

Sonal Kumar

☎ +1 240 501 3161 ✉ sonalkum@umd.edu  [realsonalkumar](https://www.linkedin.com/in/realsonalkumar)  [sonalkum](https://github.com/sonalkum)  sonalkum.github.io

EDUCATION

University of Maryland, College Park 2023 – Present
PhD in Computer Science, Advised by Prof. Ramani Duraiswami & Prof. Dinesh Manocha - 3.86/4 Maryland, USA

CHRIST (Deemed to be University), Bangalore 2017 – 2021
Bachelor of Technology in Computer Science and Engineering - 9.7/10 Bangalore, India

Publications

[TAC: Timestamped Audio Captioning](#)

Sonal Kumar, Prem Seetharaman, Ke Chen, Oriol Nieto, Jiaqi Su, Zhepei Wang, Rithesh Kumar, Dinesh Manocha, Nicholas J. Bryan, Zeyu Jin, Justin Salamon

ICML 2026 (Under Review)

[MMAU-Pro: A Challenging and Comprehensive Benchmark for Holistic Evaluation of Audio General Intelligence](#)

Sonal Kumar, Šimon Sedláček, Vaibhavi Lokegaonkar, Fernando López, Wenyi Yu, Nishit Anand, and others.

AAAI 2026

[Audio Flamingo 3: Advancing Audio Intelligence with Fully Open Large Audio Language Models](#)

Arushi Goel, Sreyan Ghosh, Jaehyeon Kim, **Sonal Kumar**, Zhifeng Kong, Sang-gil Lee, Chao-Han Huck Yang, Ramani Duraiswami, Dinesh Manocha, Rafael Valle, Bryan Catanzaro

NeurIPS 2025 (Spotlight)

[MULTIVOX: Benchmarking Voice Assistants for Multimodal Interactions](#)

Ramaneswaran Selvakumar, Ashish Seth, Nishit Anand, Utkarsh Tyagi, **Sonal Kumar**, Sreyan Ghosh, Dinesh Manocha

EMNLP 2025

[SILA: Signal-to-Language Augmentation for Enhanced Control in Text-to-Audio Generation](#)

Sonal Kumar, Prem Seetharaman, Justin Salamon, Dinesh Manocha, Oriol Nieto

WASPAA 2025

[Synthio: Augmenting Small-Scale Audio Classification Datasets with Synthetic Data](#)

Sreyan Ghosh*, **Sonal Kumar***, Zhifeng Kong, Rafael Valle, Bryan Catanzaro, Dinesh Manocha

ICLR 2025

[MMAU: A Massive Multi-Task Audio Understanding and Reasoning Benchmark](#)

S Sakshi*, Utkarsh Tyagi*, **Sonal Kumar***, Ashish Seth*, Ramaneswaran Selvakumar, Oriol Nieto, Ramani Duraiswami, Sreyan Ghosh*, Dinesh Manocha

ICLR 2025 (Spotlight)

[Visual Description Grounding Reduces Hallucinations and Boosts Reasoning in LVLMS](#)

Sreyan Ghosh*, Chandra Kiran Reddy Evuru*, **Sonal Kumar***, Utkarsh Tyagi, Oriol Nieto, Zeyu Jin, Dinesh Manocha

ICLR 2025

[ProSE: Diffusion Priors for Speech Enhancement](#)

Sonal Kumar*, Sreyan Ghosh*, Chandra Kiran Reddy Evuru*, Utkarsh Tyagi, Oriol Nieto, Zeyu Jin, Dinesh Manocha

NAACL 2025 (Oral)

[GAMA: A Large Audio-Language Model with Advanced Audio Understanding and Complex Reasoning Abilities](#)

Sreyan Ghosh*, **Sonal Kumar***, Ashish Seth, Chandra Kiran Reddy Evuru, Utkarsh Tyagi, S Sakshi, Oriol Nieto, Ramani Duraiswami, Dinesh Manocha

EMNLP 2024 (Oral)

[EH-MAM: Easy-to-Hard Masked Acoustic Modeling for Self-Supervised Speech Representation Learning](#)

Ashish Seth, Ramaneswaran S, S Sakshi, **Sonal Kumar**, Sreyan Ghosh, Dinesh Manocha

EMNLP 2024 (Oral)

[A Closer Look at the Limitations of Instruction Tuning](#)

