Xiong Xiao

Senior research scientist
Nanyang Technological University

+65 9489 1326
xiaoxiong@ntu.edu.sg
http://www3.ntu.edu.sg/home/xiaoxiong
50 Nanyang Drive, Singapore 637553

Highlights

- Supervising PhD students in speech recognition and signal processing areas.
- Leading projects for robust speech information indexing and retrieval.
- Developing deep learning tool for signal processing: https://github.com/singaxiong/SignalGraph.

Experience

2016 2013 **SignalGraph - A Matlab Based Deep Learning Toolkit**

Developing a deep learning toolkit (https://github.com/singaxiong/SignalGraph)
- Supports directed acyclic graph (DAG) network topology and GPU computing.
- Supports mainstream network types, such as DNN, CNN, and LSTM.
- Used in more than 10 publications in the past 2 years.
- implements more than 60 types of layers for signal processing and speech recognition.

2016 2015 **CHiME 2015 and 2016 Speech Separation and Recognition Challenge**

Leading a team of 5 to work on improving beamforming for speech recognition. The task is to recognize noisy speech recorded by 6 microphones mounted on an ipad. We reduced the word error rate (WER) from 11.7% (conventional beamforming based on signal processing) to 8.9% by combining signal processing and deep learning.

Aug 2015 Jul 2015 **2015 Jelinek Summer Workshop on Speech and Language Technology**

Worked with speech processing researchers from Microsoft and other organizations on a computational network that integrates both beamforming and acoustic modeling of speech recognition. The beamforming network is an end-to-end speech recognition system where the parameters in the beamforming and acoustic modeling modules are optimized jointly to reduce speech recognition cost function. Significant improvement has been obtained on real recordings of meeting rooms.

2014 2013 **REVERB Challenge 2014**

Applied DNN to the dereverberation of speech signals. Proposed a sequential cost function that improves the temporal structure of predicted spectrogram. Achieved significant reverberation reduction on both simulated and real reverberant speech.

2016 2014 **IARPA OpenKWS 2014/2015/2016**

Participated in the joint team SINGA (NTU, I^2R, GeorgiaTech), coordinated the NTU effort. The task is to find the occurene of a text-base keyword (a word or a phrase) in a speeh database. The SINGA team achived the No. 4 performance for year 2014 and 2015 and No. 5 for year 2016, although we have less resource than the sponsored teams.

2015 2014 **MediaEval Query by Example Search on Speech (QUESST) 2014/2015**

Lead a team of 5 persons. The task is to find occurence of spoken queries in a speech database. My team achieved No.2 and No.1 performance in the 2014 and 2015 evaluations, respectively. After the evaluation, I am working on an speech indexing method based on fixed dimensional vector embedding, which can scale up to large speech database.

2015 2015 **Automatic Speaker Verification Spoofing and Countermeasures**

The task is to detect whether an utterance is spoken by human or synthesized by a machine. The purpose is to protect speaker verification systems from spoofing attacks. Investigated several networks (DNN, CNN, LSTM) and features (magnitude and phase). Obtained <1% equal error rate for 10 types of common spoofing attacks.

Sep 2008 – MARCH 2008 | Visiting Scholar in Georgia Institute of Technology (GeorgiaTech)
Worked with Prof. Chin-Hui Lee and Dr. Jinyu Li on discriminative acoustic modeling using soft margin based objective function. Applied the technique on Aurora-2/3/4 benchmarking tasks and obtained significant improvements in WER in noisy test conditions.

Current – 2009 | Speech Research Projects Within My Team
Work with team members to develop automatic speech recognition systems for Southeastern Asian languages, keyword search prototypes with GUI, speaker verification and language recognition systems, robust array processing methods, etc.

WORK EXPERIENCE

Current – FEB 2011 | Senior Research Scientist (Permanent Position)
at TEMASEK LABORATORIES, NANYANG TECHNOLOGICAL UNIVERSITY (NTU), SINGAPORE
Managing research projects and participation in international evaluations, supervising PhD students and interns, performing research in various speech enhancement tasks, including speech recognition, single/multi-channel speech processing such as beamforming, keyword search with text or audio queries, discrimination of natural and synthetic speech.

JAN 2011 – FEB 2009 | Research Fellow
at SCHOOL OF COMPUTER ENGINEERING, NTU, SINGAPORE
Building speech recognition systems for Malay language. Performing research in robust speech recognition, such as histogram equalization for feature normalization and sparse acoustic feature/model combination.

