

CapAura: A Proxemic Interactive Infrastructure on Everyday Furniture through Capacitive Architecture Unit

Jianan Zhang
sonan13zjn@kmd.keio.ac.jp
Keio University Graduate School of
Media Design
Yokohama, Japan

Kohei Kobayashi
ko5884@kmd.keio.ac.jp
Department of Imaging Arts and
Sciences, Musashino Art University
Kodaira, Tokyo, Japan

Hanrui Peng
phanrui@kmd.keio.ac.jp
Keio University Graduate School of
Media Design
Yokohama, Japan

Hideaki Nii
nii@kmd.keio.ac.jp
Keio University Graduate School of
Media Design
Yokohama, Japan

Takatoshi Yoshida
yoshida@kmd.keio.ac.jp
Keio University Graduate School of
Media Design
Yokohama, Japan

Mauricio Sousa
mauricio@kmd.keio.ac.jp
Keio University Graduate School of
Media Design
Yokohama, Japan

Kouta Minamizawa
kouta@kmd.keio.ac.jp
Keio University Graduate School of
Media Design
Yokohama, Japan



Figure 1: CapAura turns an ordinary floor carpet and everyday conductive furniture into a room-scale capacitive sensing field. (A) Electrodes integrated into the floor modules and ceiling modules, with conductive furniture acting as passive electric field extenders, enable room-scale proxemic sensing. (B) When a user touches or hovers near a surface, our system senses this interaction. (C) The system jointly detects Position, Identity, Movement, and Orientation (PIMO) without cameras or worn devices. (D) Detected proxemic events drive spatial hotkey interactions, turning the living space into an ambient spatial interface.

Abstract

We propose CapAura, a proxemic infrastructure that enables users to interact with the living space. We achieve this by deploying a ceiling-to-floor capacitive space and utilizing conductive furniture in between as passive electric-field extenders, transforming

surface-limited floor sensing into volumetric, whole-body awareness with zero instrumentation on the user and no modification to the furniture itself. We develop CapAura into a prototype that enables room-scale proxemic sensing across multiple scales, detecting people's Position, Identity, Movement, and Orientation—without cameras or worn devices. From a sparse 3×5 capacitive grid, our prototype achieves 29.7 cm MPJPE (Mean Per Joint Position Error) across 32 body joints, 99.61% within-session user-identity recognition, 66.1% activity recognition, and 73.61% hand-orientation recognition. Beyond passive sensing, our user study (N=16) shows that furniture proximity and touch can serve as ambient interaction triggers—perceived as functionally effective ($M=4.44$, $p=.002$) and



This work is licensed under a Creative Commons Attribution 4.0 International License.
UIST '26, Detroit, MI, USA
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2856-3/2026/11
<https://doi.org/10.1145/3830398.3830590>

novel and exciting ($M=4.79$, $p<.001$)—turning the living space into a spatial interface that reaches beyond the user to their immediate, touchable surroundings.

CCS Concepts

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Computer systems organization** → *Embedded and real-time systems*.

Keywords

touch sensing; pose estimation; activity recognition; user identification; capacitive sensing; smart environments

ACM Reference Format:

Jianan Zhang, Kohei Kobayashi, Hanrui Peng, Hideaki Nii, Takatoshi Yoshida, Mauricio Sousa, and Kouta Minamizawa. 2026. CapAura: A Proxemic Interactive Infrastructure on Everyday Furniture through Capacitive Architecture Unit. In *The 39th Annual ACM Symposium on User Interface Software and Technology (UIST '26)*, November 02–05, 2026, Detroit, MI, USA. ACM, New York, NY, USA, 17 pages. <https://doi.org/10.1145/3830398.3830590>

1 Introduction

Our living space is far from static: it actively shapes our lives as we, in turn, interact with it, engaging in a reciprocal dialogue between human and environment [16]. Through this bidirectional process, components within this space—furniture, floors, and ceilings—transcend mere functionality to form aura-like proxemic boundaries around them at room scale. A room that responds to its users must first perceive these boundaries—who is present, how near they are, and when they touch or are in proximity. Most sensing systems behind smart rooms and proxemic interaction devices, however, are vision-based [5, 23, 30, 40, 60]: they watch the room from a distance, inferring touch, proximity, and this aura-like boundary through indirect calculation—and often failing outright when furniture or the body itself occludes the view. We wanted to make this previously conceptualized aura-like boundary [15, 49] real with a proxemic infrastructure: a sensing field that can feel the people inside—their presence and proximity, not just their touch. To realize it, we chose capacitance—the same technology behind our touchscreens—as our sensing medium.

Prior approaches each reveal a piece of the puzzle while leaving gaps. Vision-based sensing is accurate but raises privacy concerns and perceives the body indirectly [18, 29, 58, 67]. Wearable devices such as IMU- or capacitive-garment-based trackers [32, 36, 61, 63] can sidestep these issues but impose an instrumentation burden on users. Pressure-sensor floors [8, 12, 39] offer a compelling passive alternative, yet they are expensive and confined to flat surfaces—useful, but we wanted our aura to reach further. Capacitive sensing directly detects touch and proximity through electric-field coupling, yet it has traditionally been confined to small, dedicated surfaces [4, 11, 21, 22, 24, 50, 64]. A line of work has pushed beyond this limit: TileTrack observed that conductive furniture passively extends the field sensed by its floor tiles [56], Mirage showed that off-body electric-field sensing can infer presence and richer states such as activity and gesture [44], and Wall++, though still acting on an instrumented wall surface, scaled capacitive sensing to a whole

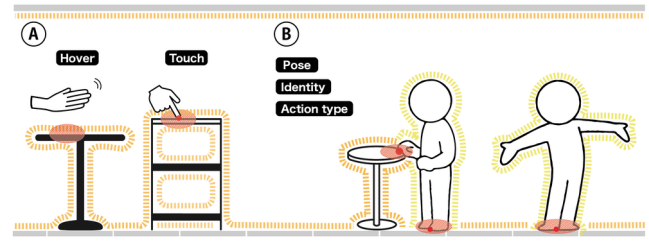


Figure 2: The concept of CapAura, which envelops a living space in a proxemic interactive field. (A) The system directly senses user interactions, including touch and hover events. (B) Unmodified conductive furniture and the floor passively extend the sensing field into the user’s proximity, enabling the estimation of human position, identity, and action type.

room [65]. Our approach stands on this line of legacy works but differs in configuration: rather than treating capacitance as a property of an instrumented surface, we turn the entire room volume into the sensing space through a floor-to-ceiling mutual-capacitance field—in which conductive furniture, though completely untouched, becomes a natural extension of the field, turning everyday objects into sensors.

To answer our research question of how to surround our living space with an interactive aura, our CapAura approach constructs a mutual-capacitance interface between electrodes integrated into the floor and ceiling (Figure 1A). In this configuration, conductive furniture positioned between the floor and ceiling electrodes—itsself entirely unmodified—serves as passive probes that carry the sensing field beyond the floor surface, effectively casting the capacitive sensing layer onto the furniture’s own conductive surfaces, as shown in Figure 2A. This design elevates the field-extension effect that TileTrack observed incidentally on its floor tiles [56] into the structuring principle of our volumetric, room-scale sensing space. Our prototype excels at detecting touch, which we term Hand Orientation Detection, corresponding to the Orientation dimension of the proxemic interaction framework proposed by Ballendat et al. [5, 23].

Beyond detecting touch, we believe the same aura can sense other dimensions of people. Just as a person’s presence can be sensed at a distance—their position, movement, and identity inferred without contact—our system performs analogous inferences. Following Ballendat et al.’s framework [5, 23] and its toolkit realization [40], we designed recognition tasks for the three remaining proxemic dimensions: Position, Identity, and Movement. As shown in Figure 2B, the sensing field cast onto the furniture and floor conforms naturally to human lifestyles, bringing capacitive proximity closer to the body and enabling more precise estimates across all four PIMO dimensions.

Based on our hardware, we collected data and trained neural networks as our recognition module. For Position, we investigate the system’s 3D pose estimation performance across 32 body joints. For Orientation, Identity, and Movement, we assess the model’s ability to classify touch and hover events, identify distinct users, and recognize different activity types, respectively.

To explore use cases, we designed a spatial hotkey system: users can record and trigger personal spatial hotkeys on their chosen furniture, demonstrating how detected proxemic events can drive ambient interactions that map the physical surroundings into a spatial interface. Combining our conceptual approach, hardware, recognition module, and applications, we name our proof-of-concept system CapAura. Our system makes the following contributions to the UIST community:

- We introduce CapAura, a proxemic infrastructure that envelops the living space in a capacitive sensing field with no cameras and no worn devices.
- We form a ceiling-to-floor mutual-capacitance space using floor and ceiling electrodes; any conductive furniture—itsself unmodified—passively joins it as part of our sensing medium, delivering volumetric PIMO sensing at room scale.
- We show that the system supports PIMO recognition at room scale through purpose-built neural architectures, and conduct a user study ($N = 16$) demonstrating that furniture-based spatial hotkey interaction is perceived as both functionally effective and novel.

2 Related Work

2.1 Proxemic Interaction in Living Spaces

The concept of the living space as an interactive medium has a rich history in the HCI and UIST communities. Prior work such as AmbientROOM [30] and HaptoRoom [54] has explored how living environments can be transformed into ambient, responsive interfaces. Building on Hall’s proxemics theory [26], Ballendat et al. [5, 23] operationalized it into four key dimensions of interaction—Position (an entity’s location and distance from others), Orientation (the direction an entity faces), Movement (changes in position and orientation over time), and Identity (unique identification of entities)—and Marquardt et al. [40] further developed a software toolkit for prototyping such interactions. While many proximity-aware interfaces rely on visual methods, touch-aware environments have also been explored. Wall++ [65] embeds capacitive sensors into walls, and InDepth [62] uses force sensing in display windows to detect hand orientation directly. Where Wall++ reaches room scale by coating a dedicated wall surface, our work takes the paradigm off the surface entirely: electrodes in the floor and ceiling form a volumetric sensing space that covers all four PIMO dimensions—without modifying furniture or requiring cameras.

