ProtoCall: Preemptive Physical Feedback for Sequential Skill Augmentation

Kyungyeon Lee
University of Maryland
College Park, Maryland, USA
kylee@umd.edu

Kirti Kishore
University of Maryland
College Park, Maryland, USA
kiki1@umd.edu

Adrien Ramsamy
University of Maryland
College Park, Maryland, USA
aramsamy@umd.edu

Atsushi Kikumoto
University of Maryland
College Park, Maryland, USA
akikumot@umd.edu

Jun Nishida
University of Maryland
College Park, Maryland, USA
jun@umd.edu

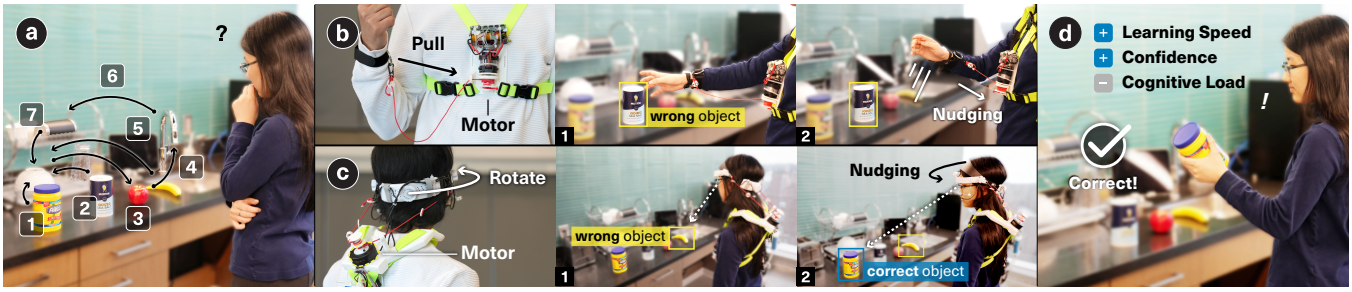


Figure 1: (a) Hesitations and procedural errors during sequential tasks, such as cooking and assembly tasks, disrupt the motor learning. We designed and evaluated three physical feedback strategies to support the smooth step-transition: (b) preemptive reaching nudging (hand actuation), (c) preemptive attentional nudging (head actuation), and post-error nudging. (d) These achieved improved task performance and skill learning, along with enhanced subjective experiences during sequential tasks.

Abstract

Mastering sequential skills is essential for everyday and professional tasks such as cooking and assembly. However, executing actions in the correct order and with appropriate timing remains challenging, particularly during transitions between consecutive steps. Existing approaches, including haptic guidance, error-based feedback, or post-error correction, either increase cognitive load, reduce users' agency, or intervene too late to prevent errors. We developed ProtoCall, a custom wearable physical feedback system, combined with a Vision–Language Model (VLM) for interaction inference, and designed three feedback strategies targeting different stages of the error process: 1) preemptive attentional nudging (attention-level guidance by head actuation), 2) preemptive reaching nudging (action-level prevention of reaching), and 3) post-error nudging. We conducted a user study with 16 participants on a sequential step-transition task, revealing trade-offs between physical feedback strategies across augmentation, retention, and user experience.

CCS Concepts

• **Human-centered computing** → **Haptic devices**.

Keywords

Haptics, Sequential Skill, Error, Body Actuation, Vision Language Model

ACM Reference Format:
Kyungyeon Lee, Kirti Kishore, Adrien Ramsamy, Atsushi Kikumoto, and Jun Nishida. 2026. ProtoCall: Preemptive Physical Feedback for Sequential Skill Augmentation. 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, 15 pages. https://doi.org/10.1145/3830398.3830584

1 Introduction

Everyday and professional activities, such as cooking, assembly, rehabilitation, and musical performance, require executing a sequence of actions with precise order and timing. Developing this proficiency depends on coordinating smooth transitions between consecutive steps, enabling fluent and efficient motor performance and learning [15].

Within sequential behaviors, each transition between steps incurs a transition cost, a brief period of uncertainty in which the learner must retrieve and initiate the next action, disrupting fluency and slowing overall performance [28]. For example, a cook who has just finished chopping vegetables may momentarily lose track of the next step (e.g., prepare the pan or sauce first) and, in that

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 979-8-4007-2856-3/26/11
https://doi.org/10.1145/3830398.3830584

uncertainty, grab the wrong tool or skip a preparation step. Such errors are particularly detrimental for novice learners. They increase cognitive demands, disrupt skill acquisition, and undermine motivation through repeated failures [20].

Prior research has explored various approaches to support *kinematic motor tasks*, such as correct timings, trajectories, and/or gestures. This has been achieved by *physical guidance*, delivered through exoskeletons [40, 58, 79], electrical muscle stimulation (EMS) [12, 32, 46, 55–57, 82], or vibrotactile cues [43, 52]. However, these approaches concentrate on administering physical guidance without considering when and why errors occur, which can reduce learners' sense of agency [50, 90]. *Error-based feedback* [49, 60, 88, 89], and *post-error correction* [13] that intervene after errors occur, encourage orienting to mistakes and post-error adaptation that leads to superior learning efficacy. Yet repeated exposure to explicit errors may negatively affect motivation and perceived competence [11, 35]. To provide more context-aware guidance in *sequential tasks*, several assistive tools have been proposed. Scenarios include cooking [65], household task execution [45], LEGO assembly [24, 78, 85, 97], sports [36], and musical instrument learning [83], using LLM- or VLM-based systems since they require contextual understanding of a user's behavior. While these systems can deliver rich contextual feedback, they provide guidance through a visual or auditory channel that does not provide somatosensory feedback necessary to reinforce motor skill acquisition [26, 98].

Taken together, these approaches leave a gap along two dimensions: existing physical feedback systems engage the motor system directly but intervene reactively, while context-aware systems provide rich guidance but rely on non-physical channels, leaving the motor system unengaged. No prior work has combined preemptive timing with physical feedback in sequential tasks, which we address in this paper.

2 Our Approach: Preemptive Physical Nudging for Step-Transition Support in Sequential Task



Figure 2: Physical Feedback Strategies

In order to provide step-transition support, we designed and evaluated the following three physical feedback strategies that prevent errors at different stages (Figure 2), revealing their trade-offs in error prevention, immediate retention, and user experience.

2.1 Preemptive Attentional Nudging (PAN)

Feedback Design: As shown in Figure 3, when a user's head is directed toward an incorrect object, the system activates haptic guidance to modulate the user's head orientation, thereby redirecting attentional focus toward a relevant and correct object.

Rationale: This design is based on the external focus theory [93] that emphasizes directing attention toward task-relevant objects. Human gaze shifts during visually guided tasks are mainly associated with horizontal head rotation, and most head movements



Figure 3: Preemptive Attentional Nudging

during daily activities occur within a limited range, typically under 50 deg [14]. Because cooking tasks primarily involve interacting with ingredients and tools arranged horizontally across a countertop workspace, this study focuses on yaw rotation (left-right). Importantly, our feedback does not control gaze itself but only influences head orientation. Since eye position is strongly correlated with head orientation [22], we employed head orientation actuation, similar to prior work showing that directly actuating the head using EMS can influence a user's spatial and visual attention [80].

It is worth noting that PAN's head actuation inherently encodes more accurate directional information, while requiring the same form factor and cost: pulling the head toward the target object simultaneously redirects gaze, providing correct-answer guidance that the other two methods cannot. This asymmetry is a consequence of wire-based actuation at different body locations. Encoding spatial direction via arm actuation would require a multi-DOF exoskeleton, making matched preemptive guidance across conditions physically impractical. The implications of this asymmetry for PAN vs. PReN comparisons are discussed in Section 8.

2.2 Preemptive Reaching Nudging (PReN)

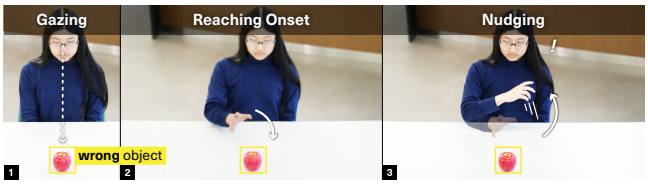


Figure 4: Preemptive Reaching Nudging

Feedback Design: As shown in Figure 4, when a user gazes at an incorrect object and initiates a hand reach, the user's hand is pulled, causing the grasp to abort before the contact.

Rationale: Recent work demonstrated preemptive error prevention in a finger-based tapping task [39], which anticipates and suppresses erroneous movements before execution. This approach is now applied and examined in spatial arm movements involving multi-joint reaching trajectories and greater muscle recruitment.

2.3 Post-Error Nudging (Post)

Feedback Design: As shown in Figure 5, immediately after the user touches an incorrect object (error), the user's hand is pulled, physically interrupting the ongoing interaction with the object.

Rationale: Following prior work on reactive post-contact correction [13], we included a Post-Error Nudging condition as a baseline for reactive guidance. While prior work has shown the benefits

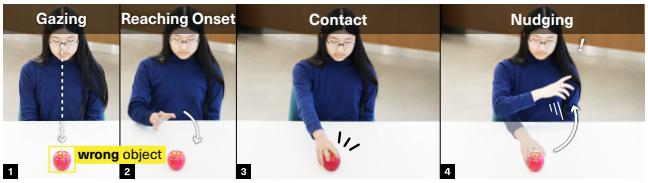


Figure 5: Post-Error Nudging

of preemptive correction in micro-motor tasks [39], this condition allows us to examine whether these advantages generalize from finger-based motor tasks to spatial, multi-object sequential tasks.

2.4 Physical Feedback and Inference System

To realize the three feedback strategies described above, we developed *ProtoCall*, a system that combines wire-based actuation with VLM-based interaction inference. ProtoCall supports sequential task transitions by detecting embodied precursors of cognitive uncertainty through multimodal sensors (i.e., a first-person-view (FPV) video, a gaze tracker, and an IMU sensor on the wrist), and impending wrong-object commitment by delivering preemptive physical feedback through our custom wearable devices. Our work extends findings from recent work exploring the use of multimodal foundation models for skill acquisition [65, 99] into preemptive physical feedback. **Wire-based Actuation:** The HCI community has adopted wire-based force-feedback mechanisms, which enable silent, low-latency, and backdrivable actuation [21, 27, 69, 72]. Based on these benefits, we designed and implemented a unified wire-based body actuation system using reel puller mechanisms to realize the three error feedback strategies. **Dual-loop Interaction Inference:** To balance context-aware inference with low-latency feedback, ProtoCall employs a dual-loop architecture: a *Slow Loop* for task-level inference via a VLM (e.g., tracking the current protocol step and identifying forbidden objects), and a *Fast Loop* for real-time motion detection, object recognition, and haptic triggers.

2.5 Hypothesis

H1: Preemptive feedback (PAN, PReN) will demonstrate higher performance than No-Haptic-Feedback (NoF). Preemptive feedback prevents errors before execution, reducing error accumulation that can disrupt motor learning and decrease motivation [20, 70]. Recent work showed that detecting and correcting errors before execution significantly improved task accuracy and learning efficiency compared to reactive feedback [39].

H2: Preemptive feedback (PAN, PReN) will yield higher performance than Post-Error Nudging (Post). Intervening before an incorrect action occurs reduces corrective actions and sequence disruption compared to feedback delivered after an error. Prior work also showed that preemptive error prevention leads to higher performance than Post [39].

H3: PReN will result in higher performance than PAN. PReN provides somatosensory feedback [4] during reaching, increasing error awareness and enabling immediate motor correction.

