


# Adaptive Modality Systems in Extended Reality: Optimizing Input-Output Bandwidth Through Context-Aware Gaze-Hand Switching

Mohammad Dastgheib<sup>1</sup>   
University of California, Riverside  
[mohammad.dastgheib@email.ucr.edu](mailto:mohammad.dastgheib@email.ucr.edu)

Fatemeh Pourmahdian   
[fatemepourmahd@gmail.com](mailto:fatemepourmahd@gmail.com)

2025-12-08

<sup>1</sup>Corresponding author: [mdast003@ucr.edu](mailto:mdast003@ucr.edu)

## Abstract

Extended Reality (XR) systems require users to engage their entire body for interaction, imposing significant ergonomic and cognitive demands. Current XR interfaces force a binary choice between controller-based interaction (suffering from “Gorilla Arm” fatigue) and gaze-based interaction (plagued by the “Midas Touch” problem and precision limitations). This manuscript introduces the theoretical foundation and methodological implementation of an Adaptive Modality System that dynamically switches between manual (hand-pointing) and ocular (gaze-based) input modalities based on real-time assessments of task difficulty and user cognitive state. The system aims to optimize the “Input-Output Bandwidth” of the human-computer loop by leveraging Fitts’s Law predictions and adaptive UI interventions. Through a rigorous within-subjects experimental design utilizing the ISO 9241-9 multi-directional tapping task with physiologically-accurate gaze simulation, we investigate whether context-aware adaptive systems yield higher throughput than static unimodal systems and reduce workload on the NASA-TLX.

**Keywords:** Extended Reality, Virtual Reality, Adaptive Interfaces, Gaze Interaction, Fitts’s Law, Human-Computer Interaction, Multimodal Interaction, Gaze Simulation

# 1 Introduction

The transition from two-dimensional graphical user interfaces (GUIs) to three-dimensional spatial computing represents one of the most profound shifts in the history of human-computer interaction (HCI). Extended Reality (XR)—an umbrella term encompassing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR)—promises to liberate digital information from the confines of flat screens, embedding it directly into the user’s physical environment. However, this liberation imposes new and significant demands on the human sensorimotor system. Unlike the desktop metaphor, where interaction is mediated by low-effort devices like the mouse and keyboard, XR requires the user to engage their entire body. The “input space” is no longer a mousepad but the physical volume surrounding the user, and the primary pointing devices are the user’s own hands and eyes.

While this embodied interaction offers unprecedented intuitive potential, it is fraught with ergonomic and cognitive challenges. The “Gorilla Arm” syndrome, a phenomenon where prolonged mid-air arm extension leads to rapid musculoskeletal fatigue and pain, remains a critical barrier to the long-term adoption of gestural interfaces. Conversely, gaze-based interaction, which leverages the speed of the human oculomotor system, suffers from the “Midas Touch” problem—the inherent ambiguity between looking for perception and looking for action—and lacks the fine motor precision required for granular manipulation tasks (Jacob, 1990).

Current XR systems typically force a binary choice: the user must either commit to a controller-based paradigm, accepting the physical fatigue, or a gaze-based paradigm, accepting the lack of precision and the potential for inadvertent triggers. This rigid dichotomy ignores the dynamic nature of human attention and the varying demands of different tasks. A high-precision manipulation task may require the stability of a hand controller, while a rapid visual search task is best served by the saccadic speed of the eye.

This manuscript introduces the theoretical foundation and methodological implementation of the xr-adaptive-modality-2025 platform, a novel research framework designed to investigate Adaptive Modality Systems in XR. Extended Reality (XR) interfaces increasingly integrate multiple input modalities (e.g., eye-gaze and hand pointing) to enhance usability, but optimizing performance across modalities remains challenging. Human perceptual-motor limitations—from sensor noise

to cognitive fatigue—can hinder precise and rapid interactions in XR. For example, gaze-based pointing is fast but prone to jitter and Midas touch issues, while hand controllers offer precision at the cost of speed.

To address these issues, we explore adaptive UI interventions that dynamically adjust to the user’s state and input modality: specifically, a **declutter mechanism** for gaze-based selection and a **width inflation** (target expansion) mechanism for hand-based selection. These adaptations aim to mitigate modality-specific weaknesses (gaze distraction and hand targeting difficulty) under varying cognitive load. By dynamically switching between manual (hand-pointing) and ocular (gaze-based) input modalities based on real-time assessments of task difficulty and user cognitive state, this system aims to optimize the “Input-Output Bandwidth” of the human-computer loop. The research leverages the ISO 9241-9 standard for evaluating non-keyboard input devices to rigorously quantify performance (Soukoreff & MacKenzie, 2004).