2.2 Room-Scale Human Sensing

Room-scale human sensing approaches fall broadly into vision-based, wearable, and ambient infrastructure categories [2, 25, 53]. Vision-based methods leverage deep learning for robust pose estimation [25, 58, 67], user identification [20, 51], and activity recognition [48, 55], but they perceive the body indirectly—contact with a surface must be inferred from estimated 3D joint positions rather than sensed directly. Additionally, they raise privacy concerns, require line-of-sight, and fail when bodies occlude surfaces of interest, which is a fundamental problem when sensing human-furniture contact. As a result, most advanced vision-based touch sensing

methods [37, 42, 43] require mobile VR/AR headsets, increasing the burden on users.

Wearable sensing offers a practical alternative. IMU-based systems [32, 36, 61, 63] achieve full-body pose estimation from sparse consumer devices, demonstrating that high-accuracy tracking is achievable without dedicated infrastructure. While accurate, such approaches still require the user to carry or wear instrumented devices. Non-visual ambient methods—including temperature [27], WiFi [59], acoustic [31], and pressure sensing [12, 39]—offer privacy-preserving alternatives but still cannot directly sense contact proximity or orientation between a user and a surface.

Capacitance-based sensing bridges this gap: it directly detects electrical field coupling, naturally distinguishing touch, hover, and no-contact states without cameras or worn hardware [44, 56, 65]. Our system, CapAura, extends this principle from surfaces to the room’s volume—sensing hand Orientation on conductive furniture surfaces while also enabling Position, Identity, and Movement detection throughout the space.

2.3 Capacitance Human Sensing

Capacitance-based human sensing, the principle behind the touchscreens on our smartphones, measures changes in capacitance to detect human presence and movement by leveraging the human body’s conductive nature. This technology is favored in Human-Computer Interaction for its low cost, minimal power consumption, and versatile sensor design possibilities. Capacitance-based human sensing can be effective for detecting localized proxemic interactions, including action and gesture recognition [1, 7, 13, 14, 35, 46, 50, 66], user identification [28], and seat occupancy detection [9, 21].

Efforts to extend the capacitance detection range have involved instrumenting existing room elements such as floors [10, 19, 52, 56], walls [65], wearable devices [38, 66, 68], clothing [32, 63], and furniture [9, 11, 21, 22, 64]. However, these approaches have significant limitations: they often require extensive modification of furniture, are confined to less interactive surfaces such as floors and walls, or are limited to restricted contact surfaces such as fingertips and clothing. Even when such systems infer richer states—a grasp on an instrumented object [50], a gesture or lean against a coated wall [65]—the sensing remains tethered to the surface itself, reaching only whoever touches or stands beside it, rather than the presence, identity, and actions of people throughout the room.

Furthermore, detecting 3D human poses at a room scale remains a significant challenge for capacitive sensing. Key limitations include a restricted effective sensing range. As noted by Bian et al. [6], the inherent nature of capacitive sensors makes it hard to focus their field, leading to ambiguity, and signal strength diminishes rapidly with distance. These constraints have largely hindered the application of capacitive methods for volumetric sensing, pointing to a research gap for an ambient, volumetric capacitive sensing approach.

3 CapAura

In this section, we present CapAura, a proxemic infrastructure to detect users within a room by embedding capacitive electrodes into existing room elements.

3.1 Concept

Our concept is to give a room a sensing aura—a field that makes everyday surfaces aware of human presence and activity at a distance, not just at the point of contact. To achieve this, we chose capacitance as our sensing modality: the human body naturally disturbs nearby electric fields, making capacitive sensing a passive, unobtrusive way to detect touch and hover without any worn devices or cameras.

We realize this aura by embedding capacitive electrodes into the floor and ceiling. When conductive furniture is placed on top, it acts as a passive electric-field extender [56], propagating the sensing field onto its own surfaces. This means the aura does not stay flat on the floor—it climbs the furniture, turning a 2D surface sensor into a room-scale, volumetric sensing field. Non-conductive furniture does not extend the field, but as long as some conductive furniture is present, the system achieves volumetric coverage spanning all four PIMO dimensions.

3.2 Capacitive Modeling and Sensing Principle

We model the room as an analog circuit: the floor and ceiling act as the two electrodes, the air between them is the dielectric, and the human body is treated as a conductor. When a user touches or approaches conductive furniture, an additional coupling path forms between the body and the floor electrode, reducing the effective impedance of the propagation path and producing a distinct change in the received amplitude—this is what allows the system to distinguish contact and proximity events. Since each floor module is assigned a different carrier frequency, the waveform received at the ceiling can be separated per module in the frequency domain, and the 3×5 grid of independent modules provides the spatial resolution needed to localize users. The full equivalent-circuit derivation is provided in the Appendix.

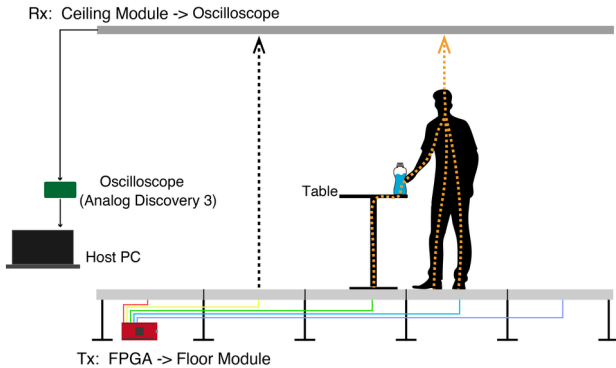


Figure 3: System Architecture for CapAura

3.3 Hardware Implementation

Our proof-of-concept interface covers an area of 2.5 by 1.5 meters, with the floor divided into 0.5 by 0.5 meter modules arranged in a 3×5 grid, as shown in Figure 4.

As shown in Figure 3, each floor module is a conductive metal plate (Figure 4A), individually wired to an FPGA-based transmitter

that assigns each module its own frequency-multiplexed sinusoidal carrier. Two sheets of consumer-grade aluminum foil mounted on the ceiling serve as the receiving antenna (Figure 4B), connected to an Analog Discovery 3 receiver [17] that acquires the superimposed waveform and separates each module’s response in the frequency domain. Full construction details of each module and signal specifications are provided in the Appendix.

A key advantage of this design is that all components are off-the-shelf: conductive metal tiles, consumer aluminum foil, standard wiring, and a commercially available measurement device. No custom fabrication is required, keeping both cost and installation effort low. The system can also be extended to support more channels if needed, which would slightly reduce the data acquisition rate but offers the potential to increase spatial resolution.

3.4 Software Implementation

The software pipeline processes the raw signals received by the Analog Discovery 3 [17]. The received waveform is transformed into the frequency domain using a Fast Fourier Transform (FFT). The frequency resolution of the spectrum is 12.21 Hz, while our floor modules’ carrier frequencies are spaced 400 Hz apart—ensuring each module’s signal can be cleanly separated. From this spectrum, the amplitude at each module’s assigned carrier frequency is extracted. The system updates these amplitude readings at approximately 12 Hz. A 4-sample moving average is then applied to smooth out short-term fluctuations.

Optionally, a baseline calibration can be performed using signals recorded with no person present, suppressing static background coupling and environmental offsets. The processed output is visualized as a color-coded heatmap, as shown in Figure 5, which serves as the raw input for downstream tasks including our neural network models and spatial hotkey application.

3.5 Recognition Module Implementation

Our recognition module consists of three neural network architectures, all taking the 3×5 grid heatmap as input: a Keypoint Detection Model for Position (3D skeleton estimation), a Classification Model whose shared structure serves the three classification dimensions—Identity, Movement, and Orientation—through task-specific output heads, and a Two Person Detection Model for multi-user tracking. The Keypoint Detection Model’s spatial feature extractor follows IntelligentCarpet [39]; our contribution lies in the temporal extension: because each tile’s signal is a baseline-shifted periodic waveform, a single frame is unstable near the contact threshold, so we add a Transformer-based module over 30-frame windows that recovers a stable offset trajectory before mapping to pose. This spatial-temporal structure is dictated by the shape of the signal itself (inherently tile-array $\times$ time); exploring lighter-weight models for datasets of our scale remains future work. Training details are provided in the Appendix.

Keypoint Detection Model (3D Pose Estimation). We adapt the encoder-decoder structure from IntelligentCarpet [39] and extend it with a TransformerEncoder [57] tailored to the $(B, 30, 3, 5)$ input format. The encoder processes each of the 30 frames through a shared 2D CNN backbone, producing a 256-dimensional feature vector per frame, which is then passed to a TransformerEncoder

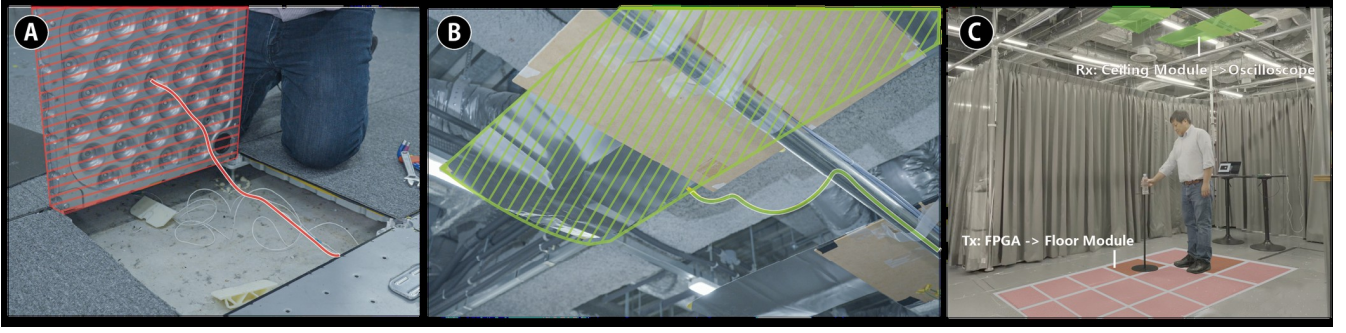


Figure 4: Finished Hardware Prototype: (A) Floor module. (B) Ceiling antenna. (C) Our prototype comprises 3×5 Tx electrodes on the floor and 1×2 Rx electrodes on the ceiling.

