Shashi Kumar

+41-76-565-1196 | shashi.kumar@epfl.ch | shashi66.github.io
shashikumar4 | shashi.kumar@idiap.ch | Google Scholar

Summary

Third year Ph.D. student at EPFL & Idiap Research Institute working on efficient adaptation of foundation models across speech and language, with publications at NeurIPS, EMNLP, ICASSP, and Interspeech. Before the Ph.D., I worked in industry training and deploying production ASR systems at scale (120K+ audio hours/month).

Research Interests

My research spans automatic speech recognition, SpeechLLMs, and efficient adaptation of foundation models, including multitask learning and cross-modal transfer. I am also interested in reasoning, compositionality, and generalization in large language models.

Education

Aug 2023–Present **Doctor of Philosophy (Ph.D.)**
École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland &
Idiap Research Institute, Martigny, Switzerland
Advisors: Prof. Andrea Cavallaro, Prof. Petr Motlíček
Expected graduation: Aug 2027.

2013–2017 **Bachelor of Technology (B. Tech.)**
Indian Institute of Technology (IIT) Guwahati, India
Subject: Electronics and Communication Engineering (ECE)

Selected Publications

For a full publication list, see my [Google Scholar] profile.

2025 **S Kumar**, Y Kaloga, J Mitros, P. Motlicek, and I Kodrasi. "Latent Space Factorization in LoRA.", Neural Information Processing Systems (NeurIPS), 2025. [Paper] [code]
Proposed FVAE-LoRA, a parameter-efficient fine-tuning method that separates task-relevant and residual features, improving robustness and performance across image, text, and audio tasks.

2025 Y Kaloga*, **S Kumar***, P Motlicek, and I Kodrasi. "A Differentiable Alignment Framework for Sequence-to-Sequence Modeling via Optimal Transport.", arXiv preprint arXiv:2502.01588. [Paper] [code]
Introduced a novel differentiable sequence-alignment framework based on 1D optimal transport, enabling the model to learn a single alignment and perform ASR in an E2E manner.

2025 **S Kumar**, et al. "Performance evaluation of slam-asr: The good, the bad, the ugly, and the way forward.", SALMA Workshop, ICASSP, 2025. [Paper] **[Best Paper Award]**
Evaluated SLAM-ASR across domain shifts and speech perturbations, identifying major robustness gaps and providing practical guidance for building more reliable LLM-based ASR systems.

2024 **S Kumar**, et al. "TokenVerse: Towards Unifying Speech and NLP Tasks via Transducer-based ASR.", EMNLP, 2024. [Paper]
Proposed a transducer-based framework that unifies transcription, speaker change detection, end-pointing, and named entity recognition in a single model, outperforming cascaded pipelines.

2026 S Burdisso, E Villatoro-Tello, **S Kumar**, et al. "Reducing Prompt Sensitivity in LLM-based Speech Recognition Through Learnable Projection.", ICASSP, 2026. [Paper] [code]

2025 **S Kumar**, et al. "XLSR-Transducer: Streaming ASR for Self-Supervised Pretrained Models.", ICASSP, 2025. [Paper]

Experience

May 2022–Apr 2023 **Senior ML Engineer** → **Lead ML Engineer**, *Level AI*, Bangalore, India

- Led development of low-resource conversational ASR, improving WER on internal benchmarks by integrating domain adaptation and entity-aware training into transducer-based models.
- Built production VAD and speaker diarization pipeline; reduced missed-speech rate and improved diarization error rate over the existing vendor system.
- Scaled ASR deployment to 120K+ audio hours/month across multiple customer tenants.

Jun 2017–Apr 2022 **Software Engineer** → **Lead Software Engineer**, *Samsung Research*, Bangalore, India

- Developed hybrid and end-to-end ASR systems for accented English and far-field settings; contributed to Bixby release-cycle improvements for ASR and contextual recognition.
- Proposed a joint VAE-based speaker normalization method for far-field and whisper speech (published at Interspeech 2021), deployed in the internal ASR pipeline.

May–July 2016 **Research Intern**, *FeltSo*, Hyderabad, India

- Developed aspect-level sentiment analysis models for e-commerce product reviews.

May–July 2015 **Research Intern**, *Chubu University*, Nagoya, Japan

- Developed a robust defect detection and classification system for PCBs, achieving state-of-the-art result among non-referential approaches.

Technical Skills

Core	Python, C++, PyTorch, Hugging Face (Transformers, PEFT, TRL, Datasets), fairseq, NeMo, SpeechBrain, Kaldi
Train at Scale	PyTorch DDP, FSDP, DeepSpeed, Slurm, multi-GPU workflows
Tools & Infra	Docker, Kubernetes, Git, Weights & Biases, Hydra, ONNX
Serving Systems	FastAPI, Flask, gRPC, Kafka, Triton Inference Server

Professional Activities

Challenges DiSPLACE-M Challenge 2026: Team ranked **1st in 3 of 4 tracks**, including diarization and Hindi Devanagari ASR for medical conversations. [Paper]

Interspeech DiCOVA Challenge 2021: Ranked 5th out of 29 teams in a COVID-19 detection task using cough sounds (AUC 83.93). [Paper]

Presentations Presented at NeurIPS 2025, Interspeech 2025, and EMNLP 2024.

Academic Reviewer for Interspeech 2025, ARR cycles.

Awards

Best Paper Award, SALMA Workshop at ICASSP 2025.

Idiap PhD Paper Award, for "Latent Space Factorization in LoRA" (NeurIPS 2025).