Sreyan Ghosh*, C. K. Evuru*, **Sonal Kumar***, Ramaneswaran S, D. Aneja, Z. Jin, R. Duraiswami, Dinesh Manocha

ICML 2024

[LipGER: Visually-Conditioned Generative Error Correction for Robust Automatic Speech Recognition](#)

Sreyan Ghosh, **Sonal Kumar**, Ashish Seth, Purva Chiniya, Utkarsh Tyagi, Ramani Duraiswami, Dinesh Manocha

Interspeech 2024 (Oral)

AV-RIR: Audio-Visual Room Impulse Response Estimation

Anton Ratnarajah, Sreyan Ghosh, **Sonal Kumar**, Purva Chiniya, Dinesh Manocha

CVPR 2024

CompA: Addressing the Gap in Compositional Reasoning in Audio-Language Models

Sreyan Ghosh*, Ashish Seth*, **Sonal Kumar***, Utkarsh Tyagi, C. K. Evuru, Oriol Nieto, Dinesh Manocha

ICLR 2024

RECAP: Retrieval-Augmented Audio Captioning

Sreyan Ghosh*, **Sonal Kumar***, Chandra Kiran Reddy Evuru, Ramani Duraiswami, Dinesh Manocha

ICASSP 2024

Span Classification with Structured Information for Disfluency Detection in Spoken Utterances

Sreyan Ghosh, **Sonal Kumar**, Yaman Kumar Singla, Rajiv Ratn Shah, S. Umesh

Interspeech 2022

Do Vision-Language Models Understand Compound Nouns?

Sonal Kumar*, Sreyan Ghosh*, S Sakshi, Utkarsh Tyagi, Dinesh Manocha

NAACL 2024

ASPIRE: Language-Guided Augmentation for Robust Image Classification

Sreyan Ghosh*, C. K. Evuru*, **Sonal Kumar**, S. Sakshi, Utkarsh Tyagi, Dinesh Manocha

ACL 2024

ABEX: Data Augmentation for Low-Resource NLU via Expanding Abstract Descriptions

Sreyan Ghosh, Utkarsh Tyagi, **Sonal Kumar**, Chandra Kiran Reddy Evuru, Ramaneswaran S, S Sakshi, Dinesh Manocha

ACL 2024

CoDa: Constrained Generation based Data Augmentation for Low-Resource NLP

Chandra Kiran Reddy Evuru*, Sreyan Ghosh*, **Sonal Kumar**, Ramaneswaran S, Utkarsh Tyagi, Dinesh Manocha

NAACL 2024

DALE: Generative Data Augmentation for Legal NLP

Sreyan Ghosh¹, C. K. Evuru*, **Sonal Kumar**, S. Sakshi, Utkarsh Tyagi, Dinesh Manocha

EMNLP 2023

CoSyn: Detecting Implicit Hate Speech in Online Conversations Using a Context Synergized Hyperbolic Network

Sreyan Ghosh, Manan Suri, Purva Chiniya, Utkarsh Tyagi, **Sonal Kumar**, Dinesh Manocha

EMNLP 2023

ACLM: A Selective-Denoising based Generative Data Augmentation Approach for Low-Resource Complex NER

Sreyan Ghosh*, Utkarsh Tyagi*, Manan Suri, **Sonal Kumar**, S Ramaneswaran, Dinesh Manocha

ACL 2023

BioAug: Conditional Generation based Data Augmentation for Low-Resource Biomedical NER

Sreyan Ghosh*, Utkarsh Tyagi*, **Sonal Kumar***, Dinesh Manocha

SIGIR 2023

RESEARCH EXPERIENCE

GAMMA Lab, University of Maryland

College Park, Maryland

Research Assistant

October 2022 – Present

- As a part of GAMMA Lab at UMD, worked on Data Augmentations for low resource NLP problems such as Complex NER, Bio-Medical NER and Legal NER. Also working on Audio and Speech Processing along with Audio-Language research. Published at ICML, ICLR, EMNLP, ICASSP, ACL, NAACL, SIGIR, Interspeech.
- Advised by [Prof. Dinesh Manocha](#)

PIRL Lab, University of Maryland

College Park, Maryland

Research Assistant

August 2023 – Present

- As a part of PIRL Lab at UMD, I am working on decoding imagined speech information from human brain's MEG signals and Audio and Speech Processing.
- Advised by [Prof. Ramani Duraiswami](#)

¹*Denotes co-first authors

INDUSTRY EXPERIENCE

Google

Student Researcher

Seattle, USA

September 2025 – Present

- Working as a Student Researcher with the Google XR team.