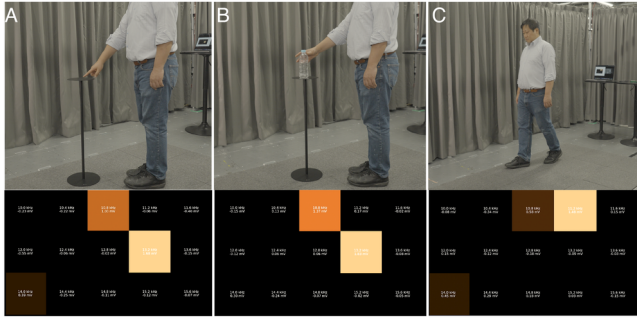


Figure 5: Software demonstration on various user-initiated capacitance events. (A) Contacting with Conductive Furniture (B) Contacting Integrated Conductive Water Bottle on Furniture (C) Walking.

with 4 layers and 8 attention heads to capture temporal dependencies. The decoder passes the Transformer’s output through a 3D convolutional upsampling network, producing per-joint volumetric heatmaps of shape $(B, 32, 3, 5)$ over the world coordinate space ($X \in [-1.25, 1.25]$ m, $Y \in [0, 2.0]$ m, $Z \in [0, 1.5]$ m). A differentiable SpatialSoftmax3D operation converts each joint’s heatmap into a continuous 3D coordinate prediction, yielding the final output of shape $(B, 32, 3)$. The model (~ 35 M parameters) is trained with a combined loss:

$$\mathcal{L} = 100 \cdot \mathcal{L}_{\text{heatmap}} + 30 \cdot \mathcal{L}_{\text{keypoint}} \quad (1)$$

where $\mathcal{L}_{\text{heatmap}}$ is the mean squared error between predicted and target volumetric heatmaps, and $\mathcal{L}_{\text{keypoint}}$ is the L2 distance between predicted and ground-truth 3D joint coordinates.

Classification Model (Identity, Movement, Orientation). For classification tasks, we implement a Recurrent Convolutional Neural Network (RCNN) with input shape $(B, 1, 30, 3, 5)$, where the single channel treats the temporal sequence as a 3D volume. Three convolutional layers (3×3 kernels, 1×1 padding) with batch normalization and ReLU [45] extract spatial features per frame, which are then fed frame-by-frame into a single-layer LSTM with 128 hidden units. A dropout layer ($p = 0.5$) and a task-specific linear head complete the network. The model is trained with cross-entropy loss. The output

dimension C varies by task: $C = 15$ for user identification, $C = 8$ for movement recognition, and $C = 3$ for orientation classification.

Two Person Detection Model (Multi-User Tracking). For two-person scenarios, we extend the Keypoint Detection Model by doubling the output dimension to $(B, 64, 3)$, representing 32 joints for each of two simultaneous users. The model is fine-tuned from the single-user checkpoint to leverage learned single-user representations.

4 Evaluation

We evaluate our system across the four dimensions of the PIMO proxemic framework [5]—Position (3D pose estimation), Movement (activity recognition), Identity (user identification), and Orientation (contact-state classification)—plus multi-user tracking.

4.1 Dataset

4.1.1 Participants and Data Collection. We recruited 16 participants, comprising 8 male and 7 female participants (1 preferred not to disclose gender; age range 23–56 years); full participant demographics are provided in the Appendix. The study was conducted in a room equipped with the CapAura system (3×5 grid, $2.5 \text{ m} \times 1.5 \text{ m}$). Ground-truth 32-joint skeletal pose was captured at approximately 8 Hz using a Microsoft Azure Kinect DK [41], which has been validated as a reliable reference for body-pose studies [3, 34]. Capacitive data was sampled at approximately 12 Hz, and synchronized frames were accepted within a ± 50 ms tolerance window.

4.1.2 Action List. Participants performed activity types organized into four scenario groups:

- **Furniture Interaction Scenarios (F):** F1—lifting conductive furniture; F2—pulling conductive furniture.
- **No Furniture Scenarios (N):** N1—walking; N2—sitting on the floor; N3—lunging.
- **With Furniture Scenarios (W):** W1—walking near furniture; W2—sitting beside furniture; W3—lunging near furniture.
- **Furniture Hand Orientation Scenarios (O):** O1—no-touch, hand away from furniture surface; O2—hover, hand within close proximity without contact (2–5 cm); O3—on-touch, hand in direct contact with furniture surface.

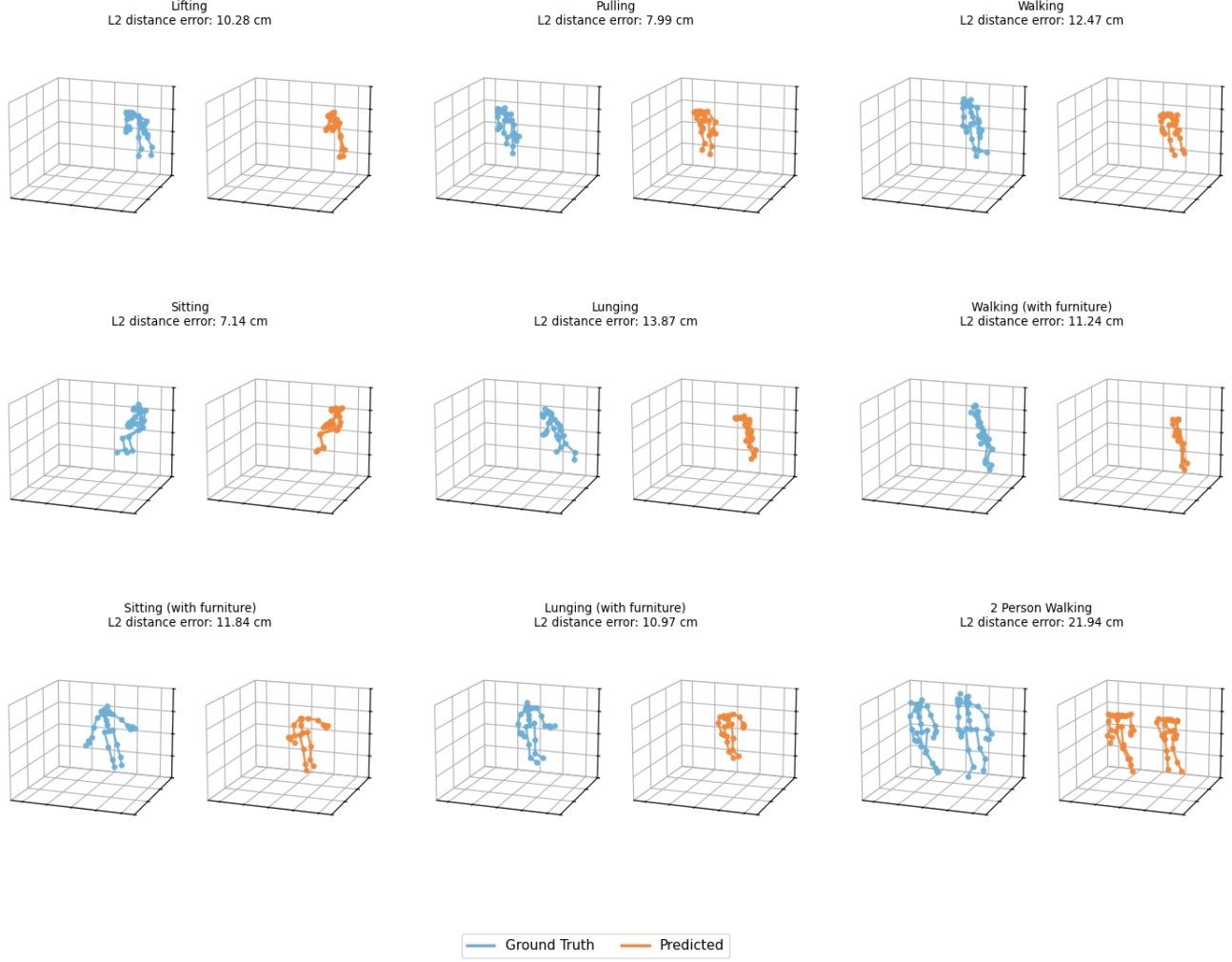


Figure 6: Predictions for 3D ambient pose estimation across action types.

4.1.3 Dataset Statistics and Splits. Each synchronized grid frame, together with the 29 frames preceding it, formed a 30-frame input sequence of shape (30, 3, 5). The full dataset was split *within session* into train / validation / test partitions at 70% / 15% / 15%, with 30-frame buffers between partitions to prevent temporal leakage. This yielded approximately 98,118 training samples, 16,351 validation samples, and 16,431 test samples.¹

4.2 Position: 3D Pose Estimation

We report Mean Per Joint Position Error (MPJPE)—the average Euclidean distance between predicted and ground-truth 3D joint coordinates across all 32 keypoints. Our Keypoint Detection Model achieves an overall MPJPE of **29.7 cm** on the test set (Table 1).

Furniture-present scenarios show lower error (28.8 cm) than those without furniture (32.1 cm), with the same pattern across all body regions (Table 2). Conductive furniture acts as a passive

Scenario	Val (cm)	Test (cm)
Overall	28.4	29.7
Furniture Present (F, W, O)	29.0	28.8
Without Furniture (N)	26.9	32.1

Table 1: MPJPE (cm) by scenario group. F, W, and O are furniture-present sessions; N is the No-Furniture condition.

electric-field extender [56], propagating our system’s field upward into the torso and arm space. Face/Head joints show the largest absolute error in both conditions (40.1 cm vs. 41.2 cm), as these joints are furthest from the floor-mounted sensing surface. Figure 6 visualises our best predicted skeletons per activity class.

Notably, our system operates at a coarse 3×5 grid resolution with 50 cm × 50 cm cells, which by itself imposes a localization

¹Our code and full dataset are available at <https://github.com/Sonanzhang/CapAura>.

Body Region	Furniture Present	Without Furniture	Overall
Torso	22.6	26.8	23.9
Arms	29.0	32.4	30.0
Lower Body	23.0	27.3	24.3
Face/Head	40.1	41.2	40.4

Table 2: Per-body-region MPJPE (cm) by furniture condition. Detailed aggregation and joint-group definitions are provided in the Appendix.

uncertainty of up to approximately 35 cm.² An MPJPE of 29.7 cm falls within this resolution-imposed uncertainty range, suggesting that sensing resolution has become a major limiting factor. A finer floor module grid would directly reduce this uncertainty and is the most direct path to improved pose estimation accuracy.

Compared to prior systems, our approach achieves competitive accuracy despite operating at significantly lower resolution. Prior capacitance-based infrastructure systems are predominantly limited to 2D localization: CapLoc [19] and CapFloor [10] report errors of 20–30 cm on flat grids; Wall++ [65] reports up to 140 cm on vertical surfaces. The pressure-based IntelligentCarpet [39] reports ≈ 30 cm error on a 12 $\times$ 12 grid for 3D pose estimation. Our system achieves comparable accuracy (29.7 cm) at a coarser 3 $\times$ 5 resolution, while additionally providing identity, movement, and orientation recognition.

