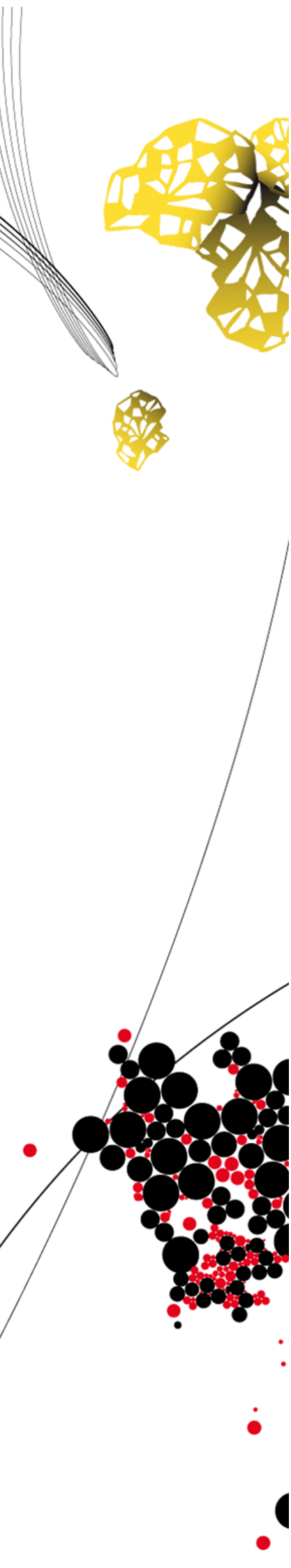


UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering, Mathematics & Computer Science

Beyond Buttons: Exploring the Design Potential of Wizard of Oz Elicited Hand Gestures for Fighting Game Control

Wahaj Ahmad

M.Sc. Interaction Technology Thesis
February 2026

Supervisors:
dr.ir Dennis Reidsma
dr. G.W.J. Bruinsma
D.P. Davison PhD

Human Media Interaction (HMI-EEMCS)
Faculty of Behavioural, Management and Social Sciences (BMS)
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Abstract

Gesture-based game control is often presented as a more natural alternative to traditional controllers, because it can connect what a player does physically with what happens on screen. In practice, however, gesture systems face recognition limits, segmentation issues, fatigue, and strict timing demands. These constraints often push designers toward small pre-defined gesture sets or choices based on intuition rather than evidence. This thesis studies the design potential of hand gestures for fast, time-critical gameplay using Wizard-of-Oz (WoZ) prototyping, which allows gesture interaction to be tested as if recognition already works while keeping the focus on human design and performance.

I use a fighting game context as a high-pressure testbed and run two WoZ-based experiments with a small participant pool. In Experiment 1, participants design and perform hand gestures for a set of fighting game actions. I analyze how gestures take shape when the mapping between the body and the avatar is indirect, and whether players converge on shared gesture solutions or keep gestures personal. In Experiment 2, I stress-test each participant's own gesture set under a speed ramp to examine how reliability and usability change as time pressure increases. I also study the felt experience of using gestures in a playable loop, with attention to intuitive physical interaction, sense of control, and factors that disrupt engagement.

The results show that WoZ elicitation captures more than a list of gesture shapes. Participants often approach the task as designing a repeatable command language rather than acting out avatar moves. Convergence appears in multiple forms, including cases where gestures remain diverse but still share stable intent cues. Under speed pressure, breakdowns are driven mainly by timing limits and throughput rather than loss of meaning. The immersive potential of gesture control appears strongest when mappings feel predictable and controllable, but it becomes fragile when interaction adds strain, delay, or attention overhead. These findings are limited to this study and its small sample, but they demonstrate that WoZ prototyping can provide actionable design evidence about gesture meaning, structure, and performance constraints before committing to full recognition.

Acknowledgements

I would like to sincerely thank my supervisors, Dennis Reidsma, Daniel Davison, and Guido Bruinsma, for their guidance, constructive feedback, and continued support throughout this thesis. Their questions and critiques strengthened both the direction and clarity of this work.

I am also grateful to my study advisor, Erik Bong, for his patience and practical support during periods of delay, and for helping me navigate administrative and visa-related challenges. His assistance allowed me to stay focused on completing this research.

To my wife, Tasbiha Asim, I owe more than words can express. Her unwavering support, encouragement, and belief in me carried me through the most difficult phases of this journey. Without her, this thesis would not have been completed.

I would also like to thank my friends, who stood by me during challenging times, listened to my frustrations, and reminded me why I started this journey in the first place. Their presence made the difficult periods manageable.

To my parents and my sisters, thank you for your patience, trust, and constant belief in me. Your support, even from afar, gave me stability and strength.

I dedicate this work to international students who quietly struggle with loneliness, displacement, and mental health challenges while pursuing their dreams far from home. The process of uprooting one's life and rebuilding it in a new country requires courage that often goes unseen.

I also dedicate this work to the people of Palestine, whose resilience, hope, and determination to build meaningful lives despite immense hardship serve as a reminder that perseverance and dignity endure even under the most difficult circumstances. Their strength has been a source of perspective and motivation throughout my own journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In most modern games, players control a virtual avatar by means of acting on tangible input devices. Typically, players operate bimanually, with one hand performing continuous tasks like moving a mouse and the other managing discrete tasks like pressing keyboard buttons (Leganchuk et al. 1998). These inputs work well, but they also force players to learn an arbitrary mapping between a physical action (pressing plastic buttons) and an in-game effect (a character move) (Isbister & Schaffer 2008, Preece et al. 2015). These movements rarely match the avatar's actions, such as walking, jumping, swimming, or climbing, which would involve full-body movement in real life. Even simple actions like punching or throwing an object typically require at least the use of the shoulder muscles to be performed, but are instead triggered by pressing a button with just a finger. While physical controllers allow for a wide range of in-game actions, they often lack a one-to-one correspondence between the player's input and the avatar's movements. Instead, the avatar performs pre-scripted actions as soon as the correct button is pressed (Ionescu et al. 2011). Even within the same genre, the same button can mean different actions across games, which turns control into a learned convention rather than something that is immediately readable from the body movement itself (Juul 2012). For fast-paced games, that convention still needs to support speed, precision, and confidence, because any delay or ambiguity in control shows up directly in performance and enjoyment (Claypool & Claypool 2006, Rogers et al. 2015).

Gesture-based control is often presented as an answer to this gap. Instead of translating intentions through a small set of abstract inputs, players can act using movements that already have meaning in everyday communication and action (Kendon 2004, McNeill 1992). Motion and body-based controllers have also shown that physical involvement can change how players feel during play, including their sense of engagement and presence (Bianchi-Berthouze

et al. 2007, Pasch et al. 2009*a*, Pietschmann et al. 2012). Controller type can shape enjoyment and motivation as well, which suggests that input is not only a technical layer but also part of the player experience (Birk & Mandryk 2013, Rogers et al. 2015).

At the same time, gesture control is not "free." A real system must recognize movement reliably across different bodies, styles, lighting conditions, camera viewpoints, and occlusions (Al-Shamayleh et al. 2018, Cheng et al. 2016, Mohamed et al. 2021).

Recognition pipelines also struggle with continuous streams of movement, where segmentation and transitional motion can blur what counts as the intended gesture (Yasen & Jusoh 2019, Alyami et al. 2024).

These constraints often push designers toward a small, pre-defined gesture set that is easier to detect and less risky to deploy (Sagayam & Hemanth 2017, Pisharady & Saerbeck 2015).

Even when tracking hardware improves, mid-air interaction can introduce its own usability costs, such as arm fatigue under repetition (Hincapié-Ramos et al. 2014).

Because of these practical limits, there is a gap between the promise of "natural" gesture control and what players actually end up doing in a playable system. A design team can guess what gestures might work, but gesture form is not trivial. Elicitation research repeatedly shows that some commands invite shared, population-level proposals while others stay diverse, even when the task is the same (Wobbrock et al. 2009, Ruiz et al. 2011, Villarreal-Narvaez et al. 2020). Proposals can also be shaped by prior exposure to existing interfaces, which means that what looks like an "intuitive" gesture may partly reflect legacy habits (Morris et al. 2014). If the goal is to understand what gesture-based control can offer, then the first step is not only to implement recognition, but to study the gesture design space itself: what people produce, what tends to converge, what stays personal, and which gestures remain usable when the pace increases (Vatavu 2019, Tsandilas 2018).

Despite the growing interest in gesture-based control as a more natural and engaging alternative to traditional game controllers, there is limited empirical understanding of how players actually design, perform, and sustain gestures in fast-paced, time-critical gameplay. In practice, gesture-based systems are often constrained by recognition challenges, fatigue, and usability concerns, which pushes designers toward small, pre-defined gesture sets or decisions based on intuition rather than evidence. As a result, there is a gap between the promise of expressive, body-based interaction and the reality of gesture control that remains learnable, responsive, and reliable under performance pressure. In particular, it is unclear which hand gestures players tend to

converge on, which remain diverse or personal, and how gesture form holds up when interaction speed and precision become critical.

Wizard-of-Oz (WoZ) prototyping offers a way to study these questions without committing to a full recognition system. Instead of relying on an automatic classifier, a human operator can interpret the user's action and trigger the system response in real time, which allows the interaction concept to be tested as if it already works (Höysniemi et al. 2004, Mai et al. 2011). This approach has been used directly for vision-based action games, where WoZ makes it possible to explore how players invent and perform gestures in a playable loop before the sensing and recognition problems are solved (Höysniemi et al. 2004, 2005). In other words, WoZ can isolate the design and human-performance questions (what gestures people choose, and what those gestures demand) from the engineering questions (how accurately a model can detect them).

