

POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

The Rooms - The Embodiment of Interactive Physical and Virtual Environments

Laurea Magistrale in Computer Science and Engineering
Ingegneria Informatica

Author: Daniel Mauricio Ruiz Suarez

Advisor: Prof. Andrea Bonarini

Academic year: 2025-2026

---

1. Introduction

The Physical Room is a research project first developed at Politecnico di Milano and published at ACM Multimedia 2024 [3], with complementary thesis work by Bonanno [1], Esfandiar [2], and Gerosa [4]. The original project established the core hypothesis: a physical space equipped with responsive robotic organs and mediated through VR embodiment can be perceived as a living social entity. Two roles define the interaction: a VR user who controls the room as its body (the Entity), and a Visitor who explores it from within. Our constant goal was for the Entity to experience the room as their own body, while the Visitor is able to perceive the room and all the devices within it as a single, coherent entity.

The ACM paper explicitly identified the need for "further research to understand how such an avatar can be more naturally embodied" [3]. A first version of the physical room (V1) was developed before [1] and proved the concept could resonate with visitors, but the implementation could not sustain the embodied experience it aimed for due to design issues and technical complications. The VR layer used an embodied control schema [4] where the Entity's arm movements were tracked and mapped to abstract ray-like pointers, but the overall experience was undermined by hardware failures (five of seven devices broke during testing).

By the time this work began, the physical installation had deteriorated further: unmounted and dispersed devices, broken components, poorly constructed and managed wiring secured with tape instead of proper soldering, inconsistent mechanical assemblies, a disorganized codebase with many repeated scripts across devices, and a non-functional VR scene with all bindings undone and no documentation to reassemble it.

This thesis then, that started from a non-functional baseline, answers that call through the V2 redesign of both, the Physical Room and the VR Room, a complete re-engineering of the system architecture, network protocol, hardware, firmware, and the VR's approach, design, concept, integration and implementation. The central question is not whether a room can be conceived as an avatar (Bonanno, Esfandiar and Gerosa already demonstrated that possibility) but what engineering foundation is required to make that conception work reliably, responsively, and safely enough for genuine embodied experience to emerge in practice.

2. Physical Room Evolution

The V2 Physical Room was built on the same skeleton as the V1 version, this time with a stronger focus on design, construction and integration, including a padded floor consistent with the room's overall appearance and feel. It houses seven devices across four types, each fulfilling a distinct embodied role in the room's non-anthropomorphic anatomy:

- **Tendos (Puller, 3 units):** Ceiling-mounted rope motors representing the room's cardiovascular and muscular system. V1 used open-loop 35 kg RC servos with PWM control and no position feedback. V2 (Figure 1 (a)) replaces these with STS3215 closed-loop serial bus servos providing encoder feedback (4096 ticks per revolution), a redesigned 45 mm spool, and a timed pull model that estimates rope speed from commanded motor units, and implemented four behavioral modes (IDLE, ALERT, SHRIEK, DANGER) using the LEDs and audio output (all modes are automatic, but ALERT and SHRIEK can also be VR-triggered via controller buttons).
- **Palpus (Probe, 2 units):** Pan-tilt assemblies representing the room's attention and perception. V1's touch detection used a balloon-based pressure sensor with a Kalman filter that was prone to noise and false positives. V2 refines the overall wiring and construction (Figure 1 (b)), and implements a three-state touch machine (TOUCH_IDLE, TOUCH_ACTIVE, TOUCH_COOLDOWN) with dynamic threshold scaling and three response patterns.
- **Cor (Sentinel, 1 unit):** A proximity-aware corner device representing boundary awareness. V1 used four HC-SR04 ultrasonic sensors with a spray mechanism, servo-driven petals, two LED stripes, and a touch-sensitive core in a single device design. V2 (Figure 1 (c)), keeps two of the four sensors, fused through a validation pipeline, repurposes the dual WS2812B LED strips with three proximity zones (SAFE, ALERT, DANGER) map to escalating LED patterns, and the touch-sensitive core.
- **Pulmo (Breather, 1 unit):** A new device that externalizes the entity's controller breathing, representing the room's respiratory system. A wearable chest band (Figure 1 (d)), with a 50 kg load cell sampled through an HX711 24-bit ADC at 10 Hz. And an adaptive threshold state detection classify each breath as INHALE, HOLD, or EXHALE. A separate servo receiver node drives the room's outer membrane in synchrony with the entity's respiratory cycle.



