UCLA Samueli
Electrical & Computer Engineering

ECE M202A (12/09/2019) Final Presentation

MREarable: Integrating Spatial Audio in a Mixed Reality Environment through Earable Sensor Modalities

Team Members: **Swapnil Sayan Saha, Vivek Jain and Siyou Pei**

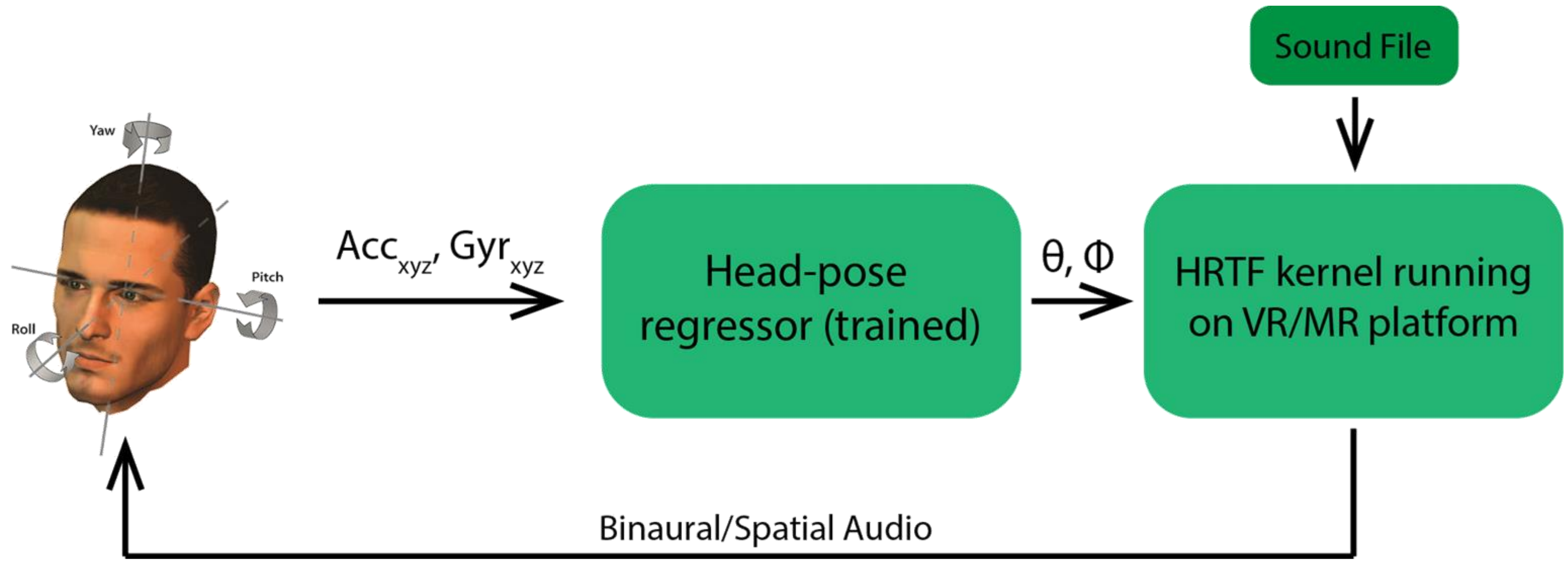
Networked & Embedded Systems Laboratory

Problem Statement and Importance

- User space restrained to VR coverage; requires bulky/expensive VR headsets and computationally intensive tracking algorithms.
- Not suitable for MR context.
- Head trackers require redundant sensors (e.g. cameras) for accurate tracking.
- Latency issues.

- **Use single inertial sensor for input and feedback.**
- **Head tracking and binaural audio generation in MR context.**
- **Improve inertial head-tracking performance using analytical and ML approach.**
- **Binaural rendering within human localization range and negligible latency.**
- **Benchmark earable sensor performance.**

Overall Project Goals and Specific Aims



Generation of a head-pose / HAR dataset using earable IMU

Headpose estimation using earable 6DOF IMU (KF + ML)

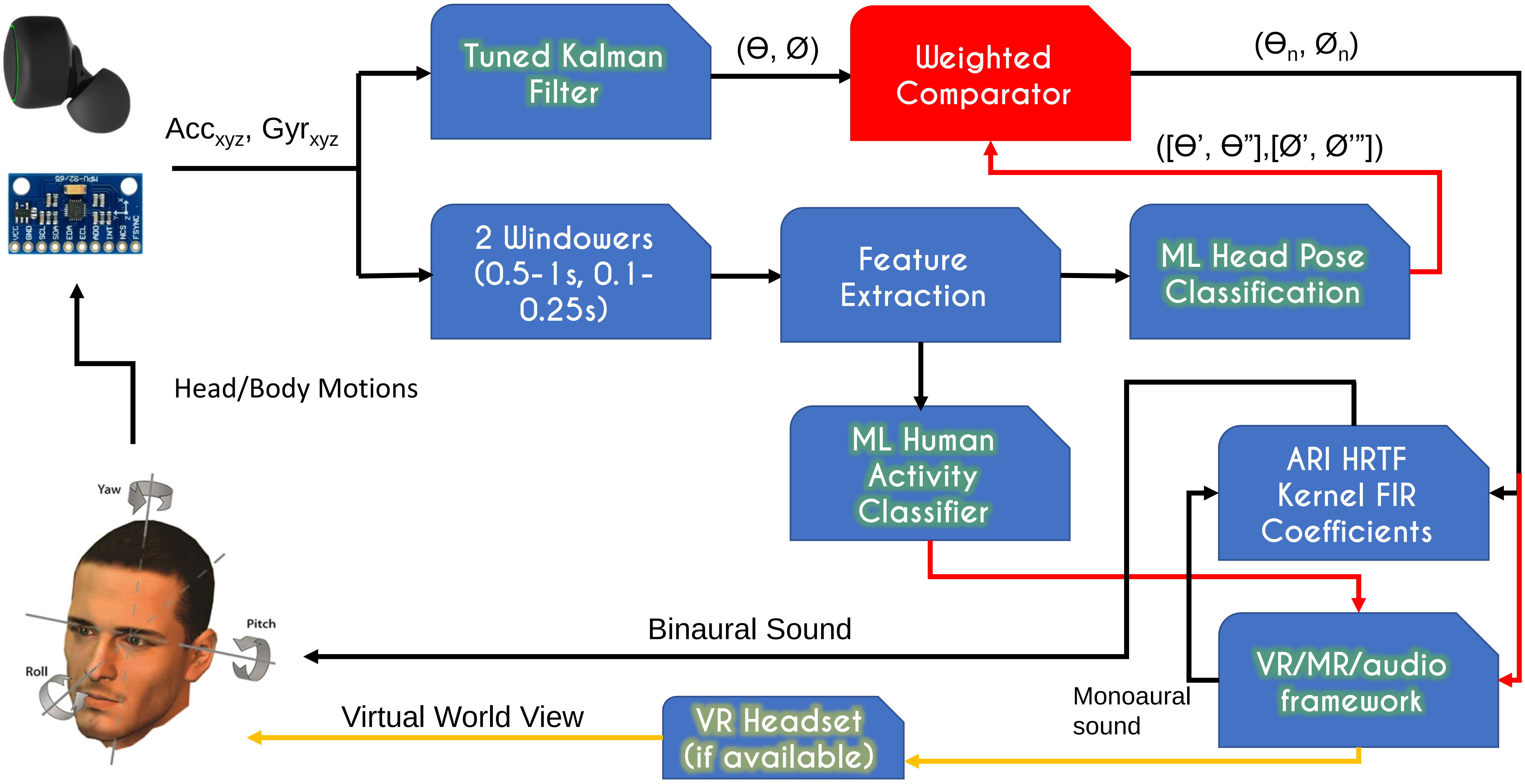
Generation of spatial audio based on inertial data

