

Focus Field

A7: Nasir Crownie, Cole Franklin, Omar Ghabayen
18-500 Capstone Design, Spring 2026
Electrical and Computer Engineering Department
Carnegie Mellon University

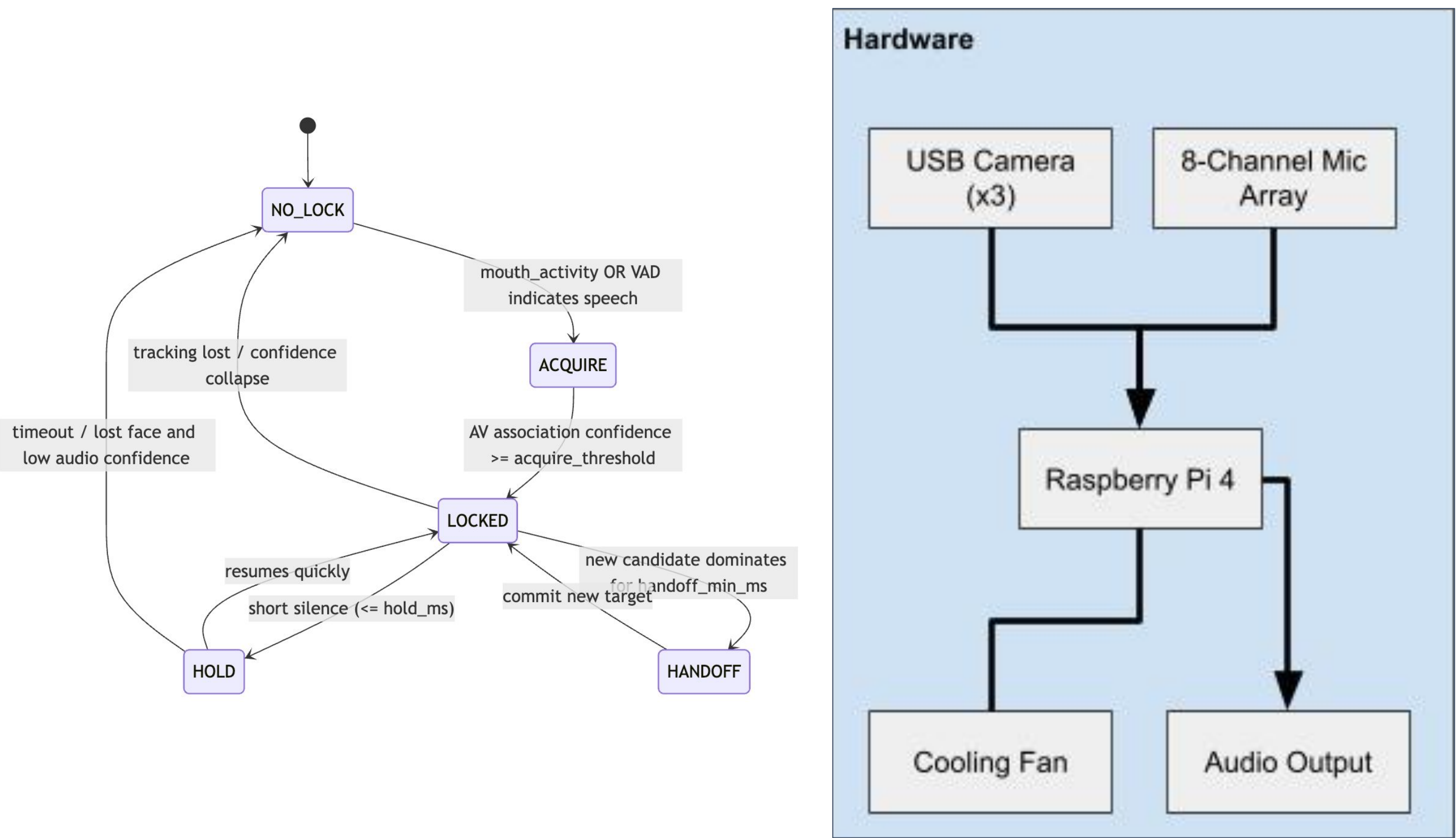


Product Pitch

In a shared space, where your microphone hears everything — FocusField hears only you. FocusField is a tabletop sensor pod that locks onto an intended speaker using a vision and microphone array fusion. The system outputs a single clean audio stream for meetings or calls — no cloud required. It solves the "loudest wins" problem with laptop and room mics by using direction-of-arrival audio and camera-assisted speaker locking. The system met its target lock accuracy requirement (91.4% vs $\geq 90\%$ target) and setup time (1.5 min vs ≤ 5 min target), and delivered a 20%+ ASR intelligibility improvement over a standard laptop mic baseline

