

DOBI: Dynamic Opportunistic Body Input via Spare Joint Recruitment for Hands-Free XR

Rachel Kim
School of Computing, KAIST
Daejeon, Republic of Korea
rachel02@kaist.ac.kr

Xun Qian
Google
San Jose, California, USA
me@xun-qian.com

Sang Ho Yoon
Graduate School of Culture
Technology, KAIST
Daejeon, Republic of Korea
sangho@kaist.ac.kr

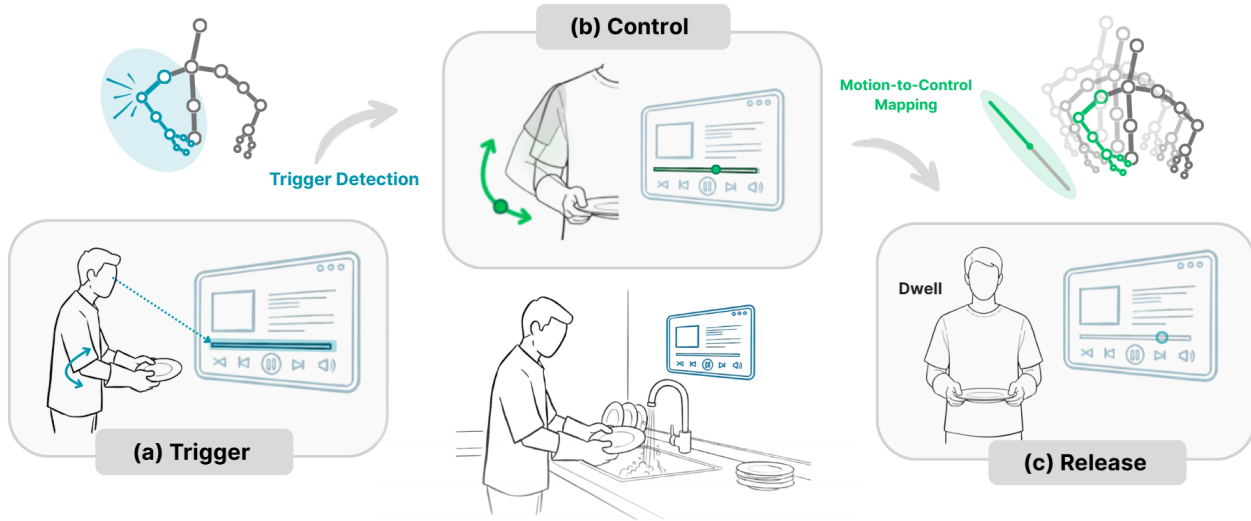


Figure 1: Overview of Dynamic Opportunistic Body Input (DOBI) in a hand-busy context. While washing dishes, the user (a) looks at a UI target and performs a brief trigger gesture with an available spare-body movement (e.g., elbow flick), (b) controls the UI element through motion along the region’s principal axis, and (c) ends the interaction with steady-state motion (e.g., holding still).

Abstract

Extended Reality (XR) systems are often most useful when users are engaged in ongoing physical tasks, yet current interaction techniques still largely assume the hands are available. We present opportunistic body input, an interaction paradigm that redirects continuous XR control to whichever available body region remains free in the moment. To investigate how users naturally coordinate these spare-body movements, we conducted an elicitation study across six hand-busy scenarios. We found that while users’ preferred spare body regions shift dynamically based on physical constraints, the resulting spontaneous movements share a consistent, low-dimensional kinematic structure organized around a dominant principal axis. Building on these findings, we present DOBI (Dynamic Opportunistic Body Input), a real-time XR technique that uses gaze to target a UI element, a brief trigger gesture

to identify the recruited spare body region, and the region’s subsequent motion to drive continuous 1D control. A 1D Fitts’ law study establishes the baseline motor performance of this paradigm across four distinct body regions, achieving throughputs up to 2.62 bits/s with an overall 5.0% error rate, and a dual-task usability study shows that DOBI supports reliable, low-effort control ($SUS = 84.2$) while users remain engaged in realistic hand-busy activities.

CCS Concepts

• **Human-centered computing** → **Gestural input**; **Empirical studies in interaction design**; **Mixed / augmented reality**.

Keywords

XR Body Input, Interaction Technique, Hands-Free Interaction



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/10.1145/3830398.3830670>

ACM Reference Format:

Rachel Kim, Xun Qian, and Sang Ho Yoon. 2026. DOBI: Dynamic Opportunistic Body Input via Spare Joint Recruitment for Hands-Free XR. 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, 14 pages. <https://doi.org/10.1145/3830398.3830670>

1 Introduction

Extended reality (XR) promises a future in which digital content can seamlessly be integrated into everyday activities. To realize this vision, the way users control and manipulate UI elements, such as adjusting sliders, scrolling, or rotating a 3D object, should remain situated and seamless, preserving the continuity of our ongoing tasks. In practice, however, most current XR interactions assume that the hands are always available, relying on mid-air hand gestures such as raycasting, pinching, and direct manipulation [45]. While effective in controlled settings, these techniques often become awkward and fatiguing [27] in the very moments when XR could be most useful, particularly when the hands are already occupied by an ongoing physical task.

To support hands-free UI manipulation, prior work has explored many alternatives to hand input, including voice [18, 21], gaze [51, 53], EMG [54], and facial expressions [32, 36]. Together, these efforts demonstrate that meaningful interaction can be performed without the use of hands. However, each modality constitutes a single, fixed input channel and therefore inherits its own contextual limitations. For instance, voice input may be socially inappropriate in public settings [13] and degrades significantly in noisy environments [10]; gaze is effective for pointing but suffers from the Midas touch problem [30] and induces fatigue during continuous control [28, 39]; EMG signals are sensitive to electrode shift and require per-session calibration, limiting practical robustness [14]; and facial expression techniques are confined to a small, discrete command vocabulary [12] that can be socially conspicuous in public [50].

To address these limitations, the body's skeletal movement offers a particularly rich resource. The body comprises many independently movable joints (shoulders, elbows, wrists, knees, ankles, or the head) [7, 62], so that when some are engaged by a physical task, others inevitably remain free. People already exploit this capacity informally, tapping a phone with an elbow when hands are wet or nudging a door with a shoulder while carrying boxes.

Prior systems have successfully established the feasibility of leveraging body regions as expressive input sources, including the wrist [5, 17, 70], forearm [19, 22, 64], torso [33, 60], and feet [46, 47, 59]. Collectively, these works provide foundational evidence that the body is a viable interaction medium. However, these techniques were predominantly designed as static, single-location interfaces that permanently anchor controls to a predefined anatomical site. If that region becomes occupied by a physical task, the interaction channel is lost. Adaptive frameworks such as Human I/O [38] have recognized that available input channels shift with situational context, and BodyScape [62] has mapped the body's interaction design space across regions. Yet no existing system provides a run-time mechanism for accepting an arbitrary spare-body movement, identifying the active joint, and deriving a control mapping on the fly.

These observations motivate an opportunistic framing of a novel body input technique in XR. Prior work in tangible interaction has described opportunistic controls as interfaces that leverage the physical affordances of the environment [25]. We apply the same principle to the body itself. Here, we use opportunistic body input to map interactions to the spontaneous movements of spare body regions, rather than restricting them to a fixed body region.

Realizing this concept of opportunistic body input requires empirical grounding. It remains an open question (1) what movements users naturally choose when physically constrained, and (2) whether the resulting articulations share sufficient kinematic regularity to support a general-purpose control mapping.

To explore this design space, we conducted an elicitation study in which participants physically enacted six hand-busy scenarios and improvised ad-hoc articulations to achieve seamless and continuous control. Our analysis suggests that while preferred body regions vary across contexts, sampled control poses within a chosen articulation are often well approximated by a dominant first principal component. Building on these observations, we present DOBI (Dynamic Opportunistic Body Input), a real-time XR interaction technique that operationalizes opportunistic body input (Figure 1). DOBI combines gaze targeting with a brief trigger gesture to identify the recruited body region. Then, the system translates motion along the movement's principal axis directly into continuous control of the UI element. Through a controlled 1D Fitts' law study and a dual-task usability study, we demonstrate that DOBI affords effective continuous control, establishing the baseline motor performance across four motions and showing that it remains usable and low-effort while users are engaged in realistic hand-busy activities.

In summary, we make the following contributions:

- (1) **Opportunistic Body Input.** An interaction paradigm that redirects continuous UI control to whichever body region is free in the moment.
- (2) **Empirical Characterization of Spare-Body Motion.** A formal analysis of spontaneously generated movements, identifying a consistent, low-dimensional kinematic structure in pose space that persists across diverse body regions and physical contexts.
- (3) **The DOBI Technique.** A real-time XR interaction technique utilizing a trigger-conditioned pipeline to map spare body movements into continuous 1D UI control signals.
- (4) **Baseline Performance & Usability.** A 1D Fitts' law study establishing the baseline motor performance (throughput and linearity) across four distinct body regions, and a dual-task usability study demonstrating that DOBI remains reliable and low-effort under realistic hand-busy conditions.

2 Related Work

2.1 Hands-Free Interaction in XR

When hands or visual attention are occupied, prior work has turned to a range of alternative input channels: voice and non-verbal audio [18, 42, 58, 73, 74], gaze [1, 51, 53], facial or oral gestures [32, 36, 65], and muscle activity sensed through EMG [15, 56]. These techniques demonstrate that meaningful hands-free interaction is achievable, each offering a distinct combination of discrete activation and continuous control, with pointing treated as a special case of the latter.