## 2 Background and Related Work

### 2.1 The Sensorimotor Implications of Spatial Input

To design an effective adaptive system, one must first deconstruct the physiological mechanisms of the component modalities. The human motor system and the visual system evolved for distinct evolutionary purposes: the hand for manipulation and the eye for information acquisition. Forcing either organ to perform the function of the other introduces friction.

**Manual Pointing in XR:** Manual input in XR, whether through held controllers or optical hand tracking, mimics the act of physical pointing. This interaction style benefits from proprioception—the body’s innate sense of limb position—which allows for high-precision corrections without visual attention. However, the biomechanical cost is substantial. In a 1:1 mapped XR environment, reaching a virtual object requires a corresponding physical motion. Frequent large-amplitude movements lead to fatigue in the deltoids and trapezius muscles. As fatigue sets in, the signal-to-noise ratio of the motor system degrades; the hand begins to tremor, increasing the effective target width required for accurate selection and reducing the overall throughput of the interaction.

**Gaze Interaction:** The eye is the fastest motor organ in the human body. Saccades—rapid,

ballistic movements of the eye—can reach velocities exceeding 900 degrees per second (Bahill & Stark, 1979), making gaze an incredibly efficient modality for target acquisition. However, the eye is fundamentally an input organ, not an output device. Using gaze for selection introduces several critical issues: (1) **The Midas Touch**—in the physical world, we can look at an object without interacting with it, but in a gaze-controlled interface, looking becomes equivalent to touching, requiring “dwell time” mechanisms that slow interaction (Jacob, 1990); (2) **Microsaccades and Jitter**—even when “fixated,” the eye performs microsaccades to refresh the retinal image (Martinez-Conde et al., 2004), meaning a gaze cursor is inherently noisy, making selection of small targets frustrating without smoothing algorithms; (3) **Saccadic Suppression**—during rapid eye movements, the visual system suppresses input to prevent motion blur (Bridgeman et al., 1975), creating a “blind” phase that makes the initial phase of gaze targeting effectively open-loop.

## 2.2 Signal Processing and Cognitive Load Theory

To study these dynamics controllably, our platform employs a generative simulation rather than raw sensor input. This simulation introduces three key constraints derived from oculomotor physiology: (1) **Sensor Lag**—a first-order lag (linear interpolation) mimics the processing latency (30–70 ms) typical of video-based eye trackers (Saunders & Woods, 2014); (2) **Saccadic Blindness**—the cursor is frozen during high-velocity movements ( $>120$  deg/s), simulating the lack of visual feedback during a saccade (Bridgeman et al., 1975); (3) **Fixation Jitter**—Gaussian noise is injected at low velocities to mimic fixational drift and tremor (Martinez-Conde et al., 2004), ensuring that the “cost” of gaze interaction is accurately represented even when using a mouse proxy.

The efficiency of an interface is not measured solely by speed or accuracy, but by the cognitive resources consumed. Cognitive Load Theory (CLT) distinguishes between intrinsic load (task difficulty) and extraneous load (interface difficulty). The proposed Adaptive Modality System posits that the optimal input modality minimizes extraneous load: for distant, large targets, the motor cost of reaching is high (extraneous physical load), making gaze interaction superior; for close, small targets, the precision requirement is high, and gaze interaction imposes high extraneous load due to jitter, making manual interaction superior. Quantifying cognitive load in real-time is challenging, and subjective measures like the NASA-TLX are retrospective (Hart, 2006). For this study, we rely primarily on subjective workload measures (NASA-TLX) and performance metrics to assess

the efficacy of adaptive interventions.

### 2.3 Adaptive Intervention Mechanisms

We implemented two modality-specific adaptive interventions. **For gaze interaction**, a declutter mechanism draws on visual attention guiding techniques in XR, related to diminished reality (DR) (Herling & Broll, 2010). When the user operates in gaze mode, the interface fades out non-target objects, mitigating peripheral distraction. This aligns with foveated rendering principles, where systems leverage the human visual system’s foveal focus to prioritize content at the gaze point (Patney et al., 2016). By decluttering the periphery, we hypothesize that gaze-based targeting becomes faster and less cognitively demanding.