In this thesis, I use fighting games as a high-pressure testbed rather than as the main topic. Fighting games make command input, timing, and error consequences very visible, which makes them useful for stress-testing any proposed control scheme (Mattiassi 2019). The broader motivation is not limited to fighting games. Many interactive systems depend on rapid, repeatable commands under time constraints, where a control method must stay learnable, physically manageable, and responsive (Claypool & Claypool 2006, Nielsen 1993). A gesture set that looks expressive at low speed may still fail when the interaction demands quick turn-taking and consistent execution (Heitz 2014).

Finally, gesture-based control is often discussed as a route to richer engagement because it links what the player does physically with what happens on screen (Bianchi-Berthouze et al. 2007, Pasch et al. 2009*a*). Core models of immersion and flow place value on involvement, challenge balance, and a stable sense of control (Brown & Cairns 2004, Ermi & Mäyrä 2005, Csikszentmihalyi et al. 2014). If a gesture controller increases physical and mental involvement but also introduces delay, confusion, or strain, then it may undermine the same experience it aims to improve (Birk & Mandryk 2013, Rogers et al. 2015). This makes it important to study gesture control as an integrated design problem: gesture form, convergence patterns, performance under speed, and the felt experience that comes from using the gestures in an interactive loop.

This thesis therefore asks:

What is the potential of hand gestures collected through Wizard of Oz prototyping in the design of a gesture-based control system for fighting games?

1.1 Main Research Question and Sub-Questions

The main research question of this thesis is:

> **What is the potential of hand gestures collected through Wizard of Oz prototyping in the design of a gesture-based control system for fighting games?**

I treat *potential* as practical. It is not about whether gestures look expressive in isolation. It is about whether WoZ elicitation produces design input that is actually useful for building a real control vocabulary under timing pressure, repetition, and indirect mappings.
I answer the main question by splitting "potential" into four connected parts. Each part targets a different kind of evidence that a designer would need before committing to implementation.

Sub-Question 1: How do hand gestures look like when performed to control fighting game actions, even when no direct body-to-avatar mapping exists?

This sub-question is the foundation. Before asking whether gestures converge, survive speed, or support immersive potential, I first need to understand what people actually do when they are asked to invent gestures for actions that are not straightforward to mimic. Sub-question 1 captures the gesture vocabulary itself, but more importantly, it captures the mapping logic participants rely on when the mapping is indirect. This is where I can observe whether participants behave like they are acting, or whether they behave like they are designing a small input language they expect to repeat. That distinction matters because repeatable control gestures have different constraints than expressive one-off performance.

Sub-Question 2: Does player behavior reveal a convergent core set of control gestures, or a broad, divergent set?

Even if gestures look motivated and usable for one person, a control system still has to decide what to standardize and what to keep flexible. Sub-question 2 therefore tests whether WoZ elicitation reveals shared defaults that could support standard mappings, or whether it reveals structured diversity that would push a design toward personalization or multiple acceptable variants. I treat convergence as multi-level because agreement can appear as a shared whole gesture, or as shared intent with varied execution. This matters

directly for the main research question because it tells us what kind of design material WoZ actually yields: a single default gesture, a stable template, multiple competing solutions, or only component-level anchors.

Sub-Question 3: How usable and reliable are user-elicited gestures when performed at high speeds, as required by fast-paced fighting game mechanics?

A gesture can be intuitive and still fail if it cannot be executed quickly and consistently enough. Sub-question 3 therefore treats speed as a filter on potential. It tests whether user-elicited gestures remain performable when the pacing ramps up, and it distinguishes between two different failure types: meaning errors (wrong gesture) versus timing misses (right gesture, but too late for the window). This sub-question ties back to the main research question because it tells us which parts of the elicited vocabulary can realistically survive in a time-constrained control loop, and whether the main constraint looks like gesture confusion or throughput limits.

Sub-Question 4: To what extent do user-elicited hand gestures show qualities associated with potential immersion in fighting games?

The main research question is not only about feasibility and performance. It is also about what kind of experience gesture control could enable. In a WoZ study with predefined tasks, it would be unrealistic to claim full immersion as a stable gameplay state. Sub-question 4 therefore stays careful and focuses on indicators that relate to immersive potential, using three constructs as a guide: intuitive physical interaction, mental imagery and embodiment, and sense of control. This sub-question answers the main research question by showing whether the elicited gestures support a feeling of directness and involvement, and by identifying the conditions under which that layer weakens (for example, when speed pressure makes the control loop noticeable).

1.2 Scope

This thesis studies the *design potential* of user-elicited hand gestures collected through a Wizard of Oz (WoZ) setup. I treat WoZ as a way to explore the interaction and the control vocabulary early, without requiring a finished real-time recognizer. The core unit of analysis is the gesture mapping as it is *invented, performed, and sustained* inside an interactive loop.

I use a fighting game as a testbed, but the intent is broader. Fighting games put strong pressure on input because they involve fast timing windows, repeated actions, and a dense action space. Those pressures make it easier to surface what breaks first in a gesture-based scheme, and what remains usable. The findings should therefore be read as design signals about gesture control under time pressure and repetition, not as claims limited to one specific title. Several things are intentionally out of scope. I do not build or benchmark a full computer-vision recognition model for these gestures, and I do not claim technical feasibility in a deployment-ready sense. I also do not run full competitive matches. Both experiments use a controlled task structure, including prompt-driven segments, which makes the results most suitable for understanding gesture mapping behavior and speed constraints rather than real match strategy. Finally, because the participant pool is small (N=10) and the move set is limited, I treat the results as patterns observed in this study, not population-level truths.

1.3 Contributions

Within the scope above, this thesis makes four practical contributions:

1. **A descriptive account of what "natural" hand gestures look like under indirect mapping pressure.** I show how participants build meaning when actions do not have a direct body-to-avatar analogue, and how they compress gestures toward repeatability, distinctness, and low effort rather than full reenactment.

2. **A multi-level view of convergence that separates shared defaults from structured diversity.** I analyze whether participants converge on a dominant gesture family, split into stable alternatives, or diverge in full form while still converging on component-level intent (especially spatial intent).

3. **Evidence about speed limits as a timing and throughput problem, not only a "gesture meaning" problem.** Using a speed ramp task, I show how reliability drops once the available time becomes smaller than the end-to-end action loop, and why "late-correct" becomes a useful signal of performability under time pressure.

4. **A theory-guided evaluation of immersive potential that stays honest to a WoZ setting.** I do not claim to measure full immersion.

Instead, I analyze immersion-related indicators (intuitive physical interaction, mental imagery/embodiment, and sense of control) and connect them to where the gesture loop stays transparent versus where it becomes noticeable and fragile.

These contributions feed into a consolidated set of design deliverables and implications later in the thesis.

1.4 Thesis outline

This thesis is structured as follows:

- **Chapter 1** introduces the motivation for gesture-based control and positions WoZ elicitation as a design-study method. It also presents the research question, sub-questions, scope, and contributions.
- **Chapter 2** reviews background literature that supports the framing of gesture elicitation, WoZ prototyping, and immersion-related concepts.
- **Chapter 3** details the WoZ setup and user study design across two experiments.
- **Chapters 4–7** answer **Sub-question 1–Sub-question 4** in order, each with methodology, results, and discussion.
- **Chapter 8** consolidates the design deliverables and implications that follow from the four sub-answers.
- **Chapter 9** integrates the findings to answer the main research question directly.
- **Chapter 10** discusses limitations and future work.
- **Chapter 11** concludes the thesis.

Throughout the report, the terms *'sub-question'* and *'SQ'* are used interchangeably.

Chapter 2

Background Literature

This chapter provides the conceptual background for the thesis by discussing related work on gestural input, Wizard of Oz prototyping, user elicitation studies, and immersion and flow theories. The reviewed literature establishes the perspectives and constraints that guide the design exploration presented later.

2.1 Gesture input as a design material

Gesture input sits between two worlds. In everyday communication, hand gestures often support speech and thinking (Kendon 2004, McNeill 1992). In interactive systems, gestures become intentional commands that must be recognized, learned, and performed repeatedly (Wobbrock et al. 2009, Ruiz et al. 2011). This shift matters because a good gesture is not only expressive. It must also be doable, distinct from other gestures, and stable under pressure (Villarreal-Narvaez et al. 2020).

Movement based game interfaces have often been discussed as a way to increase physical involvement and make control feel more natural (Bianchi-Berthouze et al. 2007, Pasch et al. 2009*b*). At the same time, this style of input has practical costs. It can increase effort, trigger fatigue, and punish inconsistency (Hincapié-Ramos et al. 2014). So gesture input creates a design tradeoff. It can raise engagement, but it can also raise performance demands (Pasch et al. 2009*a*, Rogers et al. 2015).

In this thesis, I treat gestures as a *vocabulary*. The user does not invent one move in isolation. They invent a set that must work together (Wobbrock et al. 2009). This framing makes questions about structure unavoidable. Some commands naturally invite similar gestures, while others stay diverse (Ruiz et al. 2011, Villarreal-Narvaez et al. 2020). That diversity is not al-

ways a failure. It can signal that the command is abstract, or that multiple metaphors fit (McNeill 1992, Ortega & Özyürek 2020).