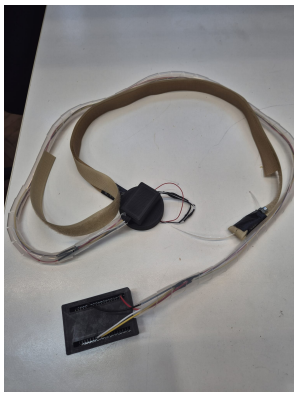
(a) Tendos: Puller device.



(b) Palpus: Probe device.



(c) Cor: Sentinel device.



(d) Pulmo: Breather device.

Figure 1: V2 Physical Room devices across four types.

The firmware architecture was also redesigned. V1 had independent Arduino sketches per device with duplicated networking code, LED control, and sensor processing. V2 organizes firmware as a set of shared headers under `shared_libraries/src/`: WiFi and UDP communication, reusable sensor denoising, and per-device configuration headers that bundle network settings, hardware pin mappings, and calibration thresholds.

3. Virtual Room Evolution

The V1 VR interface was developed by Gerosa [4], the device selection method was proximity-based (entering a virtual collider zone around a point granted control) and it already successfully implemented position-tracking of the entity's arm to control the selected device. A human silhouette visual effect indicated the Visitor's presence. While this approach successfully encouraged communication attempts from most participants, the sense of body ownership and control remained mixed, which was an important challenge to address in this revision.
The core design insight for V2 is that *sustained bodily participation* produces a stronger sense of agency than *abstract control*. In V1, arm movement triggered organ motion during selection, but discrete actions reverted to button presses. In V2, for all controllable devices, both position-tracking and joystick based controls are available, but all defaulting to position-tracking control for immersiveness. the user's arm displacement continues to drive actuators directly, but now an IK (Inverse Kinetic) tentacle makes the connection continuously visible through a dynamic bone chain and a traveling light wave that signals the connection. The user can consciously aim a controller at a device; the tentacle grows following their hand direction to the connection point. A gravity-assisted latch with a 0.5-second magnetic grace period prevents accidental disconnection, and a body-capsule check prevents selection inside the user's personal space for comfortable handling.

Tendos: The tentacle compresses vertically to mirror the rope's physical position using the encoder telemetry, providing continuous visual confirmation that the virtual pull has a physical consequence (Figure 2).

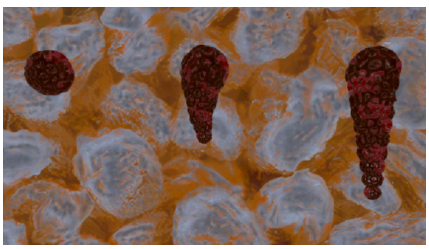


Figure 2: Tendos VR representation: (This is an isolated image of the device outside the room for clarity).

Palpus: The tentacle tip traces that orbit in space, providing visual anticipation of where the physical servos are heading. A cubic response curve provides finer precision near center (Figure 3).



Figure 3: Palpus VR representation: (This is an isolated image of the device outside the room for clarity).

Cor: Not directly controlled in VR. Its LED and audio reactions are driven autonomously by physical proximity and touch telemetry from its ultrasonic sensors (Figure 4).

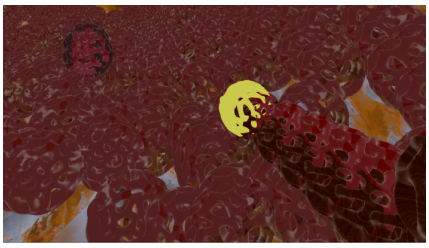


Figure 4: Cor VR representation: (In scene view).