However, each modality is tied to a single fixed channel and therefore inherits contextual limitations: voice is unreliable in noisy or socially sensitive settings [35, 52], gaze is effective for pointing but fatiguing for sustained continuous control [29, 39], and facial or EMG-based signals offer compact but specialized input. Adaptive systems such as Human I/O [38] have made this fragility explicit

by cataloging which channels remain available across everyday situations. Our work departs from this single-channel paradigm by treating the body as a pool of channels from which users recruit dynamically based on current availability.

2.2 Body Movement as Input

While prior work has explored the body as a touchable surface, such as acoustic sensing on the arm [22], textile controls in clothing [33], and touch on the ear [34], cheek [68], or lap [41], these systems typically still require a free hand to perform input and therefore do not directly address hand-busy interaction.

More relevant work treats body *movement* as the input signal. Prior systems have used the wrist and forearm motion for scrolling and shortcut control [5, 17], torso movement for eyes-free gestures [60], and lower-limb input such as foot taps [46], toe flexion [47], and gait gestures [59]. Frameworks such as BodyScape [62] further characterize how different body regions can support interaction. Collectively, these works confirm that the body offers a rich input space beyond the hands.

These systems each demonstrate that a specific body region can serve as an effective input channel. Our work asks a complementary question: rather than optimizing interaction for one predetermined region, can a single system support whichever region happens to be available? This requires both a mechanism for dynamic region recruitment and a control mapping that generalizes across regions: extracting axis and range at runtime from the user’s own motion, and assigning polarity implicitly from movement direction, rather than specifying any of these at design time.

2.3 Opportunistic and Context-Adaptive Interaction

Our conceptual framing draws on research in opportunistic and context-aware interaction, where systems adapt to the resources available in the current situation. In tangible and ubiquitous computing, prior work has repurposed nearby objects and surfaces as spontaneous controls [11, 23–26], demonstrating that interaction need not be bound to predefined physical interfaces. In XR, adaptive systems have responded to changing modality availability [38] and layout or content optimization based on the user’s situation [37], as well as ergonomic constraints such as comfort, fatigue, or grasp state [2, 8, 9, 16, 27, 72].

These systems adapt interaction to external context—nearby objects, social situation, or the user’s ergonomic state. Human I/O [38] extends this to input itself, but at the level of which whole channel (e.g., voice, gaze, body) is available. We extend this logic one level further, to the body’s own internal availability: rather than adapting *what* is displayed, *where* an interface is placed, or *which channel* is active, we adapt *which region within the body* is active, treating spare upper-body regions as opportunistic resources that can be recruited and released as the user’s task demands change.

3 Design Elicitation Study

The concept of opportunistic body input assumes that users can intuitively repurpose spare limbs to control continuous values. To explore this, we conducted a design elicitation study to examine

which spare-body movements users naturally produce during hand-busy tasks, and whether these elicited articulations share a universal kinematic structure for continuous control.

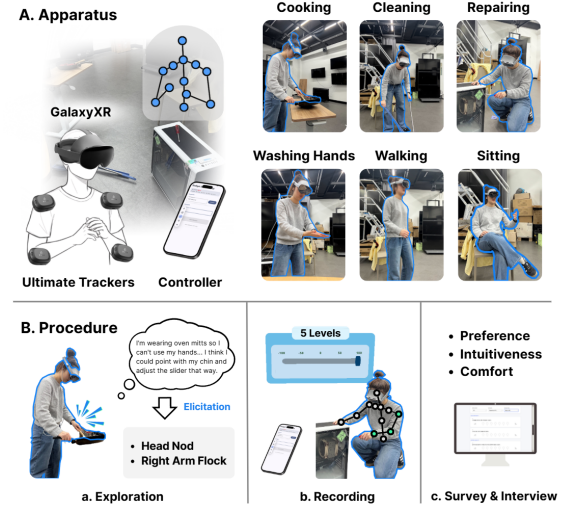


Figure 2: Design elicitation study. (A) Tracking apparatus and six everyday hands-busy scenarios. (B) Procedure involving a. think-aloud exploration, b. kinematic recording, and c. survey and interview.

3.1 Participants and Apparatus

We recruited 12 experienced XR users (> 10 hours of HMD usage) to ideate and gain insight into spare body movements ($M = 26.1$, $SD = 4.0$; 7 male, 5 female; all right-handed). In order to record these movements, participants wore a GalaxyXR headset and four VIVE Ultimate Trackers strapped to both shoulders and elbows, allowing full upper-body skeletal tracking without external base stations. The experimenter operated a companion mobile application to operate the headset remotely, allowing participants to keep their hands engaged in the ongoing activity throughout the session.

3.2 Scenarios and Task

Hand-busy situations refer to everyday contexts where the user’s hands are occupied by a physical task limiting their ability to interact with devices [55]. Drawing on prior taxonomies of situational impairments [38, 66], we selected six everyday scenarios that vary in posture, mobility, and physical constraint on the hands (Figure 2A): **Cooking, Cleaning, Repairing, Washing Hands, Walking, and Sitting**. To ensure realistic immersion, each scenario included:

Physical Props: tools for given contexts (e.g., a frying pan and spatula for cooking, or a screwdriver and desktop computer for repairing)

Contextual Applications: real applications overlaid in the FOV, providing targets for the control (e.g., a YouTube recipe video, music player)

While engaged in the primary physical activity, participants used a think-aloud protocol to propose and execute spare-body

movements for a concurrent control task, such as adjusting a virtual parameter (e.g., a slider).

3.3 Procedure

Prior to the main trials, the instructor guided participants through a warm-up session. To mitigate legacy bias, participants were encouraged to explore the full range of motion of their head, shoulders, elbows, and wrists, helping them recognize various body parts as potential interaction surfaces beyond the hands.

Following this warm-up, participants completed the six scenarios in a randomized order. Each scenario followed a three-step procedure (Figure 2B):

Step 1: Exploration. Using the provided tools and applications, participants acted out the primary tasks within the scenario. Through think-aloud exploration, they identified specific “hand-busy moments” where they needed to interact with the system but could not use their hands. Within that context, participants ideated all spare-body movements they felt were available and natural for controlling a UI element.

Step 2: Recording. Once a movement was finalized, we recorded its kinematic trajectory. To observe the movement’s directionality, polarity, and operational intervals, participants were asked to enact the articulation across five discrete intended control levels: from Level -2 (e.g., maximum decrease) to Level +2 (e.g., maximum increase), passing through Level 0 (the neutral resting posture). When the user held each pose, the experimenter remotely captured the full upper-body skeletal frame using the companion mobile application.

Step 3: Subjective Assessment. Participants rated seven upper-body regions (head, bilateral shoulders, elbows, and wrists) for intuitiveness and comfort using 7-point Likert scales, followed by a semi-structured interview to capture their selection rationale.

3.4 Findings

In total, the elicitation study yielded 437 free-form spare-body input designs across the 6 scenarios, 426 of them with valid recordings (11 were removed due to tracking loss). Because opportunistic body input assumes that users can improvise a workable control movement on the spot, we analyzed this data to understand how users naturally recruit their bodies under hand-busy constraints, and to determine whether the resulting movements can be formalized into a general-purpose control mapping.

F1. Preferred input region varies by scenario. Prior body-centric interaction research has established that multiple body locations can serve as input surfaces [7, 62], and individual systems have demonstrated viable input from the wrist [5], forearm [22], torso [60], and feet [46]. However, these systems each commit to a single, fixed body region at design time. Our results show that no single region should be fixed this way. Intuitiveness ratings (Figure 3) revealed that even the globally highest-rated region (right wrist, $M = 4.78$) ranked first in only two of six scenarios, dropping substantially in the others. Instead, the top-rated region changed according to the given task (Figure 3; Friedman tests, all $p < .01$). As P3 noted, “Wrist rotation feels easier, but when my hands are

Cooking	2.2	2.0	3.4	6.3	5.3	3.7	4.9
Cleaning	2.1	2.4	5.6	4.1	6.1	3.8	2.7
Repairing	2.6	2.6	2.9	3.9	4.0	6.5	4.5
Washing Hands	2.0	2.2	4.2	5.2	5.6	2.9	3.9
Walking	4.9	3.4	2.9	4.6	3.6	3.8	6.6
Sitting	5.8	4.4	3.6	4.2	3.6	4.4	6.1
	L-Wrist	L-Elbow	L-Shoulder	Head	R-Shoulder	R-Elbow	R-Wrist

Intuitiveness
7
1

Figure 3: Mean intuitiveness ratings (7-point Likert) for seven upper-body regions across six activity scenarios.

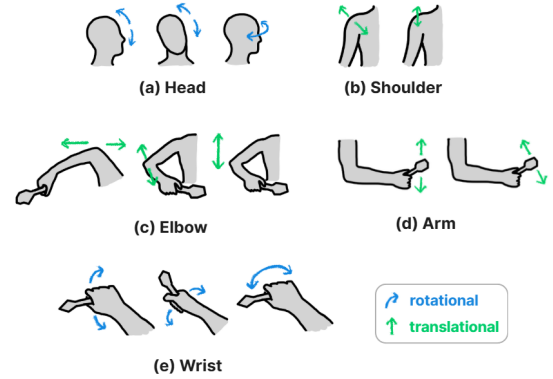


Figure 4: Representative articulation families elicited in the study across upper-body regions. Blue arrows indicate dominant rotational motion; green arrows indicate dominant translational displacement.

wet, using my elbow feels more intuitive.” This provides empirical motivation for dynamic region recruitment: the user, not the system, must initiate which region to use.

