, "Joint decoding of tandem and hybrid systems for improved keyword spotting on low resource languages", in *Proc. Interspeech*, 2015.
42. **C. Zhang**, P.C. Woodland, "Standalone training of context-dependent deep neural network acoustic models", in *Proc. ICASSP*, 2014. **(best paper)**
43. K.M. Knill, M.J.F. Gales, S.P. Rath, P.C. Woodland, **C. Zhang**, S.-X. Zhang, "Investigation of multi-lingual deep neural networks for spoken term detection," in *Proc. ASRU*, 2013.
44. **C. Zhang**, Y. Liu, Y. Xia, C.-H. Lee, "Discriminative dynamic Gaussian selection with enhanced robustness and performance for multi-accent speech recognition," in *Proc. ICASSP*, 2012.
45. **C. Zhang**, Y. Liu, C.-H. Lee, "Detection-based accented speech recognition using articulatory features," in *Proc. ASRU*, 2011. **(panel member paper)**
46. **C. Zhang**, Y. Liu, Y. Xia, T.F. Zheng, J. Olsen, J. Tian, "Reliable accent specific unit generation with dynamic Gaussian mixture selection for multi-accent speech recognition," in *Proc. ICME*, 2011. **(top 15% rated paper)**
47. **C. Zhang**, Y. Liu, T.F. Zheng, "Asymmetric acoustic model for accented speech recognition," in *Proc. APSIPA ASC*, 2011.

## **AWARDS & GRANTS**

IEEE SLT 2021, best student paper	2021
IEEE ASRU 2019, best student paper	2019
Research funding, EPSRC Natural Speech Technology Project	2013 - 2016
IEEE ASRU 2015, best paper candidate	2015
Research funding, DARPA BOLT Program	2015
IEEE ICASSP 2014, IBM spoken language processing best student paper award (about 5/600)	2014
Research funding, iARPA Babel Program	2013
Cambridge International Scholarship, Cambridge Overseas Trust (about 1/30)	2012
Excellent master graduate award, Dept. of Computer Sci. & Tech., Tsinghua Univ.	2011
Excellent academy award, State Key Lab. of Intelligent Tech. & Sys., Tsinghua Univ. (17/170)	2011
Google excellence scholarship (2/180)	2011
NCMMSC 2011, best student paper award (3/96)	2011
IEEE ASRU 2011, panel member paper award (19/97)	2011
IEEE ASRU 2011, student travel grant	2011
IEEE ICME 2011, top 15% rated paper, student travel grant	2011
Games of the XXIX Olympiad, outstanding volunteer (6/40)	2008

## **INVITED TALKS**

<b>Speech recognition with speaker diarisation for multi-speaker meeting transcription</b>	<b>2021.1</b>
--	---------------

- Zoom
- Some recent work on speech processing and beyond** **2020.12**
- John Hopcroft Centre, Shanghai Jiaotong University
- Speaker diarisation using deep neural networks** **2020.11**
- Google AI, Facebook AI
- Speaker diarisation with deep learning** **2020.10**
- Shenlanxueyuan Open AI Course, Institute of Automation, Chinese Academy of Sciences
- Efficient recurrent and feedforward neural network architectures for acoustic modelling** **2018.8**
- Tsinghua University, Tianjin University, Renmin University
- State-of-the-art deep learning for acoustic modelling — an HTK perspective** **2018.2**
- JD.com AI Research & Platform
- Multi-accent speech recognition for personalised speech assistant** **2017.6**
- Apple Siri Speech Team
- Tandem and hybrid speech recognition systems based on a generic ANN extension in HTK** **2016.3**
- Tsinghua University
- HTK 3.5 based DNN system development and research** **2015.12**
- Microsoft Speech Team
- CRF models and their applications in stock market value prediction** **2011.4**
- China International Capital Corporation

### **SELECTED PROJECTS**

Tsinghua University Overseas High-level Scholars Invite and Support Plan	2018
Deep Neural Networks for HTK (Cambridge Univ. Engineering Dept. funded)	2017 - 2021
Natural Speech Technology (EPSRC & BBC funded)	2013 - 2016
DARPA BOLT Program	2014
iARPA BABEL Program	2013
State Scholarship Fund of China Scholarship Council, 1010811231	2010 - 2012
Natural Science Foundation of China, 60975018	2009 - 2011
Telephony Bank 95533, China Construction Bank	2009 - 2010
Joint Research Grant of Nokia-Tsinghua Joint Funding	2008 - 2010

### **SELECTED SERVICE**

Session chair, Interspeech 2020	2020
Co-lead guest editor, IEEE JSTSP - Deep Learning for Multi-Model Intelligence	2019 - 2020
Regular journal reviewers for TASLP, JSTSP, JBHI, TOMM, CSL, TAPSIPA, EURASIP JASMP	
Regular conference reviewers for ICASSP, Interspeech, ACL, EMNLP, NAACL, AAAI	
Sponsorship division officer, Chinese Students & Scholars Association in Cambridge	2012 - 2013
Reporters' standing seat assistant, Games of the XXIX Olympiad	2008
Reporters' standing seat assistant, The XXIX Olympic Synchronised Swimming Qualification	2008
TV broadcasting assistant, The 16th FINA Diving World Cup	2008

### **TEACHING EXPERIENCE**

Demonstration, Speech recognition (MLSALT & MIML), Cambridge University	2015 - 2020
Teaching assistant, The principle of signal processing, Tsinghua University	2010
Teaching assistant, C++ programming design and training, Tsinghua University	2010