H4: PAN will lead to more positive subjective ratings than PReN. Because the PAN occurs before erroneous reaching and requires minimal physical intervention to the end-effector (hand), it may reduce perceived intrusiveness while improving comfort and confidence. PAN may also better preserve the sense of agency. Prior work on wire-based force feedback at the end effector shows that forcibly altering or diverting the user's movement trajectory can reduce the sense of agency, making the PReN more likely to disrupt users' perceived control over their actions [16].

H5: PAN will better preserve motor chunking stability than PReN. As procedural sequences are learned, discrete actions consolidate into motor chunks that are sensitive to mid-execution interruption [84]. Because PReN intervenes physically during the reach, it risks disrupting chunk flow; PAN redirects attention toward the correct target prior to movement without altering the action. We therefore predict that PAN will produce more stable chunk formation than PReN.

2.6 Contributions

This work makes the following contributions:

- Designed and evaluated three physical feedback strategies targeting different stages of sequential error, providing empirical findings on learning performance and user experience.
- Designed and implemented a real-time feedback system for step-transition support that combines VLM-based interaction inference with preemptive feedback.
- Provided general insights and design guidelines using a trade-off between PAN and PReN, along with potential applications.

3 Related Work

3.1 Sequential Motor Learning and Step Transition Challenges

Sequential motor learning involves acquiring temporally ordered actions through repeated practice, gradually transforming effortful, step-by-step execution into more fluent performance [15]. A key challenge in this process lies in managing transitions between steps, where errors such as hesitation, incorrect sequencing, or premature actions frequently occur—making timely intervention critical for accelerating learning [28]. As practice progresses, learners begin to organize actions into more stable patterns, reducing cognitive effort and improving execution efficiency [9]. Timely interventions that support this structuring process have been shown to yield faster performance and greater robustness to disruption [91].

3.2 Attention in Motor Execution and Learning

Attention plays a critical role in motor skill acquisition. In sequential learning, misdirected or excessive focus can increase cognitive demand and disrupt automaticity [31]. In contrast, directing attention toward task-relevant external objects improves performance quality, promotes more automatic movements, and enables more efficient muscle coordination [71, 94, 96]. It is also associated with reduced cognitive load and less interference with implicit control [31]. Prior work has explored guiding attention through

visual [19, 68], auditory [76], vibrotactile [64, 77], and EMS feedback [80]. While these methods can improve accuracy by highlighting relevant information, persistently outsourcing attentional control to the interface may reduce retention, transfer, and robustness of learned skills [61, 90]. This suggests a need for feedback that orients attention at critical moments. In this work, we introduce a preemptive attentional nudging approach that directs attention toward task-relevant cues and evaluate its effectiveness.

3.3 Physical feedback for motor learning in HCI

Error information plays a central role in motor learning, and many HCI systems support movement refinement through post-error visual or haptic feedback [8, 18, 41, 42, 92]. Visual approaches typically highlight deviations from desired movements using overlays or avatars [23, 44, 62, 73, 75], while haptic feedback provides embodied guidance that has shown benefits for rehabilitation and skill training [50, 53].

Haptic guidance has been implemented using several actuation modalities. Exoskeletons provide accurate force guidance but are often bulky and constrain natural movement [40, 58, 79]. EMS offers lightweight direct actuation but requires calibration and may cause discomfort [32, 56, 57, 80]. Vibrotactile interfaces are unobtrusive but provide relatively coarse guidance [43, 52]. Wire-based mechanisms enable fast, backdrivable, and precise mechanical guidance [6, 7, 17, 21, 72]. We therefore adopt a wire-driven design that avoids the calibration requirements and discomfort associated with EMS while providing precise, low-latency guidance.

Recent work [39] further introduced preemptive physical feedback that intervenes before an error occurs, improving both performance and user experience. Our work extends this paradigm from finger-based tapping to arm/head guidance for spatial object-selection, investigating how different styles of preemptive physical feedback generalize beyond finger-based sequential tasks.

4 System Implementation

ProtoCall consists of four components: 1) a head actuator, 2) a hand actuator, 3) wearable sensors, and 4) a VLM-based control software.

4.1 Wire-based Head Actuator

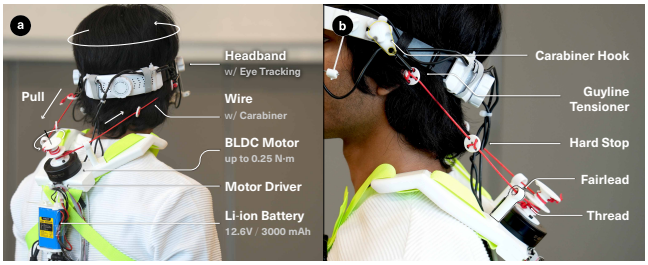


Figure 6: Head actuator based on a capstan drive. A motor pulls the wires to rotate the head: (a) back and (b) side view.

We designed a wire-based head actuator (Figure 6) driven by a high-torque BLDC gimbal motor (IPower, GM5208-12; up to 0.25 N·m), enabling yaw rotation of up to ±30 degrees. The actuation mechanism is based on a capstan drive, in which a single BLDC gimbal motor is mounted centrally and aligned with the neck's yaw axis. Two independent Dyneema wires (2 mm diameter, 152 mm each) are routed on either side of a central flange on a 3D-printed capstan drum (ϕ_{capstan} = 17 mm, ϕ_{flange} = 36 mm, total height 38 mm) coupled to the motor shaft. The drum features three flanges of 2 mm thickness each, one at each end and one at the midpoint, creating two isolated winding channels of 16 mm each, one per wire. Each wire is anchored at one end to its respective winding channel and at the other end to the corresponding side of the headset latch. This arrangement creates an antagonistic pair: when the motor rotates in one direction (e.g., right), the right-side wire winds onto its drum half and pulls the headset, while the left-side wire unwinds and slackens, ensuring the slack wire does not resist the motion. A microcontroller (Seeed Studio, Xiao nRF52840 Sense) with a built-in 6-DOF IMU (STMicroelectronics, LSM6DS3) is attached to the headband and is used to track and control head orientation with a PID closed-loop (k_p = 0.5, k_i = 0.06, k_d = 0.05). The motor is controlled by the microcontroller via a motor driver board (SimpleFOC Mini [74]) and powered by a 12.6V/3000mAh battery (PALO, XZ02). Two Velcro straps are used to attach the actuator unit to the shoulder. The total weight is 508 g and the total material cost of the device is 155.34 USD.

4.2 Wire-based Hand Actuator

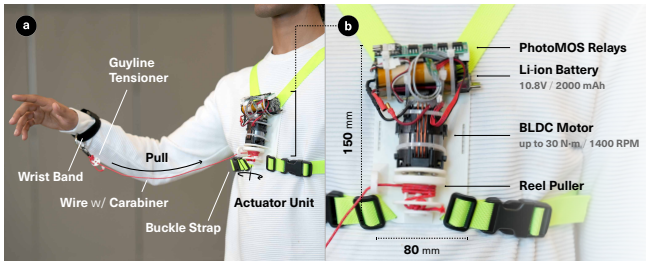


Figure 7: Hand Actuator: (a) a user's wrist is pulled by a wire, which is attached to (b) a reel puller and a BLDC motor.

We designed a wire-based hand actuator (Figure 7) using a high-torque BLDC motor with gears extracted from an electric drill (HOTO, Cordless Brushless Electric Drill), powered by a 10.8V, 2000 mAh Li-ion battery. The actuation mechanism uses a 3D-printed reel puller (ϕ = 45 mm, height = 29 mm, PLA) mounted directly on the gearhead output shaft. One end of a Dyneema wire (2 mm diameter, 600 mm length) is anchored to the reel puller, and as the motor rotates, the wire winds around the reel, shortening the effective length and generating a pulling force. The other end of the wire is attached to a carabiner clipped onto a Velcro strap wrapped around the wrist, transmitting the pulling force directly to the arm. Two Velcro straps are used to attach the actuator unit to the chest. The output torque is adjustable between 0 and 30 N·m with a maximum speed of 1400 RPM. In the study, the torque was set to 25 N·m. The BLDC motor is controlled by emulating trigger switch behavior using four PhotoMOS relays (Panasonic, AQV255) and a microcontroller (Seeed Studio, Xiao nRF52840). The total weight is 510 g, and the total material cost of the device is 191.08 USD.

Safety Considerations. Because ProtoCall physically actuates the head and hand, we incorporated safeguards at both the system and study levels. Before each session, arm-actuation torque was calibrated for each participant to be clearly perceptible but comfortable, while the head-actuator wire length and actuation magnitudes were adjusted based on pilot testing and prior literature [14] to remain noticeable without being excessive or startling. The head actuator also includes a mechanical hard-stop: the headset latch withstands up to 203 N before failure, substantially exceeding the actuator's maximum pulling force of 20 N, providing a large safety margin even in the event of a control or software fault. During the study, the experimenter could immediately cut power to both actuators using a hand-held emergency-stop switch.

4.3 Eye and Motion Tracking

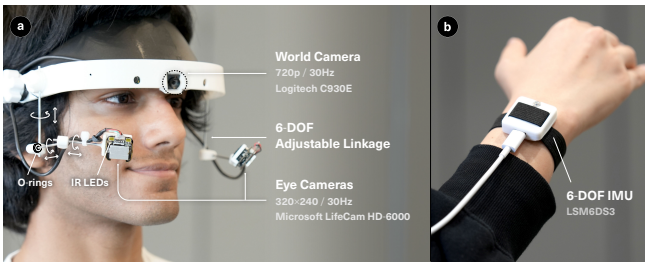


Figure 8: (a) A custom binocular eye-tracking module based on Pupil Core. (b) A wrist IMU for reaching onset detection

We designed a custom binocular **eye-tracking module** based on Pupil Core [33] (Figure 8a), which captures a user's gaze at 30 Hz. It consists of a world camera (Logitech, C930E), two eye cameras (Microsoft, LifeCam HD-6000; an IR filter is removed and two IR LEDs are added), and a 3D-printed base with 6-DOF eye camera linkages. Those are attached to a headband from off-the-shelf head-mounted magnifying glasses (UPC: 726679582257, YOCTOSUN), which has a knob on the back side to adjust to fit one's head size. Pupil Capture software [1] is used for calibration and real-time streaming via ZMQ Sub Socket. The total weight is 332 g. A microcontroller (Seeed Studio, Xiao nRF52840 Sense) equipped with a **6-DOF IMU** (STMicroelectronics, LSM6DS3) is mounted on the wrist (Figure 8b). The system detects reaching onset from the IMU stream by thresholding (0.05) the jerk magnitude (sample-to-sample change in 3-axis acceleration). To suppress spike noise, we apply a short moving average over the most recent jerk magnitudes (N=5).

4.4 Inference Architecture

The system adopts a Fast/Slow Loop architecture (modeled after VLA models [2, 59]), to balance low-latency haptic feedback with computationally intensive interaction inference. It runs locally on the NVIDIA Jetson Thor (JetPack 7.1 with Jetson Linux 38.4) to enable low-latency on-device inference [3].