MENTORSHIP EXPERIENCE

Co-supervised with Assoc. Prof. Eng Siong Chng and Prof. Haizhou Li
at Nanyang Technological University, Singapore:

2009-2015	Van Hai Do (PhD): Crosslingual acoustic modeling for low resource languages.
2009-2015	Duc Hoang Ha Nguyen (PhD): Robust acoustic modeling in adverse environments.
2016-NOW	Chenglin Xu (PhD): Deep learning for microphone array processing.

Co-supervised with Prof. Lei Xie at Northwestern Polytechnical University, China:

2015-NOW	Jia Yu (PhD): Story boundary detection and topic modeling.
2016-NOW	Jinyong Hou (PhD): Sequential data embedding for speech information retrieval.
2016-NOW	Sining Sun (PhD): Deep learning for speech enhancement and recognition.

EDUCATION

JULY 2010 | Doctor of Philosophy, **Nanyang Technological University (NTU)**, Singapore
Thesis: "Robust speech features and acoustic models for speech recognition"
Advisors: Assoc. Prof. Eng Siong CHNG and Prof. Haizhou LI

JULY 2004 | Bachelor degree in COMPUTER ENGINEERING, NTU, Singapore
Graduated with the first class honours | Dean's list for 2003 and 2004

PROGRAMMING SKILLS

Advanced: C, Bash, Matlab, LaTeX
HTK and KALDI speech recognition toolkit.

Selected Publications

Robust Speech Recognition and Signal Processing

Conferences

- **Xiong Xiao**, Shinji Watanabe, Hakan Erdogan, Liang Lu, John Hershey, Michael L. Seltzer, Guoguo Chen, Yu Zhang, Michael Mandel, Dong Yu, "**Deep beamforming networks for multi-channel speech recognition**", in Proceedings of ICASSP 2016.
- **Xiong Xiao**, Shengkui Zhao, Thi Ngoc Tho Nguyen, Douglas L. Jones, Eng Siong Chng, Haizhou Li, "**An expectation-maximization eigenvector clustering approach to direction of arrival estimation of multiple speech sources**", in Proceedings of ICASSP 2016.
- **Xiong Xiao**, Shengkui Zhao, Xionghu Zhong, Douglas L. Jones, Eng Siong Chng, Haizhou Li, "**Learning to Estimate Reverberation Time in Noisy and Reverberant Rooms**", in proceedings of InterSpeech 2015.
- **Xiong Xiao**, Shengkui Zhao, Xionghu Zhong, Douglas L. Jones, Eng Siong Chng, Haizhou Li, "**A learning-based approach to direction of arrival estimation in noisy and reverberant environments**", in proceedings of ICASSP 2015.
- Shengkui Zhao, **Xiong Xiao**, Zhaofeng Zhang, Thi Ngoc Tho Nguyen, Xionghu Zhong, Bo Ren, Longbiao Wang, Douglas L. Jones, Eng Siong Chng, Haizhou Li, "**Robust Speech Recognition Using Beamforming With Adaptive Microphone Gains and Multichannel Noise Reduction**", in Proceedings of ASRU 2015.
- **Xiong Xiao**, Jinyu Li, Eng Siong Chng, Haizhou Li, "**Feature Compensation Using Linear Combination of Speaker and Environment Dependent Correction Vectors**", in proceedings of ICASSP 2014.
- **Xiong Xiao**, Shengkui Zhao, Duc Hoang Ha Nguyen, Xionghu Zhong, Douglas L. Jones, Eng Siong Chng, Haizhou Li, "**The NTU-ADSC systems for Reverberation Challenge 2014**", in proceedings of Reverberation Challenge Workshop 2014.
- **Xiong Xiao**, Eng Siong Chng, Haizhou Li, "**Attribute-based histogram equalization (HEQ) and its adaptation for robust speech recognition**", in proceedings of Interspeech 2013.
- **Xiong Xiao**, Eng Siong Chng, Haizhou Li, "**Temporal filter design by minimum KL divergence criterion for robust speech recognition**", in proceedings of ICASSP 2013.
- **Xiong Xiao**, Jinyu Li, Eng Siong Chng, Haizhou Li, "**Lasso Environment Model Combination for Robust Speech Recognition**", in Proceedings of ICASSP 2012.
- **Xiong Xiao**, Eng Siong Chng, Haizhou Li, "**Joint Spectral and Temporal Normalization of Features for Robust Recognition of Noisy and Reverberated Speech**", in Proceedings of ICASSP 2012.
- **Xiong Xiao**, Jinyu Li, Eng Siong Chng, and Haizhou Li, "**Maximum likelihood adaptation of histogram equalization with constraint for robust speech recognition**", in Proceedings of ICASSP 2011.
- **Xiong Xiao**, Jinyu Li, Eng Siong Chng, and Haizhou Li, "**Feature Normalization Using Structured Full Transforms for Robust Speech Recognition**", in Proceedings of Interspeech 2011.
- **Xiong Xiao**, Jinyu Li, Eng Siong Chng, Haizhou Li, and Chin-Hui Lee, "**A Study on Hidden Markov Model's Generalization Capability for Speech Recognition**", in Proceedings of ASRU 2009.
- **Xiong Xiao**, Eng Siong Chng, and Haizhou Li, "**Evaluating the temporal structure normalization technique on the Aurora-4 task**", in the proceedings of InterSpeech 2007.
- **Xiong Xiao**, Eng Siong Chng, and Haizhou Li, "**Normalizing the speech modulation spectrum for robust speech recognition**", in the proceeding of the ICASSP 2007.