F2. Elicited movements can be summarized by a small set of articulation families. Gesture elicitation studies have shown that users converge on shared gesture sets for discrete commands such as pinch-to-select or swipe-to-scroll [49]. These studies primarily focus on categorical gesture types and measure inter-participant agreement. We extend this methodology from discrete gestures to continuous, analog body movements.

Although participants were free to devise any movement, design outputs clustered into 11 recurring articulations (Figure 4). Participants varied widely in *which* region they recruited (F1), but not in *how* they moved a given region once selected. This convergence is not a matter of participants agreeing on a named gesture; rather, each body region’s anatomy admits only a small number of natural, controllable motions (e.g., a wrist can rotate or flex, but has little else to offer as a continuous 1D signal), so independent participants converge on the same few articulations by default. The practical implication is that the system does not require open-ended motion recognition: once a body region is recruited, its movement can be drawn from this small, predictable set of anatomically favored articulations.

F3. User-designed movements are kinematically well approximated by a dominant 1D axis. During think-aloud sessions, we

observed that users expressed control intent in two distinct ways. When using distal joints (head and wrist), participants described rotating the joint in place (e.g., P11: “I’ll twist my wrist to turn the volume up”). When using proximal joints (elbow and shoulder), participants instead focused on displacing the limb through space (e.g., P4: “I’ll drag my elbow like a pointer sideways”)—even though this apparent translation is still produced by rotating a single joint further up the arm.

To capture both strategies in a single representation, we model each joint’s state as a 6D vector combining its 3D relative rotation (via logarithmic map) and 3D position delta: $\mathbf{v}_j = [\mathbf{r}_j, \mathbf{d}_j]^T \in \mathbb{R}^6$. For a recruited kinematic subtree S of K joints, these are concatenated into a unified pose vector $\mathbf{x} \in \mathbb{R}^{6K}$ (Figure 5A).

Applying PCA to these pose vectors confirmed that the elicited movements are kinematically well-suited for linear 1D extraction, though we note that sampling only five discrete intended levels per articulation constrains how finely this linearity (R^2) can be estimated, and treat this as a limitation of the current analysis (Section 7). Beyond concentrating variance on a single component (η_1 : median 97.6%, IQR: 95.0–98.9%), the movements follow straight trajectories in pose space (median $S = 0.95$) and maintain strict proportionality between intended and projected control levels (median $R^2 = 0.98$; Figure 5B). Together, these three properties (dimensionality, straightness, and linearity) indicate that users naturally default to simple, monotonic articulations when designing continuous control, and that a single linear projection over the kinematic subtree is sufficient to extract a usable control signal. This stability is not limited to the discrete elicitation setting: the same PCA-based axis extraction also produces proportional, continuous control in our real-time Fitts’ law evaluation ($R^2 = 0.74$ – 0.97 ; Section 5.1), indicating that the extracted axis remains reliable under continuous, freely-timed motion.

This approach builds on the established use of PCA for extracting low-dimensional control spaces from body signals [6, 49]. However, whereas prior work requires a dedicated calibration session to fit a per-user, per-region mapping, our data shows that the 1D structure emerges spontaneously from the user’s self-chosen movement. A single component captures nearly all variance across the body regions without region-specific modeling, suggesting that the control axis could be extracted from a short motion sample alone.

Decomposing the PC1 loading vector into per-joint contributions reveals that rotation universally drives the control axis, but for multi-joint chains (elbow and shoulder), downstream joints still contribute meaningfully rather than negligibly, suggesting that including the full subtree may improve signal fidelity compared to tracking only the root joint.

F4. Movement range is volatile; polarity is consistent. While the control axis structure is stable (F3), the articulation range is not. Wrist movements produced the largest median PC1 span (1.09), more than double that of head movements (0.45), with substantial variation across users and contexts. As P5 noted: “When I’m walking outside, I don’t want to wave my arm around, a small wrist twist is enough.” In contrast, *polarity agreement*—whether users agreed on which direction of movement should increase versus decrease the control value—was near-universal (Head 94%, Shoulder 97%, Elbow 95%, and Wrist 92%), indicating shared mental models for

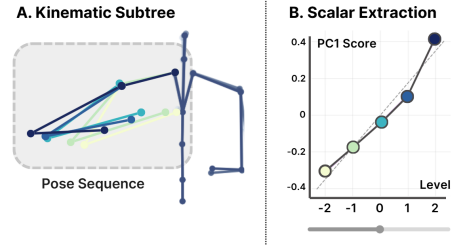


Figure 5: Example of Pose Linearization. (A) Pose vectors are extracted exclusively from the recruited kinematic subtree across five intended control levels (from -2 to +2). (B) Projecting these multi-joint poses onto their first principal component (PC1) reveals a highly linear 1D trajectory, confirming the movement can be mapped directly to a continuous control signal.

directional mapping. This motivates adaptive gain with implicit polarity assignment.

Prior body-based systems typically calibrate both range and polarity in a dedicated setup phase [6]; based on this contrast, DOBI treats the two properties differently: gain is calibrated dynamically per-activation to track the volatile articulation range, while polarity is assigned implicitly from movement direction, without a separate calibration step.

Taken together, these findings suggest a specific strategy for opportunistic body input. Region choice should remain user-initiated because preferred input regions shift across contexts (F1). Once a region is chosen, the resulting movement can often be parameterized from a short motion sample using a dominant 1D axis over the recruited kinematic subtree (F3). Because movement range varies more than directional mapping across users and contexts (F4), per-activation calibration should focus on gain while preserving a consistent polarity convention. These observations motivate DOBI’s system design goals.

3.5 Design Goals

The elicitation study suggests two complementary requirements for DOBI: flexibility in which body region is used, and consistency in how a recruited movement can be parameterized. Together with the need to avoid accidental activation during ongoing activity, these findings motivate the following design goals.

DG1. Enable on-the-fly body region recruitment. F1 showed that the most suitable region shifts with task constraints that are difficult for a system to predict. The interaction should therefore allow users to recruit whichever region they find available, rather than being bound to a predetermined one.

DG2. Require a deliberate activation trigger. Because opportunistic input coexists with ongoing physical activity, it is highly vulnerable to false positives (the Midas Touch problem). The system requires a clear and deliberate activation trigger.

DG3. Extract the control axis directly from the user’s motion. F2 showed that users converge on a small set of shared articulations, and F3 confirmed that each can be reduced to a single linear control

dimension via PCA over the full kinematic subtree. The system can therefore extract the dominant axis from the user’s motion and further generalize across body regions without joint-specific modeling.

DG4. Calibrate gain dynamically while implicitly assigning polarity. As F4 showed, the movement range is volatile, so the system normalizes gain based on the amplitude observed during activation. Movement polarity, by contrast, is assigned implicitly from the direction of the user’s motion, since users exhibit near-universal agreement on directional mapping.

DG5. Provide spatially congruent, body-centric feedback. Unlike mid-air gestures that demand visual attention to a UI, body input leverages proprioception for eyes-free control. However, since the axis is extracted dynamically (DG3), the system must provide body-anchored feedback that instantly visualizes the active region and indicates which physical direction increases the control value.

4 The DOBI System

Based on our design goals, we implemented DOBI (Dynamic Opportunistic Body Input), a real-time XR interaction technique that turns spare-body movements into continuous control signals. The interaction unfolds in three primary stages (Figure 6):

(1) Trigger: To initiate interaction, the user looks at a target UI element and performs a deliberate double-repetition trigger gesture with their chosen spare-body movement.

(2) Control Execution or Discard: Upon successful detection, DOBI instantly displays a mirrored skeleton in the user’s FOV, highlighting the recruited limb and visualizing the corresponding motion and range (DG5). The user can move the joint along this axis to continuously manipulate the UI value, or, if the trigger was accidental, they can discard the interaction by dwelling at the skeleton for 0.3s.

(3) Commit & Release: Once the user reaches the desired UI value, they simply stop moving. Dwelling at the target posture for 0.3s commits the final value and automatically disengages the control, allowing the user to resume their hand-busy task without an explicit exit command.

For both discarding and committing, we choose a 0.3s dwell threshold adapted from prior works in gaze interaction [40, 48]. Figure 7 illustrates the backend technical pipeline powering this interaction lifecycle, from trigger detection to the actual control execution. In the following subsections, we detail this system implementation.

4.1 Gaze and Trigger

To start the control, the user should perform a quick trigger motion to initiate DOBI. The trigger motion serves for two purposes: 1) to inform the system of the transition from hand input to DOBI, and 2) to hint about the control motion.