4.4.1 ***Fast Loop: Gaze, Object Recognition, and IMU.*** The Fast Loop handles real-time perception by synchronizing gaze input with egocentric video and extracting visual regions relevant to the user's attention, along with the IMU input. The Fast-Gen

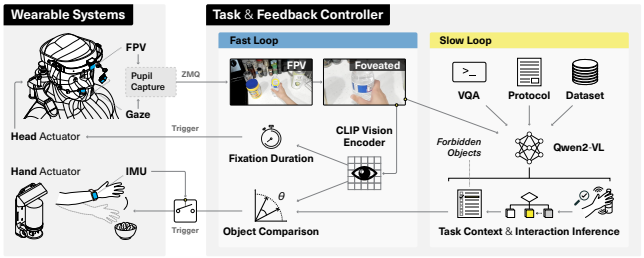


Figure 9: Inference Architecture, consisting of the fast and slow loops to enable preemptive feedback

objects dictionary was prepared by recording task videos specific to the study's objects and setup, and extracting 100 reference images representing the relevant objects and interaction states. The system synchronizes the 30 FPS egocentric video feed with 30 Hz gaze data. To mitigate saccadic jitter and artifactual noise, we implemented an adaptive Exponential Moving Average (EMA) filter equipped with a dynamic α parameter. This jump-limit logic dampens high-frequency noise during visual fixations, yet permits rapid coordinate updates during saccades (i.e., shifts exceeding a normalized threshold of 0.2). To approximate human attentional focus while reducing computational overhead, the system extracts a 224×224 pixel region centered on the user's gaze, creating a foveated crop (modeled after [37, 65, 99]). This region represents the user's current focus of attention and serves as the primary visual input to the VLM for object recognition and step recognition. Traditional autoregressive token generation in VLMs introduces substantial latency, making it unsuitable for frame-level object tracking. To address this limitation, a recognition paradigm that bypasses lexical generation is implemented. Instead, the system computes cosine similarity between the embedding (modeled after [34]) of the current foveated crop, $e_{\text{crop}} \in \mathbb{R}^d$, and a precomputed dictionary of candidate object embeddings, $\mathcal{D} = \{d_1, \ldots, d_K\}$. A CLIP model (ViT-B/32) [63] was used for visual feature extraction. The recognized object index $\hat{k}$ is given by: $\hat{k} = \arg\max_{k \in 1,\ldots,K} \frac{e_{\text{crop}} \cdot d_k}{|e_{\text{crop}}||d_k|}$.

4.4.2 ***Slow Loop: Interaction Inference.*** The Slow Loop manages interaction inference and task progression at 5 Hz. It maintains the evolving task context and verifies whether the user's actions satisfy the current protocol before advancing to the next step. The process uses an open-source VLM (4-bit quantized Qwen2-VL-7B-Instruct [87]) for feasible local execution. Based on the current protocol step, the Slow Loop determines task context using a predefined system prompt (See Appendix A), generating a list of *Target Object* and *Forbidden Objects*. This context is shared with the Fast Loop to prioritize relevant objects and trigger verification only during meaningful interactions. When a hand interaction is detected via IMU, the Slow Loop performs semantic verification via VQA with structured spatial and temporal queries (e.g., whether the user's hand is holding or reaching toward the target). If the user attends to non-task or background objects, no feedback is triggered. The system also monitors haptic events and IMU signals to regulate feedback timing and prevent premature or excessive feedback.

Preemptive Attentional Nudging is triggered when the system detects a gaze fixation toward a prohibited object. Fast-Gen object recognition first identifies a candidate object in the gaze-centered region. If the gaze persists for more than 800 ms, the motor activates and guides the user's head toward the correct object.

Preemptive Reaching Nudging is triggered when the system detects a potential erroneous reach toward a prohibited object. Fast-Gen object recognition first identifies a candidate object in the gaze-centered region. The IMU mounted on the wrist detects reaching motion using the jerk-based threshold (θ = 0.05). Feedback is issued only when both conditions occur simultaneously: the user gazes at a prohibited task object and initiates a reaching motion. To reduce false positives from transient saccades, the system applies a temporal filter over a 1 s gaze window and confirms the wrong-object state only if it persists for more than 60% of the window.

Post-Error Nudging is triggered upon detection of contact with a prohibited object. Fast-Gen object recognition first identifies a candidate object in the gaze-centered region; once the user reaches toward it, VQA validates whether hand-object contact (error) has occurred. If contact with a prohibited object is confirmed, haptic feedback is delivered.

5 Technical Evaluations

5.1 Hand Actuation Latency

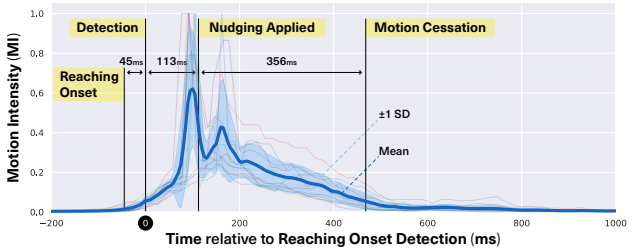


Figure 10: Time-series profile of hand motion intensity

Purpose and Setup. We measured a time-series profile of a hand motion while receiving preemptive reaching nudging to verify that the feedback provision is sufficiently fast to prevent reaching. The built-in IMU attached to the wrist was used to measure the Motion Intensity Index, which is a scalar measure of movement intensity, computed as the within-trial normalized mean of the gravity-compensated acceleration magnitude and the angular velocity magnitude. The measurement was performed 11 times.

Results. As shown in Figure 10, the reaching onset was detected 45.5 ms after motion onset on average (SD = 17.5 ms). The reaching motion was redirected **113 ms** after detection (SD = 32 ms), moving the hand in the opposite direction, which came to a complete stop 356 ms after nudge application (SD = 48 ms). This verifies that our preemptive reaching nudging feedback can effectively prevent a user's reaching to an object before contact, as it usually takes approximately 500 ms after reaching onset to complete a reach [86].

For comparison, EMS (35 V; 800 μs pulse width; 50 Hz frequency) applied to the biceps achieved 146 ms (SD=45 ms), significantly slower than our wire actuator (U=103, p=.025). While both approaches work for hand reach intervention (e.g., [47]), EMS also introduces stimulation discomfort, per-user calibration, and contraindications for users with implanted medical devices [38, 67, 81].

5.2 Head Actuation Force and Accuracy

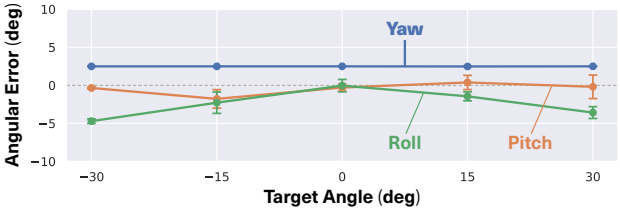


Figure 11: Head Actuation Accuracy: Target Angle vs. Error

Purpose and Setup. We measured the accuracy and force of the head actuation by using the built-in IMU attached to the headband and a digital force gauge (Beslands, AMF-500). Five trials were performed per measurement.

Results. The actuation **force** was 20.0 N on average (SD=2.28 N). The actuation **latency** was 861.1 ms (SD = 291.1 ms) and 1098.8 ms (SD = 284.0 ms) for $\Delta 15^\circ$ and $\Delta 30^\circ$ rotations, respectively. The actuation **error** was 2.49 degrees on average (SD=0.01 deg) on the yaw axis (left-right), as shown in Figure 11. The mean yaw error corresponds to a lateral displacement of 26–30 mm at 60–70 cm, which is well within the inter-object spacing of 250 mm (< 12%) in our study setup. Thus, the yaw error is unlikely to cause misidentification of the intended attention target. The consistent yaw error suggests a systematic offset rather than a nonlinear mechanical error. This could be minimized through careful calibration of the wire lengths. While the roll and pitch axes show smaller mean signed errors (-0.45 deg and -2.41 deg, respectively), they exhibit larger standard deviations (0.86 deg and 0.78 deg, respectively). This is likely because no restoring force is applied along these axes. Nevertheless, participants were able to correctly identify target objects in the study. The EMS-based head actuation system reported mean errors of 4.87–9.13 deg (SD = 2.37–4.58 deg; horizontal targets only) [80], demonstrating higher accuracy and lower variance with our setup.

5.3 Inference Accuracy and Latency

Purpose and Setup. We conducted an offline benchmark to evaluate the latency and recognition reliability of our inference pipeline, compared with a standard VLM prompting strategy [87]. The **standard prompting** processes the full-frame camera input at every frame, whereas the proposed **dual architecture** introduces the lightweight gating mechanism (i.e., hand detection or IMU motion threshold) that conditionally triggers VQA with the foveated crop when physical interaction is detected. Both methods were evaluated using the same pre-recorded dataset, consisting of 300 evaluation events derived from a study task (See Section 6.1). Each event captures a user interaction in which the user reaches toward either the target or the prohibited objects. The system running on

the NVIDIA Jetson Thor determines whether the observed object and behavior are correct or incorrect.

Results. With the dual architecture, the average latency in the **fast loop** was **20.3 ms** (SD=4.2 ms), compared to 999 ms (SD=12.5 ms) in the prompt-based. The average latency in the **slow loop** was **235.7 ms** (SD=6.9 ms) compared to 1503.7 ms (SD=16.2 ms) in the prompt-based. Recognition performance (F1 score) was 90.7% (ProtoCall) and 73.0% (prompt-based), and the system successfully triggered preemptive intervention in 93 out of 100 trials (93.0%) in ProtoCall. Critically, the end-to-end system latency (up to 400 ms) falls within the human visuomotor response window (up to 500–1000 ms) [48]. The slow loop latency (235.7 ms) is shorter than the combined duration of gaze initiation (up to 200 ms) [25] and visually guided reaching (up to 400–700 ms) [30], indicating that contextual inference completes before reaching execution.

6 User Study: Comparison on the Task, Learning Performance, and Experience

We compared two preemptive nudging strategies against post-error nudging and no feedback to examine their effects on task performance, motor learning, and subjective experience. The study was approved by our Institutional Review Board (anonymous).

6.1 Setup and Task

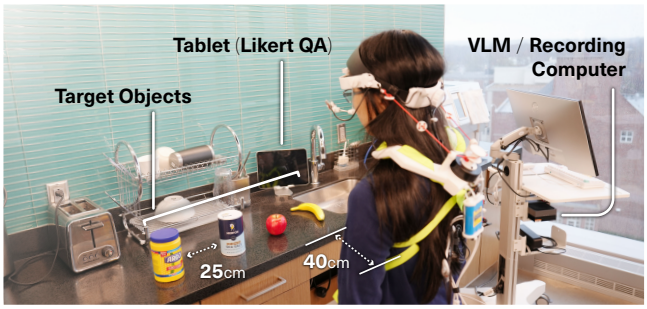


Figure 12: Task setup: Participants wear the wearable system, and the workspace includes ingredients and seasonings.

The study setup is shown in Figure 12. Two fruits (apple, banana) and two seasonings (salt, soy milk) were placed on a table within the participant's reach. To manage timing and data collection, we prepared custom software that records sensor inputs and system events while fully automating the study flow. This ensured consistent timing between feedback-based learning phases. Participants performed a sequential everyday task, cooking-themed sequential object selection, with each sequence consisting of 15 steps. Using these items, participants followed a predefined sequence by visually locating (gaze), reaching for, and selecting the target object at each step, with items potentially repeated across steps. This task was designed to require continuous step transitions and object-selection decisions. The full sequence is provided in Appendix B.

6.2 Conditions

We employed a within-subjects design: **1) No-Haptic-Feedback (NoF)** where participants wore all devices, received the same visual demonstration, but performed the learning phase without any haptic feedback, **2) Post-Error Nudging (Post)**, **3) Preemptive Reaching Nudging (PReN)**, and **4) Preemptive Attentional Nudging (PAN)**. In all conditions, participants wore all devices and received identical visual demonstrations. Only the delivery of haptic error feedback differed across conditions. Condition order was counterbalanced across participants using a Latin square. Four unique sequences with randomized item orderings were prepared, one per condition. Each participant received 12 trials in total.