Gesture control in games under fast timing demands

Gesture controllers in games often mix natural motion with physical controls to keep interaction flexible without losing precision (Ionescu et al. 2011, Foottit et al. 2014). But fast-paced genres expose the hard part quickly, because gestures must stay accurate and responsive when players act under tight timing (Teixeira et al. 2006, Paliyawan et al. 2015). In *GeFighters*, the authors used a gesture-controlled fighting game to study gesture interaction in a context that demands short and reliable response times (Teixeira et al. 2006). Work on a universal Kinect interface for fighting games also highlights how much effort goes into mapping body motions to stable inputs that existing games can accept (Paliyawan et al. 2015). Together, these examples frame fighting games as a useful stress test for gesture control, especially for latency and reliability.

2.2 Wizard of Oz prototyping as a way to study gesture control

Wizard of Oz prototyping is a method where a person simulates system intelligence so participants can interact with a concept that is not yet technically built (Kelley 1984, Dahlbäck et al. 1993). In interaction research, WoZ helps isolate questions about experience and behavior from questions about implementation (Dahlbäck et al. 1993). It also allows rapid iteration when the real system would be expensive, unstable, or not feasible yet (Kelley 1984).
WoZ is especially useful for gesture based systems because recognition technology can constrain what researchers are able to test. If a system cannot reliably detect handshape, motion, or timing, users will adapt around the failure and the study stops being about the intended concept (Dahlbäck et al. 1993). WoZ can remove that bottleneck and keep the focus on what users would do if the system worked (Kelley 1984).
At the same time, WoZ creates its own risks. The wizard can accidentally add delay, consistency, or flexibility that a real system would not have (Riek 2012). Participants can also form beliefs about what the system understands, and those beliefs shape their strategy (Dahlbäck et al. 1993). So WoZ does not eliminate bias. It changes where the bias comes from (Riek 2012). This is why WoZ work benefits from careful reporting of what the wizard controlled, how feedback was shown, and what timing constraints existed (Riek 2012).

2.2.1 Höysniemi's Wizard of Oz action games for children

This thesis is strongly inspired by Höysniemi's work on vision based action games for children (Höysniemi et al. 2004, 2005). Their WoZ setup allowed children to play a gesture controlled game while a hidden operator handled recognition and system responses (Höysniemi et al. 2004). This let them study children's intuitive gestures without waiting for robust computer vision (Höysniemi et al. 2004). Their follow up work described the kinds of gestures children naturally produced in this setting and discussed how gesture choices linked to what the action meant to the child (Höysniemi et al. 2005).
Two ideas from this line of work matter directly for my thesis. First, gesture invention reflects meaning making, not only motor preference (Höysniemi et al. 2005). Children often chose gestures that represented the action in a simple way, even when the gesture was not a literal body to avatar copy (Höysniemi et al. 2005). Second, WoZ made it possible to test a full play loop early, which helped reveal usability issues like effort, misunderstanding, and mismatch between gesture style and game pacing (Höysniemi et al. 2004). My work adapts this logic to a fighting game testbed.

2.3 User defined gesture elicitation

User defined gesture elicitation studies ask participants to propose gestures for commands, usually under a structured task (Wobbrock et al. 2009). This approach treats users as a source of mappings rather than only as evaluators of designer made gestures (Wobbrock et al. 2009). It also produces data that can be analyzed for consensus, diversity, and underlying metaphors (Ruiz et al. 2011, Villarreal-Narvaez et al. 2020).
Elicitation studies often reveal that agreement depends on the command type. Concrete actions tend to invite more shared gestures than abstract system functions (Ruiz et al. 2011). They also show that consensus can appear at different levels. Participants may disagree on exact form but still share the same spatial intent or action metaphor (Vatavu 2019). This is a key reason to analyze gestures beyond "same or different" at the whole gesture level (Vatavu 2019, Tsandilas 2018).
Elicitation results are shaped by study constraints. If prompts, examples, or existing controller habits are too strong, participants may reproduce familiar inputs instead of inventing (Morris et al. 2014). This legacy bias is a known issue when participants bring prior interface knowledge into a new input space (Morris et al. 2014). Elicitation also creates a learning problem. A gesture

might feel intuitive when invented, but still be hard to recall later if it lacks structure or if the set is too dense (Wobbrock et al. 2009, Villarreal-Narvaez et al. 2020).

2.4 How to describe hand gestures in a way that supports analysis

If gestures are treated as a vocabulary, the study needs a descriptive language that captures what differs between gestures and what stays stable. Gesture research and sign language research both treat hand movement as structured, not random (Kendon 2004, Stokoe 2005). This supports a component based description approach.

2.4.1 Physical form descriptors

A practical gesture description often breaks form into components such as handshape, movement, orientation, and location (Stokoe 2005, Sandler 2012). These dimensions are widely used in sign language phonology and have also informed gesture analysis approaches in HCI (Sandler 2012). Using these components helps separate "what the gesture is trying to express" from "how the person executed it" (Vatavu 2019).

Handshape matters because it carries functional roles like "button press," "point," "grip," or "fist impact," even when the motion stays small (Stokoe 2005, Tennant & Brown 1998). Direction and motion primitives matter because people often express action using compact movement units such as taps, flicks, traces, pushes, and holds (Kendon 2004). Palm orientation matters because it changes how a gesture reads, especially when users treat the hand as a proxy for an object or a surface (Sandler 2012).

A component view also makes it easier to talk about "simplification under pressure." Under speed or fatigue, people often reduce travel distance, drop secondary parts, or compress motion into shorter variants (Hincapié-Ramos et al. 2014). If the analysis only tracks whole gesture labels, this kind of adaptation becomes hard to describe (Vatavu 2019).

2.4.2 Representational strategies and meaning

Gesture meaning does not come only from physical form. People also choose a representational strategy, and that choice shapes what the gesture is doing communicatively. McNeill's gesture types help describe this at a high level. An **iconic** gesture shows aspects of an action or object, a **metaphoric**

gesture uses a physical image to express an abstract idea, a **deictic** gesture points to a target or direction, and a **beat** gesture marks emphasis or rhythm (McNeill 1992). These types also combine in practice, so a single gesture can point while also depicting motion, or depict an action while also marking timing (McNeill 1992). In my study context, this matters because users can still share intent even when they do not share exact form. Two people might both aim for "attack forward," but one might point forward while another might shape a fist and thrust. The type labels make that difference readable (McNeill 1992).

Müller's modes add a second layer that is more about *how* the hand depicts meaning. Instead of only asking "is this iconic or deictic," the modes describe the depiction mechanics, like **enacting** an action with the hand, **molding** or **shaping** an imagined object, **tracing** a path or outline in space, **holding** a form in place as a stable representation, or **embodying** something by treating the hand as the thing itself (Müller 2014). This distinction helps in game gestures because many inputs are compact and symbolic, not full body reenactments. A quick flick can work as a schematic depiction of an attack, while a traced line can stand for movement direction, even if the player never tries to mirror the avatar's body (Müller 2014). Silent gesture studies also show that people can communicate structured meanings without speech, and that the gesture form adapts to what needs to be conveyed (Ortega & Özyürek 2020, Janke & Marshall 2017). For game control, this means the player's hand often acts like a small model of the intended in game action. Sometimes it is literal, like "my fist is the punch," and sometimes it is schematic, like "a trace shows a dash" (McNeill 1992, Müller 2014, Ortega & Özyürek 2020). Using McNeill's types and Müller's modes together lets me compare gestures at the meaning and depiction level, which becomes important in SQ2 when whole gesture forms stay diverse but the underlying representational strategy still converges.

2.5 Agreement, convergence, and structure in gesture sets

Many elicitation studies quantify convergence by measuring how often participants propose the same gesture for the same command. Wobbrock et al. (2009) formalized this with the guessability protocol and an agreement score that is computed from the distribution of proposed gestures per referent, so higher scores reflect stronger consensus in the set. This works well when gestures naturally collapse into a few clearly repeated forms, and when the

"same gesture" decision is straightforward.
But agreement scores can hide structure. Tsandilas (2018) argues that agreement depends heavily on the analyst's grouping choices, and that different clustering decisions can change the apparent level of consensus even if the underlying behavior stays the same. Tsandilas (2018)' point is not that agreement is useless, but that agreement alone can over-simplify what is happening when people share intent but vary in how they execute it.
Dissimilarity-based approaches address this by treating gestures as points in a similarity space and estimating consensus through distances rather than only counting exact matches. Villarreal-Narvaez et al. (2020) proposes a dissimilarity-consensus approach that uses pairwise comparisons to capture "loose clustering," which is useful when gestures show many variants that still belong to the same neighborhood. This approach is also a better fit when you want to talk about partial convergence, like "people agree on direction and motion, but not on handshape." Systematic reviews of elicitation work also emphasize that convergence is not guaranteed across all commands, and that some commands reliably invite diversity depending on context, constraints, and user population (Villarreal-Narvaez et al. 2020).
In this thesis, I treat convergence as multi-level for this reason. Whole gestures can remain diverse, while intent-level components such as spatial direction or motion primitive still converge enough to form a usable structure. This matters for design because a recognizer can sometimes accept variation in detailed articulation while still enforcing consistency at the intent layer (Tsandilas 2018).