**For hand-based interactions**, our adaptive strategy is width inflation—dynamically expanding the effective size of targets when the cursor approaches. This concept is inspired by “expanding targets” research (McGuffin & Balakrishnan, 2005) and the “Bubble Cursor” (Grossman & Balakrishnan, 2005), which demonstrate that even slight, transient enlargement significantly improves pointing performance. Our implementation uses a hysteresis threshold to prevent flicker, acting as a “safety net” that compensates for motor tremor under fatigue. The expansion occurs only when the system detects the user is aiming at the target (cursor nearing the target center), to prevent overlapping targets in the dense circle layout.

## 3 Research Objectives

The primary objective of this research is to validate the efficacy of the `xr-adaptive-modality-2025` platform. The study is guided by three core research questions:

**RQ1 (Performance):** Does a context-aware adaptive system yield higher Throughput ( $TP$ ) than static unimodal systems?

**RQ2 (Workload):** Can adaptive modality switching significantly reduce “Physical Demand” and “Frustration” (NASA-TLX) compared to traditional interaction?

**RQ3 (Adaptation):** Do adaptive interventions (declutter, width inflation) significantly improve performance and reduce workload compared to non-adaptive conditions?

## 4 Theoretical Framework

This section establishes the mathematical and psychological models that underpin the system’s logic, specifically focusing on Information Theory as applied to human movement and the neuroergonomics of attention.

### 4.1 Fitts’s Law: The Information Capacity of the Human Motor System

Fitts’s Law (1954) is robustly established as the governing dynamic of pointing tasks. It frames movement not as a physical event, but as an information transmission task. The “difficulty” of a target is measured in bits, representing the amount of information the motor system must process to resolve the movement successfully. We adopt the **Shannon Formulation**, the standard for ISO 9241-9 compliance, as it better models the information entropy of the task and avoids negative ID values for very close targets:

$$ID = \log_2 \left( \frac{D}{W} + 1 \right)$$

Movement Time ( $MT$ ) is then modeled as a linear function of this difficulty:

$$MT = a + b \cdot ID$$

Where the intercept  $a$  represents non-informational additive components (e.g., reaction time, system latency) and the slope  $b$  represents the rate of information processing. The reciprocal of the slope,  $1/b$ , is often referred to as the **Index of Performance (IP)**, or bandwidth, measured in bits per second.

### 4.2 The Speed-Accuracy Tradeoff in 3D

In XR, the “Width” ( $W$ ) of a target is ambiguous. To standardize performance across users with varying error rates, the `xr-adaptive-modality-2025` platform utilizes the **Effective Width** ( $W_e$ ). Unlike nominal width,  $W_e$  is derived from the spatial distribution of selection endpoints.

Consistent with ISO 9241-9, we calculate  $W_e$  using the standard deviation of selection coordinates

projected onto the task axis. For a target at position  $P_{target}$  and a starting position  $P_{start}$ , the task axis vector is  $\vec{v} = P_{target} - P_{start}$ . The projected error  $x_i$  for a trial  $i$  with endpoint  $P_{hit}$  is computed as the scalar projection:

$$x_i = \frac{(P_{hit} - P_{target}) \cdot \vec{v}}{\|\vec{v}\|}$$

The effective width is then calculated as  $W_e = 4.133 \times \sigma_x$ , where  $\sigma_x$  is the standard deviation of these projected errors. This effectively normalizes the error rate to 4%, allowing us to calculate **Throughput** ( $TP$ ), a unified metric of speed and accuracy:

$$TP = \frac{ID_e}{MT} = \frac{\log_2(D/W_e + 1)}{MT}$$

This metric is critical for comparing Modality A (Gaze) vs. Modality B (Hand). Gaze might have a lower  $MT$  due to saccadic speed, but if its jitter results in a massive  $W_e$  (low accuracy), the overall Throughput will be lower.

### 4.3 Modeling Movement and Decision Processes

To decompose the underlying processes of movement execution and decision verification, we employ a **Hybrid Analysis Framework** that separates the ballistic trajectory from the final selection decision.