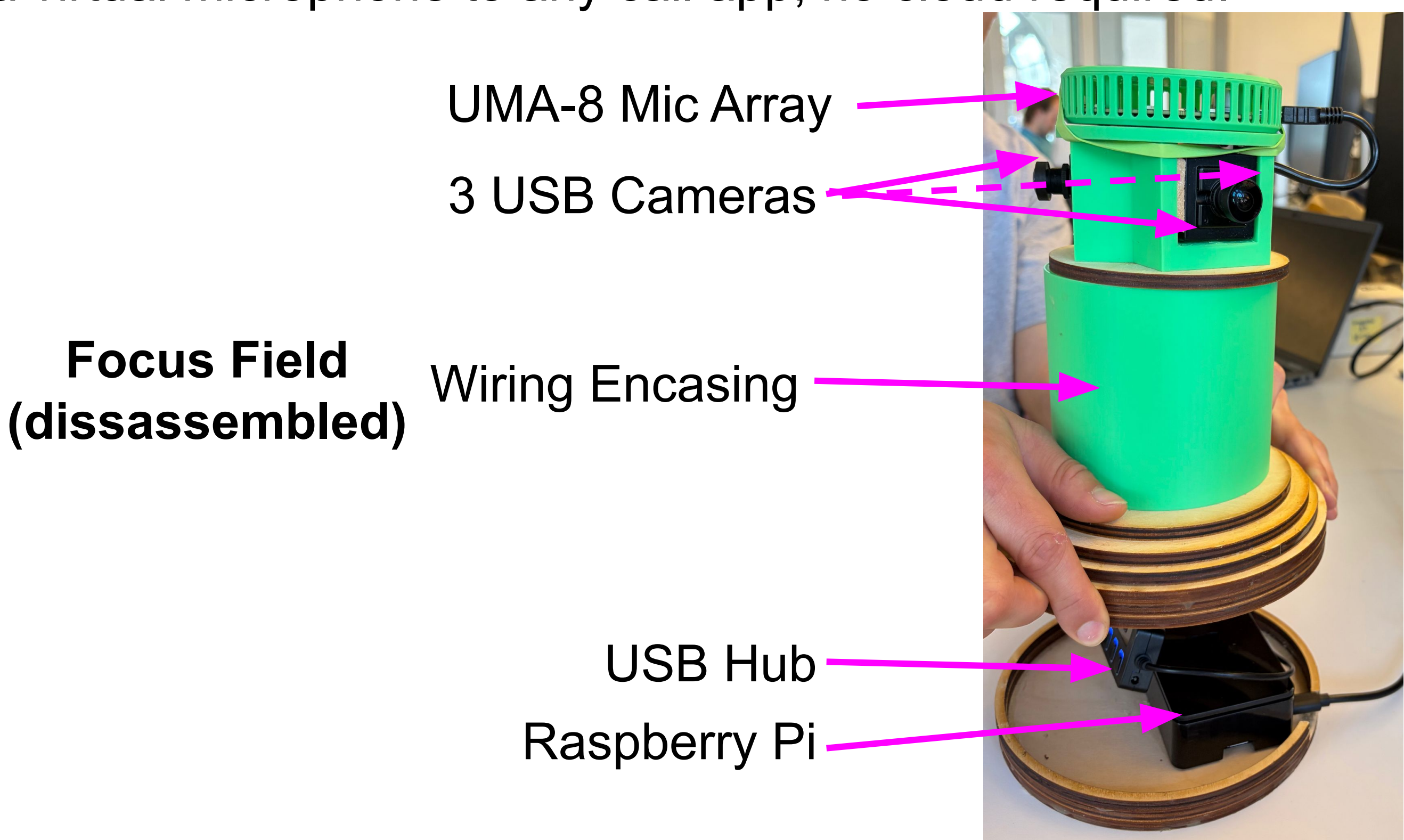
System Architecture

The hardware pipeline flows from the sensor inputs through the Pi 4 to audio output. The lock state machine governs speaker selection, entering ACQUIRE on detected speech, transitioning to LOCKED once AV confidence clears the threshold, and handing off only when a stronger candidate dominates for long enough to prevent rapid switching.