6.3 Evaluation

6.3.1 ***Task Performance.*** The primary metric is the **number of errors**, defined as any touch of an incorrect (non-target) object, which we compare across the four conditions to examine whether haptic assistance reduces errors during both the learning and evaluation phases. We also conducted an exploratory **chunking trend analysis** to examine how motor learning unfolds across the sequence. During the evaluation phase, we analyzed the proportion of errors at each sequence position to track how learning progresses across the task. We further examined the longest consecutive correct-step run within each 15-step sequence, where longer runs reflect more integrated and stable motor representations [5]. Additionally, we examined **completion time**, defined as the average duration to successfully complete each procedural step, to assess whether different feedback modalities not only correct motor errors but also accelerate task execution.

6.3.2 ***Likert-scale Questionnaires and Interview.*** After completing the evaluation phase of each trial, participants were asked to complete **7-point Likert-scale questionnaires** that included the following subjective aspects: (a) agency, (b) confidence, (c) comfort, (d) cognitive demand, (e) error awareness, and (f) anxiety. The questionnaire statements are detailed in Appendix C.

6.4 Procedure

The study procedure consists of calibration, demonstration, learning, evaluation, and Likert-scale questionnaires.

6.4.1 ***Calibration and Prior Training Session.*** Before the study, we conducted eye-tracking and motor calibration for each participant. Eye-tracking calibration ensured accurate gaze estimation via a standard multi-point procedure, while motor calibration ensured participant comfort, ensuring that feedback was delivered at the appropriate moment relative to action execution. This process took approximately 15 minutes. We then conducted a 15-minute prior training session to minimize bias and familiarize participants with the procedure. Participants practiced a sample sequence with the same structure as the main study, following the identical trial procedure until they fully understood the task.

6.4.2 ***Trial.*** Each participant performed 3 trials per condition. A different sequence was used for each condition, and all sequences were matched in difficulty. Participant performance was recorded throughout. Each trial consisted of the following four sessions:

(1) **Demonstration:** Participants observe a video demonstrating the target food-preparation sequence, which lasts 45 seconds.
(2) **Learning with feedback:** Participants practice learning and memorizing the sequence under one of the four conditions.
(3) **Evaluation without feedback:** Participants replayed the learned sequence without any feedback. No feedback is provided. This phase is time-limited to 60 seconds per step and serves as an immediate retention test [39, 91].
(4) **Likert-scale questionnaires:** After each trial, participants complete Likert-scale questionnaires assessing their experience. (See Appendix C for the questionnaire statements).

6.5 Participants

We recruited 16 participants (7 self-identified as female, 9 as male; Mean = 23.8 years, SD = 3.0). The user study lasted approximately 90 minutes, including training and calibration, and participants received 30 USD as compensation.

6.6 Results

6.6.1 ***Power Analysis.*** Based on pilot data (n = 4), we assumed a medium effect size ($f = 0.25$). With $\alpha = 0.05$ and power = 0.80, the required sample size was 12 for our within-subjects design; our final sample (n = 16) exceeded this.

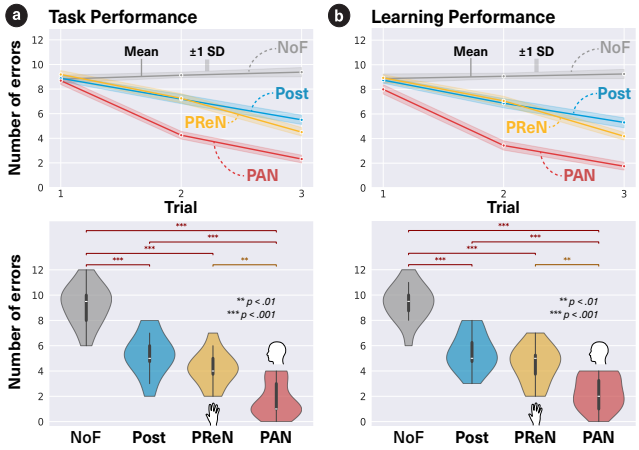


Figure 13: Trial-wise error counts and Trial 3 error distributions (a) with and (b) without error feedback.

6.6.2 ***Performance.***

Task Performance. Figure 13 (a) shows the average number of commission errors made during the learning phase with feedback ($M_{\text{NoF}} = 27.19$, $SD = 3.21$; $M_{\text{Post}} = 20.94$, $SD = 2.21$; $M_{\text{PReN}} = 20.19$, $SD = 3.35$; $M_{\text{PAN}} = 13.19$, $SD = 2.10$). Shapiro–Wilk tests confirmed normality for all conditions (all $p > .10$). A one-way repeated measures ANOVA revealed a significant effect of feedback type on error count, $F(3, 45) = 84.14$, $p < .0001$, $\eta_g^2 = .773$. Post-hoc pairwise comparisons using Benjamini–Hochberg (BH) correction indicated that PAN produced significantly fewer errors than both PReN, $t(15) = 6.50$, $p < .001$, and Post, $t(15) = 10.91$, $p < .001$. PReN and Post did not differ significantly from each other, $t(15) = 0.82$, $p = .428$. Besides, all feedback conditions (PAN, PReN, Post) produced significantly fewer errors than the NoF (all $p < .001$).

Learning Performance. Figure 13 (b) shows the average number of errors made during the evaluation phase without feedback ($M_{\text{NoF}} = 27.31$, $SD = 2.98$; $M_{\text{Post}} = 21.56$, $SD = 2.13$; $M_{\text{PReN}} = 20.94$, $SD = 3.36$; $M_{\text{PAN}} = 15.25$, $SD = 2.02$). Shapiro–Wilk tests confirmed normality for all conditions (all $p > .10$). A one-way repeated measures ANOVA revealed a significant effect of feedback type on error count, $F(3, 45) = 67.45$, $p < .0001$, $\eta_g^2 = .730$. Post-hoc pairwise comparisons using BH correction indicated that PAN produced significantly fewer errors than both PReN, $t(15) = 5.31$, $p < .001$, and Post, $t(15) = 9.62$, $p < .001$. PReN and Post did not differ significantly from each other, $t(15) = 0.65$, $p = .524$. Additionally, all three feedback conditions (PAN, PReN, Post) produced significantly fewer errors than the NoF (all $p < .001$).

6.6.3 ***Completion Time.*** We analyzed total completion time of the last trial of the evaluation phase across conditions ($M_{\text{NoF}} = 57.56$, $SD = 8.98$; $M_{\text{PAN}} = 68.31$, $SD = 14.05$; $M_{\text{PReN}} = 56.81$, $SD = 8.18$; $M_{\text{Post}} = 58.69$, $SD = 8.07$). A one-way repeated-measures ANOVA revealed a significant main effect of feedback condition on completion time ($F(3, 45) = 5.24$, $p = .004$, $\eta_g^2 = .18$; Greenhouse–Geisser $\varepsilon = 1.00$). Post-hoc pairwise comparisons with BH correction showed that PReN resulted in significantly shorter completion times than PAN ($p = .002$). PAN also showed longer completion times than NoF ($p = .035$). Differences between PReN and Post were not significant ($p = .335$).

6.6.4 ***Learning Trajectory Across Trials.*** To account for learning effects across repetitions, we fitted a mixed-effects model with Trial (1–3) as a continuous predictor, Condition as a categorical factor, and a random intercept per participant: $\textit{Errors} \sim \textit{Condition} \times \textit{Trial} + (1 \mid \textit{Participant})$. All conditions began with comparable error counts at Trial 1 (NoF: $M = 8.81$; Post: $M = 8.88$; PReN: $M = 9.19$; PAN: $M = 8.69$). Trial showed a significant learning effect overall ($\beta = -1.73$, $p < .001$), confirming that participants reduced errors across repetitions. Critically, the $\textit{Condition} \times \textit{Trial}$ interaction was significant ($\chi^2(3) = 95.10$, $p < .001$), indicating that conditions differed substantially in their rate of error reduction. Specifically, NoF showed virtually no improvement across trials ($\beta_{\text{Trial}} = 0.28$, $p = .086$), whereas Post ($\beta_{\text{Trial}} = -1.69$), PReN ($\beta_{\text{Trial}} = -2.34$), and PAN ($\beta_{\text{Trial}} = -3.19$) all exhibited significant error reductions (all $p < .001$). By Trial 3, this divergence had produced substantial differences in mean errors: NoF ($M = 9.38$), Post ($M = 5.50$), PReN ($M = 4.50$), and PAN ($M = 2.31$).

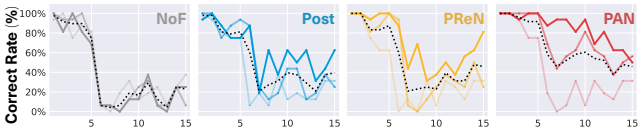


Figure 14: Step-wise accuracy, with Trial 1–3 (light to dark lines) and the mean across trials (a black dotted line).

6.6.5 ***Step-series Error Analysis.*** As shown in Figure 14, PAN produced the fewest errors across most sequence positions, whereas

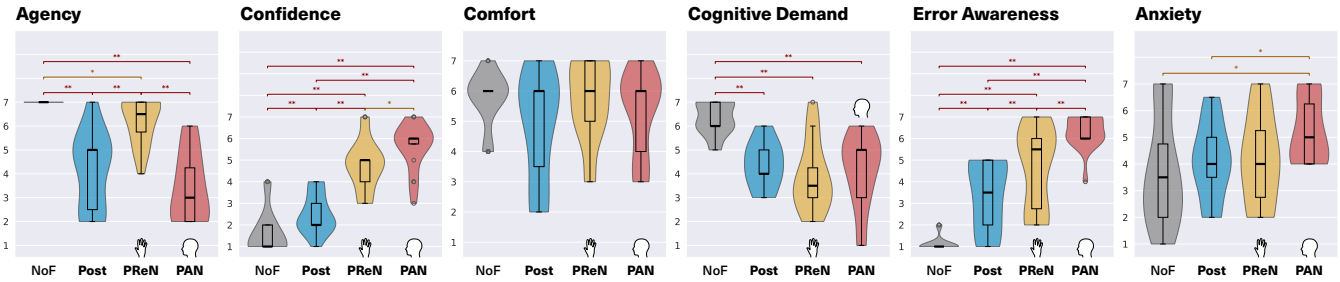


Figure 15: The violin plots show the distribution of subjective ratings across the four conditions, with overlaid box-and-whisker plots indicating the median, interquartile range, and spread of the data. Asterisks indicate significances: $^{*}p < .05$, $^{**}p < .01$.

NoF consistently yielded the highest error rates. In the early portion of the sequence (positions 1–6), both PAN and PReN achieved near-zero error rates, while NoF and Post already showed non-trivial error rates by position 5 (0.31 and 0.25, respectively). In the latter portion of the sequence (positions 7–15), all conditions produced more errors; however, PAN maintained substantially lower error rates throughout, and PReN also consistently outperformed both Post and NoF, which exhibited error rates exceeding 0.60 across most later positions. We also analyzed step-wise correctness using a mixed-effects logistic regression with condition, sequence position, and their interaction as fixed effects and participant as a random effect. With BH correction, in the later positions (7–15), PAN achieved significantly higher correctness (50.5%) than PReN (31.9%), Post (31.5%), and NoF (15.5%; all $p < .001$), indicating that PAN better maintained performance during the more error-prone portion of the sequence.

6.6.6 ***Head–Gaze Coordination.*** To examine whether PAN's head actuation successfully redirected gaze toward the cued target, we analyzed gaze and head IMU data during the learning-with-feedback phase. Across all confirmed PAN triggers ($n = 382$), gaze reached the cued target by the end of the head cue in 85% of cases, realigning within a median of 650 ms ($SD = 78$ ms) of actuation onset—faster than the head completed its rotation and consistent with the eyes leading the head [22]. To test whether PAN's benefit depends on gaze following the cue, we cross-tabulated the 57 non-following triggers against the actual object-selection record (independent of gaze). The correct object was still selected in 30 of 57 cases (53%), above both the 25% chance level and the 40% no-feedback baseline. Thus, PAN improved correct selection over no feedback even when gaze did not reach the target, indicating that the directional head guidance conferred a residual benefit beyond gaze redirection alone.

