



Basic Information

Yanmin Qian
Ph.D.
IEEE
Senior Member

• Mobile: +86-18516056597 E-mail: yanminqian@sjtu.edu.cn Zip Code: 200240
• Affiliation: SpeechLab, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
• Contact: Room 3-515, SEIEE Building, No. 800 Dongchuan Road, Minhang District, Shanghai
• Personal homepage: <https://x-lance.sjtu.edu.cn/en/members/yanmin-qian>

Work Experience

2022-Present **Shanghai Jiao Tong University, Department of Computer Science and Engineering**
• Full Professor

2017-2021 **Shanghai Jiao Tong University, Department of Computer Science and Engineering**
• Associate Professor

2015-2016 **Cambridge University, Department of Engineering, Machine Intelligence Laboratory**
• Research Associate

2013-2016 **Shanghai Jiao Tong University, Department of Computer Science and Engineering**
• Assistant Professor

Education

2007-2013 **Tsinghua University, Department of Electronic Engineering**
• Ph.D. Candidate in Electronic Engineering

2003-2007 **Huazhong University of Science & Technology, Department of Electronic and Information Engineering**
• B.E in Electronic and Information Engineering

Publications & Patents & Books

- Total **200+** published or accepted papers till now: **37** Journal Papers and **179** International Conference Papers
- **157** Papers are published on the top-level Journal and International Conference on Speech and Signal Processing, including IEEE T-ASLP, Speech Communication, ICASSP, InterSpeech, ASRU, etc
- **16** IEEE Transaction, **5** Speech Communication; **64** ICASSP, **60** InterSpeech, **12** ASRU
- Papers are cited **12,000+** times (Google Scholar), **9** papers ranked as ESI top 1%, **20** papers ranked as ESI top 3%
- **85** China National Invention Patents are applied, **52** has been granted; **3** USA Patents are applied, **3** granted
- **2** co-authored Books
- **2** co-authored Books chapters

Research Interests

- Speech & Language understanding and human computer interaction
- Automatic speech recognition and translation
- Speaker and language recognition
- Speech separation and enhancement
- Natural language understanding
- Deep learning and machine learning
- Multimedia processing
- GPU and SOC based fast speech processing systems

Projects

- **Research on Brain-like Auditory Frontend Model and System, supported by the Ministry of Science and Technology of China (PI, 18,742,000 ¥)**
- **Speech Signal Processing, Analysis and Recognition, supported by the NFSC (PI, 2,000,000 ¥)**
- **Research on Fast and Efficient Adaptation in Deep Learning Speech Recognition, supported by the NSFC (PI, 540,000 ¥)**
- **Structured Deep Learning Study for the Robust Speech Recognition in the Heterogeneous Noisy Scenario, supported by the NSFC (PI, 220,000 ¥)**

-
- Shanghai Sailing Program, supported by the Shanghai Government (PI, 200,000 ¥)
 - Multi-talker Speech Recognition for Cocktail Party Problem, supported by the Tencent Corporation (PI, 150,000 ¥)
 - High Performance Speech and Speaker Recognition System, supported by AVIC (PI, 500,000 ¥)
 - Deep Neural Network based denoising technology, supported by Baidu (PI, 100,000 ¥)
 - Deep Learning for Noise Robust Speech Recognition, supported by Shanghai Jiao Tong University (PI, 100,000 ¥)
 - Speech Objective Recognition and Content Transcription under Complex Environment, supported by the NSFC (Co-PI, 2,510,000 ¥)
 - Joint SJTU-AISpeech Laboratory, supported by AISpeech Corporation (Co-PI, 5,000,000 ¥)
 - Big Data Driven Natural Language Understanding, QA and Translation, supported by the National Key Research and Development Program of China (involved, ~50,000,000 ¥)
 - Cloud Service Platform for Service Robot, supported by the National Key Research and Development Program of China (involved, ~28,000,000 ¥)
 - Natural Speech Technology, supported by UK-EPSRC (involved, ~9,000,000 \$)
 - Babel, supported by USA-IARPA (involved, ~20,000,000 \$)
 - Speech Recognition Technology Under the Low-Data-Resource Conditions, supported by the NSFC (involved, 830,000 ¥) and the PhD Research and Innovation Fund of Tsinghua University (involved, 40,000 ¥)
 - Kaldi Speech Recognition Toolkit Development and Research
 - Large Vocabulary Continuous Speech Recognition System and Spoken Term Detection System Development and Research, Supported by the China 863 Projects, NSFC Projects and the Projects from China's Ministry of National Defense
 - Multilingual Speech Recognition Research, Supported by the Interdisciplinary Fund Support by School of Information Science and Technology in Tsinghua University (involved, 100,000 ¥)
 - Speech Recognition SOC System Development Under the Low-Hardware-Resource Condition, The SOC system is applied in the 2008 Olympic mascots, and win the High-tech Olympics Advanced Award (involved)

