SAURABH KATARIA

Ph.D. candidate, Center for Language and Speech Processing, Johns Hopkins University

saurabhkataria10@gmail.com (preferred) ◇ skatari1@jh.edu

https://saurabh-kataria.github.io (Homepage) ◇ Google Scholar ◇ Research Gate ◇ LinkedIn ◇ Github

EDUCATION

Doctor of Philosophy (Ph.D.), Electrical Engineering, *Johns Hopkins University*, USA   Aug'17 - '23 (expected)
Dept. of Electrical and Computer Engineering
<u>Advisors</u>: Prof. Najim Dehak, Prof. Jesus Villalba
<u>Thesis title</u>: "Robust Speaker Recognition using Perceptual and Adversarial Speech Enhancement"
Center for Speech and Language Processing (CLSP), *Human Language Technology Center of Excellence* (HLTCOE)

Master of Science in Engineering (M.S.E.), *Johns Hopkins University*, USA   Aug'17 - May'20
Dept. of Electrical and Computer Engineering (obtained while pursuing Ph.D.)   GPA: 3.8/4

B.Tech. - M.Tech. dual degree program, *Indian Institute of Technology (IIT) Kanpur*, India   Jul'12 - Jun'17
• **Master of Technology (M.Tech.)**, Electrical Engineering   GPA: 9.0/10
<u>Thesis title</u>: "Unsupervised and Supervised Saliency Estimation Methods for Affective Speech"
<u>Advisors</u>: Prof. Tanaya Guha, Prof. Rajesh Hegde
• **Bachelor of Technology (B.Tech.)**, Electrical Engineering   GPA: 8.4/10
• **Minor in Artificial Intelligence**, Computer Science and Engineering

EXPERIENCES

Research Intern, *Tencent America*   May'20 - Aug'20
<u>Mentor</u>: Dr. Shi-Xiong (Austin) Zhang, <u>Leader</u>: Dr. Dong Yu   Seattle, USA (remote)

- Proposed multi-modal multi-channel speaker verification system which utilizes target speaker direction
- Conducted large-margin fine-tuning studies to show improvement on real and simulated datasets

Student Researcher, *Frederick Jelinek Memorial Summer Workshop (JSALT)* [link]   May'19 - Aug'19
<u>Venue</u>: École de Technologie Supérieure (ÉTS), <u>Organizer</u>: CLSP, Johns Hopkins University   Montreal, Canada
<u>Mentors</u>: Prof. Jesus Villalba, Prof. Najim Dehak, <u>Leader</u>: Dr. Paola Garcia

- Developed filter-bank feature domain speech enhancement solutions for speaker recognition and diarization

Research Intern, *INRIA*   May'16 - Aug'16
<u>Mentor</u>: Dr. Antoine Deleforge, <u>Leader</u>: Dr. Rémi Gribonval (Team PANAMA)   Rennes, France

- Generated large scale room impulse responses using simulation of rooms with varied absorption coefficients
- Predicted source direction-of-arrival angles and wall coefficients from binaural signals using generative models

Research Intern, *New York University*   May'15 - Aug'15
<u>Mentor</u>: Prof. Siddharth Garg   New York City, USA

- Studied the de-anonymization problem in social network graphs by matching two Erdos-Renyi graphs

TECHNICAL SKILLS

Programming languages: Python, MATLAB, C++
<u>Libraries</u>: scikit-learn, PyTorch, fairseq
<u>Softwares, IDE</u>: MATLAB, Visual Studio Code, Pycharm, Spyder
<u>Misc</u>: Git, Linux, bash, Latex, Jupyter Notebook, wandb

PRE-PRINTS/IN SUBMISSION MANUSCRIPTS

1. **Kataria, S.**, Villalba, J., Moro-Velázquez, L., Żelasko, P. and Dehak, N. (2022), *Time-domain speech super-resolution with GAN based modeling for telephony speaker verification.* (*In submission* to IEEE/ACM Transactions of Speech, Audio, and Language journal.) [pdf]

FIRST-AUTHOR PEER-REVIEWED PUBLICATIONS (LEAD WORK)