4.3 Movement: Activity Recognition

The Classification Model achieves a test accuracy of **66.1%** across the eight movement classes, while accuracy in distinguishing furniture-present scenarios (F and W) from No-Furniture scenarios (N) at a broader binary level reaches **98.8%**. Figure 7 shows the full per-class confusion matrix.

4.4 Identity: User Identification

The Classification Model achieves a within-session test accuracy of **99.6%** for 15-participant identification (one of the 16 participants was excluded because their identity-calibration recording was incomplete). Per-action accuracy is consistently high across all eight classes (range: 99.2–100.0%), confirming that individual electric-field coupling signatures are stable across activity contexts within a single recording session. Because training and test data are drawn from the same recording session, this accuracy represents within-session performance; cross-session and cross-day identity generalization were not evaluated, and we discuss this limitation further in the Discussion.

4.5 Orientation: Contact-State Classification

We evaluate three-class contact-state classification (O1: No-touch, O2: Hover, O3: On-touch) on the held-out orientation test set. The system achieves an overall accuracy of **73.6%** (Table 3).

²Each floor module measures 50 cm $\times$ 50 cm. Within a module, the system can only assume the user stands at its center, so grid resolution alone introduces a localization error of at most half the module diagonal (≈ 35 cm), and ≈ 19 cm on average.

Table 3: Per-class accuracy (%) for contact-state orientation classification.

Class	Description	Accuracy (%)
O1	No-touch	72.4
O2	Hover	55.3
O3	On-touch	93.1

On-touch (O3) achieves the highest accuracy (93.1%), consistent with direct conductive contact producing a more distinctive capacitive event than either non-contact state. We examine the remaining No-touch/Hover ambiguity below.

Compared to vision-based touch sensing methods [37, 42, 43], which report on-touch accuracies of 94.9% (EgoTouch), 99.1% (ShadowTouch), and 98.0% (EclipseTouch), our 93.1% on-touch accuracy is competitive. However, these methods require users to remain relatively stationary and rely on mobile worn devices such as AR/VR headsets. In contrast, our system detects touch events while users move freely through the room—touching furniture in the course of natural activity rather than in controlled, static postures. We consider 93.1% on-touch accuracy under dynamic, unconstrained conditions a strong result for an infrastructure-based approach with no worn hardware.

4.6 Multi-User Tracking

To assess scalability beyond a single occupant, we evaluate the Two Person Detection Model on a two-person held-out test set (315 samples). The system achieves a combined MPJPE of **33.0 cm**, compared to **29.7 cm** for single-person tracking—a degradation of 3.3 cm attributable to the increased ambiguity of simultaneously attributing overlapping floor signals to two individuals.

Notably, a combined MPJPE of 33.0 cm falls within the *intimate zone* (<45 cm) as defined by Hall’s proxemics theory [26], indicating that the system can spatially resolve two users at a level of precision meaningful for proxemic interaction design.

The model was fine-tuned from the single-person checkpoint on only 1,605 two-person training samples, demonstrating that learned single-person capacitive representations transfer effectively to multi-occupant settings with minimal additional supervision.

4.7 Failure Modes

Position. One observed pose-estimation failure is *facing-direction ambiguity*: the model predicts a plausible skeleton at approximately the correct location but reverses the direction in which the body is facing. The representative current-model case in Appendix Figure 11 differs from the ground-truth facing direction by 179.7°. This ambiguity is consistent with front- and back-facing configurations producing similar observations on the coarse 3 $\times$ 5 grid.

Movement. Action-level errors are concentrated among activities with similar whole-body configurations in furniture-present settings. W3 (Lunging near furniture) is classified correctly in 41.6% of samples and is frequently confused with W1 (Walking near furniture; 20.5%) and F1 (Lifting; 17.1%). Appendix Figure 12 illustrates

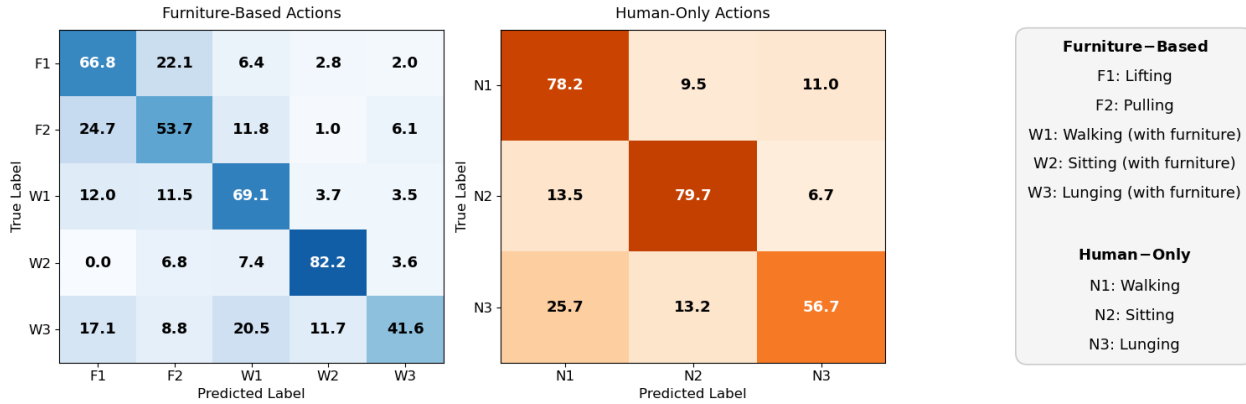


Figure 7: Confusion matrix for the action classification task.

how lunge depth and simultaneous hand–furniture interaction can make these segments resemble walking or lifting.

Orientation. Contact-state errors occur primarily at the boundary between No-touch and Hover: 27.6% of No-touch samples are classified as Hover, while 42.7% of Hover samples are classified as No-touch. In contrast, On-touch remains comparatively robust at 93.1%, indicating that the principal limitation is separating two weak non-contact states rather than detecting direct contact.

5 Application: Spatial Hotkey System

5.1 Concept

We demonstrate CapAura through our Spatial Hotkey System—a proof-of-concept application that turns conductive furniture surfaces into a set of personalized, body-triggered shortcuts. The system takes the calibrated heatmap output from the CapAura software pipeline as real-time input, allowing users to record and assign hotkeys to any capacitive event detected across the floor modules.

The concept is analogous to keyboard hotkeys: just as a key combination triggers a predefined command, a specific physical interaction with a piece of furniture triggers a user-defined action. Unlike keyboard shortcuts, however, Spatial Hotkeys require no dedicated hardware and no intentional button press—the furniture itself becomes the key. This is what distinguishes Spatial Hotkeys from prior room-scale triggering on instrumented surfaces [65]: they operate on everyday objects a person already owns, rather than on a purpose-built interactive surface. Users can define hotkeys in two ways: actively, by deliberately designating a specific interaction (e.g., touching the edge of a table to turn on the air conditioning), or passively, by embedding triggers into existing daily habits (e.g., reaching toward a bookshelf to automatically turn on a lamp).

5.2 Study Design

Each participant was provided with conductive furniture and invited to record three personal Spatial Hotkeys. During the recording phase, the participant performed a chosen interaction—such as touching a specific spot on a table or reaching toward a shelf—and assigned it to any action of their choice. The system captured the corresponding capacitive activation order as a hotkey reference

template. During the activation phase, the system continuously monitored the floor modules; when a recognized activation order was detected, the associated action was triggered.

This design intentionally leverages everyday furniture interactions as triggers, embedding hotkeys into existing habits without requiring the user to learn new gestures or operate dedicated devices.

5.3 User Study

To evaluate the Spatial Hotkey System’s usability and perceived value, we conducted a user study with $N = 16$ participants (P3–P18; 8 male, 7 female, 1 preferred not to disclose; age range 23–56, $M = 27.7$). Each participant recorded three personal hotkeys on the provided conductive furniture and used them in a series of interaction tasks. Participants then completed a 14-item questionnaire using 6-point Likert-scale items, followed by a semi-structured interview.³

5.3.1 Questionnaire Results. We tested whether each item differed significantly from the scale midpoint (3.5) using one-sample t -tests. To avoid conflating enthusiasm for the concept with judgments of how well the system actually worked, we report the items in two groups: *Novelty/Excitement* (Q6, Q13, Q14) and *Functional Performance* (the remaining 11 items). Per-participant group scores were computed as the mean of the constituent items (Table 4); full results for all 14 items are provided in the Appendix (Table 9).

Both groups significantly exceeded the neutral midpoint, though Novelty/Excitement was rated significantly higher than Functional Performance (paired $t(15) = 2.21, p = .043$)—participants’ enthusiasm for the concept partly outpaced their assessment of its current performance. Within the functional items, ease of use (Q11) was the highest-rated item overall, while the non-significant results for Q5 (felt like a natural body extension) and Q8 (hand orientation accuracy) indicate that the sense of embodiment was not yet fully established and that participants were realistically aware of the system’s recognition limits—consistent with our technical finding

³At the time of the study, the system was presented to participants under its earlier name, “Capacitive Skin”; the questionnaire and interview materials therefore use the word “skin” where this paper now uses “aura.” The system was renamed CapAura during revision.

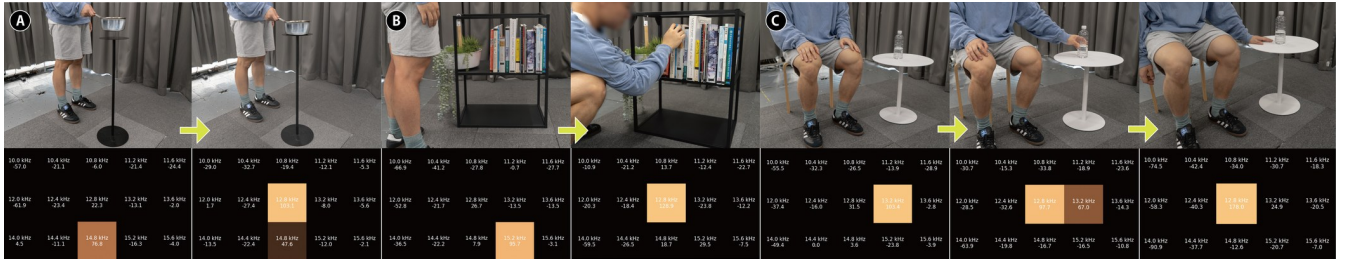


Figure 8: Common daily activities detected by the CapAura system. Changes in activated floor modules are recorded and used to trigger the Spatial Hotkey System: (A) cooking with a pot, (B) retrieving a book from a shelf, and (C) sitting and drinking at a table.