2.6 Performance constraints: speed, latency, and fatigue

Gesture control does not only need to make sense. It must also fit within timing constraints. Fighting games highlight this because actions are chained quickly and mistakes have immediate consequences.
Human performance under time pressure follows classic speed accuracy tradeoffs (Heitz 2014). When time shrinks, people tend to simplify actions, commit earlier, or accept more errors (Han et al. 2013). For gesture input, this can show up as smaller motion, fewer components, or more reliance on default gestures (Hincapié-Ramos et al. 2014).
Latency matters because it changes whether players feel in control. Even small delays can reduce performance and enjoyment in timing sensitive games (Claypool & Claypool 2006). Response time guidelines in interface design

also treat delay as a driver of perceived system quality and user confidence (Nielsen 1993). For gesture systems, latency interacts with recognition stability. If the system response feels inconsistent, players start to adapt their behavior to get a response instead of playing naturally (Claypool & Claypool 2006).
Fatigue matters because gesture input uses the body as a repeated actuator. Movement based interfaces can increase physical effort and change how long users can comfortably play (Pasch et al. 2009*b*, Bianchi-Berthouze et al. 2007). Sustained mid air input has been linked to consumed endurance effects, where users gradually reduce amplitude or change posture to cope (Hincapié-Ramos et al. 2014). These constraints are not side issues. They directly shape what kinds of gestures can remain reliable and enjoyable.

2.7 Immersion and controller experience

The concepts of immersion and flow are pivotal in understanding the player experience in video games, yet the definitions and interpretations offered by various scholars reveal both consensus and contradictions. Because of that, it helps to treat immersion as something that can build in layers, or show up through different components, rather than as one single on/off state. Brown & Cairns (2004) describe immersion as a progression from engagement, to engrossment, to total immersion, where players move deeper as they overcome barriers like learning effort and control friction. Ermi & Mäyrä (2005) also treat immersion as multi-part, but they frame it through three components: sensory immersion, challenge-based immersion, and imaginative immersion. They built this model from observations of players and then assessed it using self-report questionnaires across different games. Together, these models support a practical idea for this thesis: even if total immersion is hard to claim with confidence, I can still study conditions that make immersion more likely, like whether interaction feels responsive, whether challenge becomes overwhelming, and whether the player can stay mentally in the game loop (Michailidis et al. 2018, Brown & Cairns 2004).
Presence often overlaps with immersion, but many authors use it more specifically for the feeling of being there in the game space (Ermi & Mäyrä 2005, Michailidis et al. 2018). This matters for input research because controllers can strengthen or weaken that feeling by shaping how direct the player's actions feel. Work on controller naturalness shows that more naturally mapped controllers can increase reported spatial presence and enjoyment, partly because the input better matches what players already know from real-world action and mental models (McGloin et al. 2011). Studies also link input re-

alism and control fidelity to presence-like outcomes, which is why controller design often shows up as a key lever for immersive potential (Williams 2014). Flow comes from a different tradition, but game research often connects it to deep involvement, high focus, and a good challenge-skill balance (Csikszentmihalyi et al. 2014, Cowley et al. 2008). In games, this balance is not only about difficulty. It also depends on whether the game gives clear goals and feedback, and whether the player feels in control of what happens next. Sweetser & Wyeth (2005)'s *GameFlow* model makes this link explicit by treating enjoyment as structured by flow and listing elements like concentration, challenge, skills, control, clear goals, and feedback, with immersion as one of the connected outcomes. They initially checked the model by using expert reviews to compare a high-rated and a low-rated strategy game and showed that the criteria helped explain why one worked better than the other. Cowley et al. (2008) also argue that flow in games depends on how challenge and skill evolve over time and they discuss how profiling and player history can matter when we try to estimate this balance. At the same time, reviews point out that flow and immersion often blur together in practice, and that many studies still rely heavily on questionnaires, which makes it hard to claim we have measured a clean, single construct (Michailidis et al. 2018). This is another reason I treat immersion as supported by precursors rather than as something I can fully prove in a lab study.

For gesture and movement-based play, the body itself becomes part of the experience. Bianchi-Berthouze et al. (2007) compared more bodily forms of control against less bodily ones and argue that movement can increase engagement and can also change how players get engaged, including through affective involvement. Pasch et al. (2009*a*) studied movement-based sports games using interviews, questionnaires, video observation, and motion capture, and they report that players approach these games with different motivations (for example, "to achieve" versus "to relax"), which then relates to different movement strategies. Nijhar et al. (2012) tested how movement recognition precision affects exertion game experience, and they report higher immersion with higher recognition precision, while also arguing that the reasons differ depending on player motivation.

These findings fit the core framing of my sub-question 4: if recognition breaks down, gets inconsistent, or feels delayed, it can interrupt control and attention. If recognition stays consistent and the player can keep their rhythm, the controller has better potential to support immersion-related experiences.

Chapter 3

Wizard of Oz User Study Design and Setup

This chapter explains how I designed and ran my Wizard of Oz user study for hand gesture controls in a fighting game. I describe what I did, how I did it, and why I did it. I include the participant group, the physical setup, the technology stack, the Wizard of Oz simulation design, and the full rationale for the pre and post study questionnaires and interview prompts. I also explain how Experiment 2 used a speed ramp, and why I measured outcome consequences rather than using diagnostic workload questionnaires.
I designed the study around two experiments. Experiment 1 elicits and records gestures. Experiment 2 stress tests those gestures under increasing speed. I then use a post experiment interview questionnaire to connect performance and breakdowns back to felt experience.

Testbed selection and rationale

Tekken 8 and the limb based control scheme

I needed a fighting game that naturally fits a command mapping mindset. Tekken fits because its core control scheme already treats attacks as a small set of limb commands. It uses four basic attack buttons, and each button maps to a limb, which players commonly describe as left punch, right punch, left kick, and right kick. That design gives a clear, stable action vocabulary that supports gesture mapping. It also supports fast sequences, which matters for testing time pressure in Experiment 2 (DashFight 2024).
I also chose Tekken 8 because it represents the speed and tempo of modern fighting games, including moments where inputs must happen quickly and

repeatedly. This matters because my thesis does not only ask if gestures look natural in isolation. It asks if they remain workable when the pace rises.

Character selection: Paul Phoenix

I needed a character whose move set feels understandable without requiring fantasy powers or unusual mechanics. I chose Paul Phoenix because he is presented as a martial artist with a straightforward, physical fighting identity. His moves read as punches, kicks, and heavy strikes. That keeps the elicitation grounded in human movement metaphors, rather than special effects logic.
This choice also keeps the study general. I treat Paul as a testbed, not as the boundary of the contribution. If a gesture vocabulary can support locomotion, single hits, and multi hit strings for Paul, then the same design ideas can transfer to other characters and other fighting games that rely on rapid command execution.

Rationale for a Wizard of Oz approach

A working gesture recogniser would add two big problems at once. First, it would add modelling and recognition errors that can dominate user experience. Second, it would slow down iteration and reduce the flexibility of the study design.
Wizard of Oz avoids both issues. It lets me simulate a gesture controlled system while the participant believes they interact with a functioning recogniser. This is a standard approach when a system remains technically hard to implement, but the interaction idea needs evaluation. Wizard of Oz also fits early stage design because it supports fast iteration while keeping the task context realistic (Dahlbäck et al. 1993).

Move set selection

I selected 12 moves to cover a compact but meaningful set of fighting game demands. I wanted the set to include locomotion, simple attacks, heavier attacks, ambiguous embodiment, and multi hit strings.

Selected moves

The 12 moves were:

1. Forward

2. Backward
3. Sidestep In
4. Sidestep Out
5. Left punch (Tekken input 1)
6. Left kick (Tekken input 3)
7. Uppercut
8. Neutron Bomb
9. Hammer punch
10. Phoenix smash (Phoenix/Deathfist)
11. Phoenix Smasher (Hammer + Phoenix)
12. Hangover (Hammer, low kick, elbow)

Selection rationale

I chose this set for five reasons.
First, I needed locomotion referents. Forward, backward, and sidesteps represent constant movement tasks in fighting games. They also create a strong test for directional gesture logic, and they often create ambiguity when users pick symmetric gestures for opposite directions. That helps SQ2, because convergence and divergence often show up clearly in directional moves.
Second, I needed simple single hit actions. Left jab and left kick represent minimal input actions with clear labels and clear intent. They also anchor the design space, because most participants can reason about what a punch and kick means, even if the avatar mapping remains indirect.
Third, I needed heavier strikes with distinct feel. Uppercut, hammer punch, and phoenix punch let participants express power, effort, and direction. This matters for SQ1 because these moves often push users into different motion primitives and different spatial intent levels. It also matters for SQ4 because heavier attacks often connect to embodied imagination and sense of agency.
Fourth, I needed at least one referent with ambiguous embodiment. I included Neutron Bomb because it reads as a whole body move. Participants still must express it using hands only. That creates a useful mismatch between avatar action and available body channel, which directly tests SQ1.

It also helps SQ4 because it pressures mental imagery and the participant's ability to still feel like they act out the move.
Fifth, I needed multi hit strings. Phoenix Smasher and Hangover represent combined sequences that create memory load, chunking demands, and timing pressure. This supports SQ3 because strings reveal breakdowns earlier than single hits when time compresses. It also supports SQ4 because strings often affect control and flow more than isolated gestures.
This selection also aligns with fighting game command system work that treats player avatar interaction as a mapping problem between human intention and an avatar action vocabulary. That work highlights how command systems shape how players think about control, embodiment, and meaning making in fighting games (Mattiassi 2019).
The prompt icons used for each of these moves are provided in appendix H.

