Mahathir Monjur

Multimodal (Audio, mmWave, WiFi CSI, Vision) ML | Generative Models (Diffusion/LLMs)
Email: mahathir@cs.unc.edu | Phone: +1 (919) 904-5633
[Google Scholar](#) | [LinkedIn](#) | [GitHub](#)

Education

- **University of North Carolina at Chapel Hill**
  Ph.D. in Computer Science | 2021 - Present (Expected Graduation: Apr 2026)
  - **Research Interests**: Multimodal (Audio, RF, vision) Sensing, Machine Learning & Generative AI
  M.S. in Computer Science | Dec 2023

- **Bangladesh University of Engineering and Technology**
  B.S. in Computer Science | Oct 2018

Work Experience

- **Meta Reality Labs**
  Part-time Student Researcher | Wearable Audio Technologies Team (Aug 2025 - Dec 2025 | Remote)
  Full-time Intern | Wearable Audio Technologies Team (May 2025 - Aug 2025 | Redmond, WA)
  - Designed a multi-dimensional perceptual speech-quality estimator for both binaural and mono aural audio by utilizing large pretrained encoders (AST, EnCodec, and wave2vec), achieving <0.4 MAE on a 7-point MOS scale under wind, babble, and low-SNR conditions. The model was adopted as an internal perceptual metric, significantly reducing reliance on frequent, costly expert listening tests.
  - Developed a multimodal LLM that aligns spectral audio embeddings with Llama-3.1-8B, enabling natural-language quality scoring and interpretable rationale generation.
  Manager: Yuqi Deng | Team Lead: Nick Zacharov

- **Amazon**
  Applied Scientist Intern | Audio Research & Algorithm (ARA) Team
  May 2023 - August 2023 | Cambridge, MA
  May 2022 - August 2022 | Cambridge, MA
  - Developed the first multi-device audio sensing system that estimates a user's facing direction by leveraging the frequency-dependent directivity of human speech: each device extracts local audio & spatial features, exchanges embeddings with peer devices, and runs a DNN to produce per-device orientation estimates (with sound-source localization and confidence), which are then fused via a post-processing stage into a final robust estimate. Used in prototype devices, awarded a U.S. Patent ([Patent US12200449B1](#)).
  - Developed an attention-based network to generalize the estimation of sound source localization (SSL) and speaking direction across diverse microphone arrays, effectively addressing microphone heterogeneity.
  Manager: Mrudula Athi | Mentor: Md Tamzeed Islam | Team Lead: Wontak Kim

- **University of North Carolina at Chapel Hill**
  Research Assistant
  2021 - Present | Department of Computer Science
  - Leading research on multimodal sensing and generative modeling, combining audio, mmWave radar, WiFi CSI, and RGB-D for indoor activity understanding, 3d pose estimation and security.
  - Designed and implemented systems such as mmWEAVER, Wave2Vid, VoiceDirect, SoundSieve, AURA, CarFi, mmDefender etc. improving human activity recognition, localization, and pose estimation accuracy and enabling new capabilities on embedded and intermittent devices.
  - Published research papers in top-tier venues ([MobiSys], [WACV], [MASS], [EWSN], [ICHI])
  Advisor: Dr. Shahriar Nirjon

Technical Skills

- **ML/AI:** Deep learning, Audio ML, Diffusion Models, LLMs, Implicit Neural Networks
- **Audio:** Binaural audio, Spatial Audio, Localization, ASR/Compression
- **Frameworks:** PyTorch, TensorFlow, Triton, xFormers, HuggingFace
- **Languages:** Python, C++, Matlab, Node.js