#### 4.3.1 Control Theory and Submovement Models

The **Optimized Submovement Model** (Meyer et al., 1988) posits that pointing movements are composed of a primary ballistic impulse followed by  $n$  corrective submovements. The total movement time is the sum of the primary movement duration and the duration of subsequent corrections required to bring the endpoint error within the target bounds.

We quantify the “cost of control” by analyzing the velocity profile  $v(t)$  of the cursor. A submovement is mathematically identified as a zero-crossing in the acceleration profile (or a local maximum in velocity) after the initial ballistic phase. The **Submovement Count** ( $N_{sub}$ ) serves as a proxy for

the efficiency of the control loop:

$$N_{sub} = \sum_{t=0}^T \mathbb{1}(\text{is\_velocity\_peak}(t)) - 1$$

In gaze-based interaction, simulated lag and saccadic blindness force users into an intermittent control regime, theoretically increasing  $N_{sub}$ .

### 4.3.2 The Linear Ballistic Accumulator (LBA) Model

Once the cursor enters the target, users face a decision verification problem: “Is the cursor stable enough to click?” We model this using the **Linear Ballistic Accumulator (LBA)** (Brown & Heathcote, 2008). LBA treats the decision as a race between independent accumulators (e.g., “Select” vs. “Wait/Drift”).

For the  $i$ -th accumulator, evidence  $x_i(t)$  accumulates linearly over time  $t$  from a starting point  $A$  with a drift rate  $d_i$ :

$$x_i(t) = A + d_i \cdot t$$

The starting point  $A$  is drawn from a uniform distribution  $U[0, A_{max}]$ , and the drift rate  $d_i$  is drawn from a normal distribution  $N(v_i, s)$ . A decision is made when the evidence reaches a threshold  $b$ .

Unlike Drift Diffusion Models (DDM) which require substantial error rates to constrain the boundary separation (Lerche et al., 2017), LBA is robust to the low-error scenarios typical of Fitts’ tasks. It allows us to explicitly estimate the **Caution Threshold** ( $b - A$ )—the amount of evidence users require before committing to selection. We hypothesize that adaptive interventions (e.g., snapping) reduce the cognitive need for caution, mathematically reflecting a lower estimated threshold  $b$ .

## 5 Methods

The methodology described herein is designed to be exhaustive and reproducible, adhering to the highest standards of empirical HCI research. The platform **xr-adaptive-modality-2025** serves as



the technical apparatus for the study.

## 5.1 Apparatus and Participants

We developed a custom pointing testbed as a web-based application (React 18, TypeScript), allowing broad hardware compatibility for remote participants. The study was conducted on participants’ own computers using a standard mouse or trackpad.

### 5.1.1 Display Calibration and Reliability Measures

To ensure measurement validity across heterogeneous display configurations, we implemented a multi-layered approach addressing display variability. Before commencing experimental trials, participants completed a **Credit Card Calibration** procedure: participants placed a standard credit card (85.60 mm  $\times$  53.98 mm) on their screen and adjusted an on-screen rectangle to match its physical dimensions. This calibration enabled computation of pixels per millimeter (px/mm) and pixels per degree of visual angle (PPD), normalizing gaze simulation jitter to screen-space pixels and ensuring consistent perceptual difficulty across different display sizes and resolutions (MacKenzie, 1992).

To minimize measurement error, we enforced strict display requirements: fullscreen/maximized window (required before starting blocks), browser zoom locked to 100% (verified before each block using `window.visualViewport.scale`), and live monitoring during trials (trials automatically paused if settings changed). For every trial, we logged comprehensive display metadata: device pixel ratio (DPR), browser type, viewport dimensions, zoom level, fullscreen status, and tab visibility duration. Trials were excluded from analysis if zoom level  $\neq$  100%, fullscreen status = FALSE, DPR instability (change  $>$  0.1 between blocks), or tab hidden for  $>$  500ms. Participants with  $>$ 40% of trials excluded due to display violations were removed from the final analysis.