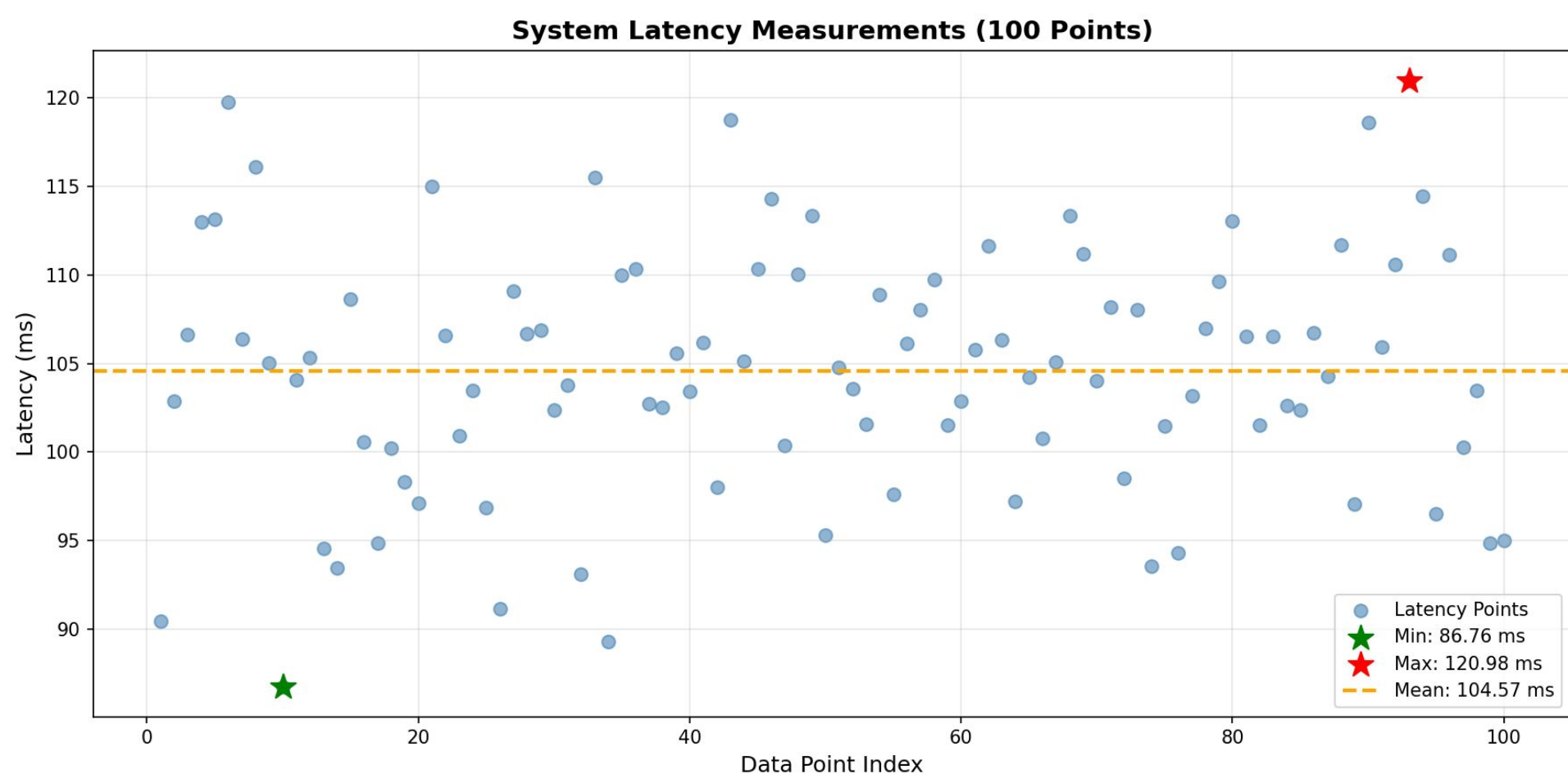
System Description

FocusField runs a fully local pipeline on a Raspberry Pi 4. A UMA-8 mic array captures audio at 48 kHz, feeding an SRP-PHAT direction-of-arrival estimator that tracks speaker angle at 10–20 Hz. Up to three USB cameras run face detection in parallel, and a fusion lock module combines both signals using an acquire/hold/handoff state machine to maintain stable speaker selection. The locked angle steers an MVDR beamformer to suppress off-axis noise, with an optional WebRTC noise suppression stage before output — exposed as a virtual microphone to any call app, no cloud required.



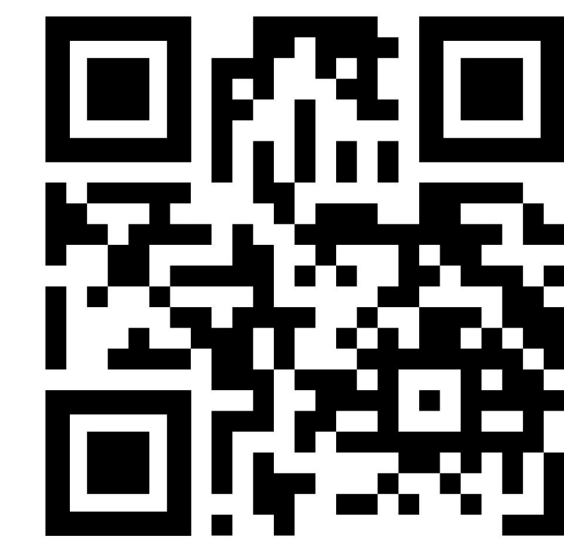
System Evaluation

Our evaluation was split into two scopes: Validation asked whether FocusField solves the right problem in a real shared-space scenario, while Verification asked whether it meets its defined requirements correctly.



Use-Case Requirements:

Metric	Target	Actual
Target Lock Accuracy	$\geq 90\%$ accuracy	91.4%
Steering Error	$\leq 15^\circ$ AAE	12.3° AAE
Interferer Suppression	≥ 8 dB at 60° (1m offset)	8.7 dB
Intelligibility	$\geq 20\%$ ASR	23.1% ASR
System Latency	≤ 100 ms	104 ms
Cost	$\leq \$600$	\$447
Setup Time	≤ 5 min	1.5 mins



<https://course.ece.cmu.edu/~ece500/projects/s26-team7/>

- The architecture is coherent, instrumented, and privacy-preserving (all processing is local).
- Met 6 of 6 quantitative targets outright.
- Lessons learned: modular design (Tier 0→2) allowed MVP delivery even when stretch goals (multi-cam robustness) weren't fully reached.
- Future work: improve beamforming latency, add gesture/hotkey override as a polished feature, and explore deployment in larger shared spaces.
- Total BOM: \$447 (well within the \$600 target).