4.1.1 Trigger Design and Data Collection. Since opportunistic spare-body movements are highly variable, the trigger mechanism must rely on a universal temporal signature that can be consistently applied across any body region. Furthermore, this signature must be easily distinguishable from everyday background motions to prevent false activations (DG2). To satisfy these requirements, we drew

inspiration from the familiar ‘double-click’ desktop metaphor and designed the trigger as a quick, double-repetition movement (e.g., a double wrist flick or double shoulder shrug). This design choice explicitly prioritizes user agency, ensuring that the system only recruits a body region when a deliberate intent is signaled, thereby mitigating the Midas Touch problem in highly dynamic, hands-busy contexts. The user performs this trigger using their intended control motion. By acting as a brief “kinematic preview,” the gesture provides the system with the spatial data required to dynamically extract the principal control axis and calibrate the input gain.

To analyze the temporal properties of this double-repetition trigger and to train a robust detection model, we collected a targeted dataset. We recruited 12 participants (6 female, 6 male; all right-handed) who completed 5-second trials across various behavioral and social hand-busy scenarios. In total, we collected 3916 samples, comprising two types of recordings:

- **Positive Samples:** Participants performed a deliberate double-repetition trigger with a prompted movement (Figure 4) in a manner that felt natural for their current physical constraint.
- **Negative Samples:** Rather than performing prompted non-trigger motions, participants continued to enact the hand-busy task itself (e.g., cooking, washing) while producing scenario-appropriate idle movements and single-repetition pulses. Because such incidental actions can share a similar temporal structure with a genuine trigger, both the training and held-out test sets contain these cases, forcing the classifier to strictly learn the distinctive two-repetition temporal signature.

4.1.2 Trigger Characterization and Parameter Selection. Temporal analysis of the positive samples revealed how trigger execution varies across different body parts (Figure 8). Wrist triggers (RW, LW) were consistently the fastest, followed by the shoulders (RS, LS) and head, while the elbows (RE, LE) produced the slowest articulations. Despite these anatomical differences, the overall duration of the double-repetition gestures exhibited a mean of 1.06 s with a 95th percentile of 1.52 s. Based on this distribution, we established a fixed rolling window of 1.5 s for our real-time detection model. This ensures the system captures the complete double-repetition event regardless of the joint used, while filtering out excessive non-trigger context.

4.1.3 ST-GCN Training and Evaluation. To detect these triggers dynamically, we implemented a Spatio-Temporal Graph Convolutional Network (ST-GCN) [69]. The network leverages both the spatial hierarchy of the human skeleton and the temporal dynamics of the 1.5 s rolling window. The model features two parallel output heads: a binary detection head, indicating whether a trigger is present, and a multi-class identification head, indicating which root joint drives the motion. The network was trained end-to-end using an Adam optimizer and a multi-task loss function on an NVIDIA RTX 5080 GPU.

We evaluated the architecture on a 20% held-out test set. As shown in Figure 9A, the binary detection task achieved an overall accuracy of 98.5%. The true positive rate (recall) of 98.6% and the exceptionally low false positive rate (1.6%) confirm that the model is highly reliable and robust against unintended activations during hand-busy tasks. For root joint identification (Figure 9B), the model

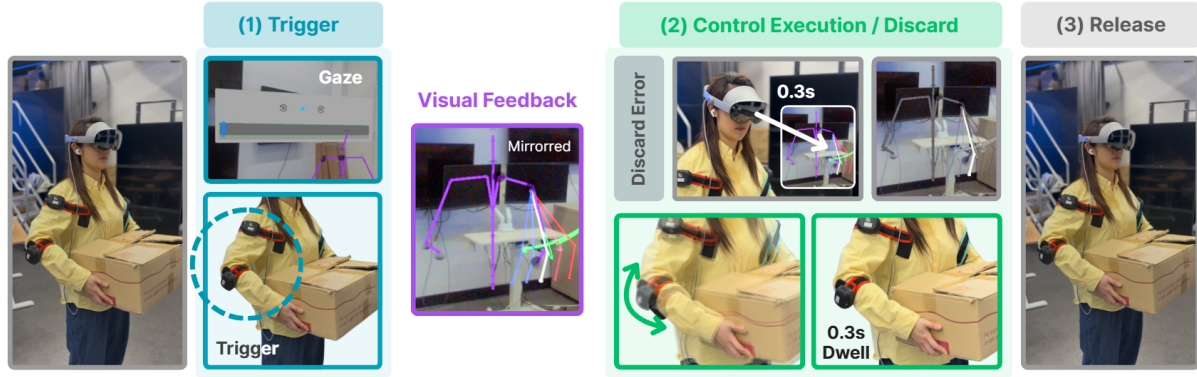


Figure 6: System Walkthrough (1) **Trigger**: The system detects a deliberate double-repetition gesture from a spare body region while the hands are occupied. (2) **Control Execution or Discard**: A mirrored skeleton visualizes the recruited limb and its motion range. The user can either manipulate the UI value through motion or discard an unintended trigger via dwell feedback. (3) **Release**: The interaction is committed and automatically disengaged through a 0.3s dwell, ensuring seamless task continuity without explicit exit commands.

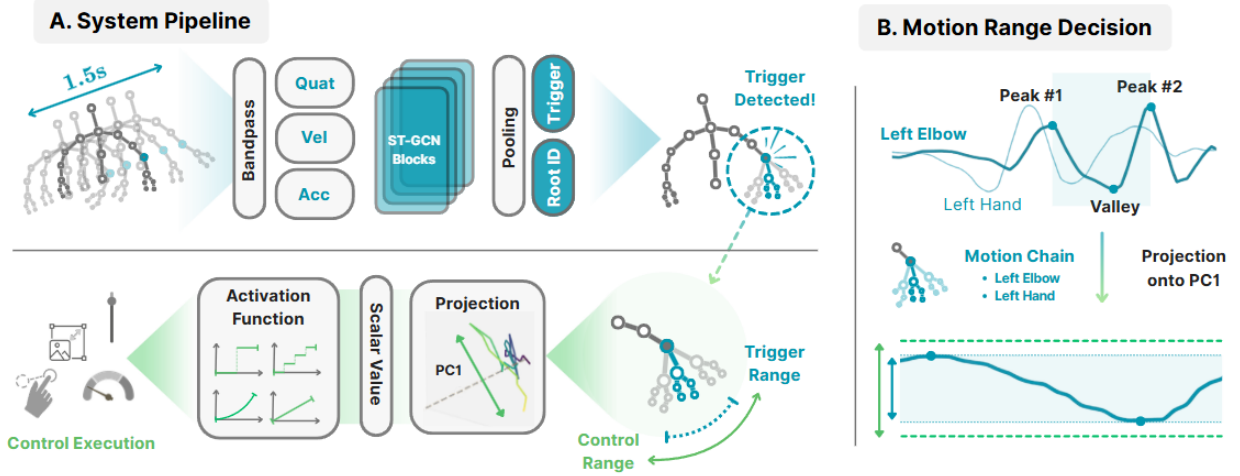


Figure 7: The DOBI interaction pipeline. (A) An ST-GCN continuously monitors a 1.5s window of upper-body skeletal data to detect a trigger and identify the active root joint. The system then extracts the kinematic subtree, projecting its motion onto PC1 to drive continuous UI activation functions. (B) Dynamic gain calibration. By analyzing temporal peaks within the trigger window and projecting the isolated motion chain onto PC1, the system automatically extracts the user’s intended spatial control range..

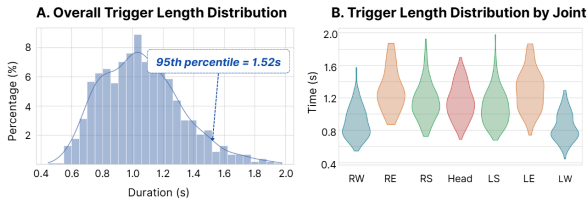


Figure 8: Trigger gesture duration: (A) overall distribution across all joints, (B) per-root joint breakdown.

achieved an average accuracy of 95.9% across all 8 classes (7 joints + 1 none class). Rare misclassifications occurred exclusively between anatomically adjacent, functionally-linked joints (e.g., Right Shoulder vs. Right Elbow).

Furthermore, the core ST-GCN model is designed to be lightweight, with 0.51M parameters, significantly smaller than the original ST-GCN architecture (3.1M) [69] and comparable to other lightweight skeleton-based models such as DD-Net (0.15M) [71]. Compared to other recent deep learning models deployed for XR usage (e.g., EclipseTouch [44] 3.3M, BlazePose 1.5–3.3M [3]), this architecture is efficient and readily deployable on modern standalone XR headset processors.

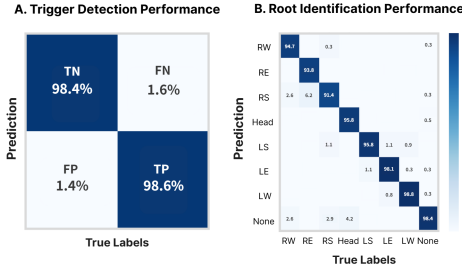


Figure 9: Quantitative performance of our model on the held-out test set. (A) The confusion matrix for the binary trigger detection task, (B) The confusion matrix for the multi-class root joint identification task.

4.2 Motion-to-Control Mapping

Once a trigger is successfully detected, the system must determine the precise motion axis and control range for continuous input. The “kinematic preview” provided during the trigger motion should be further analyzed. Figure 7B illustrates how this brief sample is transformed into a continuous control mapping.