### 5.1.2 Gaze Simulation (The “Ground Truth” Signal)

To ensure rigorous internal validity and precise control over noise characteristics, we utilized a **Physiologically-Accurate Gaze Simulation**. This approach allowed us to model the specific constraints of eye-tracking interaction—latency and jitter—with precise control over noise characteristics. The simulation transformed raw mouse input into “gaze” coordinates via three mechanisms

derived from oculomotor physiology:

1. **Saccadic Suppression & Ballistic Movement:** The cursor was “frozen” (blind) during high-velocity movements ( $>120$  deg/s) to simulate the brain’s suppression of visual input during saccades, a phenomenon known as saccadic suppression of image displacement (Bridgeman et al., 1975). This aligns with the ballistic nature of saccadic eye movements, where visual feedback is effectively open-loop until the eye settles.
2. **Fixation Jitter & Drift:** When the cursor slowed ( $<30$  deg/s), Gaussian noise (SD  $0.12^\circ$  visual angle) was injected to simulate fixational eye movements (Martinez-Conde et al., 2004), specifically the random walk characteristics of ocular drift and tremor that occur even during attempted fixation. This angular noise was normalized to screen pixels using the pixels-per-degree (PPD) calibration value, ensuring consistent perceptual difficulty across different display sizes and viewing distances.
3. **Sensor Lag:** A first-order lag (linear interpolation, factor 0.15) was applied to mimic the processing latency (typically 30–70 ms) inherent in video-based eye trackers (Saunders & Woods, 2014).

## 5.2 Input Modality Implementations

The system implemented two distinct input modalities, each with static and adaptive modes. **For hand-based trials**, participants controlled a cursor with their mouse to click the target. In static mode, this was standard 1:1 mouse pointing. In adaptive mode, the effective clickable area of the target was expanded (width scale  $> 1.0$ ) when the cursor approached the target—a “Gravity Well” effect that was invisible to the user but increased error tolerance for the final click. This implementation draws directly from the “expanding targets” technique validated by McGuffin and Balakrishnan (McGuffin & Balakrishnan, 2005). Expansion occurred only when the system detected the user was aiming at the target (cursor nearing the target center), to prevent overlapping targets in the dense circle layout.

**For gaze-based trials**, participants controlled the cursor via the simulated gaze signal described above. In static mode, selection was triggered by a **Dwell** mechanism (typically 350-500ms) or a confirmation key (Spacebar) if dwell was disabled, following the standard “dwell-to-select” paradigm

(Majaranta et al., 2006; Ware & Mikaelian, 1987). In adaptive mode, adaptation focused on **Goal-Aware Snapping** or **Dwell Time Adjustment** to counteract the injected jitter during the verification phase. For gaze declutter, the trigger was immediate at trial start in adaptive blocks—the moment a new target appeared and the user began aiming, all non-targets dimmed.

Both adaptive features (declutter and expansion) were driven by a simple rule-based policy with slight hysteresis to prevent flicker (e.g., once expanded, a target stayed expanded for the remainder of the trial). The application logged detailed event traces (gaze enter/exit target, click times, etc.), enabling us to later verify that the adaptive mechanisms functioned as intended.

### 5.3 Task and Stimuli

Participants performed a multi-directional pointing task conforming to the ISO 9241-9 standard for non-keyboard input device evaluation (ISO 9241-9, 2000). Targets were arranged in a circular layout with 8 positions (width  $W$ , amplitude  $A$ ), with one target highlighted at a time. Targets were presented with IDs ranging from approximately 2 to 6 bits, calculated using the Shannon formulation of Fitts’ Law (target width ranged from 30 px to 80 px, with corresponding distances chosen to yield the desired ID values). In half of the blocks, a “Time Pressure” condition was enforced via a visible countdown timer; failure to select within the timeout (6s) resulted in a forced error, intended to induce stress and mental workload, simulating a high-demand scenario.

### 5.4 Experimental Design

We employed a repeated-measures factorial design: all participants experienced every combination of the two input modalities (Gaze vs. Hand)  $\times$  two UI conditions (Adaptive vs. Non-adaptive)  $\times$  two workload levels (Pressure vs. No Pressure). This creates a  $2 \times 2 \times 2$  within-subjects design.

#### 5.4.1 Counterbalancing: The Williams Design

The order of modality blocks was counterbalanced using a Williams Latin square arrangement to control for learning effects (Williams, 1949). Within each modality block (lasting 10-15 minutes), adaptive and control trials were intermixed in random order or run as separate sub-blocks (counterbalanced) to avoid long streaks of one condition. Participants were not explicitly told when the system was adapting, aside from noticing the visual changes, to reduce expectancy biases.