Integration of COTS VR platform with earables

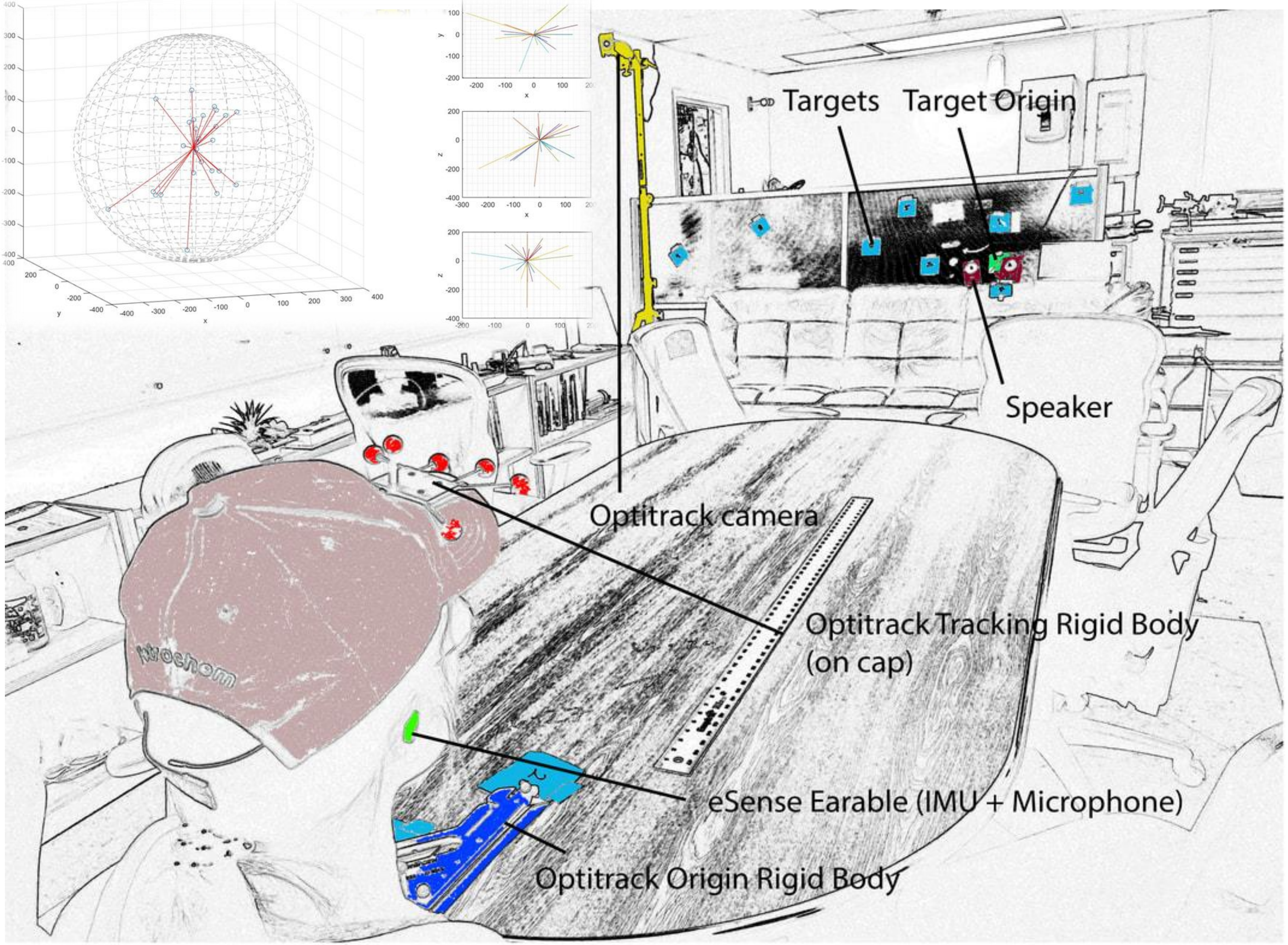
Deliverables:

- Use head-pose estimator to generate spatial audio from a stationary virtual sound source in real time (mid term)
- Integration of VR platform with eSense earables for perception-processing-feedback loop in real-time. (DONE WITH SAME SENSOR OUTSIDE EARABLES AND USING VR HEADSET)

Technical Approach



Technical Approach (contd.)



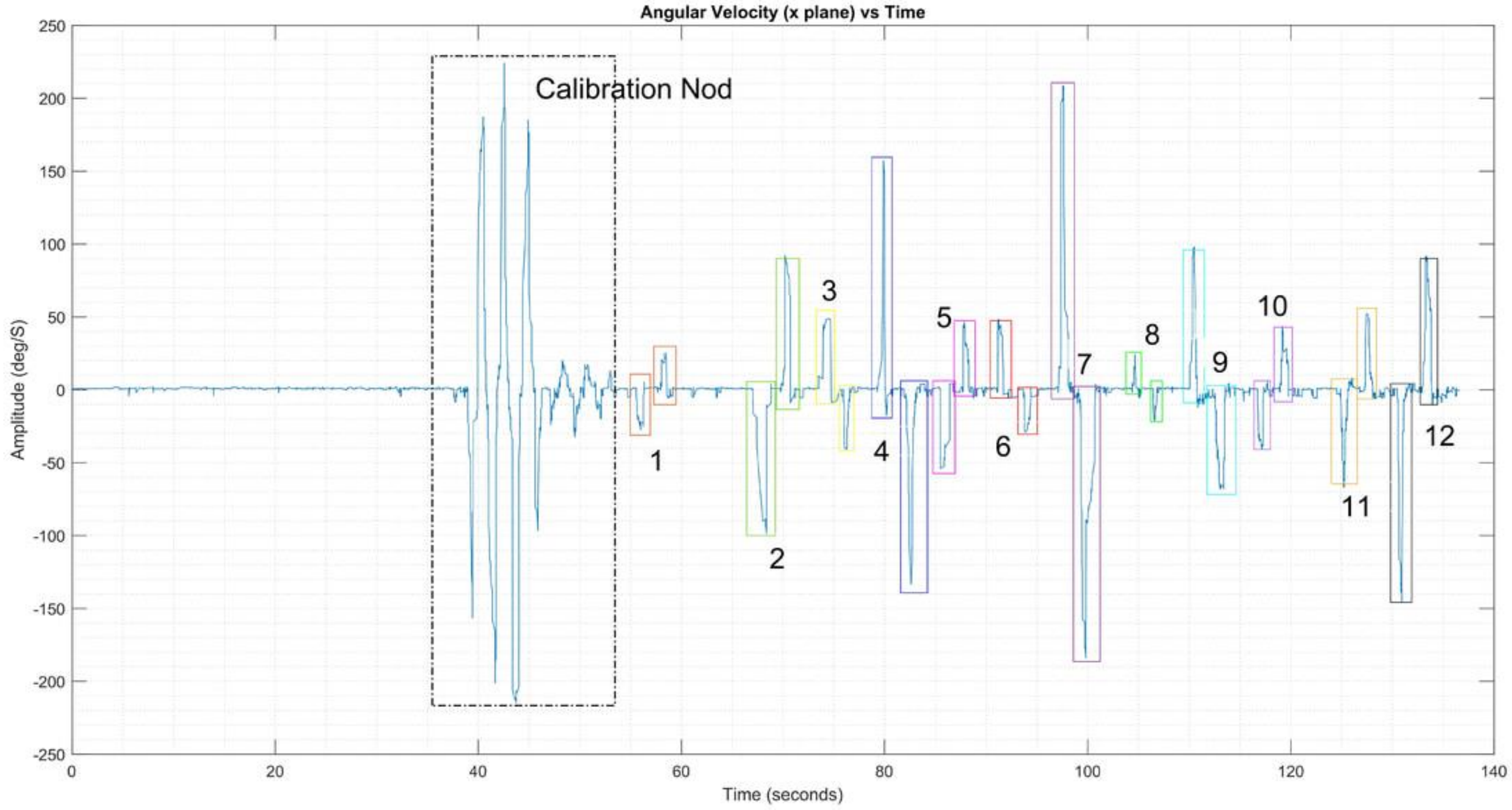
Two types of head-motion: 1. O-T-O 2. O-T1-T2-O