6.6.7 ***Motor Chunk Stability.*** PAN showed the longest correct-run lengths ($Mdn = 10.00$, $SD = 2.63$), followed by PReN ($Mdn = 6.00$, $SD = 1.71$), Post ($Mdn = 4.50$, $SD = 1.31$), and NoF ($Mdn = 4.00$, $SD = 1.36$). Post-hoc Wilcoxon signed-rank tests with BH correction revealed that PAN significantly outperformed all other conditions: NoF ($p = .001$), Post ($p = .001$), and PReN ($p = .001$). PReN also significantly outperformed both NoF ($p = .001$) and Post ($p = .005$), indicating that preemptive feedback yielded longer error-free runs than reactive feedback. In contrast, Post and NoF did not differ significantly ($p = .480$).

6.6.8 ***Likert-scale Questionnaires.*** Figure 15 summarizes the Likert-scale questionnaire results across the four feedback conditions. We conducted pairwise Wilcoxon signed-rank tests for each subjective factor and applied BH correction.

For **agency**, NoF ($Mdn = 7.0$, $SD = 0.00$) and PReN ($Mdn = 7.0$, $SD = 1.15$) received the highest ratings, whereas PAN ($Mdn = 3.0$, $SD = 1.46$) and Post ($Mdn = 2.0$, $SD = 1.68$) were rated lower. NoF was rated significantly higher than PAN ($p = .002$, $W = 0.0$) and Post ($p = .002$, $W = 0.0$). In addition, PReN was rated significantly higher than PAN ($p = .004$, $W = 10.0$) and Post ($p = .002$, $W = 2.5$).

For **confidence**, PAN ($Mdn = 6.0$, $SD = 1.09$) showed the highest rating, followed by PReN ($Mdn = 5.0$, $SD = 1.18$), Post ($Mdn = 2.0$, $SD = 0.96$), and NoF ($Mdn = 1.0$, $SD = 1.36$). NoF was rated significantly lower than PAN ($p = .001$, $W = 0.0$), PReN ($p = .001$, $W = 3.0$), and Post ($p = .002$, $W = 5.0$). PAN was rated significantly higher than PReN ($p = .021$, $W = 13.5$) and Post ($p = .001$, $W = 0.0$), and PReN was significantly higher than Post ($p = .001$, $W = 0.0$).

For **comfort**, no significant pairwise differences were found.

For **cognitive demand**, NoF ($Mdn = 6.0$, $SD = 0.70$) was rated as the most mentally demanding condition, whereas PAN ($Mdn = 5.0$, $SD = 1.54$), Post ($Mdn = 5.0$, $SD = 0.81$), and especially PReN ($Mdn = 3.0$, $SD = 1.60$) were rated lower. NoF showed significantly higher cognitive demand than PAN ($p = .003$, $W = 0.0$), PReN ($p = .003$, $W = 2.5$), and Post ($p = .003$, $W = 0.0$).

For **error awareness**, PAN ($Mdn = 6.0$, $SD = 0.79$) and PReN ($Mdn = 5.5$, $SD = 1.93$) showed the highest ratings, followed by Post ($Mdn = 3.5$, $SD = 1.59$), while NoF ($Mdn = 1.0$, $SD = 0.34$) was lowest. NoF was rated significantly lower than PAN ($p = .001$, $W = 0.0$), PReN ($p = .001$, $W = 0.0$), and Post ($p = .004$, $W = 0.0$). PAN was rated significantly higher than PReN ($p = .005$, $W = 6.0$) and Post ($p = .001$, $W = 0.0$), and PReN was also rated significantly higher than Post ($p = .001$, $W = 0.0$).

For **anxiety**, NoF ($Mdn = 3.5$, $SD = 2.10$) showed the lowest rating (indicating higher anxiety), followed by PReN ($Mdn = 4.0$, $SD = 1.57$), Post ($Mdn = 4.0$, $SD = 1.47$), and PAN ($Mdn = 5.0$, $SD = 1.24$). PAN was associated with significantly lower anxiety than NoF ($p = .042$, $W = 7.5$) and Post ($p = .047$, $W = 1.0$).

6.6.9 ***Preference.*** A majority preferred PAN (11 out of 16), while 5 favored PReN.

6.6.10 ***Qualitative Feedback.*** We analyzed the interview data using reflexive thematic analysis [10] to better understand participants' experiences learning motor sequences using ProtoCall.

Participants who preferred PReN emphasized that immediate error interception preserves sequential flow, as corrective feedback at the moment of an erroneous reach marks the error without disrupting forward sequence retrieval. P3 described this directly: *"With the hand, I found it was easier—okay, wrong, go to the next one."* In contrast, Post arrived too late, after the learner had begun retrieving the next item, causing the signal to compete with forward planning (P2): *"Even a couple of seconds—okay, that's wrong. But now what is it? And then I forgot the flow entirely."* P12 noted that delayed feedback "overlapped" with the next step, producing cascading confusion. Agency ratings under PReN were also higher than PAN, as the hand intervention left visual attention undisturbed (P7): *"With the hand I can still look around fully—it felt like a tool, not something controlling me."*

Participants who preferred PAN valued directional redirection as a means of encoding the correct target. As P5 noted, *"There are little landmarks where I know my head would turn at this point."*P9 captured the experiential contrast between PAN and PReN: *"If the hand feels like being scolded, the head feels like being gently guided—'hey, that's not it.' "* This aligns with the confidence results: PAN's directional guidance helped participants form accurate expectations, whereas PReN only indicated what was wrong. However, P10 noted that head rotation felt externally imposed, reducing control over gaze allocation, and framed it as a scaffolded learning dynamic: *"It felt like the game master was teaching me how to play—so if the master is teaching me, I'm going to get more right,"* highlighting the trade-off between agency and learning outcomes.

Participants also reflected on the physical comfort of the wearable actuators. While comfort ratings did not differ significantly across conditions, interviews revealed recurring concerns about the form factor. For the head actuator, several participants (P1, P6, P16) mentioned concerns about long-term wearability, citing neck strain from sustained use (P6): *"It was fine for now, but I feel like my neck would get sore after wearing it for a while."* For the hand actuator, the chest-mounted reel unit was occasionally reported as intrusive during reaching, as its placement and weight were noticeable (P4, P10, P13).

6.7 Discussion

6.7.1 ***H1: Error Feedback Efficacy.*** Preemptive feedback yielded significantly fewer errors than NoF in both phases, with participants also reporting higher confidence and error awareness, supporting H1 and extending the benefits of preemptive error nudging observed in finger-based tasks [39] to spatial sequential tasks. In addition, a significant Condition × Trial interaction showed that preemptive feedback, especially PAN, also accelerated the rate of learning itself across repetitions, whereas NoF showed virtually no improvement.

6.7.2 ***H2: Preemptive vs. Post.*** Contrary to H2, PReN did not outperform Post during either phase, and only PAN significantly outperformed Post, partially supporting H2. This diverges from prior finger-based sequential learning [39], where preemptive action-level intervention consistently outperformed post-error correction. We attribute this to three differences: (1) our task requires selecting among four spatial objects rather than binary finger-key mapping, increasing per-step demand; (2) prior feedback physically corrected movement, whereas PReN provides only a prompt signal (nudging); and (3) unlike prior settings where errors could be hidden, PReN does not prevent users from observing prohibited objects mid-reach, potentially increasing confusion and reducing effectiveness.

6.7.3 ***H3: PReN vs. PAN Performance.*** Contrary to H3, PAN produced significantly fewer errors than PReN during both the learning and evaluation phases. We attribute PAN's performance advantage to its ability to encode more information under a similar hardware setup: both conditions rely on a single motor and wire-based actuation mechanism. PReN delivers prompt feedback at the moment of an erroneous reach, communicating only *what not to do*, whereas PAN's head actuation simultaneously redirects gaze toward the correct target, encoding *where to go next*. This distinction was also reflected in subjective ratings: PAN yielded significantly higher confidence and error awareness than PReN, suggesting that directional guidance not only reduces errors but also builds a clearer internal model of the sequence. Additionally, head–gaze coordination analysis showed gaze reliably reached the cued target under PAN, yet even when gaze failed to follow, participants still selected the correct object above chance, indicating that PAN's advantage stems not only from gaze redirection but also from residual somatosensory guidance conveyed by the head-pull itself. However, PAN's accuracy advantage comes at a cost: it incurs substantially longer completion times due to head actuation latency (up to 860–1100 ms), whereas PReN (up to 113 ms) achieves near-baseline throughput, revealing a **speed–latency trade-off** where PAN reduces errors but slows execution, while PReN preserves speed with less guidance. This informational richness is reflected in motor chunk stability: PAN yielded longer consecutive correct runs than PReN, suggesting that directional attentional guidance supports more coherent motor sequence execution [9, 84].

6.7.4 ***H4: PReN vs. PAN Learning Experience.*** Results partially support H4. PReN preserved agency better than PAN, and this difference may also reflect differences in feedback duration: PAN applies 860–1100 ms actuation, whereas PReN delivers a prompt physical intervention (up to 113 ms), making PAN's physical imposition more prolonged and noticeable to the user. However, PAN yielded significantly higher confidence and error awareness, as its correct-target encoding allowed participants to develop spatial landmarks within the sequence that error-only signals could not provide. This aligns with external focus theory, which holds that directing attention toward task-relevant objects promotes more automatic and confident motor execution [93, 95]. This reveals each preemptive feedback's distinct benefit: PReN better preserves agency, while PAN yields higher confidence and lower anxiety during learning.

6.7.5 ***H5: PReN vs. PAN Motor Chunking Stability.*** Supporting H5, PAN produced the longest consecutive correct runs, significantly outperforming PReN. PAN's advantage arises from providing correct information without disrupting sequential processing or the action trajectory, thereby preserving the temporal coherence of chunk execution [5, 84]. Consistent with prior work [51], providing correct information can support the formation of stable motor

acquisition. This was further reflected in the step-series analysis: while PAN and PReN both achieved near-zero error rates in early sequence positions (1–6), PAN maintained significantly lower error rates in later positions (7–15), and PReN's advantage over Post diminished, suggesting that action-level interruption accumulates disruptive costs across longer sequences.

7 Design Implications and Application

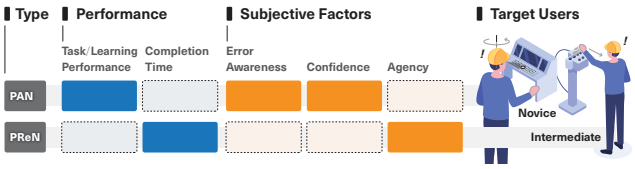


Figure 16: Design implications and potential applications

We synthesized design implications based on our user study results (Figure 16). Feedback strategies should be matched to the learner's stage while accounting for an information–speed trade-off in actuator design. **PAN** provides preemptive guidance that helps *novice* learners build spatial sequence models from scratch, yielding higher confidence and stronger motor chunking due to the richer spatial information with higher latency (860–1100 ms). Meanwhile, **PReN** delivers minimal, rapid (113 ms), action-level intervention that preserves agency and is better suited for *intermediate* learners refining partial representations. A hybrid **PAN-then-PReN** strategy may combine these benefits by first redirecting attention to the correct target, then aborting error actions for faster internalization. Extending this to applications, such as industrial safety, our systems can detect risky actions under high cognitive load and provide preemptive feedback for both novice and skilled operators, achieving enhanced productivity and mitigated workload.