4.2.1 Temporal Segmentation and Subtree Extraction. The Trigger Detection module outputs the 1.5-second rolling window in which the trigger occurred, along with the identified root joint. To isolate the exact bounds of the user’s intended motion, we perform a peak analysis within this window to identify the P_1 and P_2 peaks of the double-repetition pattern. The frames between P_1 and P_2 define the precise temporal bounds of the active trigger motion.

Using the identified root joint, we form a **kinematic subtree** containing the root and all joints distal to it in the kinematic chain (e.g., if the root is detected as the shoulder (RS), the subtree includes RS, RE, and RW). This procedure isolates the active limb and suppresses background noise from the ongoing physical task.

4.2.2 Pose-Linearized PCA and Axis Extraction. To dynamically extract the principal control axis, we perform pose-linearized PCA on the frames isolated between P_1 and P_2 . For each joint J in the subtree, we compute a 6D feature vector combining its relative rotation (q) and position (p):

$$v_J = \begin{bmatrix} r_J \\ d_J \end{bmatrix} \in \mathbb{R}^6 \quad (1)$$

where $r_J = \text{Log}(q_J \cdot q_{\text{neutral}}^{-1}) \in \mathbb{R}^3$ is the log-map of the rotation relative to a neutral pose, and $d_J = p_J - p_{\text{neutral}} \in \mathbb{R}^3$ is the position delta. The neutral pose is established as the average of the frames immediately preceding P_1 . Concatenating all K joints in the subtree yields a unified pose vector $x \in \mathbb{R}^{6K}$.

Applying PCA to these mean-centered pose vectors yields the first principal component w_1 , which defines the 1D motion axis.

4.2.3 Dynamic Gain Calibration and Signal Gating. To ensure the extracted axis is suitable for continuous UI control, the system applies two quality gates before accepting the gesture:

- (1) **Linearity Gate:** The motion must be fundamentally one-dimensional.

Using the PCA eigenvalues (λ_k), the explained variance of the

first principal component must satisfy:

$$\eta_1 = \frac{\lambda_1}{\sum_k \lambda_k} \geq 0.75 \quad (2)$$

- (2) **Range Gate:** The gesture must span a sufficient physical distance to filter out unintentional micro-twitches. We project each trigger frame x_i onto the axis by subtracting the temporal mean pose (μ), yielding $p_i = (x_i - \mu)^\top w_1$, and normalize it to $\hat{p}_i \in [-1, 1]$. The total spatial span (ρ) must satisfy:

$$\rho = \max_i \hat{p}_i - \min_i \hat{p}_i \geq 0.15 \quad (3)$$

where the 0.15 threshold was determined empirically through pilot testing.

Once both gates are satisfied, DOBI latches the targeted UI element. The spatial span (ρ) observed during the trigger establishes the input gain. All subsequent joint motion along w_1 is projected (using the same mean subtraction μ) and mapped directly into continuous scalar values.

4.2.4 UI Activation Functions. Once the motion is mapped into a normalized continuous scalar $\hat{p}_t \in [-1, 1]$, this raw 1D signal can be flexibly interpreted depending on the dimensionality and desired behavior of the target UI element. Rather than altering the core tracking pipeline, DOBI applies different activation functions to $\hat{p}_t$ to compute the final control output $u(t)$. By decoupling the raw kinematic extraction from the final UI activation, this same signal can be flexibly time-integrated or discretized to support rate control, step-based navigation, or binary thresholds. We demonstrate the versatility of these mapping strategies in our proof-of-concept application scenarios (Section 6).

5 System Evaluation

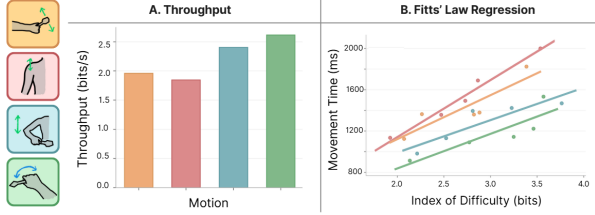
We evaluate DOBI in two complementary studies. Section 5.1 establishes DOBI’s baseline motor performance in a controlled, single-task Fitts’ law study. Section 5.2 then evaluates whether DOBI remains usable when body regions must be opportunistically recruited and switched during an ongoing physical task, in a dual-task usability study.

5.1 Fitts’ Law Study

5.1.1 Study Design. To evaluate whether DOBI’s motion-to-control mapping normalizes movement into predictable, proportional 1D control, we conducted a standard 1D Fitts’ Law reciprocal tapping task [4, 67]. We recruited 14 participants (6 female, 8 male; mean age = 25.09, SD = 2.2, all right-handed), using the same GalaxyXR and VIVE Ultimate Tracker setup as Figure 2A. Participants performed a reciprocal tapping task based on ISO 9241-9 [57], mapped to a normalized virtual slider (0.0–1.0), in a within-subjects design crossing **Input Condition** (four spare-body movements—Shoulder, Arm, Elbow, Wrist—one representative motion per category from our elicitation study, Figure 4b–e, dominant side, Latin-square counterbalanced) and **Index of Difficulty** (five levels, Table 1). Each condition comprised 5 sub-blocks of 15 trials (one per difficulty level, order randomized); participants triggered the instructed movement at the start of each block and rested between sub-blocks, yielding 4200 trials total ($14 \times 4 \times 5 \times 15$). We analyzed standard Fitts’ Law metrics [57]: Movement Time, Error Rate, Effective Width/ID

Table 1: Task parameters (A , W) and resulting nominal ID_s .

	L1	L2	L3	L4	L5
Amplitude (A)	0.30	0.45	0.60	0.70	0.80
Width (W)	0.15	0.12	0.10	0.07	0.05
ID (bits)	1.58	2.25	2.81	3.46	4.09

**Figure 10: Fitts' Law regression models for each input condition. Each data point represents the mean MT at a given ID_e level, averaged across participants.**

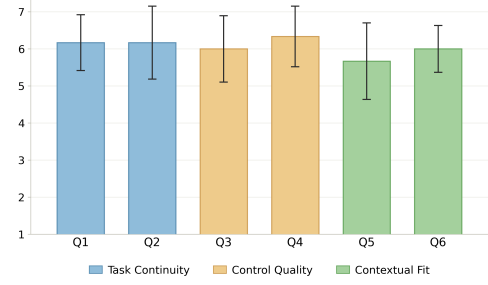
($W_e = 4.133 \times SD_x$, $ID_e = \log_2(A_e/W_e + 1)$), and Throughput ($TP = ID_e/\overline{MT}$, mean-of-means). The first trial of each sub-block was removed as warm-up, and sub-blocks with $>50\%$ error were excluded (3 of 280), following standard practice.

5.1.2 Model Fit and Linearity. All four body regions produced linear models with R^2 ranging from 0.74 to 0.97: the Shoulder was strongest ($R^2 = 0.97$, $p = .002$), followed by the Elbow ($R^2 = 0.80$, $p = .042$) and the Arm ($R^2 = 0.78$, $p = .047$); the Wrist was marginally significant ($R^2 = 0.74$, $p = .060$), likely reflecting a ceiling effect from its fast, consistent execution. These linear relationships confirm that DOBI's PCA-based mapping translates opportunistic movement into proportional 1D control consistent with standard human motor behavior.

5.1.3 Throughput and Error Rate. Across the four regions, overall throughput averaged 2.23 bits/s ($SD = 0.58$) with a 5.0% error rate ($SD = 3.5\%$)—comparable to prior hands-free/XR baselines (~ 2.47 bits/s for head pointing [20], 1.8–2.3 bits/s for gaze+pinch [63], 1.2–1.6 bits/s for foot input [61]). A repeated-measures ANOVA showed a significant effect of Input Condition on Throughput ($F(3, 21) = 5.84$, $p = .005$), with distal regions (Wrist, Elbow) outperforming proximal ones (Shoulder, Arm); Error Rate did not differ significantly ($F(3, 21) = 1.31$, $p = .30$). Effective task parameters ($A_e/A \approx 1.0$, $W_e/W \approx 0.89$ – 1.03) confirmed valid Fitts' behavior across conditions, with the Wrist showing the tightest endpoint precision (Table 2).

5.2 Dual-Task Usability Study

The Fitts' law study isolates DOBI's motor precision under single-task, fully instructed conditions. To evaluate whether DOBI remains usable when a body region must be opportunistically recruited and switched while a person is engaged in a realistic hand-busy activity, we conducted a dual-task usability study.

**Figure 11: Per-item Likert ratings from the dual-task usability study ($N = 6$). Full questionnaire items appear in Appendix A.**

5.2.1 Study Design. We recruited 6 participants for a within-subjects study. Each session began with a 10-minute tutorial in which participants practiced both DOBI and a gaze-and-pinch reference condition representative of standard hands-based XR interaction. Participants then completed three primary-task scenarios drawn from everyday hand-busy activities, cooking, carrying boxes, and walking, each performed under prescribed instructions to keep task demands consistent across participants. Four interactive widgets, video, music, weather, and maps, representative of common ambient XR applications, were placed within the participant's field of view throughout each scenario.

5.2.2 Procedure. Participants were instructed to actively engage with the four widgets using DOBI while continuing to focus on their assigned primary task, mirroring how such interactions would occur in daily use rather than as an isolated foreground activity. Throughout each session, we logged every trigger attempt and its outcome, successful activation, false positive, or a true trigger that produced an unsatisfactory control mapping and required the user to reject and re-trigger. After completing all trials, participants completed a custom 7-point Likert questionnaire (Figure 11) and the System Usability Scale (SUS), followed by a semi-structured interview to capture qualitative feedback.

