VIET ANH TRINH
vtrinh@gradcenter.cuny.edu | 551-214-6767 | tvanh512 (Skype) | New York
http://trinhvietanh.com | https://github.com/tvanh512 | https://www.linkedin.com/in/trinhvietanh/

Education

2016 - Present Ph.D. in Computer Science, Graduate Center, City University of New York (CUNY), US

- Research topics: Speech recognition, speech synthesis, feature importance, invariant representation, interpretable machine learning, continual learning, grammatical error correction, machine translation.
- Advisor: Professor Michael I Mandel

2003 - 2008 B.S. in Electronics and Telecommunications, Hanoi University of Science and Technology, VN

- Top 1% of the Electronics and Telecommunications Department

Technical skills

PyTorch, Tensorflow, Keras, ESPnet, Kaldi, NLTK, Moses, Message Passing Interface
Python, Matlab, C, C++, PHP, Java, Visual Basic, R, MySQL, HTML

Publications

V. A. Trinh and M. I. Mandel, "Directly comparing the listening strategies of humans and machines," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 312–323, 2021.

V. A. Trinh and M. I. Mandel, "Large scale evaluation of importance maps in automatic speech recognition," in *Proceedings of Interspeech*, 2020.

V. A. Trinh, B. McFee, and M. I. Mandel, "Bubble cooperative networks for identifying important speech cues," in *Proceedings of Interspeech*, 2018.

A. R. Syed, V. A. Trinh, and M. I. Mandel, "Concatenative resynthesis with improved training signals for speech enhancement," in *Proceedings of Interspeech*, 2018.

Research experience

Present Improving end to end automatic speech recognition (ESPnet) with the important speech cues

2019 Direct comparison of the listening strategies in noise between humans and machines

- Identified and compared the important time-frequency regions of human listeners and the automatic speech recognition (ASR)
- Discovered that the important time-frequency regions of the time-delay neural network - long short-term memory networks acoustic model (TDNN-LSTM AM) were more similar to those of humans than the traditional Gaussian mixture model's regions (GMM)
- The Jaccard similarity between human and GMM AM importance maps was 4.4% while human and neural network AM was 8.9%
- Recommended that the performance of ASR in noisy conditions could be improved by adapting it to attend to the same regions that humans use
- Tool: Kaldi. Dataset: AMI, CHIME-2 track 1

2019 Grammatical Error Correction for Russian

- Implemented a multilayer convolutional encoder-decoder neural network model (PyTorch/ fairseq) and augmented the available training data with the approach of minimally-augmented grammatical error correction.
- Provided benchmark results of the state-of-the-art neural machine translation (NMT) model with synthetic data generation on RU-Lang8
- Proved the usefulness of RU-Lang8 as an additional source of training data for the NMT models

2018 Bubble cooperative networks for identifying important speech cues

- Proposed a system called the Bubble cooperative network (BCN) consisting of a generator (LSTM) and a discriminator (LSTM) to identify important time-frequency regions of speech
- The BCN could obscure 97.7% the spectrogram with noise while maintaining recognition accuracy for a speech recognizer comparing a noisy test with a clean reference utterance.

- The masks predicted by BCN showed patterns similar to analyses derived from human listening tests with better generalization and less context-dependence than other approaches

2017 Concatenative analysis-by-synthesis

- Utilized pitch and intensity information to improve the performance of a feed-forward neural network unit-selection in a concatenative speech synthesizer system, which is aimed at producing a high-quality clean speech from noisy speech.

2016 Multi-channel speech enhancement

- Deployed a baseline method, which estimated the noise covariance matrix for the beamforming to improve the far-field speech recognition

Work experience

2020 **Amazon**, Applied Scientist Intern, Acoustic Modeling team, Alexa Speech, US

- Mentors: Andreas Stolcke, Brian King, Jasha Droppo and Pegah Ghahremani
- Project: NDA. Paper under internal review for publication.

2019 **Amazon**, Applied Scientist Intern, Acoustic Modeling team, Alexa Speech, US

- Mentors: Chengyuan Ma, Che-Wei Huang and Reshma Thomas
- Project: NDA. Paper under internal review for publication

2016 - Present **City University of New York**, Research Assistant, US

- Worked with my advisor in three projects: Bubble cooperative networks, multi-channel speech enhancement and concatenative analysis-by-synthesis

2011 - 2016 **Texas Instruments(TI)** Technical Business Development Engineer, Vietnam

- Provided TI solutions and integrated circuit products for clients to build electronic devices: smart phone, telecom base station, set top box, smart home devices and toy robots
- Managed TI North Vietnam sale and increased revenue by 250% in 2012, 27% in 2013, 69% in 2014, 150% in 2015 and 30% in 2016
- Conducted bi-weekly review with distributors: Avnet, Arrow, SS, WT and WPI
- Received reward letter from TI Asia President for achievement in 2016

Service

Reviewer: Association for the Advancement of Artificial Intelligence (AAAI) 2020. Subreviewer: International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018 and 2020, International Conference on Learning Representations (ICLR) 2019, International Speech Communication Association (Interspeech) 2019, Neural Information Processing Systems (NeurIPS) 2018, AAAI 2018.