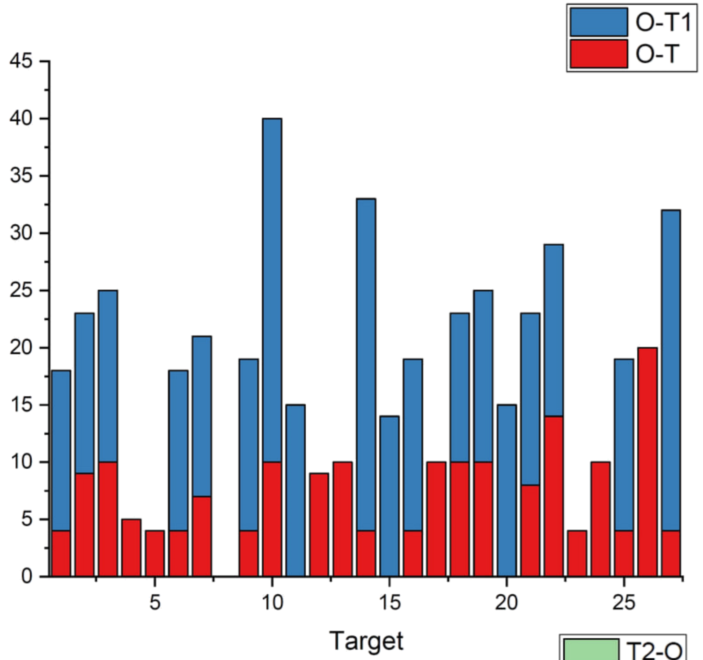
27 targets; 34 (13 + 21) distinct head-poses; 15 subjects; 9 activities

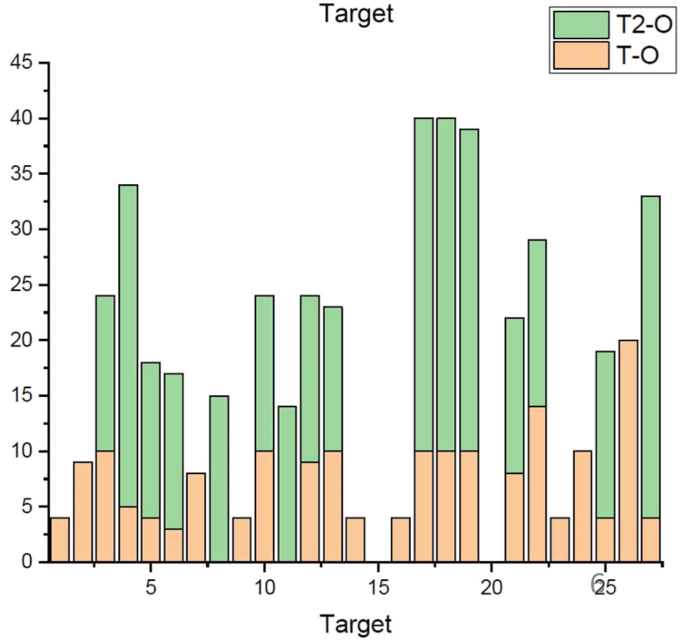
Range: +- 2g and +- 500 deg/S; Adv. & Conn. Intv: 45-55 & 20-30 mS; F_s = 100 Hz, LPF: 5 Hz.

Ground truth / Measurements: Optitrack, earable audio, Leica Disto X3 & video camera

Technical Approach (cont.)







1266 inertial frames: 356 O-T-O, 607 O-T1 and T2-O, 303 T1-T2

Two types of head-motion:
1. O-T-O 2. O-T1-T2-O

Technical Approach (cont.)

Make sliced data subject agnostic and bin into ±20 degrees.

Trial and error classification performance on 797 elevation frames

(3-fold CV; 90:10 holdout)

Performance saturation at ~42.3% using XGBoost

Use grid-search hyperparameter optimizer; performance saturation at ~75%

Data imbalance; try different SMOTEs

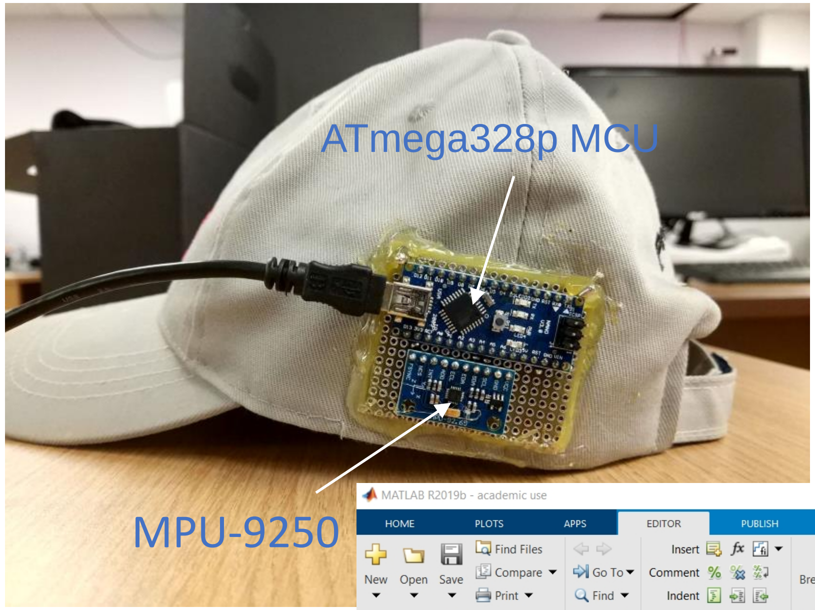
Significant performance improvement (>90%)