Figure 5 provides an overview of the V2 VR environment where the room itself renders as a responsive organism. The `BreathingRoomDriver` subscribes to Pulmo telemetry and drives deformation across the entire mycelium surface, creating a traveling wave that expands and contracts with each breath cycle. Twenty-seven specialized audio sources aim to reinforce the sense of a single body: a three-layer Pulmo audio system (hold ambient, inhale/exhale loop, resonance heartbeat), per-device state clips for Tendos, a shimmer loop for Palpus, and proximity-driven subsonic pulses for Cor. The shift from V1's abstract ray-based interface to V2's tentacle-based embodied interface represents a progression from communicating the Visitor's presence through explicit visual cues to making the Entity's own body the medium of interaction to perceive the Visitor's presence.

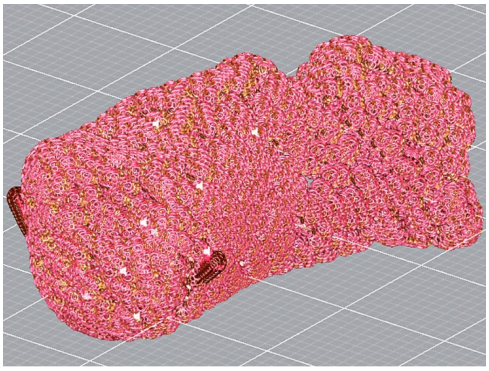


Figure 5: V2 VR environment (Side view of the whole structure): The room representation that contains all the devices, with the mycelium breathing surface and user IK tentacle connection points.

This closed loop is the system's purest expression of embodied interaction. The physical Pulmo chest band samples the Entity's breathing at 10 Hz; a hub forwards the classified state (INHALE, HOLD, EXHALE) to the VR client; `BreathingRoomDriver` maps each state to displacement targets across the mycelium's 2000 GPU-instanced capsules, creating a traveling wave that contracts and expands in synchrony with the Entity's, aiming to provide a unified sensation for Entity and scene. The room does not merely react to the Visitor with less explicit and subtle cues so as not to break the illusion of immersion; it is a complete redesign focused on making the experience as embodied as possible for the Entity's VR controller.

4. System Architecture

The V2 system follows a hub-centric three-layer architecture, as illustrated in Figure 6, replacing the V1 peer-to-peer topology. A central Python hub (`robot_hub_v2.py`) receives commands from VR clients via UDP on port 5600, translates them into device-specific 6-byte frames, and forwards them to devices on port 4210. Devices send telemetry back to the hub on port 4211, which forwards it to VR clients for real-time feedback. The hub tracks device liveness (3-second timeout), enforces safety constraints (command flood protection, value clamping), and maintains a persistent telemetry cache. Safety operates across three layers: firmware-level, hub-level and protocol-level controller.

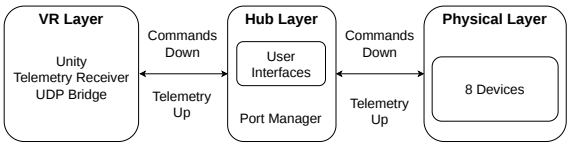


Figure 6: V2 hub-centric architecture: VR clients communicate through a central Python hub that routes commands across eight devices and forwards telemetry bidirectionally.

Communication uses `protocol_v2`, a binary protocol with a fixed 9-byte header and seven message types: HELLO, WELCOME, ROBOT_LIST, COMMAND, TELEMETRY, ACK, and ERROR. UDP was chosen over TCP because embodied feedback tolerates occasional dropped packets far better than head-of-line blocking; a lost command produces one missed frame of motion, while retransmission after 100 ms would break presence.
The hub processing overhead is 1.24 ms mean (P99 2.38 ms), adding less than 2% to the end-to-end latency budget. A 600-second endurance test with 11,448 commands across all device types showed 0.0% aggregate packet loss.
Table 1 summarizes the key architectural changes from V1 to V2.

Dimension	**V1**	**V2**
Architecture	Peer-to-peer	Hub-centric
Protocol	Per-device ad-hoc	`protocol_v2` (unified)
Firmware	Monolithic sketches	Shared libraries
Safety	Per-device relay only	Liveness + flood + clamp
Control interface	Proximity-selection + arm tracking + buttons	IK tentacle + position-tracking + joystick
Embodiment	Partial (mixed user scores)	Improved
Latency measurement	Not measured	Mean 1.24 ms
Embodiment measurement	Ad-hoc Likert + open-ended	VEQ-adapted Likert + open-ended + comparative

Table 1: V1 to V2 comparison across architectural dimensions.