5.2.3 Results. Across 186 trigger attempts from the 6 participants, 173 (93.0%) were successful, 2 (1.1%) were false positives, and 11 (6.0%) were true triggers with an unsatisfactory mapping that required a re-trigger. This low combined error rate indicates that DOBI's trigger detection and axis extraction remain robust when performed concurrently with an ongoing physical task, rather than only as an isolated, instructed motion.

Participants rated the system highly usable, with an average SUS score of 84.2 ($SD = 5.2$), well above the conventional threshold of 68 associated with above-average usability. The custom Likert ratings (Figure 11) were consistently positive across all six dimensions, highest for interaction responsiveness (Q4, $M = 6.3$), ease of operation (Q1, $M = 6.2$), and task flow preservation (Q2, $M = 6.2$), and lowest for perceived naturalness (Q5, $M = 5.7$).

5.3 Discussion

The Fitts' law results establish baseline motor characteristics under controlled, single-task conditions, and the dual-task study shows

Table 2: Comprehensive performance and movement accuracy by input condition ($N = 14$). Values are mean (SD).

Condition	Performance Metrics			Accuracy Metrics			
	TP (bits/s)	Error (%)	MT (ms)	W_e	W_e/W	A_e	A_e/A
Arm	1.96 (0.32)	5.8 (4.3)	1300 (167)	.105 (.031)	1.03	.492 (.180)	0.97
Shoulder	1.85 (0.48)	5.7 (3.4)	1435 (296)	.099 (.030)	0.99	.479 (.171)	0.93
Elbow	2.40 (0.70)	5.5 (3.5)	1224 (250)	.091 (.048)	1.01	.539 (.194)	0.95
Wrist	2.62 (0.40)	3.2 (2.7)	1138 (163)	.082 (.025)	0.89	.558 (.182)	1.02

this capability translates into reliable, well-received interaction under realistic concurrent activity. The throughput values are a baseline rather than an upper bound: they reflect instructed motions by novice users on first exposure, and DOBI reaches this range while leaving the hands free for the primary task, unlike the comparison techniques. They also reflect the control phase only; trigger and commit overhead is discussed in Section 7.

The observed gradient in motor performance also provides heuristics for interaction design: distal joints (wrist, elbow) excel at fast, fine-grained continuous control, while proximal joints (shoulder, arm) are better suited to coarse adjustments or discrete actions. Accommodating this gradient does not require altering the core motion-to-control pipeline; designers can instead tailor DOBI’s UI activation functions (Section 4.2.4) to the recruited region, for example applying a 1:1 proportional mapping for precise joints or a rate-controlled function to stabilize slower, proximal movements.

Across both studies, DOBI’s error sources were consistent and interpretable. In the dual-task study, failures traced primarily to individual differences in articulation style, some participants performed noticeably looser or tighter triggers than others, a source of error that per-user calibration of the trigger detector could likely reduce further.

The dual-task results also speak to how DOBI is used during ongoing activity. A participant’s available body region and its range of motion change continuously and are often physically limited (e.g., while holding a box); DOBI’s dynamic recruitment let participants control the UI without disrupting their posture, choosing whichever motion cost the least effort. This choice reflected both physical constraints and existing mental models (P6: “I instinctively selected the movement with the least torque”; P2: “zoom feels like I have to use rotation”), and P3 noted an attentional benefit of the trigger-then-release design: “I can control without looking, after the trigger.” Deciding which motion to use can itself add mental cost, however; rather than eliminating this cost, DOBI relocates it from a fixed, system-imposed mapping to a lightweight user choice that participants still found low-effort overall (Figure 11). This points to a design space for adaptive interfaces that suggest or pre-register suitable regions ahead of time to reduce it further.

Finally, DOBI adds to the growing space of non-hand input in XR. Alongside gaze, voice, and foot input, it offers a within-body, dynamically relocatable channel: rather than committing to one fixed modality, the choice of channel itself shifts with whatever the task leaves free. This opens further research into context-aware arbitration among non-hand channels when several are available at once.

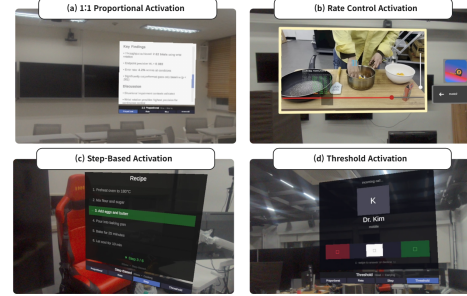


Figure 12: DOBI maps normalized spare-body movements to different UI controls: (a) 1:1 Proportional for continuous zooming, (b) Rate-Control for timeline scrubbing, (c) Step-Based for discrete navigation, and (d) Threshold Activation for binary triggers.

6 Illustrative Application Scenarios

DOBI normalizes ad-hoc movements into continuous 1D signals that can be filtered through various activation functions to support diverse UI behaviors without altering the underlying pipeline. We illustrate this flexibility through four scenarios spanning physical, spatial, and social constraints (Figure 12).

1:1 Proportional Activation: Wrist for Socially Subtle Control. In professional settings where large gestures are socially conspicuous, users can perform micro-movements of the wrist under a table to scroll through AR notes or zoom into 3D models. This leverages the wrist’s superior throughput (2.62 bits/s) and precision ($W_e = 0.082$) to provide highly precise, imperceptible analog control.

Rate-Control Activation: Shoulder for Spatially Constrained Transit. In crowded environments like subways, where range of motion is severely limited, users can control media timelines or volume via subtle shoulder shrugs. The shoulder’s high linearity ($R^2 = 0.97$) makes it ideal for an elastic joystick metaphor; the system translates a stable posture offset into a continuous rate-controlled adjustment (e.g., fast-forwarding), requiring minimal physical space.

Step-Based Activation: Elbow for Hands-Busy Domestic Tasks. When hands are messy during tasks like cooking, the user can navigate a virtual recipe dashboard using their elbow. Given the elbow’s high throughput (2.40 bits/s) and swift execution, we map vertical movements to a step-based function. This allows for deliberate, discrete jumps between recipe items or timers without the risk of overshooting.

Threshold Activation: Head for Severe Physical Encumbrance. When both hands are fully occupied (e.g., carrying a heavy box), users can manage incoming calls through discrete head tilts. By passing a predefined threshold, the movement triggers a binary "swipe" to accept or decline. This demonstrates DOBI's capacity for immediate, eyes-free binary interactions even under total limb encumbrance.

7 Limitations and Future Work

Scope of the current evaluation. Our Fitts' law study isolates DOBI's motion-to-control mapping using one pre-instructed movement per block and no concurrent physical task, establishing baseline throughput and proportionality but not evaluating free body-region switching or task continuity on its own. Our dual-task usability study (Section 5.2) begins to address this gap, showing that DOBI remains reliable and well received when users opportunistically recruit and switch body regions during realistic primary tasks. However, this study involved only 6 participants across 3 scenarios; larger-scale, longer-duration studies are needed to confirm these trends generalize across more users, tasks, and physical contexts.

Discrete sampling in the elicitation analysis. Our elicitation study sampled each articulation at five discrete control levels rather than a continuous sweep, which constrains how finely the linearity (R^2) of the extracted axis can be estimated and may bias it upward. The Fitts' law study offers complementary evidence from continuously produced motion ($R^2 = 0.74\text{--}0.97$), but a denser characterization of continuous trajectories remains future work.

From 1D control to richer continuous input. This paper focuses on whether a user-recruited spare-body movement can be normalized into a robust continuous scalar. Our elicitation analysis and PCA-based mapping suggest that this 1D formulation is a useful first step, but they do not yet establish support for higher-dimensional control. Extending DOBI to 2D UI interactions such as cursor movement would require identifying two separable and user-controllable axes, rather than simply adding a second principal component. One possible direction is to estimate a trigger-conditioned control plane from repeated motion samples and pair it with activation functions designed for planar input. Evaluating whether such axes remain independent, learnable, and stable across users, joints, and contexts remains an open question.

Scalability beyond the upper body. Our prototype and evaluations focus on upper-body joints, reflecting the capabilities of the current tracking setup and the design space explored in the elicitation study. The underlying pipeline from body-region recruitment, kinematic subtree extraction, to trigger-conditioned axis calibration may extend to other tracked kinematic chains, including the lower body. At the same time, lower-body interaction introduces challenges not examined here, including gait-induced motion, postural balance, and interference with locomotion. Future work should therefore investigate whether the same assumptions about trigger separability, linear control structure, and adaptive gain remain valid for knees, ankles, and feet in standing and walking contexts.

Trigger robustness in everyday activity. The trigger detector was evaluated on a held-out dataset of prompted triggers and collected non-trigger motions, and the dual-task usability study

further observed a low false-positive rate (1.1%) under realistic concurrent activity. Still, data collected in a lab environment may not fully reflect real-world edge cases, including unscripted background motions, atypical articulation styles, or extended sessions beyond what we tested. Future work should evaluate trigger robustness in longer, unconstrained recordings collected outside the lab and report practical deployment metrics such as false activations over time and end-to-end interaction overhead.