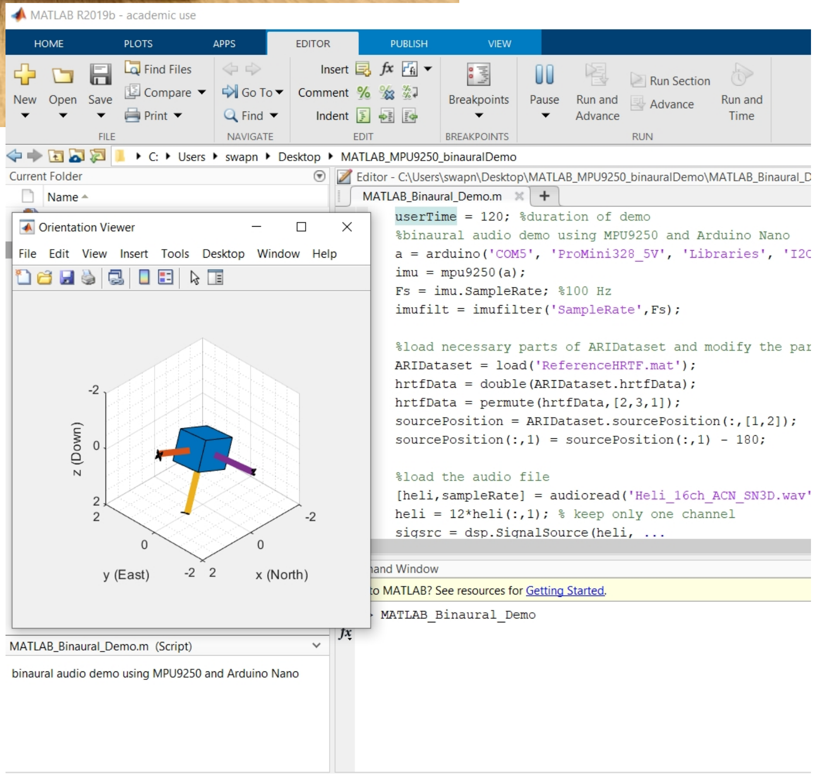
Map class to angle; calculate errors

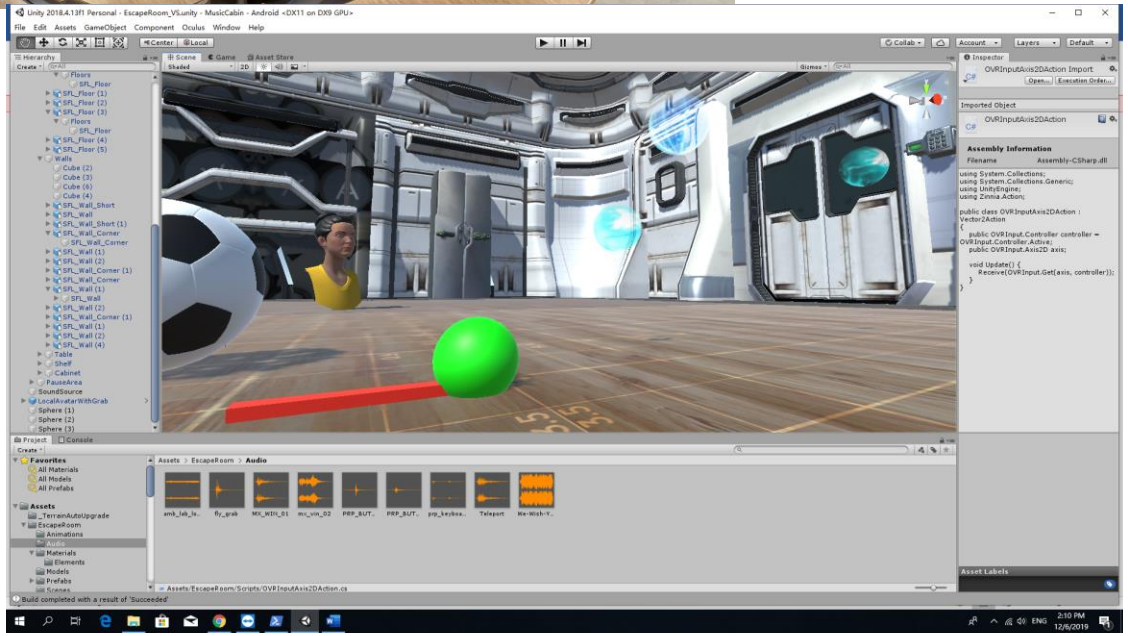
Repeat on azimuth classifier

Error = MAE and RMSE

Technical Approach (cont.)