8 Limitations and Future Work

Confounded design factors in PAN vs. PReN. Our three feedback strategies differ simultaneously in actuation location (head vs. arm), intervention timing (pre-attention, pre-reach, post-contact), and the presence of correct-object guidance. Fully disentangling these factors would require a factorial design of 8–12 conditions across multiple studies, well beyond the scope of a single paper. A key physical constraint is that wire-based arm actuation cannot encode precise spatial direction without a multi-DOF exoskeleton, making matched guidance across conditions impractical. Consequently, PAN vs. PReN reflects differences in all three factors simultaneously, whereas PReN vs. Post shares the same actuator and information content. While this paper serves as a fundamental first step in validating preemptive physical feedback, future work should introduce matched conditions to isolate each factor independently.

Absence of Directional Arm Spatial Guidance. Unlike PAN, PReN and Post could not provide directional spatial guidance. Real-time preemptive feedback requires actuators that are lightweight, backdrivable, and wearable for continuous use, whereas existing directional-guidance systems [29, 54, 66] sacrifice one or more of these properties, making them unsuitable for our setting. Consequently, PReN conveyed only error prevention (what not to do) rather than directional guidance (where to go), which likely contributed to its reduced advantage over Post in later sequence positions [42]. Future work should investigate multi-DOF wearable actuation to provide arm-level directional guidance, enabling matched comparisons that isolate the effects of body location, timing, and directional information while examining potential trade-offs with user agency.

Failure Mode Detection. Beyond the mechanical hard-stop on the head actuator, our system does not actively detect failure modes such as wire obstruction or mismatches between actuator state and the user's pose. Future work should improve robustness by integrating additional sensing (e.g., depth cameras) to detect such conditions and automatically halt actuation.

Task Generalizability. Our evaluation was conducted on a task using a fixed set of objects in a controlled tabletop environment to focus on feedback comparison. While the task captures key features of sequential tasks (step transitions, selection errors, and procedural sequencing), its generalizability to other domains, such as assembly tasks or instrument learning, remains to be validated. Our VLM-based task and interaction controller would allow for these extensions with lower adaptation cost.

9 Conclusion

Mastering sequential motor skills is essential for everyday and professional tasks, yet difficult because action errors during step transitions, such as hesitation, incorrect sequencing, and subtle execution errors, can disrupt stable motor learning. We designed and evaluated three physical feedback strategies to prevent errors at different stages: 1) preemptive attentional nudging, 2) preemptive reaching nudging, and 3) post-error nudging. To enable this real-time preemptive feedback, we developed ProtoCall, a custom wearable physical feedback system combining wire-based head and hand actuation with a dual-loop VLM architecture for context-aware, low-latency interaction inference. Following technical evaluations of each system component, we conducted a user study on a cooking-themed sequential task, which demonstrated that each preemptive strategy offers distinct benefits: PAN achieved the lowest error rates and strongest motor chunk formation, while PReN better preserved participants' sense of agency and faster task completion, revealing a trade-off between performance and agency in preemptive physical feedback design for sequential skill augmentation.

Acknowledgments

This work was supported by the JST PRESTO Program (JPMJPR25I6) and the UMD Immersive Media Design program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any funding agencies. We used large language models (ChatGPT and Claude) to assist with proofreading and system development.

A ProtoCall Prompt

Prompt For General Context Information

```
You are observing a first-person video of a user performing a multi-step cooking-themed sequential task. The camera is mounted on or near the user's head, so the user's hands may appear at the bottom or sides of the frame. Your role is to detect object interactions from the following candidate list: <object_list>
Focus only on the single most prominent object from the list that is currently being interacted with.
```

Prompt For Slow Loop

```
Analyze this egocentric cooking video frame and answer the following:
Target object: <object>
Step 1 — Locate the object: Is <object> visible in the frame? If not visible, respond: {"interaction": false, "confidence": 0.0, "reason": "object not visible"}
Step 2 — Locate the hand: Is a human hand visible in the frame? If no hand is visible, respond: {"interaction": false, "confidence": 0.0, "reason": "no hand visible"}
Step 3 — Assess contact: Evaluate the spatial relationship between the hand and <object>. Consider: - Direct contact (fingers/palm touching the surface) - Grasp (hand enclosing or gripping the object) - Indirect contact (holding a tool that contacts the object)
```

B Sample Sequence

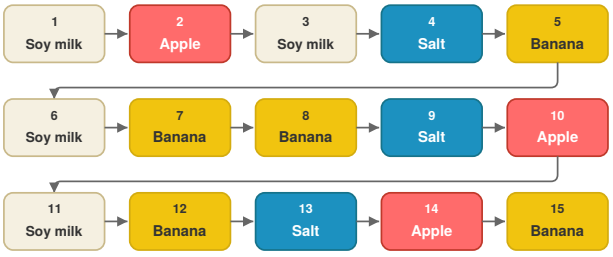


Figure 17: Sample sequence used in the study, consisting of 15 randomized steps using four different cooking items.

C Likert Questionnaire Statements

Category	Questionnaire Statements	Scale
Agency	To what extent did you feel in control of your actions?	(1) No control, (7) Full control
Confidence	How confident were you in memorizing the sequence?	(1) Not confident, (7) Completely confident
Comfort	How would you rate your physical comfort during training?	(1) Not comfortable, (7) Very comfortable
Cognitive demand	How mentally demanding did you find the task?	(1) Not at all, (7) Extremely demanding
Error Awareness	How easily could you perceive your errors while using the device?	(1) Not at all, (7) Very easily
Anxiety	How anxious did you feel while performing the task?	(1) Very anxious, (7) Not at all

Table 1: 7-point Likert-scale questionnaires for evaluating subjective experiences.

References

[1] 2014. Core - Pupil Capture - Pupil Labs Docs. https://docs.pupil-labs.com/core/software/pupil-capture/ Accessed March 9, 2026.

[2] 2025. Helix: A Vision-Language-Action Model for Generalist Humanoid Robots. https://www.figure.ai/helix Accessed April 1, 2026.

[3] 2025. NVIDIA Jetson Thor. https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-thor/ Accessed March 9, 2026.

[4] Sadiya Abdulrabba, Jessica Facchini, and Gerome Aleandro Manson. 2025. Investigating the Speed and Accuracy of Human Movement Corrections to Visual, Somatosensory, and Tactile Perturbations: Evidence for Distinct Sensorimotor Processes. *eneuro* 12, 4 (April 2025), ENEURO.0548–24.2025. doi:10.1523/ENEURO.0548-24.2025

[5] Elger L. Abrahamse, Marit F. L. Ruitenberg, Elian De Kleine, and Willem B. Verwey. 2013. Control of automated behavior: insights from the discrete sequence production task. *Frontiers in Human Neuroscience* 7 (2013). doi:10.3389/fnhum.2013.00082

[6] Alexander Achberger, Pirathipan Arulrajah, Michael Sedlmair, and Kresimir Vidackovic. 2022. STROE: An Ungrounded String-Based Weight Simulation Device. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Christchurch, New Zealand, 112–120. doi:10.1109/VR51125.2022.00029

[7] Alexander Achberger, Fabian Aust, Daniel Pohlandt, Kresimir Vidackovic, and Michael Sedlmair. 2021. STRIVE: String-Based Force Feedback for Automotive Engineering. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. ACM, Virtual Event USA, 841–853. doi:10.1145/3472749.3474790

[8] Amanda Bakkum and Daniel S. Marigold. 2022. Learning from the Physical Consequences of Our Actions Improves Motor Memory. *eneuro* 9, 3 (May 2022), ENEURO.0459–21.2022. doi:10.1523/ENEURO.0459-21.2022

[9] Krishn Bera, Anuj Shukla, and Raju S. Bapi. 2021. Motor Chunking in Internally Guided Sequencing. *Brain Sciences* 11, 3 (Feb. 2021), 292. doi:10.3390/brainsci11030292

[10] Virginia Braun, Victoria Clarke, Nikki Hayfield, Louise Davey, and Elizabeth Jenkinson. 2022. Doing Reflexive Thematic Analysis. In *Supporting Research in Counselling and Psychotherapy*, Sofie Bager-Charleson and Alistair McBeath (Eds.). Springer International Publishing, Cham, 19–38. doi:10.1007/978-3-031-13942-0_2

[11] Andy Brownback and Michael A. Kuhn. 2019. Understanding outcome bias. *Games and Economic Behavior* 117 (Sept. 2019), 342–360. doi:10.1016/j.geb.2019.07.003

[12] Chia-Yu Cheng, Yu Chen, Sitaresmi Wahyu Handani, Avijit Balabantaray, and Mike Y. Chen. 2024. Paired-EMS: Enhancing Electrical Muscle Stimulation (EMS)-based Force Feedback Experience by Stimulating Both Muscles in Antagonistic

Pairs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–7. doi:10.1145/3613904.3642841

[13] Siya Choudhary, Romain Nith, Yun Ho, Jas Brooks, Mithil Guruvugari, and Pedro Lopes. 2025. Adaptive Electrical Muscle Stimulation Improves Muscle Memory. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–11. doi:10.1145/3706598.3713676

[14] Daniel G. Cobian, Andrew C. Sterling, Paul A. Anderson, and Bryan C. Heiderscheit. 2009. Task-Specific Frequencies of Neck Motion Measured in Healthy Young Adults Over a Five-Day Period:. *Spine* 34, 6 (March 2009), E202–E207. doi:10.1097/BRS.0b013e3181908c7b

[15] Christiane Dahms, Stefan Brodoehl, Otto W. Witte, and Carsten M. Klingner. 2020. The importance of different learning stages for motor sequence learning after stroke. *Human Brain Mapping* 41, 1 (Jan. 2020), 270–286. doi:10.1002/hbm.24793

[16] Benjamin Davis, Michael Abbott, and Hannah S. Stuart. 2025. Agency-preserving robotic assistance for grasp slip recovery in body-powered prostheses. *Frontiers in Robotics and AI* 12 (Oct. 2025), 1675955. doi:10.3389/frobt.2025.1675955

[17] Nathan DeVrio and Chris Harrison. 2025. Reel Feel: Rich Haptic XR Experiences Using an Active, Worn, Multi-String Device. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–20. doi:10.1145/3706598.3713615

[18] Jörn Diedrichsen, Reza Shadmehr, and Richard B. Ivry. 2010. The coordination of movement: optimal feedback control and beyond. *Trends in Cognitive Sciences* 14, 1 (Jan. 2010), 31–39. doi:10.1016/j.tics.2009.11.004

[19] Florian Diller, Gerik Scheuermann, and Alexander Wiebel. 2024. Visual Cue Based Corrective Feedback for Motor Skill Training in Mixed Reality: A Survey. *IEEE Transactions on Visualization and Computer Graphics* 30, 7 (July 2024), 3121–3134. doi:10.1109/TVCG.2022.3227999

[20] Jaime E. Duarte and David J. Reinkensmeyer. 2015. Effects of robotically modulating kinematic variability on motor skill learning and motivation. *Journal of Neurophysiology* 113, 7 (April 2015), 2682–2691. doi:10.1152/jn.00163.2014

[21] Cathy Fang, Yang Zhang, Matthew Dworman, and Chris Harrison. 2020. Wireality: Enabling Complex Tangible Geometries in Virtual Reality with Worn Multi-String Haptics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–10. doi:10.1145/3313831.3376470

[22] Yu Fang, Ryoichi Nakashima, Kazumichi Matsumiya, Ichiro Kuriki, and Satoshi Shioiri. 2015. Eye-Head Coordination for Visual Cognitive Processing. *PLOS ONE* 10, 3 (March 2015), e0121035. doi:10.1371/journal.pone.0121035