Participants

Recruitment and sampling

I recruited ten participants using opportunistic convenience sampling from the university community. I relied on word of mouth and snowball style recruitment. Participants nominated themselves, which fits the voluntary participation model in the study information sheet.

Screening and eligibility criteria

I used these criteria:

- Age 18 to 35
- Normal or corrected vision
- No upper limb impairments
- Full, pain free range of motion in wrists, hands, and fingers

I screened participants before the session. I asked health questions about arms, wrists, hands, and fingers. I also asked them to do a brief range of motion check, such as opening and closing the hand and bending the wrist up and down. If a participant reported a relevant impairment or could not comfortably complete the check, I excluded them for safety and data validity.

Participant characteristics and what the sample represents

All participants reported right hand dominance and no recent upper limb injury in the last six months. Ages ranged from 19 to 31 years. Participants shared a broadly similar education context because they came from the university environment. At the same time, they represented diverse national backgrounds, with five nationalities represented.
Gaming familiarity varied. Self reported play time ranged from 0 to 10 hours per week for video games and from 0 to 3 hours per week for fighting games. Tekken 8 experience skewed toward novice levels. Motion control exposure also varied, with some participants reporting experience with systems like Wii, Kinect, VR, or mobile AR, and others reporting none.
I treat this group as an exploratory HCI sample. My goal here is not population estimation. My goal is to elicit gesture ideas, observe convergence patterns, and identify failure modes under speed. Small samples remain common and defensible for early stage elicitation and interaction technique work when the study aims at design understanding rather than statistical generalization.

Ethics and participant materials

Participant information

Before the session, participants received a participant information sheet. It explained the study purpose, what the session includes, the duration, the kinds of recordings, and potential risks. It also stated that participants can decline questions and withdraw at any time without giving a reason. If a participant withdrew, I would destroy their recorded data.
The information sheet also explained data handling. It stated that I would record face, hands, audio, and screen capture. It described that hand camera footage would form a de-identified gesture corpus for future machine learning research, while face and audio support think aloud analysis and would not be reused outside the study. It also described secure storage, limited access to the research team, anonymisation before publication, and retention rules. Facial data would be kept until the end of the research and then securely deleted, while hand gesture data may be retained as a gesture corpus. A copy of the Participant Information Sheet can be found in the Appendix E.

Informed consent

Participants signed an informed consent form. It asked them to confirm that they understood the study information and that they consent voluntarily. It explicitly listed video recording of face and hands, audio recording, and gameplay input logging. It also repeated the right to withdraw at any time without giving a reason. A copy of the Consent Form can be found in the Appendix F.

Debriefing

After completing the session, I provided a debrief. The goal of the debrief is transparency. I explained the Wizard of Oz nature of the setup, clarified how I simulated recognition, and reminded participants of withdrawal rights regarding their data. A copy of the Debrief Letter can be found in the Appendix G.

Physical setup and apparatus

Room layout and separation

I used a two sided Wizard of Oz apparatus. I physically separated the participant side from the researcher side so the participant could not see the control inputs. I used a rigid separator or blind to divide the room. I also used curtains to reduce visual distractions and to maintain separation.

Participant station

On the participant side, the participant sat at a table facing an external monitor that displayed Tekken 8 gameplay. I placed a clearly marked hand task space on the tabletop. I outlined it and placed it over a black mat to increase contrast. This did two things.
First, it standardized where gestures happened, which improves comparability across participants. Second, it improves hand visibility in video, which supports later gesture coding.
I asked participants to wear a black shirt and I clipped a BOYA microphone to them. This improves audio quality and keeps clothing visually consistent, which helps later coding of the video as well as gesture recognition.
I placed a small think aloud reminder card within the participant's view. This card exists because participants often fall silent during concentration.

Figure 3.1: Complete Physical Setup

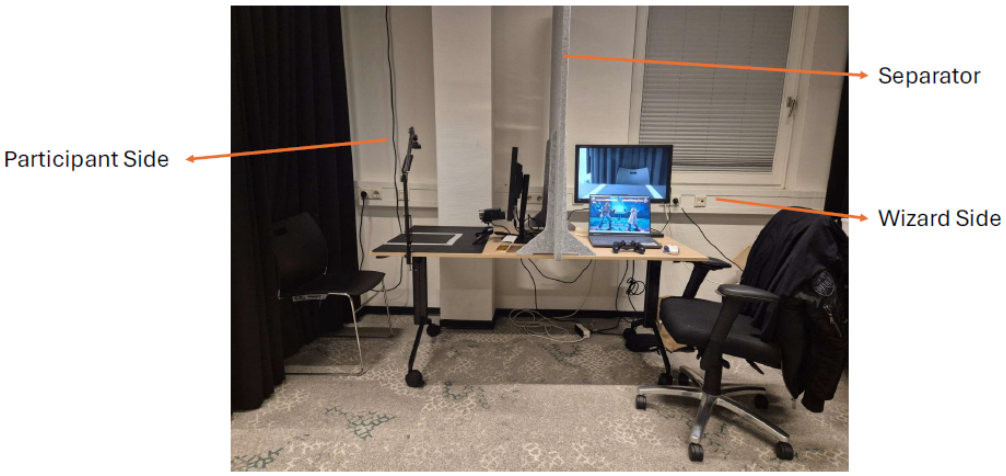


Figure 3.2: Setup without the participant blind

Figure 3.3: Participant side in focus

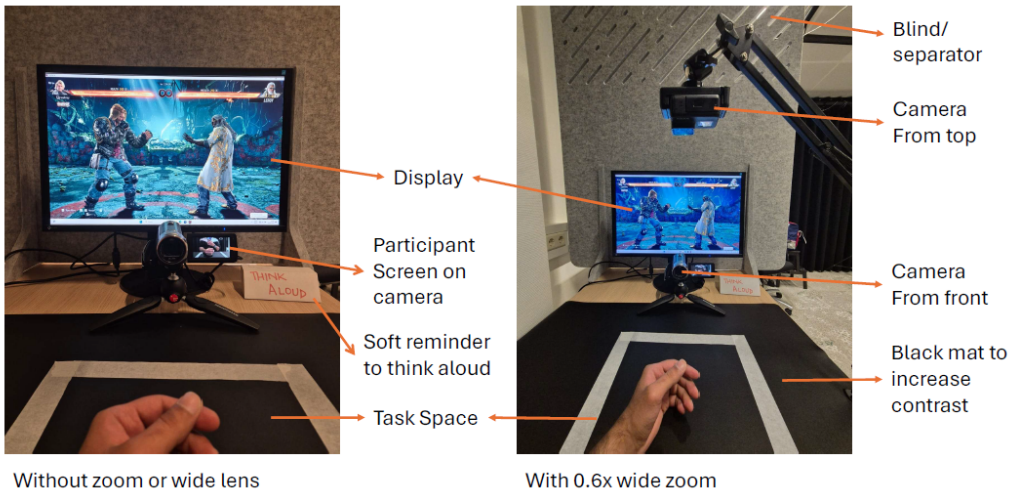


Figure 3.4: Participant's Point of View

A small visual reminder supports a more consistent verbalization behavior across the session.

Recording setup

I used two cameras.

- A front facing Panasonic HC V720 camera on a tripod. It recorded to SD card. It also provided a live HDMI feed to the wizard side, so I could interpret gestures in real time. I flipped its LCD so it would face the participant. This allowed them to confirm hand framing.
- A GoPro Hero 7 above the table. It recorded an overhead view to SD card. This view supports later analysis because it captures hand shape, finger changes, and the spatial trajectory inside the task space.

Wizard station

On the wizard side, I ran Tekken 8 on a laptop. I also ran OBS to record the gameplay and the microphone audio. I ran the Experiment 2 icon overlay on the same machine. I controlled the game using a DualShock 4 controller.
I also used a dedicated wizard monitor. It displayed the live HDMI feed from the front camera. This let me interpret hand gestures without a direct line of sight to the participant.

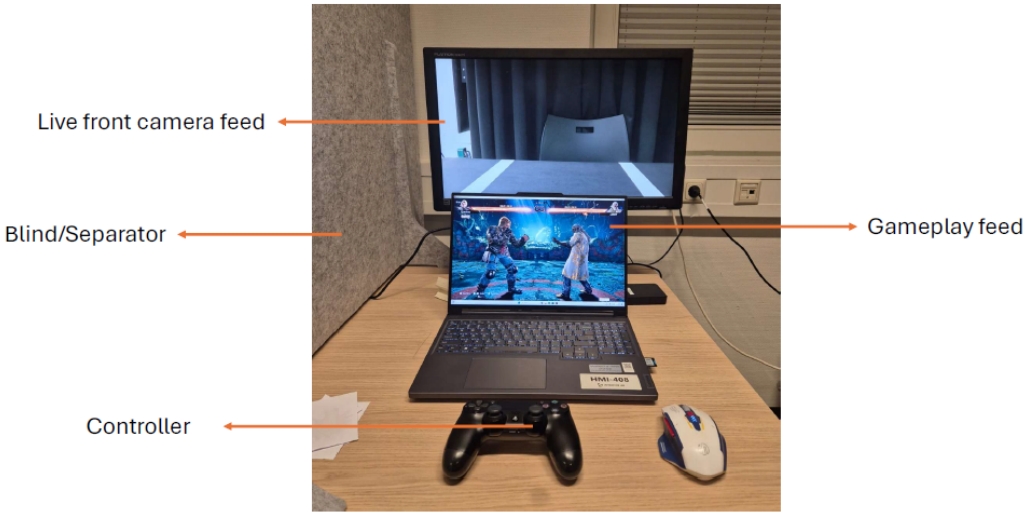


Figure 3.5: Wizard's Point of View

Setup rationale

This setup prioritizes three things.
First, it preserves the Wizard of Oz illusion. Physical separation prevents the participant from seeing my controller inputs.
Second, it supports high quality data capture. Two camera angles reduce ambiguity and reduce the chance of losing data due to occlusion.
Third, it supports future technical work. The task space constraint and the high contrast setup make the resulting recordings more useful as a gesture corpus, because the hands remain visible and spatially consistent.