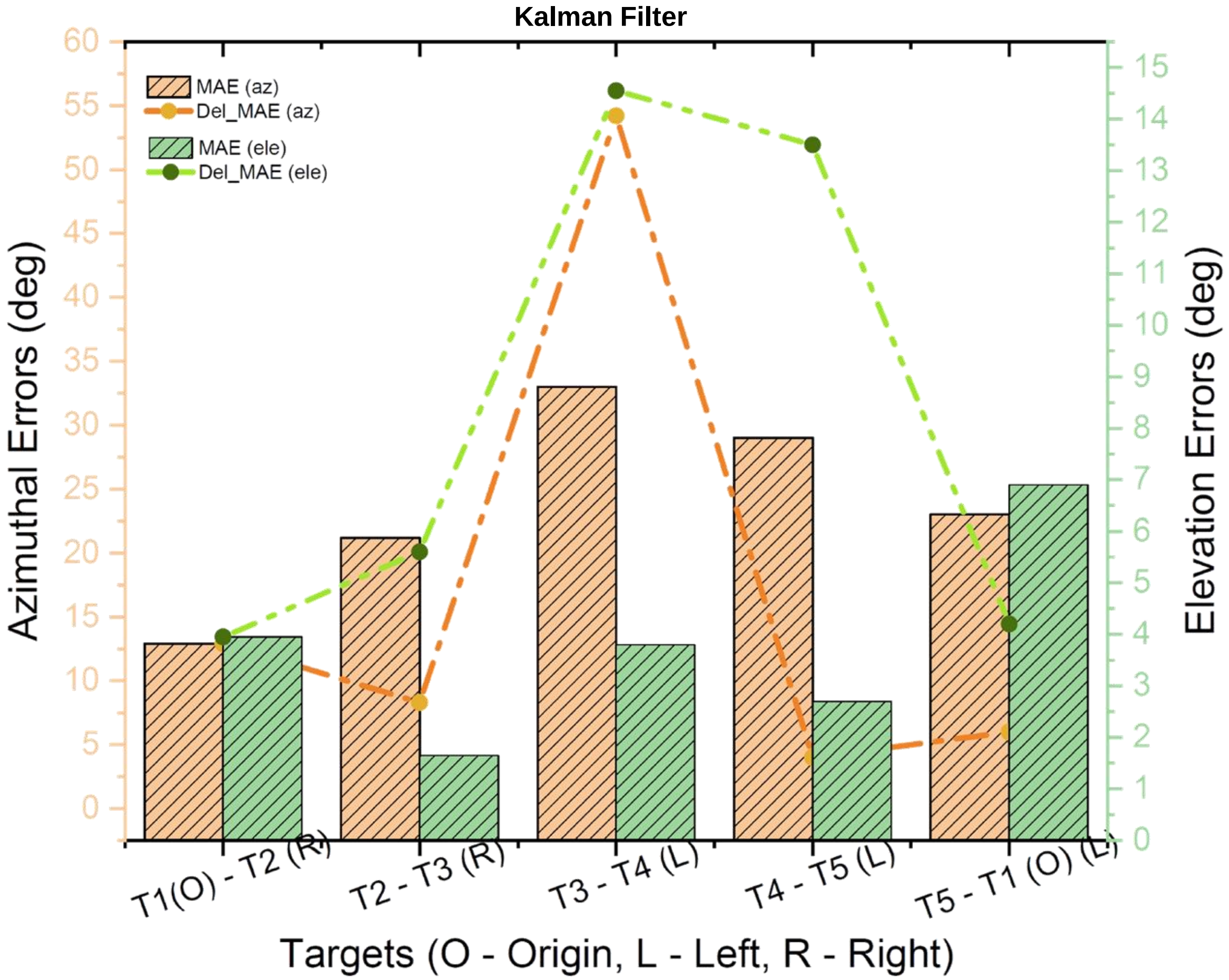


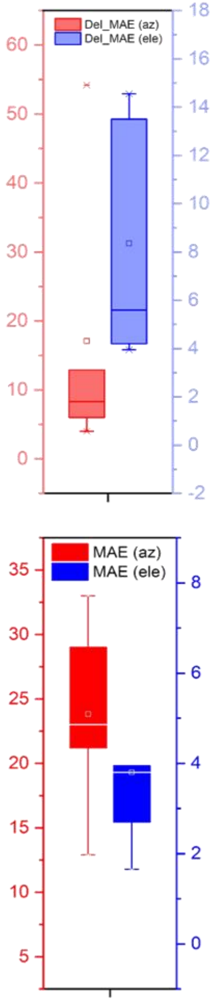


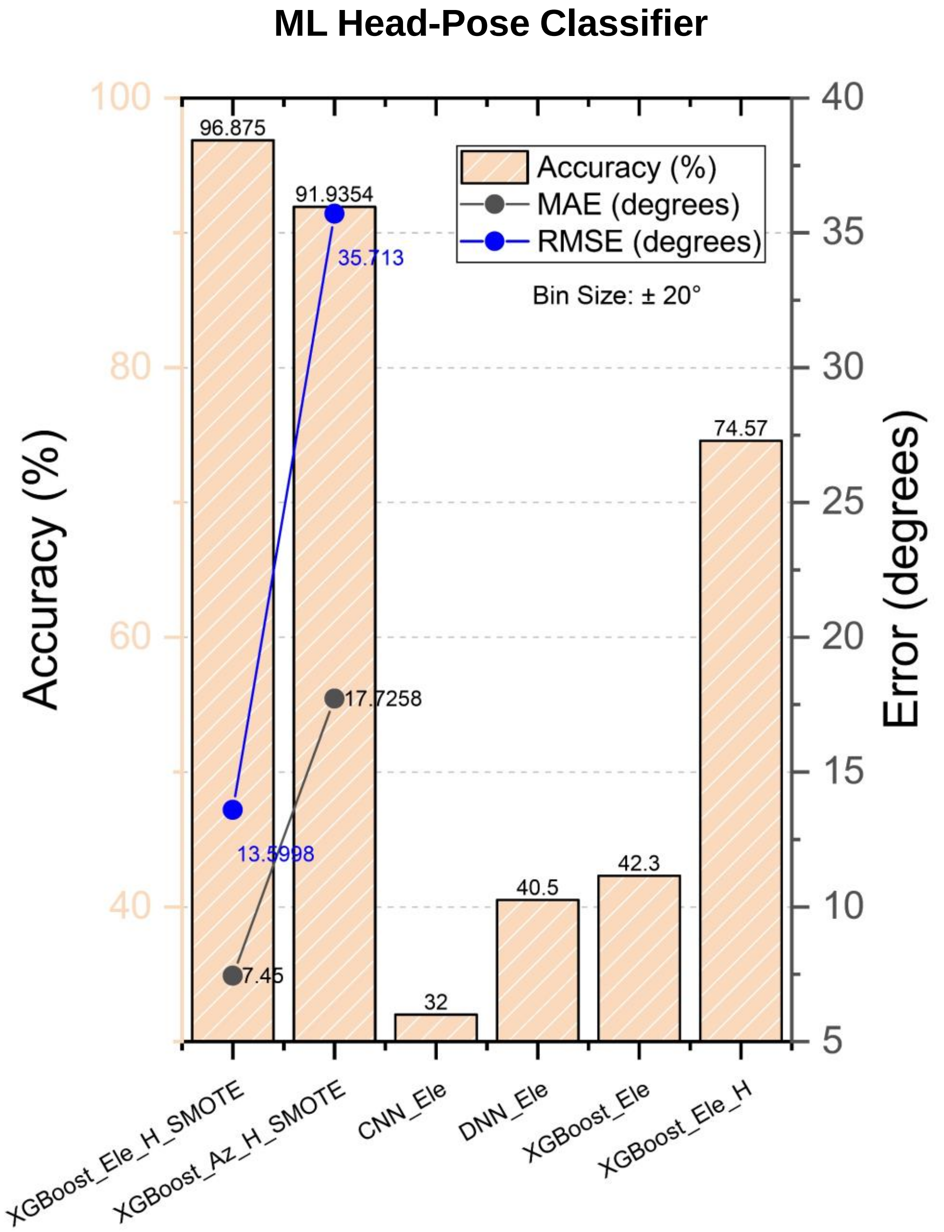
Success Metrics

- Exceed or replicate existing earable inertial head-pose tracking accuracy
- Exceed or replicate existing earable inertial activity recognition accuracy
- Achieve ML head-pose classification accuracy comparable to human sound localization range
- Negligible latency during real-time head-tracking and binaural audio rendering
- Achieve binaural rendering within human sound localization range

Key-Findings (Head-Pose)