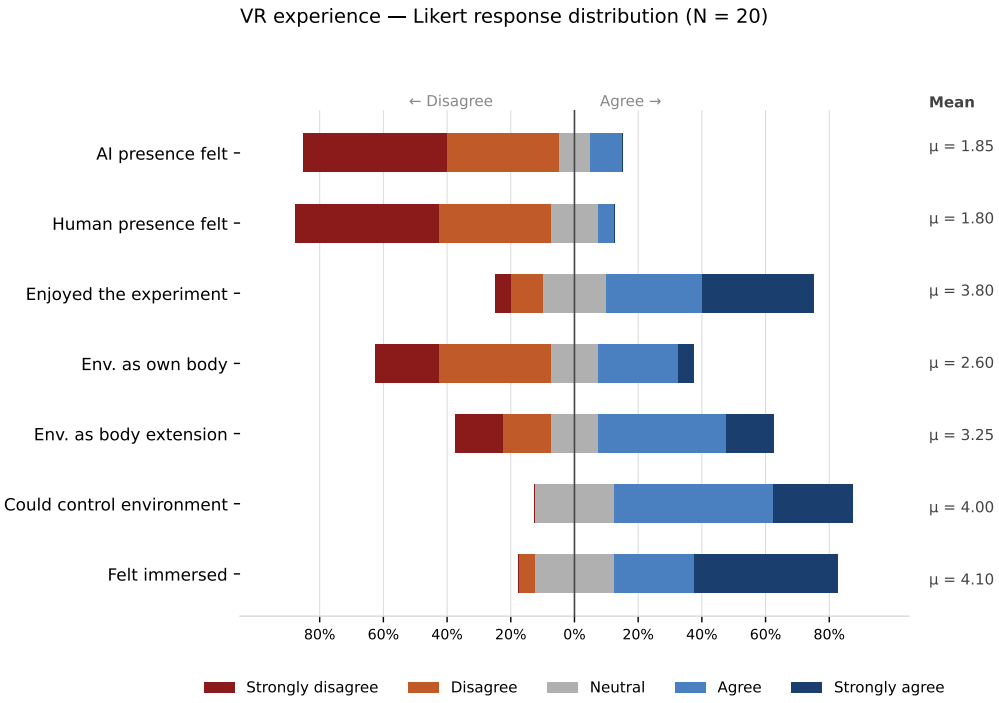


Figure 7: VR Likert scores across all dimensions (N=20): mean ratings showing high immersion and control alongside the communication asymmetry between attempted interaction (4.10) and perceived understanding (2.15).

5. Evaluation

The system was validated through technical benchmarks and a paired user study at Politecnico di Milano's AIRLab. Twenty participants (17 male, 3 female; mean age 27.5, range 22–69) across ten sessions each experienced both roles: the Entity (VR headset, controlling the room) and the Visitor (physically inside the room). After each session, participants completed a questionnaire adapted from the Virtual Embodiment Questionnaire [5], covering embodiment, immersion, control, presence, and communication.

The VR experience scored high on immersion (mean 4.10) and control (mean 4.00), confirming that the reaching-as-agency paradigm provides an effective non-anthropomorphic interface. Participants felt the environment as an extension of their body (3.25) but did not strongly identify it as their own body (2.60), consistent with the non-anthropomorphic design. The physical room excelled at exploration (4.90) and touch (4.60), with 12 of 20 participants rating overall enjoyment 4 or above.

A communication asymmetry emerged: participants attempted interaction at high rates (4.10) but felt poorly understood by the room (2.15). Understanding what the room was communicating scored the lowest in the instrument (2.15). This mirrors the pattern found in V1 [4], where participants "struggled to feel connected" despite near-universal attempts to communicate, confirming a persistent challenge of the non-anthropomorphic ambient communication paradigm. The gap also reflects the system's best-effort UDP design: the room's autonomous behaviors were interpreted as undirected rather than responsive to the individual visitor (Figure 7).