Adobe Inc.

Research Scientist Intern

San Francisco, USA

May 2025 – August 2025

- Worked as a research scientist intern at Adobe in the audio research team on Audio Generative Models.

Adobe Inc.

Research Scientist Intern

San Francisco, USA

May 2024 – November 2024

- Worked as a research scientist intern at Adobe in the audio research team on Audio Generative Models.
- Improved fine-grained acoustic controls in audio generation. The research work got integrated in **Adobe Firefly Text2Sfx** as feature.

Cisco Systems

Software Engineer 2

Bangalore, India

August 2021 – August 2023

- Responsible for automating workflow to perform bulk OS upgraded for in-production network devices with minimal impact. Led to saving of 3-5 man hours per upgrade.
- Monitored and led successful upgrade of 1000+ in production network devices.
- Led a team of 5 to build chat bot which leveraged internal data sources to solve queries related to Customer Delivery using Machine Learning techniques.
- Built Network Ready For Use (NRFU) tool to automate testing of newly deployed network saving 5-6 man hours per testing.

Software Engineer Intern

Jan 2021 – July 2021

- Worked with the Data Center team to develop automated workflows to test in production network devices.
- Received appreciation for the excellent performance in Intern Spotlight 2021

TECHNICAL SKILLS

Skills: Deep Learning, Speech Processing, NLP, Computer Vision, Multimodal Learning, Generative AI, Research and Development (R&D)

Languages: (Highly Proficient) Python, JavaScript, C++, Java, SQL

Frameworks: Pytorch, FastAI, Tensorflow, Scikit-Learn, NLTK, Flask

WORKSHOPS & CHALLENGES

JSALT Challenge and Summer School

- The 2025 Jelinek Workshop on Speech and Language Technologies (JSALT) is an eight-week residential summer research workshop. It brings together international teams to work intensively on challenging problems in speech and language engineering, machine learning, and artificial intelligence. The workshop fosters collaboration and has a lasting influence through publications, software, and data produced. We propose [Advancing Expert-Level Reasoning and Understanding in Large Audio Language Models](#) which is a workshop focused on advancing expert-level understanding and complex reasoning in audio-language models. The team, drawn from several universities and industry in the US, Europe and Asia, and with students and senior professionals from various disciplines, will allow us to achieve these goals.

DCASE 2025 Challenge - Task 5: Audio Question Answering

- Proposed [The Audio Question Answering \(AQA\)](#) task which focuses on advancing question-answering capabilities in the realm of “interactive audio understanding,” covering both general acoustic events and knowledge-heavy sound information within a single track. The AQA task encourages participants to develop systems that can accurately interpret and respond to complex audio-based questions (i.e., (A), (B), or (C) option), requiring models to process and reason across diverse audio types. Such systems could be useful in many applications, including audio-text and multimodal model evaluation, as well as building interactive audio-agent models for the audio research communities and beyond. Reproducible baselines will include one resource-efficient computing setting (i.e., single 8GB RAM); direct using enterprise API for challenge entry is prohibited.

ACHIEVEMENTS & EXTRA CURRICULAR

- **University of Maryland, College Park, January 2025** - *Recipient of NeuroPAC Fellowship*
- **Cisco Systems, March 2023** - *Recognised for outstanding performance and contribution to AI research*
- **Mentor2Go, Jan 2022 - Aug 2022** - *Mentor*
- **P&G Global Innovation Challenge, May 2020** - *Best Idea Prize*
- **CHRIST (Deemed to be University), Feb 2020** - *CHRISTITE Spirit Award* for bringing recognition to the university.
- **Neuron - AI Club, CHRIST (Deemed to be University), June 2019 - Aug 2019** - *President*
- **Hindustan Unilever BFS Technology Hackathon, Sep 2019** - *Winner*
- **Kaggle Quora Insincere Questions Classification, Feb 2019** - *Bronze Medal*
- **MLDS-2019 Republic Day Hackathon,TEG Analytics, Jan 2019** - *Winner*