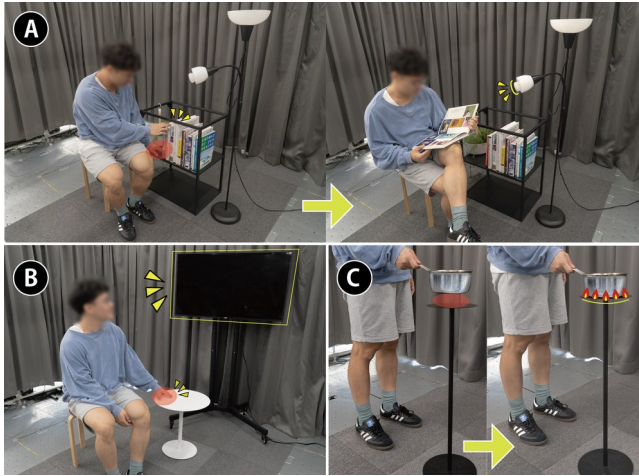


Figure 9: With our Spatial Hotkey System, a user can bind everyday furniture interactions to any action of their choice: (A) Proactive Lighting—retrieving a book automatically triggers a lamp; (B) Furniture as Remote—touching the conductive table toggles the television; (C) Intent Detection—hovering a pot over the stove activates the heating function.

that hover detection remains the most challenging classification task.

5.3.2 Interview Findings. Each brief semi-structured interview (approximately 5–10 minutes) was guided by four question areas: (1) whether the interaction felt more like controlling the system or being sensed by it; (2) how participants perceived the room-scale sensing metaphor; (3) the scenarios in which they would want to use the system; and (4) their preference between passive, habit-based triggering and active triggering, with follow-up questions where an answer invited elaboration. Interviews were audio-recorded and transcribed, and one researcher grouped similar responses. Given their brevity, we report these as findings rather than themes from a formal thematic analysis; they do not map one-to-one onto the question areas, as the metaphor question produced little substantive response while the scenarios question produced two:

Group / Item	M	SD	t(15)	Sig.
Novelty/Excitement (Q6, Q13, Q14)	4.79	0.93	5.58	$p < .001^{***}$
Q13: Exciting experience	4.94	1.06	5.41	$p < .001^{***}$
Q14: Highly inventive	4.88	1.09	5.06	$p < .001^{***}$
Q6: Room’s “skin” sensed my presence	4.56	1.21	3.51	$p < .01^{**}$
Functional Performance (11 items)	4.44	0.97	3.87	$p < .01^{**}$
Q11: Easy and uncomplicated	5.12	0.81	8.06	$p < .001^{***}$
Q8: Hand orientation felt accurate	4.06	1.39	1.62	$p = .126$
Q5: Natural extension of movement	3.75	1.53	0.65	$p = .523$

Table 4: Likert results by group ($N = 16$, scale 1–6, midpoint 3.5), with selected items shown per group. $^{***} p < .001$, $^{**} p < .01$. Full results in Table 9.

Furniture affordance shapes interaction attribution. Different furniture pieces led participants to different interaction framings—some felt they were “controlling” the system, others felt the room was “sensing” them. This distinction correlated with the physical affordance of the furniture: pieces that invite deliberate, discrete actions (tapping a table edge) evoked a sense of control, while interactions with more passive, continuous surfaces (reaching toward a bookshelf) evoked a sense of ambient awareness. The same underlying system thus produced fundamentally different interaction experiences depending on the furniture involved.

Bed scenario as compelling use case. The strongest consensus was the bedroom scenario—using proximity to the bed frame to trigger a user-defined action such as dimming lights or turning off devices, without any intentional act. Participants framed this as qualitatively different from voice commands because it required no deliberate interaction.

Interaction agency for users with limited mobility. Several participants raised a use scenario we had not anticipated when preparing the guide: the potential value for users with limited mobility. P18 noted that furniture-based ambient sensing could significantly

expand interaction agency for users who have difficulty moving—elderly users or people with disabilities who spend extended time in bed or in otherwise restricted settings could gain meaningful control over their environment with minimal effort and without any additional hardware.

Misfire tolerance as the key design variable. Participants distinguished between low-consequence actions (turning on a light) where passive triggering was acceptable, and high-consequence actions (activating a stove) where an explicit multi-step activation was preferred—such as combining two sequential touches across different furniture pieces to form a compound hotkey. This suggests a tiered interaction model: passive hotkeys for ambient actions, active confirmation for critical ones. Interestingly, P14 proposed a directional asymmetry: preferring passive triggering for turning off devices, but requiring active confirmation for turning them on, reflecting a nuanced risk calculus based on the consequence of false activation rather than the action type alone.

6 Discussion

6.1 From Capacitive Sensing to Proxemic Infrastructure

CapAura should be understood not as a substitute for precision body tracking, but as a proxemic infrastructure: a room-scale substrate for recognizing coarse spatial relationships without instrumenting the user. Whereas wearable systems such as MobilePoser [61] and SeamPose [63] target fine-grained joint motion, CapAura explores interactions based on where a person is, how they move, and whether they approach or touch surrounding furniture.

The Spatial Hotkey study illustrates one value of this configuration. Participants repeatedly identified bed-proximity lighting as compelling because approaching the bed or reaching toward a bedside surface could coincide with an action they already perform. Unlike a voice command, it need not introduce a separate command step; unlike scheduled automation, it remains contingent on present behavior. We describe this alignment as *behavioral congruence*: matching an existing behavior to a response the user would plausibly want at that moment.

The questionnaire item stating that the room’s “skin” sensed the participant’s presence and intentions was rated above the neutral midpoint ($M = 4.56$, $p = .003$). Although this does not establish long-term acceptance, it suggests that participants perceived the intended shift from using a dedicated device to interacting through the environment. This opportunity is limited to settings containing conductive structures: CapAura does not make every object an input surface, but shows how existing metal bed frames, desks, or shelving can join a shared field without object-level electronics.

6.2 Designing Around Recognition Uncertainty

Participants found the system easy to use while not being fully convinced of its accuracy. Ease of use was the highest-rated item (Q11, $M = 5.12$, $p < .001$), whereas perceived hand-orientation accuracy (Q8, $M = 4.06$, $p = .126$) and embodiment (Q5, $M = 3.75$, $p = .523$) did not significantly exceed the midpoint. Participants found the interaction easy while remaining more reserved about its

accuracy and naturalness, consistent with the technical difficulty of fine-grained movements and the No-touch/Hover boundary.

These observations motivate a hypothesis: low interaction effort may sometimes compensate for imperfect recognition when errors have limited consequences. The present study neither establishes that effort matters more than accuracy nor measures long-term tolerance of false activations. It does show that participants reasoned about consequence: passive activation was acceptable for low-stakes outcomes such as lighting, while activating a stove called for a deliberate sequence or confirmation.

This distinction suggests a consequence-aware interaction model. Low-stakes Spatial Hotkeys can trigger passively if activation is visible and easily undone. Higher-risk actions should require stronger evidence through a compound sequence, explicit confirmation, or both. Interaction effort should therefore scale with false-activation cost, treating classifier uncertainty as a design parameter rather than mapping every output directly to an action.

6.3 Scope and Limitations

All recordings came from one configured room, and train, validation, and test partitions were drawn from the same sessions, with temporal buffers used to reduce leakage. The results therefore demonstrate within-session recognition rather than cross-room or cross-day generalization. This is especially important for the 99.6% identity result because capacitive signatures may vary with hydration, clothing, footwear, humidity, and furniture placement. Deployment requires cross-session, cross-day, and cross-room evaluation.

The sensing configuration also bounds what can be inferred. The sparse 3×5 grid constrains spatial resolution, and similar body configurations or weak non-contact states remain difficult to distinguish. Greater electrode density may provide more spatial information, but also adds channels and may reduce acquisition rate or require higher-bandwidth hardware. CapAura further depends on conductive furniture; adding coatings or embedded structures to non-conductive objects would no longer preserve the same unmodified-furniture premise.

The multi-user and user-study results are initial evidence. Multi-user tracking covered two people and 315 test samples, not larger groups, crossing paths, or identity assignment under overlapping signals. The Spatial Hotkey study involved 16 participants in a short, single-room experience, and did not measure false activations, recovery, or acceptance over time. These outcomes require direct longitudinal measurement.

6.4 Toward Reliable Deployment

Camera-free sensing does not eliminate privacy concerns: inferred identity, movement, and routines remain sensitive even without images. Deployment should communicate when sensing is active, support opt-in and temporary disabling, minimize raw-signal retention, and keep inference local where possible, especially in shared rooms.

Three practical directions follow. Longitudinal operation needs dynamic baseline calibration as humidity, furniture placement, and

hardware change, potentially during unoccupied periods. Applications should expose recognition state and provide recovery proportional to consequence, from undo for lighting to confirmation for safety-critical actions. Finally, deployments should begin where conductive infrastructure and low-effort interaction are plausible—such as care environments with metal bed frames or workplaces with steel desks—while evaluating transfer, calibration burden, and social acceptance in situ. The immediate priority is reliable and accountable operation beyond the controlled room.

7 Conclusion

We have introduced CapAura, a passive capacitive infrastructure that envelops everyday living spaces in a proxemic interactive field—without cameras, worn devices, or furniture modification. By treating conductive furniture as electric-field extenders, the system achieves room-scale volumetric sensing across all four PIMO dimensions from a single floor-to-ceiling substrate. Our user study ($N = 16$) suggests that furniture-based proxemic interaction is perceived as both functionally effective and novel, and our Spatial Hotkey System shows that everyday interactions—retrieving a book, placing a pot, sitting at a table—can seamlessly trigger user-defined actions. We believe CapAura offers a foundation for future smart environments in which the sensing infrastructure is already present in every home, waiting only to be connected to the floor beneath it.