Dependence on external tracking hardware. Our prototype relies on external Vive Ultimate Trackers fused with the headset's built-in body tracking, which adds setup overhead and limits immediate deployability. Because the pipeline operates purely on the resulting skeletal stream rather than any tracker-specific signal, it can pair with less cumbersome pose-estimation methods to make deployment more practical: camera-based pose estimation from the headset's onboard cameras [3], sparse IMU-based full-body estimation from devices users already wear such as phones, watches, and earbuds [43], or neural inverse kinematics from a few body-worn sensors [31]. Combined with such approaches, DOBI could run with substantially reduced or no dedicated tracking hardware; whether its trigger detection and axis extraction stay robust under the noisier, occlusion-prone estimates these methods produce remains an open question for future work.

8 Conclusion

We presented DOBI, a trigger-conditioned XR technique that enables users to recruit an available spare bodyregion for continuous 1D control when the hands are occupied. Through an elicitation study across six hand-busy scenarios, we found that preferred body regions vary with context, while sampled poses within a chosen articulation are often well approximated by a dominant principal axis. Building on this observation, DOBI combines gaze targeting, double-repetition trigger detection, kinematic subtree extraction, and runtime PCA-based mapping to derive a continuous control signal from a short motion sample.

A controlled 1D Fitts' law evaluation established baseline motor performance across four representative body regions, with throughputs up to 2.62 bits/s and distal joints supporting faster, more precise control than proximal ones, and a dual-task usability study showed reliable, well-received interaction ($SUS = 84.2$) during realistic hand-busy activities. Together, these results suggest that spare-body motion is a promising source of on-demand XR input for hand-busy interaction. More broadly, DOBI points toward XR interfaces that adapt not only to the user's task, but also to which body regions remain available for interaction at that moment.

Acknowledgments

This work was partly supported by the National Research Foundation of Korea (NRF) grant (RS-2026-25470200, 10%), Institute of Information & communications Technology Planning & Evaluation (IITP) & ITRC (Information Technology Research Center) grant (IITP-2026-RS-2024-00436398, 40%), and the IITP grant funded by the Korea government (MSIT) (RS-2026-25507551, Development of Egocentric Data Sensing and Spatial Immersive Experience Technology, 50%).

References

- [1] Nicholas Adams, Mark Witkowski, and Robert Spence. 2008. The inspection of very large images by eye-gaze control. In *Proceedings of the working conference on Advanced visual interfaces*. 111–118.
- [2] Alejandro Aponte, Arthur Caetano, Yunhao Luo, and Misha Sra. 2024. Grav: Grasp volume data for the design of one-handed xr interfaces. In *Proceedings of the 2024 ACM designing interactive systems conference*. 151–167.
- [3] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. 2020. BlazePose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204* (2020).
- [4] Xiaojun Bi, Yang Li, and Shumin Zhai. 2013. FFitts law: modeling finger touch with fitts' law. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1363–1372.
- [5] Daniel Buschek, Bianca Roppelt, and Florian Alt. 2018. Extending keyboard shortcuts with arm and wrist rotation gestures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [6] Maura Casadio, Rajiv Ranganathan, and Ferdinando A Mussa-Ivaldi. 2012. The body-machine interface: a new perspective on an old theme. *Journal of Motor behavior* 44, 6 (2012), 419–433.
- [7] Xiang'Anthony' Chen, Nicolai Marquardt, Anthony Tang, Sebastian Boring, and Saul Greenberg. 2012. Extending a mobile device's interaction space through body-centric interaction. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. 151–160.
- [8] Yi Fei Cheng, Christoph Gebhardt, and Christian Holz. 2023. Interactionadapt: Interaction-driven workspace adaptation for situated virtual reality environments. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–14.
- [9] Yi Fei Cheng, Tiffany Luong, Andreas Rene Fender, Paul Strel, and Christian Holz. 2022. ComforTable user interfaces: Surfaces reduce input error, time, and exertion for tabletop and mid-air user interfaces. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 150–159.
- [10] Leigh Clark, Philip Doberšek, Benjamin R. Cowan, Nadia Pantidi, Omar Mubin, Cosmin Munteanu, Ewa Luger, Emma Nicol, Robert AJ Clark, and Martin Porcheron. 2019. The State of Speech in HCI: Trends, Themes and Challenges. *Interacting with Computers* 31, 4 (2019), 349–371. doi:10.1093/iwc/iwz016
- [11] Christian Corsten, Ignacio Avellino, Max Möllers, and Jan Borchers. 2013. Instant user interfaces: repurposing everyday objects as input devices. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*. 71–80.
- [12] Enrico Costanza, Samuel A. Inverso, Robert Allen, and Pattie Maes. 2007. Intimate Interfaces in Action: Assessing the Usability and Subtlety of EMG-Based Motionless Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 819–828. doi:10.1145/1240624.1240747
- [13] Aarthi Easwara Moorthy and Kim-Phuong L. Vu. 2014. Voice Activated Personal Assistant: Acceptability of Use in the Public Space. In *International Conference on Human Interface and the Management of Information (Lecture Notes in Computer Science, Vol. 8522)*. Springer, 324–334. doi:10.1007/978-3-319-07863-2_32
- [14] Erfan Eddy, Erik J. Scheme, and Scott Bateman. 2023. A Framework and Call to Action for the Future Development of EMG-Based Input in HCI. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, 1–21. doi:10.1145/3544548.3580962
- [15] Ethan Eddy, Erik J Scheme, and Scott Bateman. 2023. A framework and call to action for the future development of EMG-based input in HCI. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–23.
- [16] João Marcelo Evangelista Belo, Anna Maria Feit, Tiare Feuchtnner, and Kaj Grønbaek. 2021. Xrgonomics: Facilitating the creation of ergonomic 3d interfaces. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–11.
- [17] Jacqui Fashimpaur, Amy Karlson, Tanya R Jonker, Hrvoje Benko, and Aakar Gupta. 2023. Investigating wrist deflection scrolling techniques for extended reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [18] Markus Funk, Vanessa Tobisch, and Adam Emfield. 2020. Non-verbal auditory input for controlling binary, discrete, and continuous input in automotive user interfaces. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.
- [19] Nur Al-huda Hamdan, Ravi Kanth Kosuru, Christian Corsten, and Jan Borchers. 2017. Run&Tap: investigation of on-body tapping for runners. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. 280–286.
- [20] John Paulin Hansen, Vijay Rajanna, I. Scott MacKenzie, and Per Bækgaard. 2018. A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In *Proceedings of the Workshop on Communication by Gaze Interaction (COGAIN '18)*. ACM, 1–5. doi:10.1145/3206343.3206344
- [21] Susumu Harada, James A Landay, Jonathan Malkin, Xiao Li, and Jeff A Bilmes. 2006. The vocal joystick: evaluation of voice-based cursor control techniques. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*. 197–204.
- [22] Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 453–462.
- [23] Fengming He, Xiyun Hu, Xun Qian, Zhengzhe Zhu, and Karthik Ramani. 2024. AdapTUI: Adaptation of Geometric-Feature-Based Tangible User Interfaces in Augmented Reality. *Proceedings of the ACM on Human-Computer Interaction* 8, ISS (2024), 44–69.
- [24] Fengming He, Xiyun Hu, Jingyu Shi, Xun Qian, Tianyi Wang, and Karthik Ramani. 2023. Ubi Edge: Authoring Edge-Based Opportunistic Tangible User Interfaces in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [25] Steven J Henderson and Steven Feiner. 2008. Opportunistic controls: leveraging natural affordances as tangible user interfaces for augmented reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*. 211–218.
- [26] Anuruddha Hettiarachchi and Daniel Wigdor. 2016. Annexing reality: Enabling opportunistic use of everyday objects as tangible proxies in augmented reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1957–1967.
- [27] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1063–1072.
- [28] Teresa Hirzle, Maurice Cordts, Enrico Rukzio, and Andreas Bulling. 2020. A Survey of Digital Eye Strain in Gaze-Based Interactive Systems. In *ACM Symposium on Eye Tracking Research and Applications (ETRA '20)*. ACM, 1–12. doi:10.1145/3379155.3391313
- [29] Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-Based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 11–18. doi:10.1145/97243.97246
- [30] Robert J. K. Jacob. 1993. What You Look at is What You Get: Eye Movement-Based Interaction Techniques. *Computer* 26, 7 (1993), 65–67. Originally presented at CHI '90.
- [31] Jiayi Jiang, Paul Strel, Xuejing Luo, Christoph Gebhardt, and Christian Holz. 2024. Manikin: biomechanically accurate neural inverse kinematics for human motion estimation. In *European Conference on Computer Vision*. Springer, 128–146.
- [32] Arata Jingu, Yudai Tanaka, and Pedro Lopes. 2023. Lipio: Enabling lips as both input and output surface. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [33] Thorsten Karrer, Moritz Wittenhagen, Leonhard Lichtschlag, Florian Heller, and Jan Borchers. 2011. Pinstripe: eyes-free continuous input on interactive clothing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1313–1322.
- [34] Takashi Kikuchi, Yuta Sugiura, Katsutoshi Masai, Maki Sugimoto, and Bruce H Thomas. 2017. EarTouch: turning the ear into an input surface. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–6.
- [35] Marion Koelle, Swamy Ananthanarayan, and Susanne Boll. 2020. Social Acceptability in HCI: A Survey of Methods, Measures, and Design Strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 1–19. doi:10.1145/3313831.3376162
- [36] Yunfei Lai, Minghui Sun, and Zhuofeng Li. 2024. GazePuffer: hands-free input method leveraging puff cheeks for VR. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 331–341.
- [37] Zhipeng Li, Christoph Gebhardt, Yves Inglin, Nicolas Steck, Paul Strel, and Christian Holz. 2024. Situationadapt: Contextual ui optimization in mixed reality with situation awareness via llm reasoning. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–13.
- [38] Xingyu Bruce Liu, Jiahao Nick Li, David Kim, Xiang'Anthony' Chen, and Ruofei Du. 2024. Human i/o: Towards a unified approach to detecting situational impairments. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [39] Päivi Majaranta and Andreas Bulling. 2014. Eye Tracking and Eye-Based Human-Computer Interaction. In *Advances in Physiological Computing*. Springer, 39–65. doi:10.1007/978-1-4471-6392-3_3
- [40] Päivi Majaranta and Kari-Jouko Räihä. 2002. Twenty years of eye typing: systems and design issues. In *Proceedings of the 2002 symposium on Eye tracking research & applications*. 15–22.
- [41] Tzu-Wei Mi, Jia-Jun Wang, and Liwei Chan. 2023. Laptouch: Using the lap for seated touch interaction with hmds. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (2023), 1–23.
- [42] Yoshiyuki Mihara, Etsuya Shibayama, and Shin Takahashi. 2005. The migratory cursor: accurate speech-based cursor movement by moving multiple ghost cursors using non-verbal vocalizations. In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*. 76–83.
- [43] Vimal Mollyn, Riku Arakawa, Mayank Goel, Chris Harrison, and Karan Ahuja. 2023. Imuposer: Full-body pose estimation using imus in phones, watches, and earbuds. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing*