[23] David W. Franklin and Daniel M. Wolpert. 2008. Specificity of Reflex Adaptation for Task-Relevant Variability. *The Journal of Neuroscience* 28, 52 (Dec. 2008), 14165–14175. doi:10.1523/JNEUROSCI.4406-08.2008

[24] Markus Funk, Thomas Kosch, and Albrecht Schmidt. 2016. Interactive worker assistance: comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Heidelberg Germany, 934–939. doi:10.1145/2971648.2971706

[25] Diane C. Gooding and Michele A. Basso. 2008. The tell-tale tasks: A review of saccadic research in psychiatric patient populations. *Brain and Cognition* 68, 3 (Dec. 2008), 371–390. doi:10.1016/j.bandc.2008.08.024

[26] Graham Grindlay. 2008. Haptic Guidance Benefits Musical Motor Learning. In *2008 Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. IEEE, Reno, NV, USA, 397–404. doi:10.1109/HAPTICS.2008.4479984

[27] Mohssen Hosseini, Roberto Meattini, Andres San-Millan, Gianluca Palli, Claudio Melchiorri, and Jamie Paik. 2020. A sEMG-Driven Soft ExoSuit Based on Twisted String Actuators for Elbow Assistive Applications. *IEEE Robotics and Automation Letters* 5, 3 (July 2020), 4094–4101. doi:10.1109/LRA.2020.2988152

[28] Andrew Howes, Richard L. Lewis, and Alonso Vera. 2009. Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review* 116, 4 (2009), 717–751. doi:10.1037/a0017187

[29] Felix Huppert, Gerold Hoelzl, and Matthias Kranz. 2021. GuideCopter - A Precise Drone-Based Haptic Guidance Interface for Blind or Visually Impaired People. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. doi:10.1145/3411764.3445676

[30] M. Jeannerod. 1984. The Timing of Natural Prehension Movements. *Journal of Motor Behavior* 16, 3 (Sept. 1984), 235–254. doi:10.1080/00222895.1984.10735319

[31] E.C. Kal, J. Van Der Kamp, and H. Houdijk. 2013. External attentional focus enhances movement automatization: A comprehensive test of the constrained action hypothesis. *Human Movement Science* 32, 4 (Aug. 2013), 527–539. doi:10.1016/j.humov.2013.04.001

[32] Shunichi Kasahara, Kazuma Takada, Jun Nishida, Kazuhisa Shibata, Shinsuke Shimojo, and Pedro Lopes. 2021. Preserving Agency During Electrical Muscle Stimulation Training Speeds up Reaction Time Directly After Removing EMS. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–9. doi:10.1145/3411764.3445147

[33] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, Seattle Washington, 1151–1160. doi:10.1145/2638728.2641695

[34] Muhammad Haris Khan, Miguel Altamirano Cabrera, Dmitrii Iarchuk, Yara Mahmoud, Daria Trinitatova, Issatay Tokmurziyev, and Dzmitry Tsetserukou. 2025. HapticVLM: VLM-Driven Texture Recognition Aimed at Intelligent Haptic Interaction. In *2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, Vienna, Austria, 3727–3732. doi:10.1109/SMC58881.2025.11343053

[35] Hyosub E Kim, Darius E Parvin, and Richard B Ivry. 2019. The influence of task outcome on implicit motor learning. *eLife* 8 (April 2019), e39882. doi:10.7554/eLife.39882

[36] Kangbeen Ko, Minwoo Oh, Minwoo Seong, and SeungJun Kim. 2025. LEGOLAS: Learning & Enhancing Golf Skills through LLM-Augmented System. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–10. doi:10.1145/3706599.3720141

[37] Robert Konrad, Nitish Padmanaban, J. Gabriel Buckmaster, Kevin C. Boyle, and Gordon Wetzstein. 2024. GazeGPT: Augmenting Human Capabilities using Gaze-contingent Contextual AI for Smart Eyewear. doi:10.48550/arXiv.2401.17217 arXiv:2401.17217 [cs].

[38] Georgios Korres, Wanjoo Park, and Mohamad Eid. 2022. A Comparison of Vibrotactile Feedback and Electrical Muscle Stimulation (EMS) for Motor Response During Active Hand Movement. *IEEE Transactions on Haptics* 15, 1 (Jan. 2022), 74–78. doi:10.1109/TOH.2022.3142442

[39] Kyungyeon Lee, Jai Vaichalkar, Arnav Dadarya, Wooje Chang, Atsushi Kikumoto, and Jun Nishida. 2026. FIXical I/O: Exploring the Effects of Real-time Error Sensing and Physical Intervention on Finger-based Motor Sequence Learning. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2026). doi:10.1145/3772318.3790432

[40] Kyungyeon Lee, Daniel Yang, Kriti Singh, and Jun Nishida. 2025. Hapticus: Exploring the Effects of Haptic Feedback and its Customization on Motor Skill Learning: Tactile, Haptic, and Somatosensory Approaches. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan. doi:10.1145/3706598.3713821

[41] Li-Ann Leow, Welber Marinovic, Aymar De Rugy, and Timothy J. Carroll. 2020. Task Errors Drive Memories That Improve Sensorimotor Adaptation. *The Journal of Neuroscience* 40, 15 (April 2020), 3075–3088. doi:10.1523/JNEUROSCI.1506-19.2020

[42] Danielle Levac, Kate Driscoll, Jessica Galvez, Kathleen Mercado, and Lindsey O'Neil. 2017. OPTIMAL practice conditions enhance the benefits of gradually increasing error opportunities on retention of a stepping sequence task. *Human Movement Science* 56 (Dec. 2017), 129–138. doi:10.1016/j.humov.2017.10.017

[43] J.. Lieberman and C.. Breazeal. 2007. TIKL: Development of a Wearable Vibrotactile Feedback Suit for Improved Human Motor Learning. *IEEE Transactions on Robotics* 23, 5 (Oct. 2007), 919–926. doi:10.1109/TRO.2007.907481

[44] Klemen Lilija, Soren Kyllingsbaek, and Kasper Hornbaek. 2021. Correction of Avatar Hand Movements Supports Learning of a Motor Skill. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, Lisboa, Portugal, 1–8. doi:10.1109/VR50410.2021.00069

[45] 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 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. doi:10.1145/3613904.3642065

[46] Pedro Lopes, Alexandra Ion, and Patrick Baudisch. 2015. Impacto: Simulating Physical Impact by Combining Tactile Stimulation with Electrical Muscle Stimulation. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, Charlotte NC USA, 11–19. doi:10.1145/2807442.2807443

[47] Pedro Lopes, Patrik Jonell, and Patrick Baudisch. 2015. Affordance++: Allowing Objects to Communicate Dynamic Use. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 2515–2524. doi:10.1145/2702123.2702128

[48] Sergey Lytaev. 2022. Long-Latency Event-Related Potentials (300–1000 ms) of the Visual Insight. *Sensors* 22, 4 (Feb. 2022), 1323. doi:10.3390/s22041323

[49] Laura Marchal-Crespo, Panagiotis Tsangaridis, David Obwegeser, Serena Maggioni, and Robert Riener. 2019. Haptic Error Modulation Outperforms Visual Error Amplification When Learning a Modified Gait Pattern. *Frontiers in Neuroscience* 13 (Feb. 2019), 61. doi:10.3389/fnins.2019.00061

[50] Laura Marchal-Crespo, Mark Van Raai, Georg Rauter, Peter Wolf, and Robert Riener. 2013. The effect of haptic guidance and visual feedback on learning a complex tennis task. *Experimental Brain Research* 231, 3 (Nov. 2013), 277–291. doi:10.1007/s00221-013-3690-2

[51] J.P Maxwell, R.S.W Masters, and F.F Eves. 2003. The role of working memory in motor learning and performance. *Consciousness and Cognition* 12, 3 (Sept. 2003), 376–402. doi:10.1016/S1053-8100(03)00005-9

[52] Troy L. McDaniel, Morris Goldberg, Shantanu Bala, Bijan Fakhri, and Sethuraman Panchanathan. 2012. Vibrotactile feedback of motor performance errors for enhancing motor learning. In *Proceedings of the 20th ACM international conference on Multimedia*. ACM, Nara Japan, 419–428. doi:10.1145/2393347.2393408

[53] Jingyan Meng, Na Li, Yang Hu, Dazheng Zhao, Guoning Li, Jingyan Hu, Tao Song, Yehao Ma, Rongzhen Fu, Guokun Zuo, Liang Tao, Min Tang, Yunfeng Liu, and Changcheng Shi. 2023. Effects of Error Modulation-Based Visual and Haptic Feedback Fusion Strategies on Motor Learning and Motivation. *Applied Bionics*

and Biomechanics 2023 (July 2023), 1–14. doi:10.1155/2023/8974148

[54] Marie Muehlhaus, Alexander Liggesmeyer, and Jürgen Steimle. 2025. ExoKit: A Toolkit for Rapid Prototyping of Interactions for Arm-based Exoskeletons. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–17. doi:10.1145/3706598.3713815

[55] Arinobu Niijima, Ryosuke Aoki, Yukio Koike, and Shinji Miyahara. 2023. Motor-Skill-Transfer Technology for Piano Playing with Electrical Muscle Stimulation. In *SIGGRAPH Asia 2023 Emerging Technologies*. ACM, Sydney NSW Australia, 1–2. doi:10.1145/3610541.3614570

[56] Arinobu Niijima, Toki Takeda, Kentaro Tanaka, Ryosuke Aoki, and Yukio Koike. 2021. Reducing Muscle Activity when Playing Tremolo by Using Electrical Muscle Stimulation to Learn Efficient Motor Skills. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (Sept. 2021), 1–17. doi:10.1145/3478110

[57] Jun Nishida, Shunichi Kasahara, and Kenji Suzuki. 2017. Wired muscle: generating faster kinesthetic reaction by inter-personally connecting muscles. In *ACM SIGGRAPH 2017 Emerging Technologies*. ACM, Los Angeles California, 1–2. doi:10.1145/3084822.3084844

[58] Jun Nishida, Yudai Tanaka, Romain Nith, and Pedro Lopes. 2022. DigituSync: A Dual-User Passive Exoskeleton Glove That Adaptively Shares Hand Gestures. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. ACM, Bend OR USA, 1–12. doi:10.1145/3526113.3545630

[59] NVIDIA, :, Johan Bjorck, Fernando Castañeda, Nikita Cherniadev, Xingye Da, Runyu Ding, Linxi "Jim" Fan, Yu Fang, Dieter Fox, Fengyuan Hu, Spencer Huang, Joel Jang, Zhenyu Jiang, Jan Kautz, Kaushil Kundalia, Lawrence Lao, Zhiqi Li, Zongyu Lin, Kevin Lin, Guilin Liu, Edith Llontop, Loic Magne, Ajay Mandlekar, Avnish Narayan, Soroush Nasiriany, Scott Reed, You Liang Tan, Guanzhi Wang, Zu Wang, Jing Wang, Qi Wang, Jiannan Xiang, Yuqi Xie, Yinzhen Xu, Zhenjia Xu, Seonghyeon Ye, Zhiding Yu, Ao Zhang, Hao Zhang, Yizhou Zhao, Ruijie Zheng, and Yuke Zhu. 2025. GR00T N1: An Open Foundation Model for Generalist Humanoid Robots. doi:10.48550/ARXIV.2503.14734 Version Number: 2.