References

- [1] Karan Ahuja, Paul Strelly, and Christian Holz. 2021. TouchPose: hand pose prediction, depth estimation, and touch classification from capacitive images. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 997–1009.
- [2] Fadi Al Machot, Mohammed R. Elkobaisi, and Kyandoghene Kyamakya. 2020. Zero-shot human activity recognition using non-visual sensors. *Sensors* 20, 3 (2020), 825.
- [3] Justin Amadeus Albert, Victor Owolabi, Arnd Gebel, Clemens Markus Brahm, Urs Granacher, and Bert Arnrich. 2020. Evaluation of the Pose Tracking Performance of the Azure Kinect and Kinect v2 for Gait Analysis in Comparison with a Gold Standard: A Pilot Study. *Sensors* 20, 18 (2020), 5104. doi:10.3390/s20185104
- [4] Atika Arshad, Sheraz Khan, AHM Zahirul Alam, Ahmad Fadzil Ismail, and Rumana Tasnim. 2017. Capacitive proximity floor sensing system for elderly tracking and fall detection. In *2017 IEEE 4th international conference on smart instrumentation, measurement and application (ICSIMA)*. IEEE, 1–5.
- [5] Till Ballendat, Nicolai Marquardt, and Saul Greenberg. 2010. Proxemic interaction: designing for a proximity and orientation-aware environment. In *ACM International Conference on Interactive Tabletops and Surfaces*. 121–130.
- [6] Sizhen Bian, Mengxi Liu, Bo Zhou, Paul Lukowicz, and Michele Magno. 2024. Body-area capacitive or electric field sensing for human activity recognition and human-computer interaction: A comprehensive survey. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 8, 1 (2024), 1–49.
- [7] Sizhen Bian, Siyu Yuan, Vitor Fortes Rey, and Paul Lukowicz. 2022. Using human body capacitance sensing to monitor leg motion dominated activities with a wrist worn device. In *Sensor-and Video-Based Activity and Behavior Computing: Proceedings of 3rd International Conference on Activity and Behavior Computing (ABC 2021)*. Springer, 81–94.
- [8] Alan Bränzel, Christian Holz, Daniel Hoffmann, Dominik Schmidt, Marius Knaust, Patrick Lühne, René Meusel, Stephan Richter, and Patrick Baudisch. 2013. GravitySpace: tracking users and their poses in a smart room using a pressure-sensing floor. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 725–734.
- [9] Andreas Braun, Sebastian Frank, Martin Majewski, and Xiaofeng Wang. 2015. CapSeat: capacitive proximity sensing for automotive activity recognition. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 225–232.
- [10] Andreas Braun, Henning Heggen, and Reiner Wichert. 2012. CapFloor—a flexible capacitive indoor localization system. In *Evaluating AAL Systems Through Competitive Benchmarking. Indoor Localization and Tracking: International Competition, EvAAL 2011, Competition in Valencia, Spain, July 25–29, 2011, and Final Workshop in Lecce, Italy, September 26, 2011. Revised Selected Papers 1*. Springer, 26–35.
- [11] Andreas Braun, Stephan Neumann, Sönke Schmidt, Reiner Wichert, and Arjan Kuijper. 2014. Towards interactive car interiors: the active armrest. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. 911–914.
- [12] Wenqiang Chen, Yexin Hu, Wei Song, Yingcheng Liu, Antonio Torralba, and Wojciech Matusik. 2024. Cavatar: Real-time human activity mesh reconstruction via tactile carpets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 4 (2024), 1–24.
- [13] Gabe Cohn, Sidhant Gupta, Tien-Jui Lee, Dan Morris, Joshua R Smith, Matthew S Reynolds, Desney S Tan, and Shwetak N Patel. 2012. An ultra-low-power human body motion sensor using static electric field sensing. In *Proceedings of the 2012 ACM conference on ubiquitous computing*. 99–102.
- [14] Gabe Cohn, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. Humantenna: using the body as an antenna for real-time whole-body interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1901–1910.
- [15] Diane J Cook, Juan C Augusto, and Vikramaditya R Jakkula. 2009. Ambient intelligence: Technologies, applications, and opportunities. *Pervasive and mobile computing* 5, 4 (2009), 277–298.
- [16] Audrey Desjardins and Ron Wakkary. 2016. Living in a prototype: A reconfigured space. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 5274–5285.
- [17] Diligent Inc. 2025. *Analog Discovery 3 Specifications*. <https://diligent.com/reference/test-and-measurement/analog-discovery-3/specifications>
- [18] Guoguang Du, Kai Wang, Shiguo Lian, and Kaiyong Zhao. 2021. Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review. *Artificial Intelligence Review* 54, 3 (2021), 1677–1734.
- [19] Nathaniel Faulkner, Baden Parr, Fakhrol Alam, Mathew Legg, and Serge Demidenko. 2020. CapLoc: Capacitive sensing floor for device-free localization and fall detection. *IEEE Access* 8 (2020), 187353–187364.
- [20] Md Tahmid Hasan Fuad, Awal Ahmed Fime, Delowar Sikder, Md Akil Raihan Iftee, Jakaria Rabbi, Mabrook S Al-Rakhami, Abdu Gumaei, Ovishake Sen, Mohtasim Fuad, and Md Nazrul Islam. 2021. Recent advances in deep learning techniques for face recognition. *IEEE Access* 9 (2021), 99112–99142.
- [21] Bobby George, Hubert Zangl, Thomas Brettertklieber, and Georg Brasseur. 2010. A combined inductive–capacitive proximity sensor for seat occupancy detection. *IEEE transactions on instrumentation and measurement* 59, 5 (2010), 1463–1470.
- [22] Nan-Wei Gong, Steve Hodges, and Joseph A Paradiso. 2011. Leveraging conductive inkjet technology to build a scalable and versatile surface for ubiquitous sensing. In *Proceedings of the 13th international conference on Ubiquitous computing*. 45–54.
- [23] Saul Greenberg, Nicolai Marquardt, Till Ballendat, Rob Diaz-Marino, and Miaosen Wang. 2011. Proxemic interactions: the new ubicomp? *interactions* 18, 1 (2011), 42–50.
- [24] Tobias Grosse-Puppenthal, Yannick Berghoefer, Andreas Braun, Raphael Wimmer, and Arjan Kuijper. 2013. OpenCapSense: A rapid prototyping toolkit for pervasive interaction using capacitive sensing. In *2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 152–159.
- [25] Fuqiang Gu, Mu-Huan Chung, Mark Chignell, Shahrokh Valaee, Baoding Zhou, and Xue Liu. 2021. A survey on deep learning for human activity recognition. *ACM Computing Surveys (CSUR)* 54, 8 (2021), 1–34.
- [26] Edward T Hall, Ray L Birdwhistell, Bernhard Bock, Paul Bohannon, A Richard Diebold Jr, Marshall Durbin, Munro S Edmonson, JL Fischer, Dell Hymes, Solon T Kimball, et al. 1968. Proxemics [and comments and replies]. *Current anthropology* 9, 2/3 (1968), 83–108.
- [27] Ju Han and Bir Bhanu. 2005. Human activity recognition in thermal infrared imagery. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*. IEEE, 17–17.
- [28] Chris Harrison, Munehiko Sato, and Ivan Poupyrev. 2012. Capacitive fingerprinting: exploring user differentiation by sensing electrical properties of the human body. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. 537–544.
- [29] Sabera Hoque, Md Yasir Arafat, Shuxiang Xu, Ananda Maiti, and Yuchen Wei. 2021. A comprehensive review on 3D object detection and 6D pose estimation with deep learning. *IEEE Access* 9 (2021), 143746–143770.
- [30] Hiroshi Ishii, Craig Wisneski, Scott Brave, Andrew Dahley, Matt Gorbett, Brygg Ullmer, and Paul Yarin. 1998. ambientROOM: integrating ambient media with architectural space. In *CHI 98 conference summary on Human factors in computing systems*. 173–174.
- [31] Minhuk Jung and Seokho Chi. 2020. Human activity classification based on sound recognition and residual convolutional neural network. *Automation in Construction* 114 (2020), 103177.
- [32] Deniz Kasap, Taraneh Aminosharieh Najafi, Jérôme Paul Rémy Thevenot, Jonathan Dan, Stefano Albini, and David Atienza. 2025. VersaPants: A Loose-Fitting Textile Capacitive Sensing System for Lower-Body Motion Capture. arXiv:2511.16346 [cs.HC] <https://arxiv.org/abs/2511.16346>