Wizard of Oz implementation

Wizard role and responsibilities

I acted as both the researcher and the wizard. This means I delivered prompts and I also executed the corresponding Tekken actions. I designed the setup so I could do this without exposing the illusion.

Consistency controls and reliability

I used controller macros through the *DS4Windows* application. The goal is consistency and timing reliability. Instead of manually entering complex

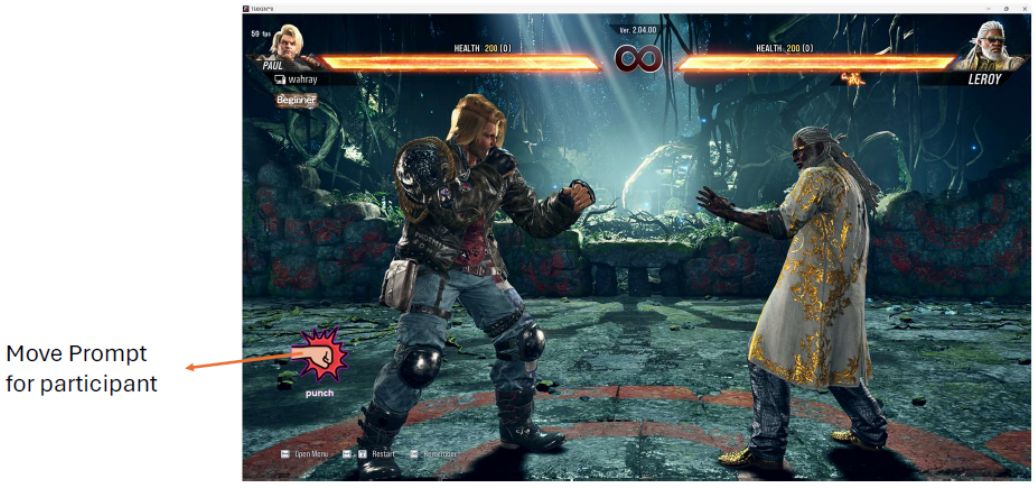


Figure 3.6: Experiment 2 Gameplay

Tekken input sequences for strings, I mapped each move to a single trigger. One button press could execute a full move, including longer multi hit strings. This reduces variation in execution timing and reduces error risk under speed. This choice also makes the measured timing more interpretable. In Experiment 2, when I log the wizard input time, I treat it as the system's moment of recognition and action dispatch. If I had to input long sequences manually, the log would mix recognition time with manual command entry time.

Simulation fidelity

The simulation has high fidelity on the game output side. Participants see real Tekken 8 gameplay with real animation and game feedback. They also see the same visual context they would see in a real game.
The simulation has limited fidelity on the recognition side. A real gesture recognizer would detect the gesture through sensing and classification. In this study, I detect the gesture through live video observation and I trigger the mapped command manually.
This design matches the core purpose of Wizard of Oz. I want to test the interaction concept and the gesture vocabulary under realistic game output. I do not want early recognition engineering issues to dominate the study (Dahlbäck et al. 1993).

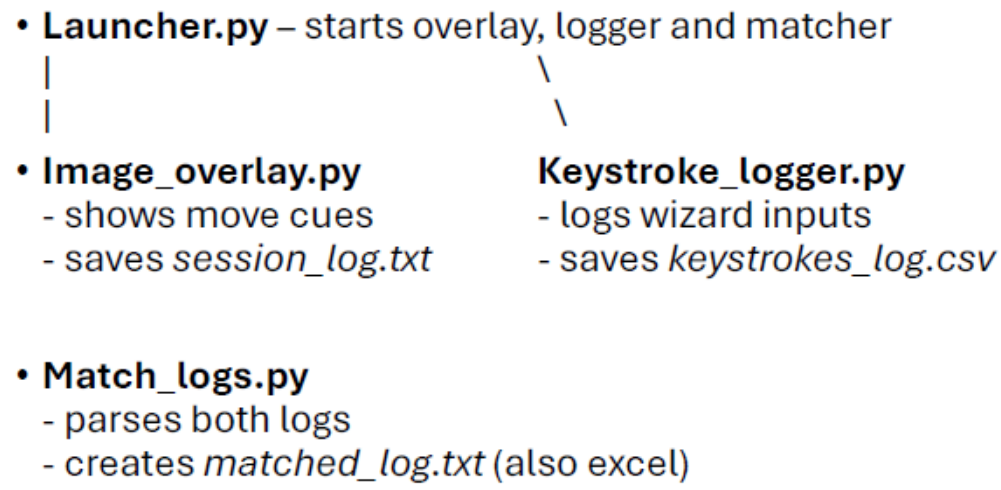


Figure 3.7: Software Overview

Software and logging infrastructure

Experiment control software

I implemented a small Python tool chain to support Experiment 2 and to produce structured logs:

- A launcher script that starts the other modules. It takes the participant number for file naming. It also takes the playing side to ensure correct left and right prompts and icons.
- An image overlay module that presents on screen icons during Experiment 2.
- A keystroke logger that records detected wizard inputs to a CSV file. It records timestamp, key identity, and the gesture label tied to that key.
- A log matching script that parses the overlay session logs and matches them with the keystroke log. The overlay logs contain icon time, move name, display duration, and the inter prompt wait time. The output is a combined log file in text and spreadsheet form, including reaction time estimates and correctness indicators based on expected versus matched input.

Prompt and input logging rationale

I needed an end to end timing trace that matches the design of the study. In Experiment 2, the prompt appears on screen, the participant perceives it, performs the gesture, and then the system responds. In my setup, the response time includes my Wizard of Oz response.
So I log:

- When each prompt appears
- Which move it represents
- How long it stayed on screen
- How long the gap was before the next prompt
- When the wizard triggered the move

From this, I can estimate reaction time as prompt onset to wizard trigger. I can also classify events as correct, incorrect, missed, or late relative to the prompt interval. This produces a performance view that matches how a real game would accept or reject input under speed.

Experiment 1: Gesture elicitation

Design goal

Experiment 1 collects the gesture vocabulary. Participants invent gestures for the 12 moves. I treat participants as users and as designers. They are users because they perform gestures to control actions. They are designers because they create the mapping between referents and gestures.
This matches the logic of user defined gesture elicitation. Elicitation studies let pandidate gestures, then let researchers analyse the resulting set for structure, convergence, and design implications (Brown & Cairns 2004).

Stimuli and prompts

For each move, I showed:

- A demonstration of the specific move in Tekken 8
- A distinct icon associated with that command

- A verbal prompt naming the action

Participants then invented a hand gesture they believed should trigger that move. They performed the gesture five times. I asked them to think aloud while doing this.

Rationale for the full move preview

Before starting the per move elicitation, I showed a demonstration of all selected moves. I added this because of what I observed in pilot testing.
When participants only see one move at a time, they often recycle an earlier gesture. Then they later backtrack and try to re tweak earlier gestures once they realise two referents need separation. Showing all moves up front helps participants form a complete design space. It encourages them to create a full gesture vocabulary with clearer boundaries.
This also aligns with elicitation study considerations reported in prior work, where participants benefit from understanding the full task set when they design a gesture set as a whole, rather than inventing each gesture in isolation (Wobbrock et al. 2009).

Mental model probing during elicitation

During the session, I asked short questions to understand why a participant chose a gesture. One key prompt was:

> Imagine explaining your gesture to a friend who has never played. How would you describe why your hand movement equals a punch or kick?

I use this kind of probe to capture metaphorical mapping and mental imagery. The words participants use often reveal the source domain they rely on, such as pushing, flicking, pressing,tters for SQ1 and SQ4 because it connects gesture form to meaning making, embodiment, and sense of control.
This also fits fighting game command system work that links player avatar interaction to meaning making and to how players mentally model action execution (Mattiassi 2019).

Supporting gesture revision

I allowed participants to revisit and reinvent gestures during Experiment 1. If a participant realised a gesture created ambiguity across two moves,

they could refine it. This supports confidence and reduces the risk that participants feel stuck with a poor early choice. It also better matches real design behaviour, where people iterate once they see conflicts.

Handling rapid repetitions and gesture mashing

In pilot testing, some participants mashed gestures, meaning they repeated gestures rapidly instead of performing a single clear instance. I did not add mashingimental condition. Instead, I observed it when it occurred. If a participant mashed, I asked follow up questions to understand why. I then guided them to perform the gesture once in a clean way so I could record it for analysis.