[60] Jiyoung Park, Dohoon Kwak, Jaewan Lim, Taeyoon Lee, Seung-gyeom Kim, and Yongjae Yoo. 2025. To Guide or to Disturb - How to Teach Dexterous Skills Using AI?. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*. ACM, Cagliari Italy, 1458–1469. doi:10.1145/3708359.3712097

[61] Pauline Ploettner, Christoph Muehlberg, Felix Psurek, Christopher Fricke, and Jost-Julian Rumpf. 2025. Divergent effects of pitch feedback on online and offline motor sequence learning. *Frontiers in Behavioral Neuroscience* 19 (Oct. 2025), 1680277. doi:10.3389/fnbeh.2025.1680277

[62] Meir Plotnik, Oran Ben-Gal, Glen M. Doniger, Amihai Gottlieb, Yotam Bahat, Maya Cohen, Shani Kimel-Naor, Gabi Zeilig, and Michal Schnaider Beeri. 2021. Multimodal immersive trail making-virtual reality paradigm to study cognitive-motor interactions. *Journal of NeuroEngineering and Rehabilitation* 18, 1 (May 2021), 82. doi:10.1186/s12984-021-00849-9

[63] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. doi:10.48550/ARXIV.2103.00020 Version Number: 1.

[64] Eitan Raveh, Jason Friedman, and Sigal Portnoy. 2018. Evaluation of the effects of adding vibrotactile feedback to myoelectric prosthesis users on performance and visual attention in a dual-task paradigm. *Clinical Rehabilitation* 32, 10 (Oct. 2018), 1308–1316. doi:10.1177/0269215518774104

[65] Jun Rekimoto. 2025. GazeLLM: Multimodal LLMs incorporating Human Visual Attention. In *Proceedings of the Augmented Humans International Conference 2025*. ACM, Masdar City, Abu Dhabi United Arab Emirates, 302–311. doi:10.1145/3745900.3746075

[66] Makoto Sato, Yukihiro Hirata, Hiroshi Kawarada, and Precision and Intelligent Laboratory, Tokyo Institute of Technology, 4259, Nagatsuda, Midori-ku, Yokohama-shi, 226 Japan. 1997. Space Interface Device for Artificial Reality – SPIDAR –. *Journal of Robotics and Mechatronics* 9, 3 (June 1997), 177–184. doi:10.20965/jrm.1997.p0177

[67] Tomonori Sawada, Hiroki Okawara, Daisuke Nakashima, Kaito Ikeda, Joji Nagahara, Haruki Fujitsuka, Sosuke Hoshino, Yuta Maeda, Yoshinori Katsumata, Masaya Nakamura, and Takeo Nagura. 2022. Constant Load Pedaling Exercise Combined with Electrical Muscle Stimulation Leads to an Early Increase in Sweat Lactate Levels. *Sensors* 22, 24 (Dec. 2022), 9585. doi:10.3390/s22249585

[68] Alessandra Semeraro and Laia Turmo Vidal. 2022. Visualizing Instructions for Physical Training: Exploring Visual Cues to Support Movement Learning from Instructional Videos. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–16. doi:10.1145/3491102.3517735

[69] Hyeon-Seok Seong, Do-Hyeong Kim, Igor Gaponov, and Jee-Hwan Ryu. 2020. Development of a Twisted String Actuator-based Exoskeleton for Hip Joint Assistance in Lifting Tasks. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, Paris, France, 761–767. doi:10.1109/ICRA40945.2020.9197359

[70] Giulia Sgubin, Manuela Deodato, and Luigi Murena. 2023. Gamification in Rehabilitation: The Role of Subjective Experience in a Multisensory Learning Context – A Narrative Review. *Gestalt Theory* 45, 1-2 (Aug. 2023), 121–137. doi:10.2478/gth-2023-0012

[71] Charles H Shea and Gabriele Wulf. 1999. Enhancing motor learning through external-focus instructions and feedback. *Human Movement Science* 18, 4 (Aug. 1999), 553–571. doi:10.1016/S0167-9457(99)00031-7

[72] Vivian Shen and Chris Harrison. 2025. Kinethreads: Soft Full-Body Haptic Exosuit using Low-Cost Motor-Pulley Mechanisms. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. ACM, Busan Republic of Korea, 1–16. doi:10.1145/3746059.3747755

[73] Yongmin Shin, Jaeseo Lim, Yonggwan Kim, Deog-Gyu Seo, and Jungjoon Ihm. 2022. Effects of virtual body-representation on motor skill learning. *Scientific Reports* 12, 1 (Sept. 2022), 15283. doi:10.1038/s41598-022-19514-9

[74] Antun Skuric, Hasan Sinan Bank, Richard Unger, Owen Williams, and David González-Reyes. 2022. SimpleFOC: A Field Oriented Control (FOC) Library forControlling Brushless Direct Current (BLDC) and Stepper Motors. *Journal of Open Source Software* 7, 74 (June 2022), 4232. doi:10.21105/joss.04232

[75] Suzanne J. Snodgrass, Darren A. Rivett, Val J. Robertson, and Elizabeth Stojanovski. 2010. Real-time feedback improves accuracy of manually applied forces during cervical spine mobilisation. *Manual Therapy* 15, 1 (Feb. 2010), 19–25. doi:10.1016/j.math.2009.05.011

[76] Tasha R. Stanton and Charles Spence. 2020. The Influence of Auditory Cues on Bodily and Movement Perception. *Frontiers in Psychology* 10 (Jan. 2020), 3001. doi:10.3389/fpsyg.2019.03001

[77] Tim Claudius Stratmann, Andreas Löcken, Uwe Gruenefeld, Wilko Heuten, and Susanne Boll. 2018. Exploring Vibrotactile and Peripheral Cues for Spatial Attention Guidance. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays*. ACM, Munich Germany, 1–8. doi:10.1145/3205873.3205874

[78] Qiming Sun and Pinata Winoto. 2019. The Design of an Intelligent LEGO Tutoring System for Improving Social Communication Skills Among Children with Autism Spectrum Disorder. In *Universal Access in Human-Computer Interaction. Multimodality and Assistive Environments*, Margherita Antona and Constantine Stephanidis (Eds.). Vol. 11573. Springer International Publishing, Cham, 172–181. doi:10.1007/978-3-030-23563-5_15 Series Title: Lecture Notes in Computer Science.

[79] Nobuhiro Takahashi, Shinichi Furuya, and Hideki Koike. 2020. Soft Exoskeleton Glove with Human Anatomical Architecture: Production of Dexterous Finger Movements and Skillful Piano Performance. *IEEE Transactions on Haptics* 13, 4 (Oct. 2020), 679–690. doi:10.1109/TOH.2020.2993445

[80] Yudai Tanaka, Jun Nishida, and Pedro Lopes. 2022. Electrical Head Actuation: Enabling Interactive Systems to Directly Manipulate Head Orientation. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–15. doi:10.1145/3491102.3501910

[81] Yudai Tanaka, Akifumi Takahashi, and Pedro Lopes. 2023. Interactive Benefits from Switching Electrical to Magnetic Muscle Stimulation. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–12. doi:10.1145/3586183.3606812

[82] Sho Tatsuno, Tomohiko Hayakawa, and Masatoshi Ishikawa. 2017. Trajectory adjustment system for learning based on electrical stimulation. In *Proceedings of the 8th Augmented Human International Conference*. ACM, Silicon Valley California USA, 1–4. doi:10.1145/3041164.3041197

[83] Issatay Tokmurziyev, Miguel Altamirano Cabrera, Muhammad Haris Khan, Yara Mahmoud, and Dzmitry Tsetserukou. 2025. LLM-Glasses: GenAI-driven Glasses with Haptic Feedback for Navigation of Visually Impaired People. doi:10.48550/ARXIV.2503.16475 Version Number: 2.

[84] Willem B. Verwey. 2015. Contributions from associative and explicit sequence knowledge to the execution of discrete keying sequences. *Acta Psychologica* 157 (May 2015), 122–130. doi:10.1016/j.actpsy.2015.02.013

[85] Yuxuan Wan, Kaichen Zhou, jinhong Chen, and Hao Dong. 2024. SCANet: Correcting LEGO Assembly Errors with Self-Correct Assembly Network. doi:10.48550/ARXIV.2403.18195 Version Number: 4.

[86] Chunji Wang, Yupeng Xiao, Etienne Burdet, James Gordon, and Nicolas Schweighofer. 2016. The duration of reaching movement is longer than predicted by minimum variance. *Journal of Neurophysiology* 116, 5 (Nov. 2016), 2342–2345. doi:10.1152/jn.00148.2016

[87] Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Yang Fan, Kai Dang, Mengfei Du, Xuancheng Ren, Rui Men, Dayiheng Liu, Chang Zhou, Jingren Zhou, and Junyang Lin. 2024. Qwen2-VL: Enhancing Vision-Language Model's Perception of the World at Any Resolution. doi:10.48550/ARXIV.2409.12191 Version Number: 2.

[88] Eva C Wentink, Eline J Talsma-Kerkdijk, Hans S Rietman, and Peter Veltink. 2015. Feasibility of error-based electrotactile and auditive feedback in prosthetic walking. *Prosthetics & Orthotics International* 39, 3 (June 2015), 255–259. doi:10.1177/0309364613520319

[89] Camille K. Williams, Luc Tremblay, and Heather Carnahan. 2016. It Pays to Go Off-Track: Practicing with Error-Augmenting Haptic Feedback Facilitates Learning of a Curve-Tracing Task. *Frontiers in Psychology* 7 (Dec. 2016). doi:10.3389/fpsyg.2016.02010

[90] Carolee J. Winstein, Patricia S. Pohl, and Rebecca Lewthwaite. 1994. Effects of Physical Guidance and Knowledge of Results on Motor Learning: Support for the Guidance Hypothesis. *Research Quarterly for Exercise and Sport* 65, 4 (Dec. 1994), 316–323. doi:10.1080/02701367.1994.10607635

[91] Carolee J. Winstein and Richard A. Schmidt. 1990. Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16, 4 (July 1990), 677–691. doi:10.1037/0278-7393.16.4.677

[92] Daniel M. Wolpert, Jörn Diedrichsen, and J. Randall Flanagan. 2011. Principles of sensorimotor learning. *Nature Reviews Neuroscience* 12, 12 (Dec. 2011), 739–751. doi:10.1038/nrn3112

[93] Gabriele Wulf. 2013. Attentional focus and motor learning: a review of 15 years. *International Review of Sport and Exercise Psychology* 6, 1 (Sept. 2013), 77–104. doi:10.1080/1750984X.2012.723728

[94] Gabriele Wulf, Markus Höß, and Wolfgang Prinz. 1998. Instructions for Motor Learning: Differential Effects of Internal Versus External Focus of Attention. *Journal of Motor Behavior* 30, 2 (June 1998), 169–179. doi:10.1080/00222899809601334

[95] Gabriele Wulf and Rebecca Lewthwaite. 2016. Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic Bulletin & Review* 23, 5 (Oct. 2016), 1382–1414. doi:10.3758/s13423-015-0999-9

[96] Gabriele Wulf, Nancy McNevin, and Charles H. Shea. 2001. The automaticity of complex motor skill learning as a function of attentional focus. *The Quarterly Journal of Experimental Psychology Section A* 54, 4 (Nov. 2001), 1143–1154. doi:10.1080/713756012

[97] Wei Yan. 2022. Augmented reality instructions for construction toys enabled by accurate model registration and realistic object/hand occlusions. *Virtual Reality* 26, 2 (June 2022), 465–478. doi:10.1007/s10055-021-00582-7

[98] Xing-Dong Yang, Walter F. Bischof, and Pierre Boulanger. 2008. Validating the Performance of Haptic Motor Skill Training. In *2008 Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. IEEE, Reno, NV, USA, 129–135. doi:10.1109/HAPTICS.2008.4479929

[99] Sotaro Yokoi and Jun Rekimoto. 2025. MermaidLLM: Dataflow Diagrams for Explainable Skill Formalization and Real-time Support with Multimodal LLMs. In *Adjunct Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. ACM, Busan Republic of Korea, 1–3. doi:10.1145/3746058.3758449