## 5.5 Participants

### 5.5.1 Sample Size and Power Analysis

Sample size was determined through a combination of power analysis and Williams design constraints. For a  $2 \times 2 \times 2$  within-subjects design with an expected medium-to-large effect size (Cohen’s  $d = 0.60$ ), power of 0.80, and  $\alpha = 0.05$  (two-tailed), a minimum of  $N = 24$  participants would be required for a simple repeated-measures design. However, to ensure proper counterbalancing with the Williams Latin square arrangement, the sample size must be a multiple of the number of treatment sequences. For our  $2 \times 2 \times 2$  factorial design, we require 8 distinct counterbalancing sequences. We selected  $N = 48$  participants, which allows for 6 complete Williams blocks (8 participants per block), ensuring balanced representation across all treatment orders while providing increased statistical power ( $\approx 0.95$  for  $d = 0.60$  in a within-subjects design) and robustness to potential data exclusions.

### 5.5.2 Recruitment

We recruited  $N = 48$  participants (target: balanced gender distribution, age range 18-35) from the university community. All participants had normal or corrected-to-normal vision and no known motor impairments. The study was approved by the Institutional Review Board (IRB) and all participants provided informed consent.

## 5.6 Procedure

Each participant completed a short training session to get familiar with gaze selection (including practice with the simulated gaze interface and dwell clicking) and hand selection. During the experiment, they performed multiple trials for each target difficulty and condition. In total, participants completed approximately  $3 \text{ (IDs)} \times 8 \text{ (directions)} \times 2 \text{ (pressure levels)} \times 2 \text{ (adaptation conditions)}$  96 trials per modality, leading to  $\sim 192$  trials per person overall.

After each block, participants filled out a NASA-TLX workload survey (rating mental demand, physical demand, etc.) and took a short break to mitigate fatigue. We also collected qualitative feedback at the end about their preferences (gaze vs hand, adaptive vs not, perceived helpfulness of declutter/expansion). The entire session lasted about 1 hour per participant.

## 5.7 Measures and Analysis Strategy

The primary performance measures were Movement Time (MT) in milliseconds (from trial start to successful selection) and Selection Accuracy (hit vs. miss rate, including specific error types: misses, timeouts, false activations). From these, we computed Throughput (TP) in bits/second for each condition by combining each trial’s effective ID ( $ID_e$ ) and MT, following ISO 9241-9’s recommended calculation using the effective target width  $W_e$  (computed from the spread of hit points) to incorporate accuracy into the difficulty measure.

Given the specific nature of the simulated noise (generative Gaussian jitter) and the high accuracy of the task, standard Drift Diffusion Models (DDM) were ill-suited due to the scarcity of error trials (~3%). Previous methodological reviews indicate that diffusion models require a sufficient error density to reliably estimate boundary separation parameters (Lerche et al., 2017). We instead adopted a **Hybrid Analysis** framework to disentangle “motor” performance from “decision” caution: (1) **Macro-Performance (Fitts’ Law)**—we modeled the overall throughput (bits/s) and the regression slope ( $b$ ) to quantify the global cost of the gaze simulation lag, following standard ISO 9241-9 throughput calculations; (2) **Path Quality (Control Theory)**—we calculated submovement counts (velocity peaks) and target re-entries to quantify the “cost of control” (how often the user had to correct their trajectory due to the simulated lag and “saccadic blindness”), drawing on the Optimized Submovement Model (Meyer et al., 1988); (3) **Decision Verification (LBA)**—we used the Linear Ballistic Accumulator (LBA) model (Brown & Heathcote, 2008) to analyze the selection phase, which is robust to low error rates and allows accurate estimation of the “Caution Threshold” users adopted to overcome fixation jitter.

Data were analyzed using repeated-measures ANOVAs and linear mixed models. We tested for main effects of Modality, Adaptation, and Pressure, as well as their interactions. Where appropriate, pairwise comparisons with Bonferroni correction were performed to probe simple effects. All statistical tests assumed  $\alpha = 0.05$ .

## 5.8 Deployment and Data Collection Infrastructure

The experimental platform was deployed as a web application (React 18, TypeScript, Vite) hosted on Vercel (<https://xr-adaptive-modality-2025.vercel.app>) to enable remote, asynchronous data col-

lection. Each participant received a unique URL containing embedded participant ID and session number parameters (e.g., `?pid=P001&session=1`). The application automatically detected these parameters on load, initializing the session with the appropriate identifier. Session state was managed client-side using browser `localStorage`, enabling participants to pause and resume sessions while maintaining progress tracking.

