Desh Raj

Website: desh2608.github.io
Email: draj@cs.jhu.edu
LinkedIn: rdesh26
GitHub: github.com/desh2608

EDUCATION

The Johns Hopkins University   Baltimore, US
Ph.D. in Computer Science   2018–Present

– Advisors: Sanjeev Khudanpur, Dan Povey

– Research interests: Speech recognition, speaker diarization

Indian Institute of Technology Guwahati   Guwahati, India
B.Tech. in Computer Science and Engineering, GPA: 9.35/10.00   2013–2017

– Advisor: Ashish Anand

– Thesis: "Relation extraction for biomedical text"

PROJECTS

Informed target-speaker ASR   JSALT 2020
Advisors: Marc Delcroix (NTT, Japan), Shinji Watanabe (JHU)

– Devised a novel constrained optimization based approach for overlap-aware diarization; improves DER over baseline system on LibriCSS data from 16.3% to 9.3% (submitted to SLT 2021)

– Experimented with training strategies for target speaker ASR – architecture of embedding usage, discriminative training, etc; cumulative improvement in WER from 27.9% to 16.6%.

– Integrated both components to recognize multi-speaker overlapping speech without explicit separation.

Integration of separation, diarization, and ASR   JSALT 2020
Collaborators: Zhuo Chen (Microsoft), Hakan Erdogan (Google)

– Created modular Kaldi pipeline integrating speech separation, diarization, and ASR.

– Implemented a novel cross-stream clustering technique to diarize multiple audio streams simultaneously — pipeline provides 37.6% relative WER improvement over single-stream approach.

– Evaluated the pipeline with different variants of each module; led efforts for manuscript submission to SLT 2021.

CHiME-6 challenge   Spring 2020
Advisors: Paola Garcia, Shinji Watanabe, Sanjeev Khudanpur

– Created baseline Kaldi recipe for the challenge: TDNN-stats based SAD (5.1% error rate on dev) and x-vector + PLDA backend for diarization ($\sim$36% DER on dev)

– Led diarization efforts for JHU team's participation in the challenge – added multi-array fusion and VB-HMM based overlap assignment to the pipeline.

– Final WER improved by 10% absolute compared with baseline – finished top 2 in "diarization + ASR" track. Prepared system description manuscript for the submission.

EXPERIENCE

Samsung Research   Bengaluru, India
Research Engineer in Advanced Technology Lab (ATL)   June 2017 –June 2018

– Context engine for conversational assistant

Microsoft India　　　　　　　　　　　　　　　　　　　　　　　Hyderabad, India
Software Developer Intern　　　　　　　　　　　　　　　　　　　　　Summer 2016

- – Cross-platform application for digital contracting system

Hanyang University　　　　　　　　　　　　　　　　　　　　　　　Ansan, Korea
Research Intern in Pattern Recognition & Fuzzy Systems Lab　　　　　　　Summer 2015

- – Advisor: Frank Chung-hoon Rhee
- – Analysis of multi-dimensional type-2 fuzzy membership functions, and application to fuzzy ART algorithm for image segmentation

Teaching

- **Teaching Assistant** at Johns Hopkins University　　　　　　　　　　Fall 2020
  Introduction to Human Language Technology (601.647/667)

Mentorship & Professional Services

- **Reviewer**: IEEE SLT 2021
- CLSP Graduate Admissions Committee: 2021
- CLSP Student Recruitment Committee: 2019, 2020
- Student Mentor for 8 freshmen at IIT Guwahati　　　　　　　　　　　2014-15

Skills

- **ML/DL Toolkits:** PyTorch, Tensorflow, Scikit-learn
- **ASR Frameworks:** Kaldi, ESPNet
- **Other:** Audacity, Git

Languages

- **Programming:** Python, C++, Bash
- **Natural:** English, Hindi, French (beginner)
- – **TOEFL:** 119/120

Achievements

- **Top 2** in the CHiME-6 challenge track 2 (diarization + ASR)　　　　　　2020
- **INAE Travel Grant** by Govt. of India (worth INR 50,000)　　　　　　　2017
- **Kalyani Research Scholarship** from Alumni Affairs, IIT Guwahati　　　　2017
- **INSPIRE Scholarship for Higher Education (SHE)** from DST, Govt. of India (refused)　2013

Publications

[1] **D. Raj**, P. Denisov, Z. Chen, H. Erdogan, Z. Huang, M. He, S. Watanabe, J. Du, T. Yoshioka, Y. Luo, N. Kanda, J. Li, S. Wisdom, and J. R. Hershey, "Integration of speech separation, diarization, and recognition for multi-speaker meetings: System description, comparison, and analysis", *IEEE SLT*, 2021.

[2] **D. Raj**, P. Garcia, Z. Huang, S. Watanabe, D. Povey, A. Stolcke, and S. Khudanpur, "DOVER-Lap: A method for combining overlap-aware diarization outputs", in *IEEE SLT*, 2021.

[3] **D. Raj**, Z. Huang, and S. Khudanpur, "Multi-class spectral clustering with overlaps for speaker diarization", *IEEE SLT*, 2021.

[4] Z.-Q. Wang, H. Erdogan, S. Wisdom, K. Wilson, **D. Raj**, S. Watanabe, Z. Chen, and J. R. Hershey, "Sequential multi-frame neural beamforming for speech separation and enhancement", *IEEE SLT*, 2021.

[5] A. Arora, **D. Raj**, A. S. Subramanian, K. Li, B. Ben-Yair, M. Maciejewski, P. Zelasko, P. Garcia, S. Watanabe, and S. Khudanpur, "The JHU multi-microphone multi-speaker ASR system for the CHiME-6 challenge", *CHiME-6 Workshop at IEEE ICASSP*, 2020.

[6] **D. Raj**, D. Snyder, D. Povey, and S. Khudanpur, "Probing the information encoded in x-vectors", *IEEE ASRU*, 2019.

[7] **D. Raj**, S. K. Sahu, and A. Anand, "Learning local and global contexts using a convolutional recurrent network model for relation classification in biomedical text", in *CoNLL*, 2017.