- Systems. 1–12.
- [44] 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.
 - [45] Pedro Monteiro, Guilherme Gonçalves, Hugo Coelho, Miguel Melo, and Maximino Bessa. 2021. Hands-Free Interaction in Immersive Virtual Reality: A Systematic Review. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2702–2713.
 - [46] Florian Müller, Joshua McManus, Sebastian Günther, Martin Schmitz, Max Mühlhäuser, and Markus Funk. 2019. Mind the tap: Assessing foot-taps for interacting with head-mounted displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
 - [47] Florian Müller, Daniel Schmitt, Andrii Matvienko, Dominik Schön, Sebastian Günther, Thomas Kosch, and Martin Schmitz. 2023. Tictactoes: Assessing toe movements as an input modality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
 - [48] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th symposium on spatial user interaction*. 99–108.
 - [49] Camilla Pierella, Farnaz Abdollahi, Ali Farshchiansadegh, Jessica Pedersen, Elias B Thorp, Ferdinando A Mussa-Ivaldi, and Maura Casadio. 2015. Remapping residual coordination for controlling assistive devices and recovering motor functions. *Neuropsychologia* 79 (2015), 364–376.
 - [50] Halley P Profita, James Clawson, Scott Gilliland, Clint Zeagler, Thad Starner, Jim Budd, and Ellen Yi-Luen Do. 2013. Don't mind me touching my wrist: a case study of interacting with on-body technology in public. In *Proceedings of the 2013 international symposium on wearable computers*. 89–96.
 - [51] Argenis Ramirez Ramirez Gomez, Christopher Clarke, Ludwig Sidenmark, and Hans Gellersen. 2021. Gaze+ hold: eyes-only direct manipulation with continuous gaze modulated by closure of one eye. In *ACM symposium on eye tracking research and applications*. 1–12.
 - [52] Julie Rico and Stephen Brewster. 2010. Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 887–896. doi:10.1145/1753326.1753458
 - [53] Tim Rolf, Jenny Gabel, Lauren Zerbin, Niklas Hypki, Susanne Schmidt, Markus Lappe, and Frank Steinicke. 2025. A Hands-free Spatial Selection and Interaction Technique using Gaze and Blink Input with Blink Prediction for Extended Reality. *arXiv preprint arXiv:2501.11540* (2025).
 - [54] T. Scott Saponas, Desney S. Tan, Dan Morris, Ravin Balakrishnan, Jim Turner, and James A. Landay. 2009. Enabling Always-Available Input with Muscle-Computer Interfaces. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology (UIST '09)*. ACM, 167–176.
 - [55] Zhanna Sarsenbayeva. 2018. Situational impairments during mobile interaction. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. 498–503.
 - [56] Xiyuan Shen, Yukang Yan, Chun Yu, and Yuanchun Shi. 2022. Clenchclick: Hands-free target selection method leveraging teeth-clench for augmented reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–26.
 - [57] R William Soukoreff and I Scott MacKenzie. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International journal of human-computer studies* 61, 6 (2004), 751–789.
 - [58] Adam J Sporka, Sri H Kurniawan, Murni Mahmud, and Pavel Slavík. 2006. Non-speech input and speech recognition for real-time control of computer games. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*. 213–220.
 - [59] Ching-Yi Tsai, Ryan Yen, Daekun Kim, and Daniel Vogel. 2024. Gait Gestures: Examining Stride and Foot Strike Variation as an Input Method While Walking. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–16.
 - [60] Vincent van Rheden, Lukas Pöckl, and Alexander Meschtscherjakov. 2024. GestureShirt: Eyes-Free Interaction while Running. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 244–248.
 - [61] Eduardo Velloso, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. Interactions under the desk: A characterisation of foot movements for input in a seated position. In *IFIP Conference on Human-Computer Interaction*. Springer, 384–401.
 - [62] Julie Wagner, Mathieu Nancel, Sean G Gustafson, Stephane Huot, and Wendy E Mackay. 2013. Body-centric design space for multi-surface interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1299–1308.
 - [63] Uta Wagner, Ken Pfeuffer, Sophie Stellmach, Stefan Grüner, Susanne Mayr, and Hans Gellersen. 2023. A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, 1–14. doi:10.1145/3544548.3581423
 - [64] Martin Weigel, Vikram Mehta, and Jürgen Steimle. 2014. More than touch: understanding how people use skin as an input surface for mobile computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 179–188.
 - [65] Graham Wilson, Jamie McCready, Euan Freeman, Florian Mathis, Harvey Russell, and Mark McGill. 2025. InterFACE: Establishing a Facial Action Unit Input Vocabulary for Hands-Free Extended Reality Interactions, From VR Gaming to AR Web Browsing. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–23.
 - [66] Jacob O Wobbrock. 2019. Situationally aware mobile devices for overcoming situational impairments. In *Proceedings of the ACM SIGCHI symposium on engineering interactive computing systems*. 1–18.
 - [67] Jacob O Wobbrock, Edward Cutrell, Susumu Harada, and I Scott MacKenzie. 2008. An error model for pointing based on Fitts' law. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1613–1622.
 - [68] Koki Yamashita, Takashi Kikuchi, Katsutoshi Masai, Maki Sugimoto, Bruce H Thomas, and Yuta Sugiura. 2017. CheekInput: turning your cheek into an input surface by embedded optical sensors on a head-mounted display. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. 1–8.
 - [69] Sijie Yan, Yuanjun Xiong, and Dahua Lin. 2018. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.
 - [70] Eunhye Youn, Taejun Kim, and Geehyuk Lee. 2024. WristMenu with Tactons: An Eyes- and Ears-Free Menu with Tactons Describing Menu Items in the Wrist Rotation Space. *International Journal of Human-Computer Interaction* 40, 9 (2024), 2314–2325.
 - [71] Xing-Yi Zhang, Fan Min, Shulin Pan, Qiong Xu, Xue-Yang Min, Guojie Song, and Ke Wang. 2024. DD-Net: Dual decoder network with curriculum learning for full waveform inversion. *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024), 1–17.
 - [72] Qian Zhou, Sarah Sykes, Sidney Fels, and Kenrick Kin. 2020. Gripmarks: Using hand grips to transform in-hand objects into mixed reality input. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–11.
 - [73] Daniel Zielasko, Sebastian Freitag, Dominik Rausch, Yuen C Law, Benjamin Weyers, and Torsten W Kuhlen. 2015. Blowclick: A non-verbal vocal input metaphor for clicking. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*. 20–23.
 - [74] Daniel Zielasko, Neha Neha, Benjamin Weyers, and Torsten W Kuhlen. 2017. A reliable non-verbal vocal input metaphor for clicking. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 40–49.

A Dual-Task Usability Questionnaire

Table 3 lists the full text of the custom 7-point Likert items used in the dual-task usability study (Section 5.2). Each item was rated from 1 (strongly disagree) to 7 (strongly agree), and “[activity]” was replaced by the participant’s current primary task (cooking, carrying boxes, or walking). The six items span three concepts—Task Continuity (Q1–Q2), Control Quality (Q3–Q4), and Contextual Fit (Q5–Q6). Per-item means and standard deviations are reported in Figure 11.

Table 3: Full text of the custom 7-point Likert questionnaire items used in the dual-task usability study, grouped by concept.

Item	Concept	Statement
Q1	Task Continuity	“I could operate the interface while [activity].”
Q2		“Interacting with the interface did not break the flow of [activity].”
Q3	Control Quality	“I could set the slider value or operate the button as intended.”
Q4		“The interface responded predictably to my movement.”
Q5	Contextual Fit	“This technique felt natural while [activity].”
Q6		“I could operate the interface comfortably with the hand constraints.”