- [33] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations (ICLR 2015)*. <http://arxiv.org/abs/1412.6980>
- [34] Gregorij Kurillo, Evan Hemingway, Mu-Lin Cheng, and Louis Cheng. 2022. Evaluating the accuracy of the azure kinect and kinect v2. *Sensors* 22, 7 (2022), 2469.
- [35] Piyawat Lertvittayakumjorn, Shanqing Cai, Billy Dou, Cedric Ho, and Shumin Zhai. 2024. Can Capacitive Touch Images Enhance Mobile Keyboard Decoding?. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–17.
- [36] Yadong Li, Shuning Wang, Yongjian Fu, Justin Chen, Xingyu Chen, Ju Ren, Xinyu Zhang, Akshay Gadre, and Ke Sun. 2025. UltraPoser: Pushing the Limits of IMU-based Full-Body Pose Estimation with Ultrasound Sensing on Consumer Wearables. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–15.
- [37] Chen Liang, Xutong Wang, Zisu Li, Chi Hsia, Mingming Fan, Chun Yu, and Yuanchun Shi. 2023. Shadowtouch: Enabling free-form touch-based hand-to-surface interaction with wrist-mounted illuminant by shadow projection. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–14.
- [38] Yu Lu, Dian Ding, Hao Pan, Yijie Li, Juntao Zhou, Yongjian Fu, Yongzhao Zhang, Yi-Chao Chen, and Guangtao Xue. 2024. Handpad: Make your hand an on-the-go writing pad via human capacitance. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–16.
- [39] Yiyue Luo, Yunzhu Li, Michael Foshey, Wan Shou, Pratyusha Sharma, Tomás Palacios, Antonio Torralba, and Wojciech Matusik. 2021. Intelligent carpet: Inferring 3d human pose from tactile signals. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 11255–11265.
- [40] Nicolai Marquardt, Robert Diaz-Marino, Sebastian Boring, and Saul Greenberg. 2011. The proximity toolkit: prototyping proxemic interactions in ubiquitous computing ecologies. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 315–326.
- [41] Microsoft. 2019. Azure Kinect DK. <https://azure.microsoft.com/en-us/products/kinect-dk/>. Accessed: May 16, 2025.
- [42] Vimal Mollyn, Nathan DeVrio, and Chris Harrison. 2025. EclipseTouch: Touch Segmentation on Ad Hoc Surfaces using Worn Infrared Shadow Casting. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–13.
- [43] Vimal Mollyn and Chris Harrison. 2024. EgoTouch: On-body touch input using AR/VR headset cameras. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–11.
- [44] Adiyani Mujibiya and Jun Rekimoto. 2013. Mirage: exploring interaction modalities using off-body static electric field sensing. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 211–220.
- [45] Vinod Nair and Geoffrey E. Hinton. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, Johannes Fürnkranz and Thorsten Joachims (Eds.). Omnipress, 807–814.
- [46] Frazer Noble, Muqing Xu, and Fakhru Alam. 2023. Static hand gesture recognition using capacitive sensing and machine learning. *Sensors* 23, 7 (2023), 3419.
- [47] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, Curran Associates, Inc., 8024–8035. <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- [48] Abhisek Ray, Maheshkumar H Kolekar, Raman Balasubramanian, and Adel Hafiane. 2023. Transfer learning enhanced vision-based human activity recognition: A decade-long analysis. *International Journal of Information Management Data Insights* 3, 1 (2023), 100142.
- [49] Fariba Sadri. 2011. Ambient intelligence: A survey. *ACM Computing Surveys (CSUR)* 43, 4 (2011), 1–66.
- [50] Munechiko Sato, Ivan Poupyrev, and Chris Harrison. 2012. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 483–492.
- [51] Yi Sun, Xiaogang Wang, and Xiaoou Tang. 2014. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1891–1898.
- [52] Xinyao Tang and Soumyajit Mandal. 2019. Indoor occupancy awareness and localization using passive electric field sensing. *IEEE Transactions on Instrumentation and Measurement* 68, 11 (2019), 4535–4549.
- [53] Thiago Teixeira, Gershon Dublon, and Andreas Savvides. 2010. A survey of human-sensing: Methods for detecting presence, count, location, track, and identity. *Comput. Surveys* 5, 1 (2010), 59–69.
- [54] Kiryu Tsujita, Takatoshi Yoshida, Kohei Kobayashi, Nobuhisa Hanamitsu, Kanghong Zhong, Arata Horie, and Kouta Minamizawa. 2025. HaptoRoom: Designing Spatial and Reconfigurable Haptic Experience using Vibrotactile Floor Interface. In *Proceedings of the Nineteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–11.
- [55] Elahe Vahdani and Yingli Tian. 2022. Deep learning-based action detection in untrimmed videos: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 4 (2022), 4302–4320.
- [56] Miika Valtonen, Jaakko Maentausta, and Jukka Vanhala. 2009. Tiletrack: Capacitive human tracking using floor tiles. In *2009 IEEE international conference on pervasive computing and communications*. IEEE, 1–10.
- [57] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017), 5998–6008.
- [58] Jinbao Wang, Shujie Tan, Xiantong Zhen, Shuo Xu, Feng Zheng, Zhenyu He, and Ling Shao. 2021. Deep 3D human pose estimation: A review. *Computer Vision and Image Understanding* 210 (2021), 103225.
- [59] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st annual international conference on mobile computing and networking*. 65–76.
- [60] Andrew D Wilson. 2005. PlayAnywhere: a compact interactive tabletop projection-vision system. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*. 83–92.
- [61] Vasco Xu, Chenfeng Gao, Henry Hoffmann, and Karan Ahuja. 2024. Mobileposer: Real-time full-body pose estimation and 3d human translation from imus in mobile consumer devices. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–11.
- [62] Takatoshi Yoshida, Junichi Ogawa, Kyung Yun Choi, Sanad Bushnaq, Ken Nakagaki, and Hiroshi Ishii. 2021. InDepth: Force-based interaction with objects beyond a physical barrier. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–6.
- [63] Tianhong Catherine Yu, Manru Mary Zhang, Peter He, Chi-Jung Lee, Cassidy Cheesman, Saif Mahmud, Ruidong Zhang, Francois Guimbretiere, and Cheng Zhang. 2024. SeamPose: Repurposing Seams as Capacitive Sensors in a Shirt for Upper-Body Pose Tracking. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–13.
- [64] Adriaan S Zeeman, Marthinus J Booyens, Giuseppe Ruggeri, and Bruno Laganá. 2013. Capacitive seat sensors for multiple occupancy detection using a low-cost setup. In *2013 IEEE International Conference on Industrial Technology (ICIT)*. IEEE, 1228–1233.
- [65] Yang Zhang, Chouchang Yang, Scott E Hudson, Chris Harrison, and Alanson Sample. 2018. Wall++ room-scale interactive and context-aware sensing. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–15.
- [66] Junyi Zhao, Pornthep Preechayasomboon, Tyler Christensen, Amirhossein H Memar, Zhenzhen Shen, Nicholas Colonnese, Michael Khbeis, and Mengjia Zhu. 2024. TouchpadAnyWear: textile-integrated tactile sensors for multimodal high spatial-resolution touch inputs with motion artifacts tolerance. In *Proceedings of the 37th annual ACM symposium on user interface software and technology*. 1–14.
- [67] Ce Zheng, Wenhan Wu, Chen Chen, Taojiannan Yang, Sijie Zhu, Ju Shen, Nasser Kehtarnavaz, and Mubarak Shah. 2023. Deep learning-based human pose estimation: A survey. *Comput. Surveys* 56, 1 (2023), 1–37.
- [68] Junyi Zhu, Tianyu Xu, Jiayu Wang, Emily Guan, JaeYoung Moon, Stiven Morvan, D Shin, Andrea Colaço, Stefanie Mueller, Karan Ahuja, et al. 2025. EI-Lite: Electrical Impedance Sensing for Micro-gesture Recognition and Pinch Force Estimation. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

A Capacitive Modeling Derivation

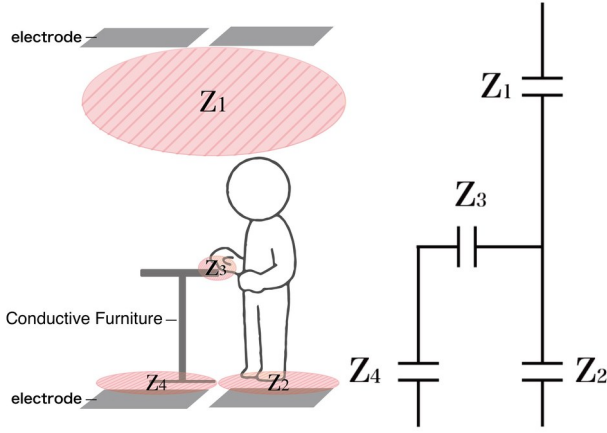


Figure 10: Principles of capacitance-based human sensing.

We define four impedances in the equivalent circuit (Figure 10): Z_1 (person–ceiling), Z_2 (person–floor), Z_3 (person–furniture), and Z_4 (furniture–floor). Each behaves as a parallel-plate capacitor:

$$C = \epsilon \frac{S}{d}$$

where ϵ is the permittivity of the medium, S the effective plate area, and d the distance between plates. As a person moves, S and d change across all four impedances simultaneously—for example, Z_1 varies with the height of the person’s raised limbs, Z_2 with foot contact area during walking, Z_3 with how close the hand is to a furniture surface, and Z_4 with the coupling between furniture legs and the floor electrode. These coordinated changes produce a unique multi-channel signal pattern for each body configuration.

The received amplitude depends on the total impedance of the propagation path between the floor and ceiling electrodes. When the user is not touching a conductive object, the signal travels through Z_1 and Z_2 in series:

$$Z_{\text{all}} = Z_1 + Z_2.$$

Because the ceiling is typically more than 1 m above the user while the foot-to-floor distance is around 1 cm, most signal variation is dominated by foot–floor coupling Z_2 .

When the user touches conductive furniture, an additional parallel path is formed through Z_3 and Z_4 , yielding:

$$Z_{\text{all}} = Z_1 + \frac{1}{\frac{1}{Z_3 + Z_4} + \frac{1}{Z_2}}.$$

This additional path reduces the effective impedance and produces a distinct change in received amplitude, allowing the system to distinguish contact events.

B Hardware Construction and Training Details

Floor Module. Each floor module consists of a conductive metal plate with a central threaded hole. A metal screw is fastened through this hole, establishing an electrical contact point on the underside

of the plate. A wire is attached to the screw and routed beneath the floor surface to the transmitter, keeping all wiring hidden. Across the full 3×5 grid, only 15 individual wires are required—one per module—minimizing wiring complexity. Electrical tape is applied at contact points between adjacent plates to prevent short circuits.

Ceiling Module. The ceiling module consists of two sheets of consumer-grade aluminum foil serving as the receiving antenna, sized to cover sufficient area for reliable signal reception (each sheet measures approximately 50×100 cm in our prototype). The foil is insulated on the back side and mounted onto an existing ceiling beam without any structural modification. A single wire connects it to the Analog Discovery 3 receiver.

Transmitter. The transmitter is implemented using an FPGA-based signal generation circuit. It generates up to 25 sinusoidal signals with carrier frequencies ranging from 10 kHz to just under 20 kHz, spaced 400 Hz apart, with a measured frequency error of no more than 8 Hz. Each output channel is assigned to one floor module.

Receiver. The receiver is implemented using the Analog Discovery 3 [17], connected to the ceiling antenna. It acquires the superimposed waveform and separates each floor module’s response by analyzing its frequency components.

Training Details. All models were implemented in PyTorch [47] and trained on an NVIDIA A100 GPU using the Adam optimizer [33] with learning rate 1×10^{-3} , weight decay 1×10^{-5} , for 50 epochs with batch size 40.

C Evaluation Aggregation Details

C.1 Scenario Aggregation

The Position evaluation uses the same scenario grouping in both MPJPE tables. *Furniture Present* pools Furniture Interaction (F), With Furniture (W), and Furniture Hand Orientation (O) sessions; *Without Furniture* contains the No Furniture (N) sessions. Overall results pool all four scenario groups. Table 5 reports the corresponding sample counts. All means are micro-averages over the included samples and joints rather than unweighted averages of scenario or body-region means.

Condition	Validation	Test
Furniture Present (F, W, O)	11,646	11,703
Without Furniture (N)	4,705	4,728
Overall	16,351	16,431

Table 5: Sample counts used for the condition-level and body-region Position results.

