

EDUCATION

The Johns Hopkins University Ph.D. in Computer Science	Baltimore, US 2018–2024
<ul style="list-style-type: none"> – Advisors: Sanjeev Khudanpur, Dan Povey – Research interests: Multi-talker speech recognition and speaker diarization 	
Indian Institute of Technology Guwahati B.Tech. in Computer Science and Engineering, GPA: 9.35/10	Guwahati, India 2013–2017

WORK EXPERIENCE

Meta AI Research Intern, AI Speech (Manager: Ozlem Kalinli)	Menlo Park, US May 2022 –August 2022
<ul style="list-style-type: none"> – Designed and implemented target-speaker ASR models to improve transducer performance in background speech and noise; obtained 19.6% relative WER reduction – Work published at IEEE ICASSP 2023 	
Microsoft Corporation Research Intern, AI Cognitive Services (Manager: Jinyu Li)	Redmond, US (remote) May 2021 –August 2021
<ul style="list-style-type: none"> – Extended Streaming Unmixing and Recognition Transducer (SURT) model for long-form meeting transcription, obtaining >20% WER reduction using dual-path encoders – Work published at IEEE ICASSP 2022 	
Samsung Research Research Engineer in Advanced Technology Lab (ATL)	Bengaluru, India June 2017 –June 2018

AWARDS

• Awarded Frederick Jelinek fellowship for 2023-2024	2023
• Selected for ICASSP Rising Stars in Signal Processing at IEEE ICASSP 2023	2023
• Recipient of the JHU+Amazon AI2AI fellowship for 2022-23	2022
• Reviewing awards: InterSpeech 2023 (top 2%), NeurIPS 2023 (top 8%)	
• JHU nominee for Microsoft Research Fellowship and Apple Scholars in AI/ML	2021
• ISCA Travel Grant (registration + membership + travel funds) for attending InterSpeech	2021
• Placed top 2 in the CHiME-6 (track 2) and DIHARD-3 challenges	2020

TEACHING

• Teaching Assistant at Johns Hopkins University <i>Information Theory (520.447/647)</i>	Fall 2021
• Teaching Assistant at Johns Hopkins University <i>Introduction to Human Language Technology (601.467/667)</i>	Fall 2020

MENTORSHIP & PROFESSIONAL SERVICES

- **Reviewer:** ICML, NeurIPS, ICLR, ICASSP, InterSpeech, ASRU, SLT, Computer Speech and Language, Speech Communications, IEEE TASLP
- **Organizer:** [CHiME-7 DASR Challenge](#), [InterSpeech 2023 tutorial](#) on next-gen Kaldi
- **Student Volunteer:** SLT (2022)
- CLSP Student Recruitment Committee (2019-20) and Graduate Admissions Committee (2021-24)

SKILLS

- **ML/DL Toolkits:** PyTorch, Scikit-learn
- **ASR Frameworks:** Kaldi, ESPNet, Lhotse, k2
- **Other:** Audacity, Git

LANGUAGES

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

PUBLICATIONS

- [22] J. D. Fox, **D. Raj**, N. Delworth, Q. McNamara, C. Miller, and M. Jett'e, "[Updated Corpora and Benchmarks for Long-Form Speech Recognition](#)", in *IEEE ICASSP*, 2024.
- [21] S. Cornell, M. Wiesner, S. Watanabe, **D. Raj**, X. Chang, P. García, Y. Masuyama, Z. Wang, S. Squartini, and S. Khudanpur, "[The CHiME-7 DASR Challenge: Distant Meeting Transcription with Multiple Devices in Diverse Scenarios](#)", in *CHiME Workshop at InterSpeech*, 2023.
- [20] D. Gao, H. Xu, **D. Raj**, L. P. G. Perera, D. Povey, and S. Khudanpur, "[Learning from Flawed Data: Weakly Supervised Automatic Speech Recognition](#)", in *IEEE ASRU*, 2023.
- [19] Z. Huang, **D. Raj**, P. Garcia, and S. Khudanpur, "[Adapting self-supervised models to multi-talker speech recognition using speaker embeddings](#)", in *IEEE ICASSP*, 2023.
- [18] **D. Raj**, J. Jia, J. Mahadeokar, C. Wu, N. Moritz, X. Zhang, and O. Kalinli, "[Anchored Speech Recognition with Neural Transducers](#)", in *IEEE ICASSP*, 2023.
- [17] **D. Raj**, D. Povey, and S. Khudanpur, "[GPU-accelerated Guided Source Separation for Meeting Transcription](#)", in *InterSpeech*, 2023.
- [16] **D. Raj**, D. Povey, and S. Khudanpur, "[SURT 2.0: Advances in Transducer-based Multi-talker Speech Recognition](#)", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023.
- [15] G. Morrone, S. Cornell, **D. Raj**, L. Serafini, E. Zovato, A. Brutti, and S. Squartini, "[Low-Latency Speech Separation Guided Diarization for Telephone Conversations](#)", in *IEEE SLT*, 2022.
- [14] **D. Raj**, L. Lu, Z. Chen, Y. Gaur, and J. Li, "[Continuous Streaming Multi-talker ASR with Dual-path Transducers](#)", in *IEEE ICASSP*, 2022.
- [13] M. Wiesner, **D. Raj**, and S. Khudanpur, "[Injecting Text and Cross-lingual supervision in few-shot learning from self-supervised models](#)", in *IEEE ICASSP*, 2022.
- [12] M. He, **D. Raj**, Z. Huang, J. Du, Z. Chen, and S. Watanabe, "[Target-Speaker Voice Activity Detection with Improved i-Vector Estimation for Unknown Number of Speaker](#)", in *InterSpeech*, 2021.
- [11] **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](#)", in *IEEE SLT*, 2021.
- [10] **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.

- [9] **D. Raj**, Z. Huang, and S. Khudanpur, “[Multi-class spectral clustering with overlaps for speaker diarization](#)”, in *IEEE SLT*, 2021.
- [8] **D. Raj** and S. Khudanpur, “[Reformulating DOVER-Lap Label Mapping as a Graph Partitioning Problem](#)”, in *InterSpeech*, 2021.
- [7] 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](#)”, in *IEEE SLT*, 2021.
- [6] M. Wiesner, M. Sarma, A. Arora, **D. Raj**, D. Gao, R. Huang, S. Preet, M. Johnson, Z. Iqbal, N. K. Goel, J. Trmal, L. P. G. Perera, and S. Khudanpur, “[Training Hybrid Models on Noisy Transliterated Transcripts for Code-Switched Speech Recognition](#)”, in *InterSpeech*, 2021.
- [5] K. Žmolíková, M. Delcroix, **D. Raj**, S. Watanabe, and J. H. Cernocký, “[Auxiliary Loss Function for Target Speech Extraction and Recognition with Weak Supervision Based on Speaker Characteristics](#)”, in *InterSpeech*, 2021.
- [4] 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](#)”, in *CHiME-6 Workshop at IEEE ICASSP*, 2020.
- [3] **D. Raj**, J. Villalba, D. Povey, and S. Khudanpur, “[Frustratingly Easy Noise-aware Training of Acoustic Models](#)”, *ArXiv*, 2020.
- [2] **D. Raj**, D. Snyder, D. Povey, and S. Khudanpur, “[Probing the Information Encoded in X-Vectors](#)”, in *IEEE ASRU*, 2019.
- [1] **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.

See [Google Scholar](#) for a complete list of publications (800+ citations, h-index=14).