In the forced-choice comparison (Figure 8), the Physical Room felt more natural (12/20), but VR provided a stronger sense of connection (12/20). An order effect was prominent: 13 of 20 participants reported that the sequence of exposure influenced their perception, with VR-first helping understand interaction logic and PR-first providing context for VR controls.

Qualitative feedback from free-form three-word descriptors reinforces this picture. The most frequent descriptors for the Physical Room were *interesting* (7), *curious* (6), *weird* (6), and *strange* (4); for VR they were *interesting* (4), *interactive*

(3), and *alien* (3). The prevalence of exploratory language confirms the room successfully evoked a sense of discovery. When asked what they liked most, participants highlighted the aesthetic environment and responsive audiovisuals. Criticisms centered on unclear interaction logic. The absence of an explicit goal was a deliberate design choice, but these responses highlight a refinement opportunity.

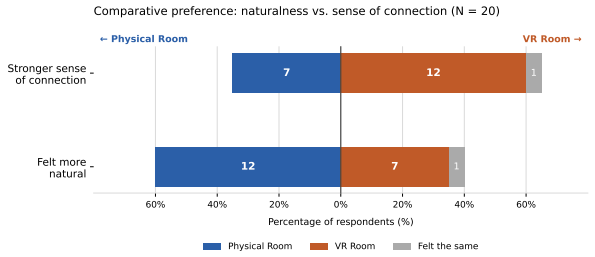


Figure 8: Comparative forced-choice results: the Physical Room felt more natural (12/20), but VR provided a stronger sense of connection (12/20).

6. Conclusions

The principal contributions of this work are:

- A **hub-centric distributed control system** that translates VR commands into device-specific frames, tracks liveness, and forwards telemetry without blocking.
- **protocol_v2**, a versioned binary communication protocol with a fixed 9-byte header supporting seven message types and extensibility for future device types.
- A **modular firmware architecture** with shared libraries (ESPUDP, KalmanFilter1D, per-device configuration) eliminating V1 code duplication.
- A reaching-as-agency VR paradigm with IK tentacle connection, position-tracking control, mycelium breathing environment, and spatial audio, achieving consistent control (4.00/5) and immersion (4.10/5) where V1's implementation produced mixed results.
- A **multi-layer safety system** at firmware (current sensing), hub (liveness detection, flood protection), and protocol (controller binding, sequence numbers) levels.

The V2 redesign demonstrates that architectural simplicity provides the foundation on which richer embodied experiences can be built. The V1-to-V2 arc also illustrates a broader principle: reliable embodiment does not require anthropomorphic form or complex artificial intelligence; a hub, a unified protocol, and modular firmware are sufficient when the architecture treats the user's body as the primary interface.

The system was validated through technical benchmarks (1.24 ms hub latency, 0.0% packet loss) and a paired user study with 20 participants, confirming reliable real-time performance and identifying concrete strengths (immersion 4.10, exploration 4.90, control 4.00) and open challenges (communication asymmetry 4.10 vs 2.15) for future iterations. The hub-centric architecture and unified protocol make the system straightforward to adapt for deeper embodied VR research, and educational contexts.

Future work should address the communication asymmetry by developing more responsive representations that let the Entity's controller perceive the visitor more accurately.

References

[1] Giuseppe Bonanno. The room: Building expressive environments through modular design for emotional and information exchange. M.sc. thesis, Politecnico di Milano, Milan, Italy, 2024.

[2] Mohammadsobhan Esfandiar. The room: Design of a space as a social being – a speculative approach. M.sc. thesis, Politecnico di Milano, Milan, Italy, 2024.

[3] Federico Espositi and Andrea Bonarini. The room: Design and embodiment of spaces as social beings. In *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24)*, Melbourne, VIC, Australia, October 28–November 1 2024. ACM.

[4] Andrea Gerosa. Being the room: Embodiment of living spaces as non-anthropomorphic avatars in virtual reality. M.sc. thesis, Politecnico di Milano, Milan, Italy, 2024.

[5] Daniel Roth and Marc Erich Latoschik. Construction of a validated virtual embodiment questionnaire. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3546–3558, 2020.