Virtual Reality (XR) Content Technology Based on Generative AI and Emotion Recognition

Team members: Zanika Hossain, Natalia Tondi, Gokul Chaluvadi, Kshitij Kokkera | **Faculty adviser:** Kostadin Damevski, Ph.D. | **Sponsor:** VCU College of Engineering

Introduction

Real-time emotion recognition in VR environments presents unique challenges requiring both speed and accuracy. This research extends recent work in multilingual speech emotion recognition by introducing pseudo labeling and optimizing for real-time performance in VR settings.

Prior Work

- Based on Osman, M., Nadeem, T., & Khoriba, G. (2023). Towards Generalizable SER: Soft Labeling and Data Augmentation for Modeling Temporal Emotion Shifts in Large-Scale Multilingual Speech. arXiv preprint arXiv:2311.08607.
- Combined 16 datasets (375 hours)
- Whisper encoder architecture
- Soft labeling approach
- Strong zero-shot performance

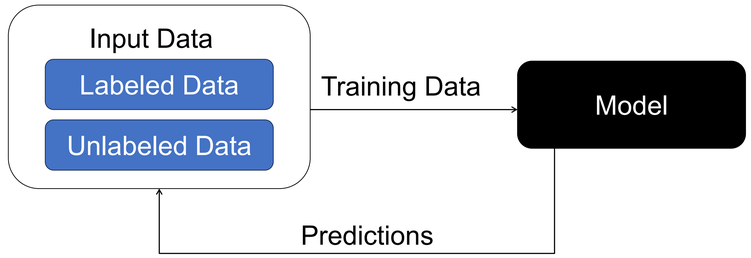
Research Challenges

- Real-time processing requirements
- VR environmental noise
- Speaker variability
- Limited labeled data
- Latency constraints

Proposed Methodology

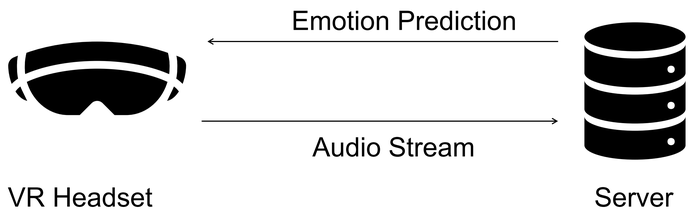
Pseudo labeling Strategy

- Initial model training on labeled data
- High-confidence prediction selection
- Iterative model refinement
- Temporal consistency checks



Real-time Processing

- Streaming audio buffer system
- Optimized inference pipeline
- Adaptive window sizing
- Efficient feature extraction



Expected Outcomes

- Real-time emotion recognition (< 100ms)
- Improved accuracy via pseudo labeling
- Robust VR noise handling
- Cross-speaker generalization

Research Timeline

- Phase 1: Implementation of base model
- Phase 2: Pseudo labeling integration
- Phase 3: Real-time optimization
- Phase 4: VR integration & testing
- Phase 5: Evaluation & refinement

Evaluation Plan

- Accuracy metrics
- Latency measurements
- Cross-dataset validation
- VR environment testing
- User experience studies