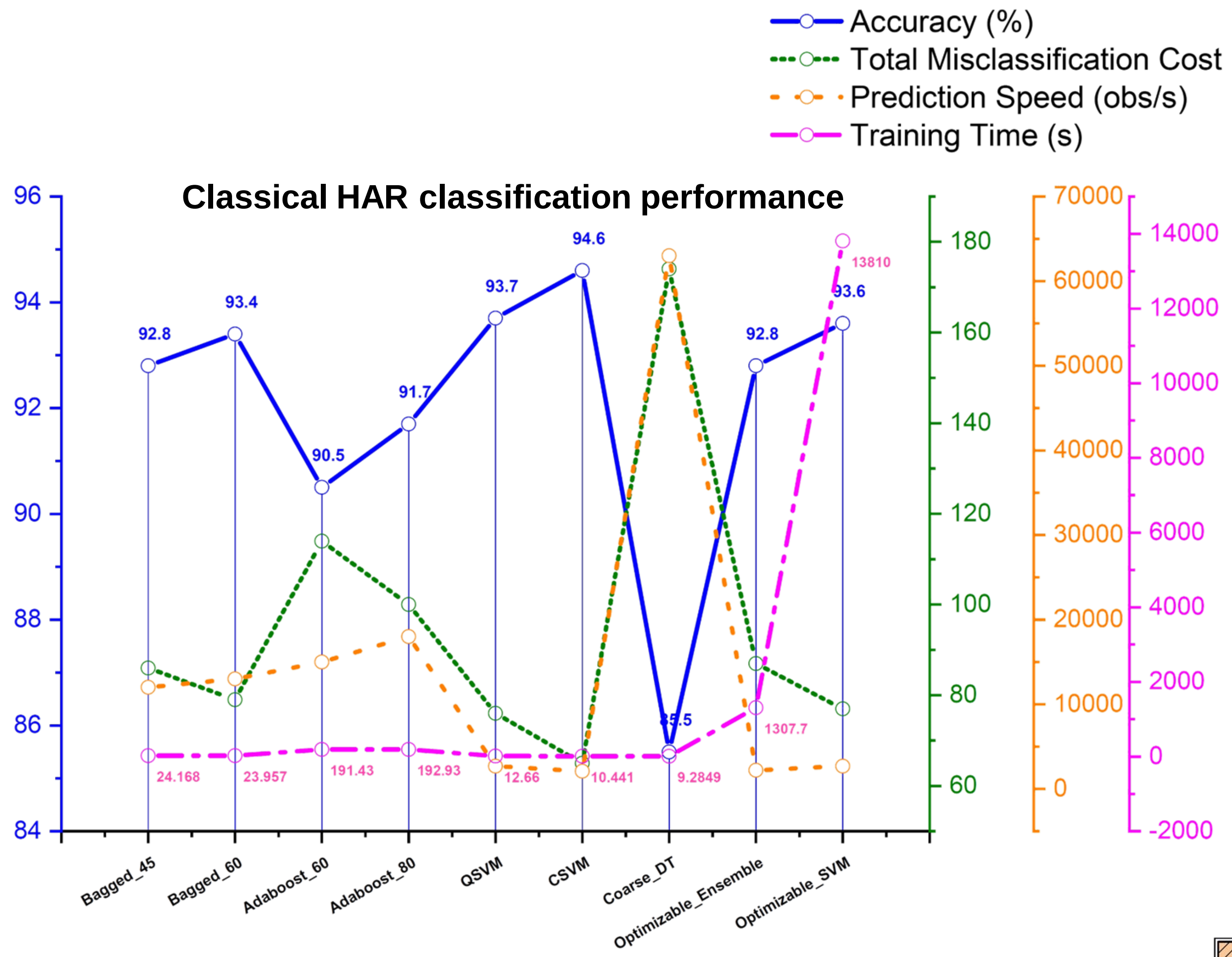


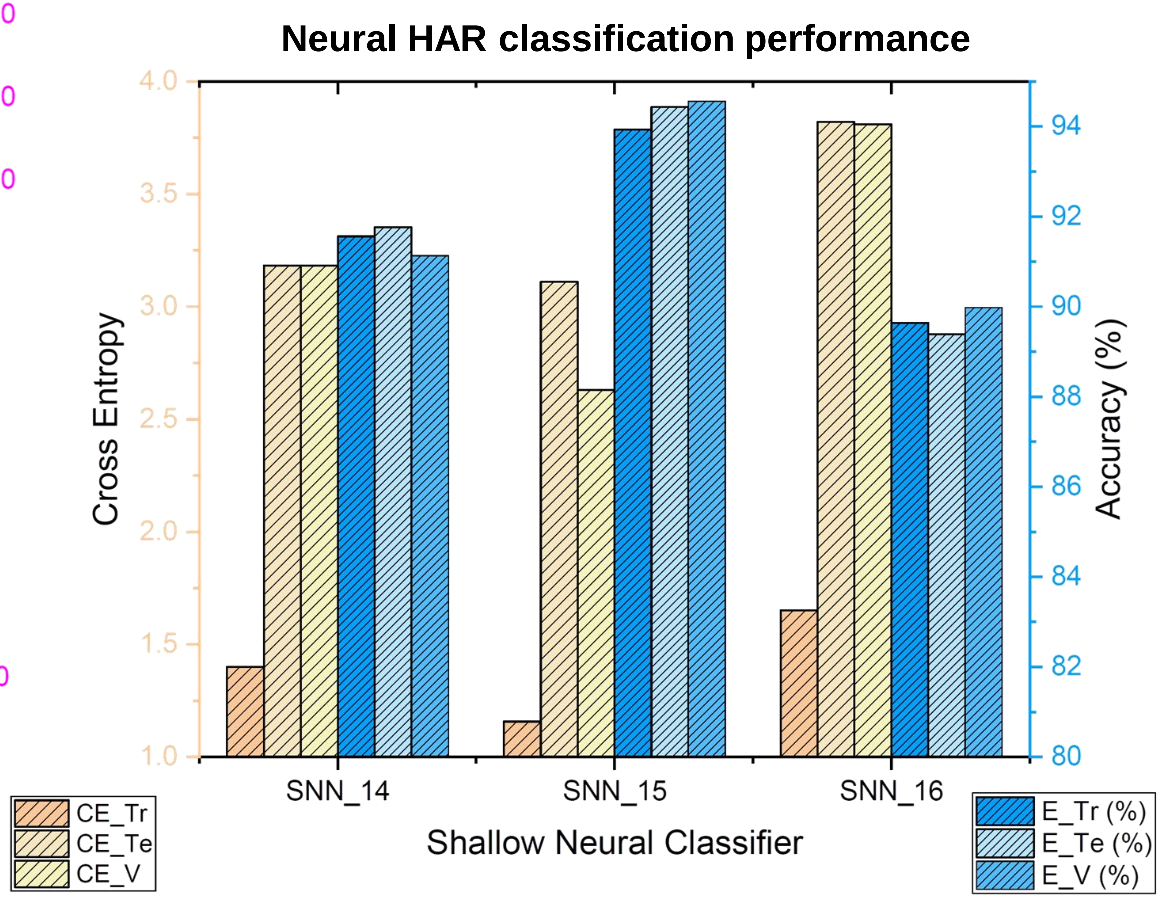




KF median delta motion errors of 8.3 and 5.6 degrees; ML mean absolute error is 17.7 and 7.5 degrees (in azimuth and elevation planes respectively).