1. **Kataria, S.**, Villalba, J., Moro-Velázquez, L., Thebaud, T. and Dehak, N., 2023. *Self-FiLM: Conditioning GANs with self-supervised representations for bandwidth extension based speaker recognition* INTERSPEECH 2023 [pdf]

2. **Kataria, S.**, Villalba, J., Moro-Velázquez, L. and Dehak, N., *Joint domain adaptation and speech bandwidth extension using time-domain GANs for speaker verification.* INTERSPEECH 2022 [pdf]

3. Joshi, S.*, **Kataria, S.***, Villalba, J. and Dehak, N., *AdvEst: Adversarial Perturbation Estimation to Classify and Detect Adversarial Attacks against Speaker Identification.* INTERSPEECH 2022 [pdf] (* = joint first authors)

4. **Kataria, S.**, Villalba, J. and Dehak, N., *Perceptual loss based speech denoising with an ensemble of audio pattern recognition and self-supervised models.* ICASSP 2021 [pdf]

5. **Kataria, S.**, Villalba, J., Żelasko, P., Moro-Velázquez, L. and Dehak, N., *Deep feature cyclegans: Speaker identity preserving non-parallel microphone-telephone domain adaptation for speaker verification.* INTERSPEECH 2021 [pdf]

6. **Kataria, S.***, Zhang, S.X.* and Yu, D., *Multi-Channel Speaker Verification for Single and Multi-Talker Speech.* INTERSPEECH 2021 [pdf]

7. **Kataria, S.**, Nidadavolu, P.S., Villalba, J. and Dehak, N., *Analysis of deep feature loss based enhancement for speaker verification.* Odyssey 2020 [pdf]

8. **Kataria, S.**, Nidadavolu, P.S., Villalba, J., Chen, N., Garcia-Perera, P. and Dehak, N., *Feature enhancement with deep feature losses for speaker verification.* ICASSP 2020 [pdf]

9. **Kataria, S.**, Gaultier, C. and Deleforge, A, *Hearing in a shoe-box: binaural source position and wall absorption estimation using virtually supervised learning.* ICASSP 2017 [pdf]

OTHER PEER-REVIEWED PUBLICATIONS (COLLABORATION WORK)

1. Advances in Language Recognition in Low Resource African Languages: The JHU-MIT Submission for NIST LRE22 (INTERSPEECH 2023) [pdf available soon]

2. Shao, Y., Villalba, J., Joshi, **S., Kataria**, S., Dehak, N. and Khudanpur, S., *Chunking Defense for Adversarial Attacks on ASR.* INTERSPEECH 2022 [pdf]

3. Joshi, S., **Kataria, S.**, Shao, Y., Zelasko, P., Villalba, J., Khudanpur, S. and Dehak, N., *Defense against Adversarial Attacks on Hybrid Speech Recognition using Joint Adversarial Fine-tuning with Denoiser.* INTERSPEECH 2022 [pdf]

4. Villalba, J., Borgstrom, B.J., **Kataria, S.**, Rybicka, M., Castillo, C.D., Cho, J., Garcıa-Perera, L.P., Torres-Carrasquillo, P.A. and Dehak, N., *Advances in Cross-Lingual and Cross-Source Audio-Visual Speaker Recognition: The JHU-MIT System for NIST SRE21.* Odyssey 2022. [pdf]

5. Villalba, J., Borgstrom, B.J., **Kataria, S.**, Cho, J., Torres-Carrasquillo, P.A. and Dehak, N., *Advances in Speaker Recognition for Multilingual Conversational Telephone Speech: The JHU-MIT System for NIST SRE20 CTS Challenge.* Odyssey workshop 2022 [pdf]

6. Nidadavolu, P.S., **Kataria, S.**, Villalba, J., Garcia-Perera, P. and Dehak, N., *Unsupervised feature enhancement for speaker verification.* ICASSP 2020 [pdf]

7. Villalba, J., Garcia-Romero, D., Chen, N., Sell, G., Borgstrom, J., McCree, A., Garcia-Perera, L.P., **Kataria, S.**, Nidadavolu, P.S., Torres-Carrasquillo, P.A. and Dehak, N., *Advances in speaker recognition for telephone and audio-visual data: the JHU-MIT submission for NIST SRE19.* Odyssey 2020 [pdf]