Open-Source Toolkit

2023	Wespeaker Speaker Embedding Learning Toolkit, released in 2023 (https://github.com/wenet-e2e/wespeaker)
2021	ESPNet-SE End-to-End Speech Enhancement and Separation Toolkit, released in 2021 (https://github.com/espnet/espnet)
2016	CUED-RNN LM Toolkit, released in 2016 (http://mi.eng.cam.ac.uk/projects/cued-rnnlm/)
2011	The Kaldi Speech Recognition Toolkit, released in 2011 (http://github.com/kaldi-asr/kaldi)

International Challenges

2022	Conversational Short-phrase Speaker Diarization Challenge, ranked 1 st of 40 teams
2022	VoxCeleb Speaker Recognition Challenge, ranked 3 rd of 30 teams in track 1
2022	VoxCeleb Speaker Recognition Challenge, ranked 3 rd of 30 teams in track 3
2022	CN-Celeb Speaker Recognition Challenge, ranked 1 st of 15 teams in the fixed track
2022	CN-Celeb Speaker Recognition Challenge, ranked 2 nd of 15 teams in the open track
2021	Children Speech Recognition Challenge, ranked 1 st of 28 teams in track 1
2020	Accented English Speech Recognition Challenge, ranked 1 st of 49 teams in track 1
2020	Accented English Speech Recognition Challenge, ranked 2 nd of 49 teams in track 2
2019	Mandarin-English Code-Mix Speech Recognition Challenge, ranked 2 nd of 28 teams in track 3
2016	BTAS 2016 Speaker Anti-spoofing Competition, ranked 3 rd of 7 teams

-
- | | |
|------|---|
| 2015 | MGB Recognition Challenge - Recognition of Multi-Genre Broadcast Data, ranked 1 st of 20 teams |
| 2015 | Automatic Speaker Verification Spoofing and Countermeasures Challenge, ranked 3 rd of 16 teams |

Selected Talks

- | | |
|------|--|
| 2019 | Robust Speech Processing on Multi-talker Mixed Speech @ Cambridge University |
| 2019 | Recent Advances in Deep Embedding Learning for Speaker Identification and Spoofing Detection @ Duke Kunshan University |
| 2017 | New Challenges and Recent Progresses in Speech Recognition @ Microsoft Research AI Summit |
| 2017 | Some Advances in Deep Learning based Acoustic Modeling for Noise Robust Speech Recognition @ Tsinghua University |
| 2016 | Speech Processing on Multi-Genre Broadcast Media @ Microsoft Research Redmond |
| 2016 | Theory and Practice: Current Status and Effective Toolkits for Speech Recognition @ China Mobile Developer Conference |
| 2015 | Far-Field Speech Recognition @ Edinburgh University |