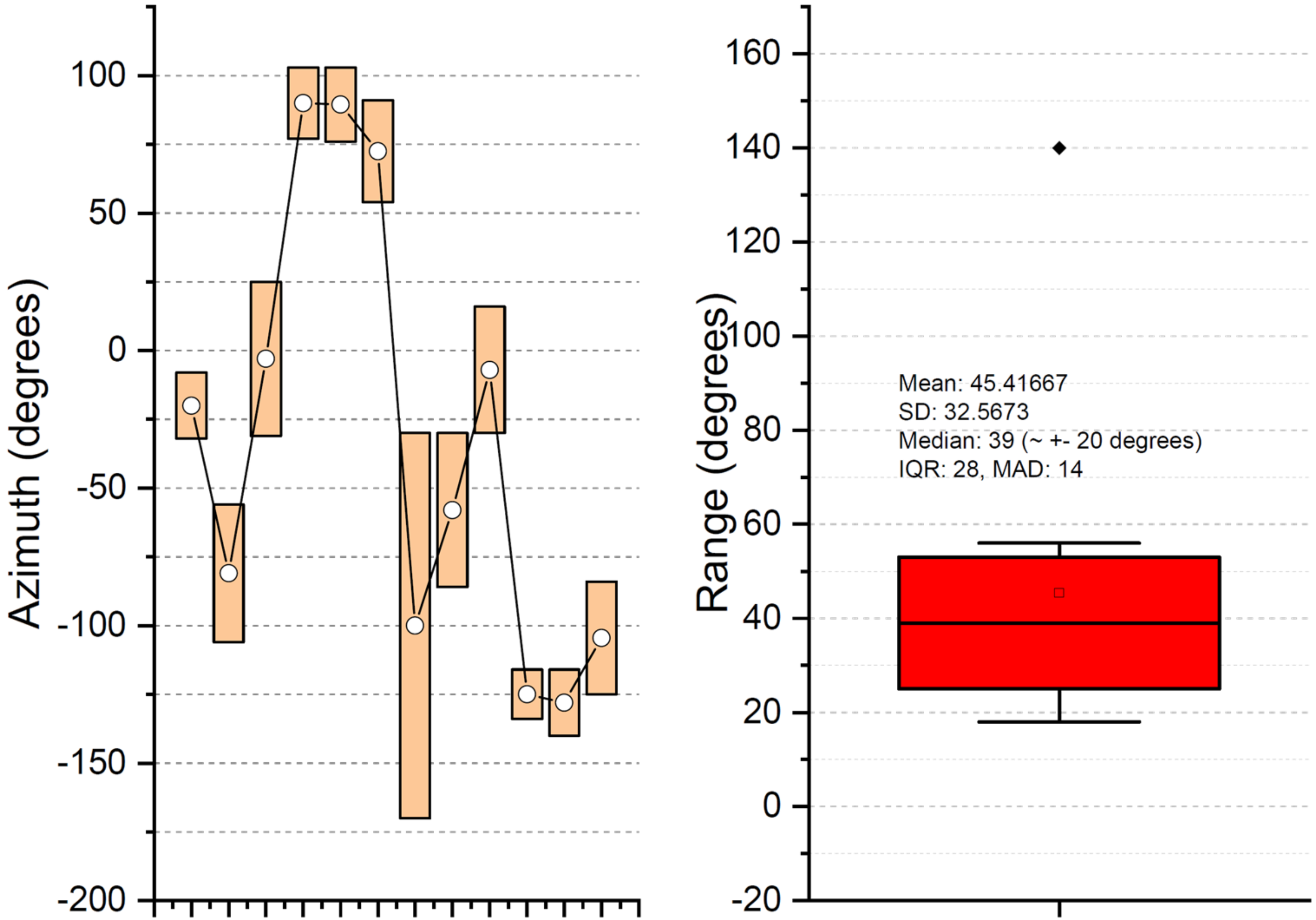
Key-Findings (HAR)





Key-Findings (Binaural Audio)

Snapshot of localization test



Individualization problem in HRTF databases

Different people have different sense of binaural sound localization

±20 degrees is standard (verified)

Implementation / Demo

Head-Tracking (using Kalman Filters) and Binaural Audio Rendering in real-time in MATLAB using custom-built hardware (same sensor config. in eSense):

https://youtu.be/UobTEpy7bVo

Head-Tracking and Binaural Audio Rendering in real-time using Oculus Quest in the Unity 3D VR Platform:

https://youtu.be/t_jXOIgn3_s

Prior Work and Relative Novelty

Head-pose estimation using inertial sensors:

- Ferlini et al. [1] proposed the fusion of 6DOF inertial data from <u>two earable sensors</u> in order to estimate final head position using quaternions, delta motions and complementary filters, achieving estimation errors as low as 5.4 degrees and 18.7 degrees for short and long movements (noisier) respectively.
- Commercial VR systems <u>simply do not depend solely on dead-reckoning algorithms</u> but use additional <u>computationally intensive safeguarding techniques</u>: time warp and occasional visual tracking (e.g. Oculus Quest has 4 cameras) [2].

- **<u>Using a single inertial sensor setup</u>, we achieved median delta motion errors of 8.3 and 5.6 degrees (in azimuth and elevation planes) via our fine-tuned Kalman Filter for <u>long (noisy) movements</u> (enough for binaural audio rendering).**
- **Our machine learning model achieved 17.7 and 7.5 degrees mean absolute error, <u>which is within binaural sound localization range of humans</u>.**

[1]. Ferlini, A. Montanari, C. Mascolo, and R. Harle, "Head Motion Tracking Through in-Ear Wearables," in *EarComp 2019, 1st International Workshop on Earable Computing; ACM Conference (UbiComp/ ISWC'19 Adjunct)*, London, United Kingdom, 2019.

[2]. C. Anthes, R. J. García-Hernández, M. Wiedemann, and D. Kranzlmüller, "State of the Art of Virtual Reality Technology," in 2016 IEEE Aerospace Conference, 2016, pp. 1–19.

Prior Work and Relative Novelty (contd...)

Activity Recognition using Earable Sensors:

- New sensor, limited work:
  - Ahad et al [1]: 6 activities; 50467 frames (90 seconds per frame); accuracy: 81.20% (KNN) and 88.33% (CNN); proposes sensor fusion for improvement.
  - Prakash et al [2]: Step counting accuracy: 95% across 5 activities; uses DTW.
  - Radhakrishnan et al. [3]: Sensor fusion in gym; 8 exercise activities; 92% accuracy (random forests).
  - Nirjon et al [4]: Sensor fusion; 7 activities; 96.8% accuracy.

- **Without sensor fusion, we achieved classification accuracy of 94.6% using C-SVM and 94.25% using SNN_15 across 9 activities (despite variable sampling rate).**

[1]. T. Hossain, M. S. Islam, M. A. R. Ahad, and S. Inoue, "Human Activity Recognition Using Earable Device," in *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers - UbiComp/ISWC '19*, London, United Kingdom, 2019, pp. 81–84.
[2]. Prakash, Z. Yang, Y. L. Wei, and R. R. Choudhury, "STEAR: Robust Step Counting from Earables," *Power*, vol. 30, no. 20, p. 10, 2019.
[3]. M. Radhakrishnan and A. Misra, "Can Earables Support Effective User Engagement during Weight-Based Gym Exercises?," in *EarComp 2019, 1st International Workshop on Earable Computing; Proceedings of the ACM Conference (UbiComp/ ISWC'19 Adjunct)*, ACM, London, United Kingdom, 2019.
[4]. S. Nirjon et al., "MusicalHeart: A Hearty Way of Listening to Music," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems - SenSys '12*, Toronto, Ontario, Canada, 2012, p. 43.

