Zhehuai (Tom) Chen

SJTU SpeechLab
Department of Computer Science and Engineering
Shanghai Jiao Tong University
3-502 SEIEE Building, 800 Dongchuan Road, Shanghai, 200240

Phone: +086 15921010742
Email: chenzhehuai@sjtu.edu.cn
chenzhehuai@foxmail.com
Skype: chenzhehuai@outlook.com

RESEARCH INTERESTS

Automatic speech recognition (ASR) with focus on weighted finite-state transducers (WFST), novel inference architecture (Decoder), discriminative training and end-to-end (E2E) System.
Keyword Spotting (KWS), Robust ASR, Parallel Computing, Language Model, Speech Synthesis.

EDUCATION

2014 - present, Ph.D. Candidate, Computer Science, Shanghai Jiao Tong University.
Supervised by Prof. Kai Yu (`http://speechlab.sjtu.edu.cn/~kyu/`).

2010 - 2014, B.E., Electronics and Information Engineering, Huazhong University of Science and Technology.

RESEARCH EXPERIENCES

GPU WFST Decoder Visiting Research Scholar (JHU & NVIDIA) 2018.1 - 2018.4

Working on an extension [1] of the Kaldi toolkit that supports WFST decoding on GPUs, supervised by Daniel Povey (`http://www.danielpovey.com`). Kaldi contributor.

The lattice based WFST decoder achieves identical results and significant speedups (**15**-fold for single sequence and **46**-fold with sequence parallelism). We submit a conference paper on this topic.

Robust ASR Research Internship (MSR, Redmond) 2017.5 - 2017.7

Research internship in *speech and dialog research group*, Microsoft Research, Redmond, supervised by Jasha Droppo (`https://www.microsoft.com/en-us/research/people/jdroppo/`).

Significantly advancing the state-of-the-art unsupervised single-channel overlapped speech recognition system and publishing a transaction and a conference paper on this topic. CNTK contributor.

Speech Recognition End-to-end ASR and Decoding Framework 2014.8 - present

1. Inference framework in CTC. The proposed PSD framework achieves **5**-fold speedup versus traditional CTC-based system and **30**-fold speedup versus HMM-based system. The framework can be extended to LF-MMI.
2. End-to-end Speech Recognition. Propose modular training strategy for direct acoustics-to-words (A2W) modeling.
3. Sequence discriminative training in KWS. Solve the search space modeling problem in KWS. Sequence discriminative training in both HMM and CTC achieves significant improvement.
4. Confidence measure in CTC. The proposed confidence measure achieves significant improvement versus the traditional method in both CTC and HMM trained models.
5. LSTM language modeling and lattice rescoring. Speed up the lattice rescoring significantly. The improvement includes model inference, history clustering and stream parallelization.

[1] `https://github.com/chenzhehuai/kaldi/tree/gpu-decoder`

6. Human-directed ASR errors are collected from confusion network. BLSTM language model is trained to estimate sentence completion scores, combined with the confusion network scores to do correction.
7. Implement On-the-fly Rescore WFST decoding method and compare it with traditional 2-pass rescore WFST decoder.
8. Design a parametric model, which can be inferenced with offline decoding records of the whole process, to tune beam dynamically by features in decoding process so as to thoroughly speedup.
9. Improve lattice quality in WFST Decoder by efficient time realignment. By this way, Confidence measure and ASR result from Confusion Network(CN) can both be improved.

Speech Synthesis Speech Synthesis using HMM & DNN 2014.3-2014.7

Develop HMM & DNN Speech Synthesis systems and analyze the performance gap between them.

PUBLICATIONS

Zhehuai Chen, Justin Luitjens, Hainan Xu, Yiming Wang, Daniel Povey, Sanjeev Khudanpur, A GPU-based WFST Decoder with Exact Lattice Generation, submit to 19th Annual Conference of the International Speech Communication Association (InterSpeech), 2018.

Zhehuai Chen, Jasha Droppo, Sequence Modeling in Unsupervised Single-channel Overlapped Speech Recognition, IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), Calgary, Canada, 2018.

Zhehuai Chen, Qi Liu, Hao Li, Kai Yu, On Modular Training of Neural Acoustics-to-word Model for LVCSR, IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), Calgary, Canada, 2018.

Zhehuai Chen, Yanmin Qian, Kai Yu, Sequence Discriminative Training for Deep Learning based Acoustic Keyword Spotting, 2018, accepted by Speech Communication.

Zhehuai Chen, Jasha Droppo, Jinyu Li, Wayne Xiong, Progressive Joint Modeling in Unsupervised Single-channel Overlapped Speech Recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 1, pp. 184-196, Jan. 2018. doi: 10.1109/TASLP.2017.2765834.

Zhehuai Chen, Yanmin Qian, and Kai Yu. A unified confidence measure framework using auxiliary normalization graph, IScIDE, 2017.

Zhehuai Chen, Yimeng Zhuang, Kai Yu. Confidence Measures for CTC-based Phone Synchronous Decoding. IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), New Orleans, USA, 2017.

Zhehuai Chen, Yimeng Zhuang, Yanmin Qian, Kai Yu. Phone Synchronous Speech Recognition with CTC Lattices. IEEE/ACM Transactions on Audio, Speech and Language Processing, vol. 25, no. 1, pp. 86-97, Jan. 2017. doi: 10.1109/TASLP.2016.2625459.

Zhehuai Chen, Wei Deng, Tao Xu, Kai Yu. Phone Synchronous Decoding with CTC Lattice. 17th Annual Conference of the International Speech Communication Association (InterSpeech), San Francisco, America, 2016.

Zhehuai Chen, Kai Yu, An Investigation of Implementation and Performance Analysis of DNN Based Speech Synthesis System. 12th IEEE International Conference on Signal Processing(ICSP), Hangzhou, 2014.

Yue Wu, Tianxing He, **Zhehuai Chen**, Yanmin Qian and Kai Yu. Multi-view LSTM Language Model with Word-synchronized Auxiliary Feature for LVCSR, CCL, 2017.

Da Zheng, **Zhehuai Chen**, Yue Wu, Kai Yu, Directed Automatic Speech Transcription Error Correction Using Bidirectional LSTM. International Symposium on Chinese Spoken Language Processing(ISCSLP), Tianjin, China, 2016.

Bo Chen, **Zhehuai Chen**, Jiachen Xu, Kai Yu. An Investigation of Context Clustering for Statistical Speech Synthesis with Deep Neural Network. 16th Annual Conference of the International Speech Communication Association (InterSpeech), Dresden, Germany, 2015.

AWARDS AND ACTIVITIES

2018, JHU Visiting Research Scholar (mentor: Daniel Povey).

2017, Microsoft Research Internship (mentor: Jasha Droppo).

2017, ICASSP 2017 student travel grant.

2016, Interspeech 2016 student travel grant.

2014 - present, AISpeech Ltd. Internship.

2013, Microsoft Young Fellows Scholarship. (Asia, total 36 undergraduates)

2013, Excellent Student (University, top 2%)

2013, the National Undergraduate Electronic Design Contest (National, Second Prize)

2013, Challenge Cup (Province, First Prize)