Journals

- Duc Hoang Ha Nguyen, **Xiong Xiao**, Eng Siong Chng, Haizhou Li, "**Feature Adaptation Using Linear Spectro-Temporal Transform for Robust Speech Recognition**", accepted by IEEE/ACM Transactions on Audio, Speech, and Language Processing.

- **Xiong Xiao**, Shengkui Zhao, Duc Hoang Ha Nguyen, Xionghu Zhong, Douglas L. Jones, Eng Siong Chng, Haizhou Li, "**Speech dereverberation for enhancement and recognition using dynamic features constrained deep neural networks and feature adaptation**", EURASIP Journal on Advances in Signal Processing, 2016(1), pp. 1-18.
- **Xiong Xiao**, Jinyu Li, Eng Siong Chng, Haizhou Li, and Chin-Hui Lee, "**A Study on the Generalization Capability of Acoustic Models for Robust Speech Recognition**", IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 6, pp. 1158-1169, August 2010.
- **Xiong Xiao**, Eng Siong Chng, and Haizhou Li, "**Normalization of speech modulation spectra for robust speech recognition**", IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 8, pp. 1662-1674, November 2008.
- **Xiong Xiao**, Eng Siong Chng, and Haizhou Li, "**Temporal structure normalization of speech feature for robust speech recognition**", IEEE Signal Processing Letters, vol. 14, no. 7, pp. 500-503, July 2007.

Spoken Term Detection With Text or Audio Queries

- Nancy Chen, Van Tung Pham, Haihua Xu, **Xiong Xiao**, Van Hai Do, Chongjia Ni, I-Fan Chen, Sunil Sivadas, Chin-Huil Lee, Eng Siong Chng, Bin Ma, Haizhou Li, "**Exemplar-Inspired Strategies for Low-Resource Spoken Keyword Search in Swahili**", in Proceedings of ICASSP 2016.
- Van Tung Pham, Haihua Xu, **Xiong Xiao**, Nancy F. Chen, Eng Siong Chng, Haizhou Li, "**Keyword search using query expansion for graph-based rescoring of hypothesized detections**", in Proceedings of ICASSP 2016.
- Haihua Xu, Jingyong Hou, **Xiong Xiao**, Van Tung Pham, Cheung-Chi Leung, Lei Wang, Van Hai Do, Hang Lv, Lei Xie, Bin Ma, Eng Siong Chng, Haizhou Li, "**Approximate search of audio queries by using DTW with phone time boundary and data augmentation**", in Proceedings of ICASSP 2016.
- Haihua Xu, Peng Yang, **Xiong Xiao**, Lei Xie, Cheung-Chi Leung, Hongjie Chen, Jia Yu, Hang Lv, Lei Wang, Su Jun Leow, Bin Ma, Eng Siong Chng, Haizhou Li, "**Language independent query-by-example spoken term detection using N-best phone sequences and partial matching**", in proceedings of ICASSP 2015.

Synthetic Speech Detection

- Xiaohai Tian, Zhizheng Wu, **Xiong Xiao**, Eng Siong Chng, Haizhou Li, "**Spoofing detection from a feature representation perspective**", accepted by ICASSP 2016.
- **Xiong Xiao**, Xiaohai Tian, Steven Du, Haihua Xu, Eng Siong Chng, Haizhou Li, "**Spoofing Speech Detection Using High Dimensional Magnitude and Phase Features: the NTU Approach for ASVspoof 2015 Challenge**", in proceedings of InterSpeech 2015.
- Zhizheng Wu, **Xiong Xiao**, Eng Siong Chng, Haizhou Li, "**Synthetic speech detection using temporal modulation feature**", in proceedings of ICASSP 2013.