Prior Work and Relative Novelty (contd…)

Fusing Head Inertial Data with Binaural Audio in VR context:

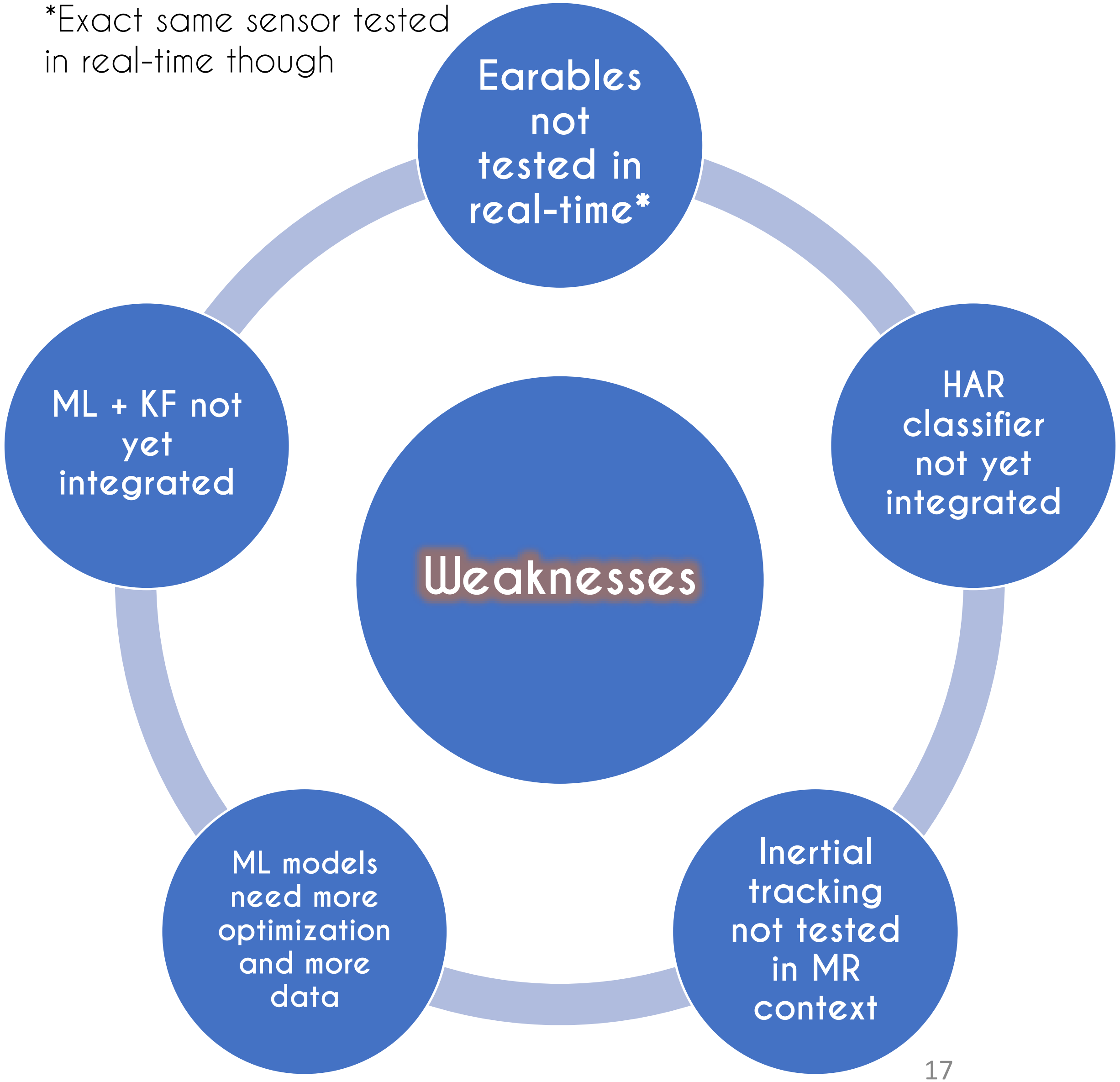
- Requires bulky, computationally hungry headsets/VR frameworks and sophisticated head-trackers [1][2] but gives high audio fidelity and low-latency.
- Audio spatializer plug-ins mostly support only VR headsets and game engines (not standalone sensors) [3][4].
- Limited to virtual space.

- **Our MATLAB demo shows that it is <u>not necessary to use VR headsets and trackers</u> for binaural audio rendering fused with head tracking without sacrificing low latency (opens possibilities in MR context).**

[1]. S. Serafin, M. Geronazzo, C. Erkut, N. C. Nilsson, and R. Nordahl, "Sonic Interactions in Virtual Reality: State of the Art, Current Challenges, and Future Directions," *IEEE Computer Graphics and Applications*, vol. 38, no. 2, pp. 31–43, Mar. 2018.
[2]. D. N. Zotkin, R. Duraiswami, and L. S. Davis, "Rendering Localized Spatial Audio in a Virtual Auditory Space," *IEEE Trans. Multimedia*, vol. 6, no. 4, pp. 553–564, Aug. 2004.
[3]. C. Jenny, P. Majdak, and C. Reuter, "SOFA Native Spatializer Plugin for Unity—Exchangeable HRTFs in Virtual Reality," Audio Engineering Society Convention 144, 2018.
[4]. D. Murphy and F. Neff, "Spatial Sound for Computer Games and Virtual Reality" in Game Sound Technology and Player Interaction: Concepts and Developments. IGI Global, 2011.

Strengths, Weaknesses and Future Directions

*Exact same sensor tested in real-time though



Member Contributions

Swapnil Sayan Saha

The Hustler

- Overseeing website/GitHub repo.
- Survey of literature pertinent to head-pose estimation and activity recognition.
- Formulation of data collection testbed and software framework at NESL.
- Preprocessing, labeling and filtering collected data.
- Training of human activity classifier.
- Developing computer framework for collection of earable data and modifying earable characteristics.
- Developing hardware for real-time head-pose estimation and binaural audio rendering (using complementary Kalman Filter)

Vivek Jain

The Hacker

- Survey of literature pertinent to integration of head-pose with binaural audio in context of VR/MR platforms
- Binning head-pose data and training of machine learning classifiers with head-pose data.
- Optimization of machine learning models for head-pose estimation.
- Implementation of binaural capabilities with the trained model.

Siyou Pei

The Designer

- Survey of literature pertinent to implementation and simulation of spatial audio dynamics.
- Implementation of spatial audio platform in Unity 3D.
- Integrating Oculus VR platform with spatial audio platform in Unity 3D.
- Integrating head-pose transfer function and binaural audio in VR platform.

THANK YOU