Research Awards

- | | |
|------|---|
| 2022 | Changjiang Scholars Program of the Ministry of Education in China (教育部长江学者奖励计划) |
| 2021 | Outstanding Youth Fund of the National Natural Science Foundation of China (国家自然科学基金优秀青年科学基金) |
| 2020 | The First Prize of the Wu Wenjun Artificial Intelligence Science and Technology Award (吴文俊人工智能科学技术奖, 自然科学奖一等奖) |
| 2019 | Speech Communication Journal Best Paper Award |
| 2019 | IEEE ASRU Best Paper Award |
| 2017 | Shanghai Jiao Tong University SMC-Chenxin Level-B Young Scholar Award (上海交通大学“SMC-晨星青年学者奖励”优秀青年教师 B 类) |
| 2016 | IEEE ISCSLP Best Student Paper Award |
| 2016 | Shanghai Science and Technology Young Scholar Award (上海市青年科技英才) |
| 2015 | Shanghai Jiao Tong University SMC-Chenxin Level-C Young Scholar Award (上海交通大学“SMC-晨星青年学者奖励”优秀青年教师 C 类) |
| 2014 | The Second Prize of the Wu Wenjun Artificial Intelligence Science and Technology Award (吴文俊人工智能科学技术奖, 科技进步奖二等奖) |
| 2014 | Excellent Tutor for Bachelor Degree Thesis in SJTU (上海交通大学优异学士论文优秀指导教师) |
| 2013 | The Second Excellent Doctoral Dissertation Award in Tsinghua University |
| 2012 | Google Grants Award in InterSpeech2012 (Total 4 PhDs around the world) |
| 2012 | Tsinghua-JiangZhen Scholarship, First Class (Total 25 students in Tsinghua University) |
| 2011 | Tsinghua-JiangZhen Scholarship, First Class (Total 25 students in Tsinghua University) |
| 2010 | Excellent PhD Academic Newcomer Award Nomination of Chinese Education Ministry |
| 2010 | PhD Research and Innovation Award of Tsinghua University |
| 2009 | Interdisciplinary Fund Support by School of Information Science and Technology in Tsinghua University |

Qualification & Activities

- IEEE Senior Member & ISCA Member
- Member of IEEE Speech and Language Processing Technical Committee
- Kaldi Group Member & Developer
- Regular reviewer for IEEE/ACM Transactions on Audio, Speech and Language, IEEE Journal of Selected Topics in Signal Processing, IEEE Signal Processing Letter, Speech Communication, Computer Speech and Language, Neurocomputing, Multimedia Tools and Applications, etc
- Regular reviewer for International conferences: ICASSP, INTERSPEECH, ASRU, SLT, ISCSLP, ChinaSip, EUSIPCO, COCODA, NCMMS, ICPR, etc

List of Patents (Partial)

1. **Yanmin Qian**, Dong Yu. Monaural multi-talker speech recognition with attention mechanism and gated convolutional networks. Applied ID: 16/050,825, Applied Date: 2018-07-31, Authorized Date: 2020-02-28.
2. Jia Liu, **Yanmin Qian**. Embedded Mandarin-English bilingual speech recognition approach. Applied ID: CN200910242406.X, Applied Date: 2009-12-10, Granted ID: CN101727901B, Authorized Date: 2011-11-09.
3. Nanxin Chen, Linting Ge, Hao Gu, Xuankai Chang, **Yanmin Qian**, Kai Yu. Text-dependent speaker recognition method based on joint deep learning. Applied ID: CN201510107647.9, Applied Date: 2015-03-12, Granted ID: CN104732978B, Granted Date: 2015-06-24, Authorized Date: 2018-05-08.
4. Kai Yu, **Yanmin Qian**, Yimeng Zhuang, Zhehuai Chen, Xuankai Chang. Customizable voice wake-up method and system. Applied ID: CN201610462976.X, Applied Date: 2016-06-23, Granted ID: CN106098059A, Granted Date: 2016-11-09, Authorized Date: 2019-06-18.
5. **Yanmin Qian**, Nanxin Chen, Kai Yu. Method and system for detecting voice spoofing attack of speakers on basis of deep learning. Applied ID: CN201610478041.0, Applied Date: 2016-06-27, Granted ID: CN105869630A, Granted Date: 2016-08-17, Authorized Date: 2019-08-02.
6. Kai Yu, **Yanmin Qian**, Xu Xiang. Speech recognition system based on acoustic model of binary neural network. Applied ID: 201710055681.5, Applied Date: 2017-01-25, Granted ID: CN106816147A, Granted Date: 2017-06-09.
7. **Yanmin Qian**, Kai Yu, Heinrich Dinkel. For voice fraud detection of the convolution-based short and long-term memory end-to-end depth neural network. Applied ID: 201710062794.8, Applied Date: 2017-01-25, Granted ID: CN106875007A, Granted Date: 2017-06-20.
8. Kai Yu, **Yanmin Qian**, Shuai Wang. The single-channel multi-speaker identification method and system. Applied ID: CN201810053962.1, Applied Date: 2018-01-19, Granted ID: CN108417201A, Granted Date: 2018-08-17.
9. **Yanmin Qian**, Wen Ding, Tian Tan. A deep mixed generating network self-adaptive method and system. Applied ID: CN201810054314.8, Applied Date: 2018-01-19, Granted ID: CN108417207A, Granted Date: 2018-08-17.
10. **Yanmin Qian**, Zili Huang, Shuai Wang. Speaker recognition network model training method, speaker recognition method and system. Applied ID: CN201810025592.0, Applied Date: 2018-01-11, Granted ID: CN108417217A, Granted Date: 2018-08-17.
11. Kai Yu, **Yanmin Qian**, Zhehuai Chen, Yongbin You. The audio data identification method and system. Applied ID: CN201810025834.6, Applied Date: 2018-01-11, Granted ID: CN108389575A, Granted Date: 2018-08-10.
12. **Yanmin Qian**, Yongbin You, Zhehuai Chen, Minghun Huang. The compressed speech recognition model optimization method and system. Applied ID: CN201810021903.6, Applied Date: 2018-01-10, Granted ID: CN108389576A, Granted Date: 2018-08-10.
13. **Yanmin Qian**, Xiaowei Jiang, Shuai Wang, Xu Xiang. The voice platform for intelligent dialogue audio training and recognition method and electronic apparatus. Applied ID: CN201711320515.X, Applied Date: 2017-12-12, Granted ID: CN108109613A, Granted Date: 2018-06-01.
14. Kai Yu, **Yanmin Qian**, Qi Liu. A method of constructing the language model method, input method and system.