Gesture ambiguity management

Some gestures become ambiguous in predictable ways. A common example is using the same motion for forward and backward. When I identified ambiguity, I recorded it. I then encouraged the participant to refine the gesture with simple support. I did this to keep engagement high and to avoid making participants feel that they failed.

Data captured in Experiment 1

Experiment 1 produced:

- Overhead hand video
- Front hand video
- Audio of think aloud commentary
- Screen recording of gameplay output
- Notes about participant explanations and revisions

This combination supports form analysis and meaning analysis. It also supports later coding and similarity analysis.

Experiment 2: Gesture performance under increasing speed

Design goal

Experiment 2 stress tests the gesture set under time pressure. The participant performs the same gestures they invented in Experiment 1. Instead of slow prompting, the system presents a stream of move icons. I gradually increase the speed until the prompt timing becomes tight.
This design targets SQ3. Fighting games often require quick reactions and quick execution. I therefore needed a manipulation that compresses the available time, forces prioritisation, and exposes breakdowns.

Prompt format

I used the same distinct icons from Experiment 1. In Experiment 2, the icons appeared on the gameplay screen as an overlay. Each icon represented one of the 12 moves.
Participants had to perform the gesture for the current icon. If they missed a gesture, I instructed them to move on rather than retry.

Speed ramp schedule

The speed ramp had two timing parameters:

- Prompt display time, meaning how long the icon stays visible
- Inter prompt gap, meaning the time between an icon vanishing and the next icon appearing

I designed the ramp as follows.

Initial sequence

The first 12 prompts were the same 12 moves in the same sequence as Experiment 1. This gives participants a familiar starting rhythm and reduces cold start confusion.

Prompt display time reduction

The first prompt displayed for 5.0 seconds. The next prompt displayed for 4.9 seconds. From 5.0 seconds down to 2.0 seconds, I reduced the display time by 0.1 seconds for every consecutive prompt.
Once the display time reached 2.0 seconds, I slowed the reduction rate. From 2.0 seconds down to 0.3 seconds, I reduced the display time by 0.1 seconds after every 3 prompts.
I chose 0.3 seconds as the lower bound because it sits near commonly cited simple reaction time scales for visual stimuli, and it acts as an extreme lower limit for a prompt driven response task (Woods et al. 2015).
In my implementation, the display time reached 0.3 seconds around the 100th to 105th prompt.

Inter prompt gap reduction

After a prompt vanished, I initially waited 2.0 seconds before showing the next prompt. I then reduced this gap by 0.1 seconds after every 3 prompts until it reached 1.0 second. After it reached 1.0 second, it no longer reduced.

Why I shaped the ramp in phases

From 5.0 to 2.0 seconds, participants still have enough time to execute most gestures without panic. So a per prompt reduction makes sense and creates a smooth ramp.
Below 2.0 seconds, small changes create large practical differences. At that point, I slow the reduction rate to avoid a too steep cliff and to keep the ramp long enough to observe adaptation behaviours, not only immediate failure.

Data captured and computed measures

Experiment 2 produced:

- Overlay prompt logs, including prompt onset time, move label, display duration, and inter prompt gap
- Wizard input logs from the keystroke logger
- Matched logs that link each prompt to the detected input and compute a reaction time estimate
- Screen recording and audio recording through OBS

- Hand videos from both cameras for later qualitative checking

The matched log supports performance classification, such as whether the participant triggered the correct command in the available interval, whether they triggered late relative to the next prompt, and whether they missed or substituted actions.

Interpreting reaction time in a Wizard of Oz setup

Reaction time in this study means prompt onset to wizard trigger. It includes:

- Participant perception time
- Participant gesture execution time
- Wizard interpretation time
- Wizard actuation time

This is not the same as a pure human reaction time measure. It is an end to end pipeline measure that approximates what a real gesture recognition system would need to achieve, including sensing and classification delay. I therefore use it as a practical indicator of playability under speed rather than as a psychometric reaction time test.

Pre session questionnaire and rationale

I used a participant questionnaire to capture context that can shape gesture design and gesture performance. I did not treat this as a hypothesis testing instrument. I treat it as explanatory context.
Below, I explain the constructs and why they matter.

Demographics and physical readiness

I captured age to control for broad motor differences. Age relates to processing speed and motor performance, especially under time pressure.
I captured dominant hand because handedness affects precision and comfort for one handed gestures.
I captured injury status because even mild discomfort can change gesture amplitude, speed, and willingness to repeat movements. I used this both as a safety filter and as a validity check.
I captured vision status because the study uses visual prompts and ieded to perceive the cues reliably.

Gaming background and controller familiarity

I asked about platform ownership and overall gaming frequency to understand baseline controller familiarity. Controller familiarity can shape how participants think about commands, how they map actions to inputs, and how they manage timing.
I asked about general gaming and fighting game experience because expertise can change attention patterns, anticipation, and speed management. Action game experience has been linked to faster processing and response speed in prior work. I do not use this as a claim about causality in my small sample. I use it as context for interpreting differences (Dye et al. 2009).

Tekken familiarity

I asked about Tekken 8 familiarity because Tekken uses an established limb based command vocabulary. Players with prior Tekken knowledge may already carry motor schemas for how moves should feel and how they chain. That can shape gesture proposals and chunking strategies (DashFight 2024).

Motion control exposure

I asked about motion control exposure to capture whether participants already have a gesture control metaphor library. Prior exposure to Wii, Kinect, VR, or AR can shape what a participant considers a valid gesture, how large they make movements, and how they treat gesture timing.

Gesture confidence scale

I included a gesture confidence scale because elicitation depends on willingness to propose ideas. Some participants feel creative and bold. Others feel anxious about inventing gestures. Confidence and self efficacy can shape the diversity and expressiveness of proposals.
This connects to self efficacy theory, where beliefs about capability influence willingness to attempt and persist in a task (Bandura 2010).

Working memory

I included a working memory measure because the study includes multi hit strings. Strings require either remembering a complex gesture or chunking multiple components into one meaningful unit.

Working memory theory treats this as a limited capacity system that supports holding and manipulating information during tasks. This matters for understanding why some participants may simplify or avoid complex gesture designs (Baddeley 2000).

Embodiment and metaphor mindset

I included an embodiment and metaphor mindset item because participants can approach gesture mapping in different ways. Some participants aim to act out the avatar. Others treat the hand as a symbolic device, like pressing buttons in mid air.
This matters because it connects directly to SQ4, where I study immersion related qualities such as intuitive physical interaction, mental imagery, and agency. It also aligns with fighting game command system work that discusses how players relate to avatars through action and imagination (Mattiassi 2019).

Post experiment interview questionnaire and rationale

I used a post experiment interview questionnaire after both experiments. I designed it to connect observed behaviour to felt experience. I also designed it so that the same immersion constructs appear in both Experiment 1 and Experiment 2 sections. This supports comparison, because participants can reflect on the same constructs under slow creation versus fast performance. All the questions used after each experiment can be found in the Appendix D.

Why I used interviews alongside logs

A purely log based view cannot capture internal states like perceived control, discomfort, frustration, or confidence. User experience definitions explicitly include perceptions and responses during and after interaction, including comfort and accomplishments. So interviews are not an optional add on here. They are required to answer SQ4, and they also support SQ3 by revealing why breakdowns happened.
This approach also matches standard practice in usability testing and games user research. Researchers often combine observation with verbal reports to understand breakdowns and to interpret performance outcomes (Simor et al. 2016).

Target experience constructs

I structured the questionnaire around three constructs:

- Intuitive physical interaction
- Imaginative mental imagery and embodiment
- Sense of control, meaning agency and natural mapping

These constructs align with established game immersion models and player experience work.
The SCI model distinguishes sensory, challenge, and imaginative immersion, and it treats imaginative involvement as a core part of how players connect winded immersion work also highlights involvement and control as central to deeper engagement.
Embodied play research argues that physical involvement can change emotional engagement and the felt connection between action and outcome.
Controller type research in HCI also links input method to enjoyment and motivation, which supports why input design can shape experience, not only performance (Laura Ermi et al. 2005).

Linking Experiment 1 questions to SQ4

In Experiment 1, participants invented gestures. So the key question is whether gesture invention felt like acting out the avatar, or like inventing abstract commands.
Question 1.1 asks if gestures felt natural and intuitive, like directly acting out the character. This targets intuitive physical interaction. It captures whether the participant experienced a bodily match between intention and action.
Question 1.2 asks whether they imagined the character performing the move, or focused only on hand motion. This targets mental imagery and embodiment. It also helps interpret why some gestures become more iconic or more symbolic.
Question 1.3 asks about sense of control and agency. It targets whether the mapping felt predictable and trustworthy, even in a simulated system.
Questions 1.4 and 1.5 capture preference and comfort. They support design implications and help identify gestures that may look good but feel bad to repeat.