Selected Publications

- **Mahathir Monjur**, Shahriar Nirjon, "*mmWEAVER: mmWave Signal Synthesis from a Photo and Activity Description*", In the IEEE/CVF Winter Conference on Applications of Computer Vision 2026 [**WACV'26**].
- **Mahathir Monjur**, et al. "*User orientation estimation.*" **U.S. Patent 12,200,449**, issued January 14, 2025.
- Shao-Jung (Louie) Lu, **Mahathir Monjur** et al., "*mmDefender: A mmWave System for On-Body Localization of Concealed Threats in Moving Persons*", In Proceedings of the 2025 International Conference on Embedded Wireless Systems and Networks [**EWSN '25**]. (**Best Paper Runner-Up**)
- **Mahathir Monjur**, Shahriar Nirjon, "*VoiceDirect: mmWave and Audio Signal Fusion for User Localization and Speaking Direction Estimation*", In Proceedings of the 2024 International Conference on Embedded Wireless Systems and Networks [**EWSN '24**].
- **Mahathir Monjur**, et al., "*Data Distribution Dynamics in Real-World WiFi-Based Patient Activity Monitoring for Home Healthcare*", in 2024 IEEE 12th International Conference on Healthcare Informatics [**IEEE ICHI'24**], Orlando, FL, USA, 2024 pp. 228-233. (2024)
- **Mahathir Monjur**, Yubo Luo, Zhenyu Wang, and Shahriar Nirjon. 2023. *SoundSieve: Seconds-Long Audio Event Recognition on Intermittently-Powered Systems*. In Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services [**MobiSys '23**].

Selected Research Projects

2025 Perceptual Evaluation of Speech Quality using Deep Learning
At Meta Reality Labs, developed a deep learning framework for multidimensional binaural speech quality estimation under realistic acoustic conditions (wind, babble, low SNR, and artifacts from modern enhancement pipelines). Leveraging large pretrained encoders (AST, wav2vec, EnCodec), the system achieves <0.4 MAE on a 7-point scale, making it a practical automated proxy for expensive human listening tests. Extending the idea, designed a multimodal LLM-based "expert listener" using an audio encoder (AST or Whisper) aligned with an LLM. Trained using structured QA pairs, the model achieves a MOS MAE of ~0.41 with strong correlation (0.86) while supporting interpretable assessments not possible with traditional regressors (SpeechQualityLLM).

2025 Wave2Vid: Video Generation from mmWave Radar and Visual Context
Proposed the first system to generate realistic human-activity videos from mmWave radar using a body-pose estimator and a HyperMotion-based diffusion video model. Achieved large gains on baseline HuPR—reducing FVD from 232.4→150.2, halving AKD (0.091→0.035), and boosting PCK@0.2 from 0.519→0.741—while preserving identity and scene via reference-image conditioning. Currently under submission at **CVPR 2026**.

2024 mmWEAVER: mmWave Signal Synthesis from a Photo and Activity Description
Developed the first framework that synthesizes realistic, environment-specific mmWave radar signals directly from an RGB-D image and an activity description using *hypernetworks* and *implicit neural representations*. The system generates high-fidelity I/Q signals 6–35× faster than simulation while improving activity recognition and pose estimation accuracy by up to 7% and 15%, respectively. Accepted at **WACV 2026** (mmWeaver).

2023 VoiceDirect: Acoustic and mmWave Signal Fusion for Enhanced Speaking Direction Estimation
Created the first standalone smart-hub system that fuses mmWave radar with microphone-array audio to accurately estimate a user's speaking direction anywhere in a room. The system achieves a median error of just 19°, outperforming single- and multi-device audio-only baselines by 16–43°, using cross-modal attention, radar-assisted localization, and pose-aware acoustic normalization. Published at **EWSN 2024** (VoiceDirect).

2023 SoundSieve: Seconds-Long Audio Event Recognition on Intermittently-Powered Systems
Designed the first system enabling seconds-long audio event recognition on intermittently powered devices by learning to identify and selectively sample only the most informative audio segments. Developed a combined offline–online pipeline with content-aware sampling and intermittence-aware classification, yielding 5%–30% higher accuracy than state-of-the-art intermittent systems. Published at **MobiSys 2023** (SoundSieve).

2022 AURA- Connecting Audio and Radio Sensing Systems to Improve Care at Home
Developed a multimodal connected-health sensing dataset and pipelines integrating WiFi CSI, wearable devices, and smart speakers to support activity tracking, passive monitoring, and daily health queries. Built end-to-end data collection, synchronization, and baseline analytics for real-world home-care scenarios. Published at different health conferences and journals (IEEE-ICHI'24, JMIR).