8. Nidadavolu, P.S., **Kataria, S.**, Villalba, J. and Dehak, N., *Low-resource domain adaptation for speaker recognition using cycle-gans.* ASRU 2019 [pdf]

9. García, P., Villalba, J., Bredin, H., Du, J., Castan, D., Cristia, A., Bullock, L., Guo, L., Okabe, K., Nidadavolu, P.S. and **Kataria, S.**, *Speaker detection in the wild: Lessons learned from JSALT 2019.* Odyssey 2020 [pdf]

10. Gaultier, C., **Kataria, S.** and Deleforge, A., *VAST: The virtual acoustic space traveler dataset.* In International Conference on Latent Variable Analysis and Signal Separation (ICA-LVA 2015) [pdf]

11. Parthasarathy, A., **Kataria, S.**, Kumar, L. and Hegde, R.M., *Representation and modeling of spherical harmonics manifold for source localization.* ICASSP 2015 [pdf]

NOTABLE RESEARCH PROBLEMS WORKED ON

Improving speech enhancement using self-supervised learning *(ongoing)*

- Developing completely unsupervised enhancement solutions while leveraging pre-trained self-supervised audio models

Unsupervised domain adaptation for microphone and telephone speech

- Tackled the task of adapting telephone to/from microphone speech for Telephony speaker recognition

Speech enhancement using deep feature losses for speaker recognition

- Proposed Perceptual Ensemble Regularization Loss (PERL) which leverages several pre-trained audio models

Detecting unknown adversarial attacks on speaker recognition
(collaboration under the project: *DARPA RED (Reverse Engineering of Deceptions)* [link])

- Developed a contrastive loss based GAN denoiser which improves signatures of unknown adversarial attacks

Defending against adversarial attacks on Automatic Speech Recognition (ASR)
(collaboration under the project: *DARPA GARD (Guaranteeing AI Robustness Against Deception)* [link])

- Developed pre-trained as well as adversarially fine-tuned denoisers for removing adversarial perturbations
- Improved upon the state-of-the-art method (Adversarial Training) on strong attacks

ACHIEVEMENTS

Johns Hopkins University ECE department graduate fellowship (2017-2018)

Best project in the course EE698M (Topics in Image and Signal Processing), 2016

Awarded 2 travel grants by Dept. of Science and Technology (DST), Govt. of India (2015) and IIT Kanpur.

PARTICIPATION IN CHALLENGES

(includes collaborations with MIT Lincoln Laboratory, Panasonic, etc.)

NIST Language Recognition Evaluation (LRE): '22, '17

NIST Speaker Recognition Evaluation (SRE): '21, '20, '19, '18

PUBLIC TALKS

"Robust Speaker Verification using perceptual and adversarial speech enhancement", Brookhaven National Lab, NY, May'23

"Conditioning bandwidth extension GAN models with self-supervision", CLSP seminar, JHU, Feb'23

"CycleGAN variants for Speech Domain Adaptation and Bandwidth Expansion", CLSP seminar, JHU, Nov'21

"Single-Channel Speech Enhancement with Deep Feature Losses for Speaker Verification", CLSP seminar, JHU, Feb'20

TEACHING/COURSE ASSISTANCE SERVICE

Machine Learning for Medical Applications (2021, EN.520.639, JHU)
Computational Modeling for Electrical and Computer Engineering (2020, EN.520.123, JHU)
Machine Learning for Signal Processing (2020, EN.520.612, JHU)
Modeling and Representation Techniques for Images (2017, EE698K, IIT Kanpur)
Image processing (2016, EE604A, IIT Kanpur)

CONFERENCE/WORKSHOP ORGANIZING

8th Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL 2020)
Program Organizing Committee, University of Maryland, College Park

REVIEWER

ICLR ('23, '21), Interspeech'23, ICASSP'23, CISS ('23), ACML ('22), Speech Communication (multiple years), NeurIPS'20