Data collection occurred entirely client-side to ensure participant privacy. The application implemented a structured CSV logging system that captured comprehensive trial-level data in real-time (23+ columns per trial), including participant metadata (ID, demographics), trial parameters (block, trial, modality, UI mode, pressure, ID, A, W), performance metrics (RT, accuracy, error type, hover duration), system metadata (browser type, DPR, timestamp), and workload measures (NASA-TLX subscales). At the completion of each session, participants exported their data via browser-based CSV download. Data export occurred entirely locally—no data was transmitted to servers during the experimental session, ensuring participant privacy. The application was built as a single-page application (SPA) with client-side routing, with the gaze simulation algorithm, adaptive policy engine, and data logging system all operating in real-time within the browser. Event-driven architecture (via an internal event bus) coordinated trial timing, data logging, and UI updates, ensuring precise temporal alignment between user actions and recorded data. All code is open-source and available for reproducibility.

## 6 Results

*[Results section to be completed after data collection and analysis. This section will include:*

- *Descriptive statistics for all dependent variables across conditions*
- *Mixed-effects model results for RQ1 (Performance)*
- *NASA-TLX analysis for RQ2 (Workload)*
- *Analysis of adaptive intervention effects for RQ3 (Adaptation)*
- *Figures showing throughput by condition and ID level*
- *Tables summarizing statistical tests]*

## 7 Discussion

*[Discussion section to be completed after results. This section will:*

- *Interpret findings in the context of the research questions*
- *Compare results to prior work on multimodal XR interaction*
- *Discuss implications for adaptive interface design*
- *Address limitations of the current study*
- *Propose future directions for adaptive modality research]*

## 8 Conclusion

*[Conclusion section to be completed. This section will:*

- *Summarize key contributions*
- *Restate main findings*
- *Highlight practical implications for XR system designers*
- *Provide closing thoughts on the future of adaptive interfaces]*

## 9 Acknowledgments

This research was conducted as an independent project. The author thanks the participants for their time and the open-source community for the tools that made this work possible.

## 10 References

- Bahill, A. T., & Stark, L. (1979). The trajectories of saccadic eye movements. *Scientific American*, 240(1), 108–117.
- Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, 15(6), 719–722.
- Brown, S. D., & Heathcote, A. (2008). The linear ballistic accumulator model of simple choice and reaction time. *Cognitive Psychology*, 57(3), 153–178.
- Grossman, T., & Balakrishnan, R. (2005). The bubble cursor: Enhancing target acquisition by

- dynamic resizing of the cursor’s activation area. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 281–290.
- Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50, 904–908.
- Herling, J., & Broll, W. (2010). Advanced self-contained object removal for realizing real-time diminished reality in unconstrained environments. *2010 IEEE International Symposium on Mixed and Augmented Reality*, 207–212.
- ISO 9241-9: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) — Part 9: Requirements for Non-Keyboard Input Devices (2000).
- Jacob, R. J. (1990). What you look at is what you get: Eye movement-based interaction techniques. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 11–18.
- Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods*, 49, 513–537.
- MacKenzie, I. S. (1992). Fitts’ law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7(1), 91–139.
- Majaranta, P., MacKenzie, I. S., Aula, A., & R  ih  , K.-J. (2006). Effects of feedback and dwell time on eye typing speed and accuracy. *Universal Access in the Information Society*, 5(2), 199–214.
- Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5(3), 229–240.
- McGuffin, M., & Balakrishnan, R. (2005). Fitts’ law and expanding targets: Experimental studies and designs for user interfaces. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(4), 388–422.
- Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Smith, J. E. K. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review*, 95(3), 340–370.
- Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., Luebke, D., & Lefohn, A. (2016). Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)*, 35(6), 1–12.
- Saunders, D. R., & Woods, R. L. (2014). Direct measurement of the system latency of gaze-



- contingent displays. *Behavior Research Methods*, 46(2), 439–447.
- Soukoreff, R. W., & MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in HCI. *International Journal of Human-Computer Studies*, 61(6), 751–789.
- Ware, C., & Mikaelian, H. H. (1987). An evaluation of an eye tracker as a device for computer input. *Proceedings of the CHI+GI '87 Conference on Human Factors in Computing Systems and Graphics Interface*, 183–188.
- Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry*, 2(2), 149–168.