Applied ID: CN201711367409.7, Applied Date: 2017-12-18, Granted ID: CN108108428A, Granted Date: 2018-06-01.

15. Kai Yu, **Yanmin Qian**, Yue Wu, Tianxing He, Zhehuai Chen. Multi-view language recognition method based on unidirectional self-tagging auxiliary information. Applied ID: CN201710561261.4, Applied Date: 2017-07-11, Granted ID: CN107452374A, Granted Date: 2017-12-08.
16. **Yanmin Qian**, Peiyao Sheng, Zhuolin Yang, Tian Tan. Speech data augmentation method and system. Applied ID: CN201810792672.9, Applied Date: 2018-07-18, Granted ID: CN108922518A, Granted Date: 2018-11-30.
17. Kai Yu, **Yanmin Qian**, Zhehuai Chen, Hao Li, Qi Liu, The training method and system for end-to-end speech recognition models, Applied ID: CN201811523980.8, Applied Date: 2018-12-13, Granted ID: CN109346064A, Granted Date: 2019-02-15.
18. Kai Yu, **Yanmin Qian**, Jiaqi Guo, Yongbin You, Joint decoding method and system for speech recognition, Applied ID: CN201811583120.3, Applied Date: 2018-12-24, Granted ID: CN109559749A, Granted Date: 2019-04-02.
19. Kai Yu, **Yanmin Qian**, Shuai Wang, Yexin Yang, The method and device for knowledge distillation, Applied ID: CN201811645776.3, Applied Date: 2018-12-29, Granted ID: CN109637546A, Granted Date: 2019-04-16.
20. Lianwu Chen, Meng Yu, **Yanmin Qian**, Dan Su, Dong Yu. The method and device for multi-talker speech separation. Applied ID: 2018109044889, Applied Date: 2018-08-09.
21. **Yanmin Qian**, Dong Yu. Adaptive permutation invariant training with auxiliary information for monaural multi-talker speech recognition. Applied ID: 15/940,246, Applied Date: 2018-03-29.
22. **Yanmin Qian**, Dong Yu. Knowledge transfer in permutation invariant training for single-channel multi-talker speech recognition. Applied ID: 15/940,197, Applied Date: 2018-03-29.