Saurabh Kumar

Google Scholar: tinyurl.com/4pyub4pd
GitHub: github.com/saurabhk0317
Email: saurabhk0317@gmail.com
LinkedIn: saurabhk0317

Education

National Institute of Technology, Patna 2016–2020
B.Tech. in Electronics and Communication Engineering
GPA: 7.42/10.00

Experience

SPIRE Lab, IISc Bangalore, India July 2021–Present
Research Associate
Advisor: Prof. Prasanta Kumar Ghosh
Projects:

- **RESPIN**: *Speech Recognition in Agriculture and Finance for the Poor in India*
  Publications: [1]–[5]
  - Developed a multimodal fusion model for joint ASR and dialect identification (DID) (**4.65% CER and 81.63% DID accuracy**), integrating speech and text embeddings for Indian dialects [2].
  - Led the first large-scale study on DID in Indian languages (**79.81% accuracy across 33 dialects**) [3].
  - Enhanced low-resource ASR via adapter-based domain adaptation, reducing **WER (10%-30%)** and **CER (3%-18%)** across 9 languages [4].
  - Proposed a Gated Multi-Encoder ASR model with frame-level gating, reducing **WER by 7.5% (Bhojpuri) and 9% (Bengali)** [5].
- **SYSPIN**: *Text-to-Speech Synthesizer in Nine Indian Languages*
  Publications: [6]
  - Developed a data filtering pipeline using Kaldi-based ASR models, improving corpus quality.

ARTPARK, IISc Bangalore, India Sep 2022–Present
Machine Learning/Signal Processing Consulting Engineer
Project:

- **VAANI**: *Capturing the language landscape for an inclusive digital India*
  - Contributing to the developement of large-scale speech and text corpora covering **773 Indian districts**.
  - Designed automated data pipelines for quality checks using signal processing and ML techniques.

Department of ECE, NIT Patna, India July 2019–June 2020, Dec 2020–June 2021
Undergraduate Student Researcher
Advisor: Dr. S. Shahanawazuddin
Publications: [7], [8]

- Investigated formant-modification and bispectrum-based front-end features to improve children's ASR under zero-resource conditions [7].
- Developed a CGAN-based voice conversion framework, transforming adult speech into child-like speech for ASR adaptation [8].

EICT Academy, IIT Guwahati, India May 2019–June 2019
Summer Intern

- Designed Verilog code for a 5-stage discrete wavelet transform cascade filter bank for FPGA-based simulation.

Mentorship & Professional Services

Reviewer: ICASSP 2024, ICASSP 2025, IJCNN 2025.
Organizer:

- **ICASSP 2024 (LIMMITS'24):** Co-organized session on Indic TTS, developed baselines.
- **ASRU 2023 (MADASR):** Led data curation, ASR baselines for Bengali, Bhojpuri.
- **Interspeech 2022 (Gram Vaani ASR):** Organized Hindi ASR challenge, developed Kaldi baselines.

Student Volunteer: ICASSP 2024, ICASSP 2025.

Awards

SLT-CODE Hackathon Winner 2022 – Won **Best Hackathon Project** Award for developing dialectal ASR systems for Bengali and Bhojpuri, presented at **SLT 2022**.

Selected Publications

[1] **Saurabh Kumar**, A. Singh, D. G, Amartyaveer, J. Bandekar, S. Murthy, S. Sharma, S. Badiger, S. Udupa, A. Nagireddi, S. Raghavan K M, R. Saxena, J. Nanavati, R. Nanavati, J. Sridharan, A. Mehta, A. Khuraishi K S, S. P. R. Mora, P. Venkataramakrishnan, G. Date, K. P, and P. K. Ghosh, "RESPIN-S1.0: A read speech corpus of 10000+ hours in dialects of nine Indian Languages", in *The Thirty-ninth Annual Conference on Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, 2025. [Online]. Available: https://openreview.net/forum?id=qL8M2dOY4L.

[2] **Saurabh Kumar**, Amartyaveer, and P. K. Ghosh, "Jointly Improving Dialect Identification and ASR in Indian Languages using Multimodal Feature Fusion", in *Interspeech 2025*, 2025, pp. 2770–2774. DOI: 10.21437/Interspeech.2025-421.

[3] Amartyaveer, **Saurabh Kumar**, S. Sharma, S. Udupa, S. Badiger, A. Singh, D. G, J. Bandekar, S. Murthy, and P. Kumar Ghosh, "Improving Dialect Identification in Indian Languages Using Multimodal Features from Dialect Informed ASR", in *Proc. ICASSP*, 2025, pp. 1–5. DOI: 10.1109/ICASSP49660.2025.10889099.

[4] S. Udupa, J. Bandekar, **Saurabh Kumar**, D. G, S. B, A. S, S. Murthy, P. Pai, S. Raghavan, R. Nanavati, and P. K. Ghosh, "Adapter pre-training for improved speech recognition in unseen domains using low resource adapter tuning of self-supervised models", in *Proc. Interspeech*, 2024, pp. 2529–2533. DOI: 10.21437/Interspeech.2024-1587.

[5] S. Udupa, J. Bandekar, G. Deekshitha, **Saurabh Kumar**, P. K. Ghosh, S. Badiger, A. Singh, S. Murthy, P. Pai, S. Raghavan, and R. Nanavati, "Gated Multi Encoders and Multitask Objectives for Dialectal Speech Recognition in Indian Languages", in *Proc. of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2023, pp. 1–8. DOI: 10.1109/ASRU57964.2023.10389624.

[6] S. Udupa, J. Bandekar, A. Singh, D. G, **Saurabh Kumar**, S. Badiger, A. Nagireddi, R. R, P. K. Ghosh, H. A. Murthy, P. Kumar, K. Tokuda, M. Hasegawa-Johnson, and P. Olbrich, "LIMMITS'24: Multi-Speaker, Multi-Lingual INDIC TTS With Voice Cloning", *IEEE OJSP*, vol. 6, pp. 293–302, 2025. DOI: 10.1109/OJSP.2025.3531782.

[7] S. Shahnawazuddin, A. Kumar, V. Kumar, **Saurabh Kumar**, and W. Ahmad, "Robust children's speech recognition in zero resource condition", *Applied Acoustics*, vol. 185, p. 108 382, 2022, ISSN: 0003-682X. DOI: https://doi.org/10.1016/j.apacoust.2021.108382.

[8] S. Shahnawazuddin, A. Kumar, **Saurabh Kumar**, and W. Ahmad, "Enhancing robustness of zero resource children's speech recognition system through bispectrum based front-end acoustic features", *Digital Signal Processing*, vol. 118, p. 103 226, 2021, ISSN: 1051-2004. DOI: https://doi.org/10.1016/j.dsp.2021.103226.

Languages and Skills

Languages

- **Programming:** Python, Bash, PHP
- **Natural:** Bhojpuri, English, Hindi, Maithili

Skills

- **ML/DL Toolkits:** PyTorch, Scikit-learn
- **ASR Frameworks:** ESPnet, Kaldi, Fairseq
- **Others:** MATLAB, Audacity, Praat, Git