Desh Raj

Website: desh2608.github.io
Email: r.desh26@gmail.com
LinkedIn: rdesh26
GitHub: github.com/desh2608

Work Experience

Meta AI — New York, US
Senior Research Scientist, Meta Superintelligence Labs — August 2025 –present

- Integrated chain-of-thought for multi-stream **full-duplex voice models** to improve reasoning accuracy by 2-3x, and proposed novel methods to enable "early thinking" (submitted to ICLR 2026).
- Improving full-duplex behavior (such as back-channels) using **online RL** methods.

Research Scientist, GenAI Foundations — February 2024 –July 2025

- LLaMA-4 speech post-training
  - Built **direct preference optimization** (DPO) recipe using techniques like rejection sampling and modality masking; improved model accuracy by 18% on factuality tasks, and reduced miss turns by 8%.
  - Designed and implemented modular **data-loader** used in all training stages, and an **evaluation framework** supporting on-the-fly tool calling with speech.
- Other speech research (published at IEEE ICASSP 2025)
  - Conceptualized **M-BEST-RQ**, the first multi-channel speech foundation model for smart glasses.
  - ASR using Speech-LLMs: improved the accuracy by 18% using multi-pass training, and latency by 3.2x using multi-token prediction.

Meta AI — Menlo Park, US
Research Intern, AI Speech — May 2022 –August 2022

- Designed and implemented **target-speaker ASR** models to improve transducer performance in background speech and noise, obtaining 19.6% relative WER reduction (published at IEEE ICASSP 2023).

Microsoft Corporation — Redmond, US (remote)
Research Intern, AI Cognitive Services — May 2021 –August 2021

- Developed transducer-based multi-talker ASR for **long-form meeting** transcription, obtaining >20% WER reduction (published at IEEE ICASSP 2022).

Education

The Johns Hopkins University — Baltimore, US
PhD in Computer Science (Advisors: Sanjeev Khudanpur, Dan Povey) — September 2018–January 2024

- Thesis: "Listening to multi-talker conversations: Modular and end-to-end perspectives"

Indian Institute of Technology Guwahati — Guwahati, India
B.Tech. in Computer Science and Engineering, GPA: 9.35/10 — August 2013–May 2017

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

Mentorship & Professional Services

- **Reviewer**: ICML, NeurIPS, ICLR, IJCAI, ICASSP, InterSpeech, ASRU, SLT, Odyssey, Computer Speech and Language, Speech Communications, IEEE TASLP
- **Organizer**: CHiME-7 DASR Challenge, InterSpeech 2023 tutorial on next-gen Kaldi
- **Teaching Assistant:** Information Theory (Fall 2021), Intro to HLT (Fall 2020)

Publications

[27] **D. Raj**, G. Keren, J. Jia, J. Mahadeokar, and O. Kalinli, "Faster Speech-LLaMA Inference with Multi-token Prediction", in *IEEE ICASSP*, 2025.

[26] A. K. S. Yadav, G. Keren, **D. Raj**, W. Zhou, J. Jia, K. Li, Y. Xu, C. Wu, J. Mahadeokar, and O. Kalinli, "Speech-N-LlaMA: Improving Speech LLMs with Multi-Pass Training", in *IEEE ICASSP*, 2025.

[25] Y. Yang, **D. Raj**, J. Lin, N. Moritz, J. Jia, G. Keren, E. Lakomkin, Y. Huang, J. Donley, J. Mahadeokar, and O. Kalinli, "M-BEST-RQ: A Multi-Channel Speech Foundation Model for Smart Glasses", in *IEEE ICASSP*, 2025.

[24] 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.

[23] A. Hussein, **D. Raj**, M. Wiesner, D. Povey, P. Garcia, and S. Khudanpur, "Enhancing Neural Transducer for Multilingual ASR with Synchronized Language Diarization", in *Interspeech*, 2024.

[22] **D. Raj**, M. Wiesner, M. Maciejewski, L. P. Garcia-Perera, D. Povey, and S. Khudanpur, "On Speaker Attribution with SURT", in *Speaker Odyssey*, 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.