C.2 Body-Region Definitions

The 32 Azure Kinect joints are partitioned into four non-overlapping anatomical regions (Table 6). The regions jointly cover every predicted joint exactly once. Region-level MPJPE is computed by pooling the Euclidean errors of all joints assigned to that region across all samples in the corresponding condition.

Region	Included joints
Torso	Pelvis, spine navel, spine chest, and neck (indices 0–3)
Arms	Left and right clavicles, shoulders, elbows, wrists, hands, hand tips, and thumbs (indices 4–17)
Lower Body	Left and right hips, knees, ankles, and feet (indices 18–25)
Face/Head	Head, nose, left and right eyes, and left and right ears (indices 26–31)

Table 6: Joint assignments used for the per-body-region MPJPE analysis.

D Failure-Mode Cases

D.1 Position: Facing-Direction Ambiguity

Figure 11 shows a representative failure from the current Keypoint Detection Model’s held-out test predictions. Although the predicted skeleton remains structurally plausible and occupies approximately the same room region, its facing direction is reversed. Facing direction is visualized as the horizontal direction perpendicular to the left–right shoulder axis. In this example, the predicted and ground-truth directions differ by 179.7° , producing an MPJPE of 42.7 cm.

D.2 Movement: Lunging Confusions

Figure 12 compares three sustained 60-frame segments whose ground-truth label is W3 (Lunging near furniture). For each segment, we identify the consecutive nine-frame interval with the lowest mean pelvis-to-floor height and visualize its average pose alongside the majority classifier output over the full segment. Height is measured after anchoring the lowest foot joint of the average pose to the floor plane. The resulting heights are 57.9 cm for the correctly classified segment, 67.7 cm for the segment classified as Walking, and 68.4 cm for the segment classified as Lifting. Both confused segments therefore exhibit shallower sustained postures than the correctly classified lunge; in addition, the participant in the Lifting-confused segment places both hands at the conductive furniture, providing a lifting-like interaction cue.

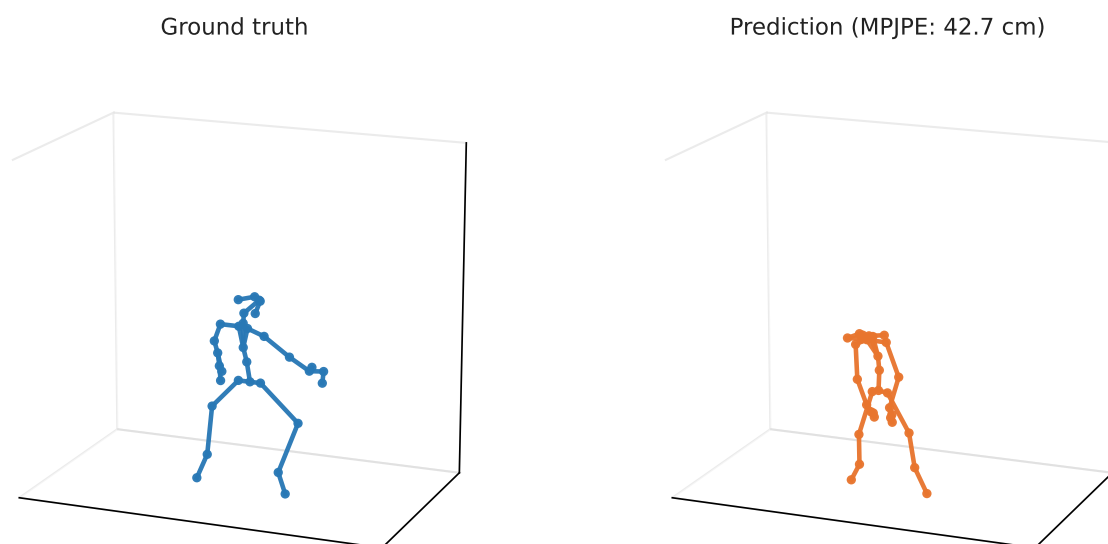
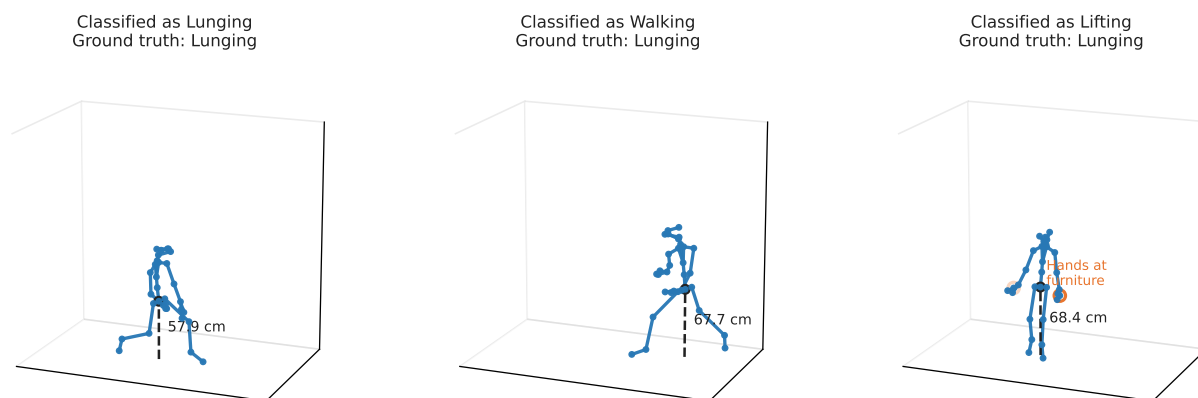


Figure 11: Representative current-model 3D pose-estimation failure. The ground-truth skeleton is shown in blue and the prediction in orange under identical axes and viewing angles; each skeleton is vertically aligned so that its lowest foot joint lies on the floor plane.



Each panel shows the average pose in the lowest consecutive 9-frame interval within its 60-frame window.

Figure 12: Representative movement-classification outcomes for three ground-truth Lunging segments. Each panel shows the average pose over the lowest consecutive nine-frame interval within its 60-frame window; orange circles highlight the hands positioned at the furniture in the segment classified as Lifting.

E Participant Demographics

The user study and interviews recruited 16 participants (P3–P18). Ages ranged from 23 to 56 years ($M = 27.7$). Gender was self-reported, with one participant electing not to disclose. Tables 7 and 8 report the aggregate composition and per-participant breakdown, respectively.

Attribute	Value
Participants (N)	16 (P3–P18)
Gender: male	8
Gender: female	7
Gender: prefer not to disclose	1
Age range	23–56 years
Age mean (M)	27.7 years
Education: postgraduate (Master’s/PhD)	13
Education: bachelor’s	3

Table 7: Aggregate participant demographics for the Spatial Hotkey user study and interviews ($N = 16$).

Participant	Gender	Age	Education
P3	Prefer not to say	25	Postgraduate
P4	Male	28	Postgraduate
P5	Male	23	Bachelor’s
P6	Female	26	Postgraduate
P7	Female	25	Postgraduate
P8	Female	28	Postgraduate
P9	Male	25	Bachelor’s
P10	Female	28	Postgraduate
P11	Male	26	Postgraduate
P12	Female	26	Postgraduate
P13	Male	25	Postgraduate
P14	Male	56	Postgraduate
P15	Female	27	Bachelor’s
P16	Male	23	Postgraduate
P17	Female	25	Postgraduate
P18	Male	27	Postgraduate

Table 8: Per-participant demographics for the user study and interviews ($N = 16$, P3–P18). “Postgraduate” denotes a Master’s or PhD degree.

F Questionnaire Results

All items used a 6-point Likert scale (1 = Strongly Disagree, 6 = Strongly Agree; midpoint = 3.5). One-sample t -tests were conducted against the scale midpoint ($N = 16$). *** $p < .001$, ** $p < .01$, * $p < .05$, ns = not significant. Item wording is reproduced verbatim as presented to participants, under the system’s earlier name (“Capacitive Skin”).

G Interview Guide

Each brief post-study interview (approximately 5–10 minutes per participant) was guided by the following four question areas. Follow-up questions were asked where a participant’s answer invited elaboration.

- (1) **Control vs. being sensed:** During the interaction, did it feel more like you were controlling the system, or more like the room was sensing you?
- (2) **Skin metaphor:** How did you perceive the idea of the room having a sensing “skin”—a layer over the floor and furniture that is aware of your presence and intentions?
- (3) **Desired scenarios:** In which situations or places in your daily life would you want to use this kind of interaction?
- (4) **Passive vs. active triggering:** Would you prefer interactions that trigger passively from your existing habits (e.g., naturally placing a hand on the table), or explicit activations requiring deliberate steps (e.g., touching several distinct spots in sequence)?

Interviews were audio-recorded and transcribed. One researcher grouped similar responses across participants; the recurring and notable points are reported as the four findings in the paper. The second question area produced little substantive response and is therefore not reported as a separate finding, while the third produced two.

Item	Question	M	SD	t	p	Sig.
Q1	I felt in control of the digital system through my interaction with the furniture.	4.69	1.20	3.97	.001	**
Q2	The system responded as I expected when I touched or hovered near the furniture.	4.62	1.09	4.14	.001	***
Q3	The way the interaction was triggered felt natural and intentional.	4.44	1.59	2.36	.032	*
Q4	The transition from passive furniture to active controller felt continuous.	4.38	1.36	2.57	.021	*
Q5	The interaction felt like a natural extension of my physical movements.	3.75	1.53	0.65	.523	ns
Q6	It felt as if the room's "skin" was sensing my presence and intentions.	4.56	1.21	3.51	.003	**
Q7	The digital response felt like it belonged to the physical action I just performed.	4.62	1.26	3.58	.003	**
Q8	The system's detection of my hand orientation felt accurate.	4.06	1.39	1.62	.126	ns
Q9	The timing of the system's response felt appropriate and instantaneous.	4.19	1.11	2.48	.025	*
Q10	The sensitivity of the furniture surface felt consistent with my expectations.	4.25	1.18	2.54	.023	*
Q11	The interaction with the system was easy and uncomplicated.	5.12	0.81	8.06	<.001	***
Q12	The system's feedback and operation logic were clear and not confusing.	4.75	1.44	3.48	.003	**
Q13	Interacting with this system was an exciting experience.	4.94	1.06	5.41	<.001	***
Q14	The system and its interaction method felt highly inventive and leading edge.	4.88	1.09	5.06	<.001	***

Table 9: Full questionnaire results ($N = 16$, scale 1–6, midpoint 3.5).