Linking Experiment 2 questions to SQ3 and SQ4

In Experiment 2, participants perform under speed. So I added questions that capture both immersion qualities and reliability consequences.
Questions 2.1 to 2.4 repeat the immersion constructs at speed and add an explicit disruption question. The disruption question targets moments where a gesture pulled the participant out of the experience. This is important because speed can break embodied imagination and agency.
Questions 2.5 to 2.7 directly target SQ3. They ask about simplification, gestures becoming impossible, and mistakes. These questions help interpret observable breakdown patterns in the logs.
Questions 2.8 to 2.10 target fatigue, strain, and anticipated bodily state after repetition. These are core SQ3 consequences because physical discomfort can become the limiting factor for a gesture control scheme even when recognition works.
Question 2.11 asks overall enjoyment of playing at high speed using these gestures. This connects SQ3 and SQ4. A gesture can be technically feasible but still unpleasant.
The extra questions capture global recall of what felt natural, what felt hard to remember, and whether participants would use the gestures in a real game. These support design implications.

SQ3 analysis strategy: observable and felt consequences

Why I focused on consequences

SQ3 asks about usability and reliability under speed. I decided to measure the outcomes of fast paced gesture play rather than diagnosing underlying causes through workload subscales.
In other words, I care about what happens to the interaction when time compresses. I care about errors, breakdowns, hesitation, gesture simplification, fatigue, frustration, and flow interruptions. These consequences directly describe whether the system stays playable.
This aligns with HCI and UX work that treats experience as holistic and situated, shaped by control, affect, and engagement. It also aligns with games research practices where observation and post task verbal reports form the core evidence for how input methods hold up during play.

Why I did not use NASA TLX, SUS, or GEQ

Diagnostic tools like NASA TLX break experience into workload dimensions. SUS focuses on perceived usability of a system in a broad product sense. GEQ focuses on game experience dimensions through a longer instrument.
My study had a different goal. I wanted to understand whether the gestures remain doable and whether they preserve control and flow under speed. I also wanted to keep the session focused and to avoid participant form fatigue. Long diagnostic questionnaires can interrupt the flow of a demanding speed task, and they can increase fatigue without directly adding the kind of evidence I need for SQ3.
NASA TLX is also designed as a workload measure. It remains useful in many contexts. But in my case, I cared more about concrete play consequences than about decomposed workload scores (Bianchi-Berthouze et al. 2007).

Transcript coding scheme for reported consequences

Some SQ3 outcomes do not show up in logs or video. Participants can feel loss of rhythm, rising stress, or subtle strain before behaviour fully breaks. So I used the post experiment transcript to extract SQ3 relevant content.
I coded felt consequences using outcome focused codes:

- Errors and breakdowns, meaning what participants felt they got wrong and why
- Speed adaptation, meaning simplification, strategy changes, or skipping
- Fatigue and comfort, meaning strain, discomfort, or sustained comfort under repetition
- Control and flow, meaning stress, loss of rhythm, or calm continuation
- Memory aids and coping, meaning chunking, verbal rehearsal, or acceptance of misses

I chose these codes because they map to the consequences that matter for fast paced interaction. They also map to known player experience ideas like agency, flow disruption, and embodied strain. I used systematic coding practice where I treat each statement as a meaning unit and I group patterns carefully across participants (*Qualitative Data Analysis* n.d.).

Summary of data collected and what it answers

Across the whole study, I collected both qualitative and quantitative data.

Qualitative

- Think aloud commentary during gesture invention
- Participant explanations of why gestures map to moves
- Post experiment interview responses about intuition, imagery, agency, disruption, and comfort

Quantitative

- Prompt timing logs from the overlay
- Wizard input logs from the keystroke logger
- Matched logs that support reaction time estimation and correctness classification

Role of participants

Participants acted as both users and designers. They acted as users when they performed gestures to trigger game actions. They acted as designers when they invented the mapping and revised it to resolve ambiguity.

Link to sub questions

- SQ1 uses the Experiment 1 gesture videos and participant explanations to describe gesture form and meaning.
- SQ2 uses the Experiment 1 gesture set to analyse convergence and divergence patterns.
- SQ3 uses Experiment 2 logs, video checks, and transcript codes to analyse reliability and consequences under speed.
- SQ4 uses the post experiment interview questions to evaluate immersion related qualities across slow and fast contexts.

Chapter 4

Sub-Question 1

4.1 Methodology

Purpose and analytic stance

The subquestion 1 asks:

> How do hand gestures look like when performed to control fighting game actions, even when no direct body-to-avatar mapping exists?

It asks how user-invented hand gestures look, and what they express when players try to control fighting game actions without a direct body-to-avatar mapping. I used the Wizard-of-Oz (WoZ) dataset to describe gesture form and to interpret the mapping logic participants seemed to rely on. I treat the findings as descriptive and interpretive. I do not use them to predict behavior outside this study, and I avoid general claims because the participant pool is small (N=10). My unit of analysis is one gesture instance that is one participant's performed gesture for one move. I combine three sources of evidence for each instance:

1. physical descriptors from video annotation (what the gesture looks like),
2. representational descriptors (what the gesture expresses and how), and
3. the participant's own think-aloud explanation (why it made sense to them).

This combination is inspired from gesture elicitation work that collects user-defined input, then analyzes both the observable form and the user's intended

meaning (Wobbrock et al. 2009). A key assumption in SQ1 is that "no direct body-to-avatar mapping" acts like a design pressure. It pushes users to find other ways to make gestures learnable and usable, especially for repeated actions. Therefore, I treat Navigation, and the Neutron Bomb move as stress-test cases: these moves are harder to represent as a one-hand bodily simulation, so they will hopefully expose how participants switch to control metaphors, proxy limbs, or simplified cues.

Data sources and preparation

I used four data sources for SQ1:

1. annotated gesture videos for all participants and moves,
2. think-aloud notes collected during gesture invention,
3. pre-questionnaire responses used as background context, and
4. Paul move metadata (move timing and my allocentric/egocentric framing labels).

The metadata helped me interpret timing, and reference-frame related comments, but I did not treat it as a variable for hypothesis testing. I annotated gestures in ELAN, which supports multi-tier time-aligned annotation for multimodal data (Sloetjes & Wittenburg 2008). For each move, I segmented a gesture interval that covers the deliberate movement used to trigger the action. I then coded that interval across multiple tiers (described below). After annotation, I exported the tiers to a master spreadsheet, one row per gesture instance. Before analysis, I normalized values to avoid counting the same label as separate categories. This mattered most for Müller modes, where I collapsed slash-separated variants by taking the first term (e.g., enact/act → enact; hold/handle → hold; mold/shape → mold; embody/spatialize → embody). I also treated multi-value codes consistently: if a gesture contained multiple motion primitives, I stored them as a comma-separated list, and I counted each listed primitive when computing frequency distributions.

Coding framework rationale

Physical descriptors (gesture appearance)

I coded physical descriptors to capture how gestures are built at the level of observable movement and configuration. This supports SQ1 because it lets

me describe "what gestures look like" without assuming a specific recognition system or sensor.

Motion primitives: I broke each gesture into basic movement components (e.g., tap, extend, compress, rotate, flick, tilt). I use primitives as a compact way to describe motion structure and to compare gestures that differ in surface details but share the same underlying movement pattern. This approach aligns with the idea (also seen in ASL descriptions and gesture literature) that movement can be represented using a small set of discrete features rather than requiring full motion capture (Tennant & Brown 1998, Madapana et al. 2020).

Spatial direction / intent: I coded the intended direction of movement (up, down, left, right, toward self, toward screen). I prioritized intent when there was no literal displacement. For example, I allowed "toward screen" even when the hand stayed in place if finger motion clearly projected forward intent. Direction is a useful descriptor because it captures what an action is oriented toward, and direction-of-motion categories are commonly used to describe gesture form (Tennant & Brown 1998, Madapana et al. 2020).

Handshape: I coded the dominant hand configuration during the gesture (e.g., fist/closed, flat hand, point, handgun, thumbs up, etc.). Handshape often acted as a category marker, especially when participants tried to keep a small set of distinct triggers. This mirrors how handshape is treated in sign descriptions: it is visually salient, naturally discrete, and therefore practical for building a clear gesture vocabulary (Tennant & Brown 1998, Janke & Marshall 2017).

Palm orientation: I coded the palm's facing direction (up, down, left, right, toward screen, toward self) to capture how participants stabilized gestures and differentiated similar movements. Palm orientation works well because it reduces a complex 3D pose into a simple, observable feature: the direction the palm faces (up, down, left, right) usefully captures how the hand is "turned," and once you know that, the orientation of the fingers/back of hand is largely implied (Tennant & Brown 1998).

Complexity: I labeled a gesture as simple if it contained a single motion primitive within the gesture interval, and compound if it combined two or more primitives. I use this rule because it makes the complexity label reproducible and directly tied to the movement structure I coded.

Duration: I measured gesture duration as the mean of the length of all the annotated gesture intervals for a specific gesture move. I then report medians and ranges rather than means because the dataset is small and durations can be skewed by outliers.

Mental model descriptors (what the gesture expresses)

Physical form alone does not explain how participants intended a gesture to map to a game action. To capture representational intent, I coded each gesture using McNeill's gesture types and Müller's modes of representation. These frameworks are complementary: McNeill's types describe what kind of meaning a gesture conveys (iconic, metaphoric, deictic, beat), while Müller's modes describe how the hands depict that meaning (e.g., enact, hold, mold, trace, embody) (McNeill 1992, Müller 2014).

McNeill's types: I coded gestures as iconic when they depicted a concrete action or trajectory, metaphoric when they expressed an abstract idea through space or shape, deictic when they pointed or anchored reference, and beat when they mainly structured speech rhythm.

Müller's modes of representation: I coded enact when the participant simulated doing an action, trace when they drew a path or outline, mold when they shaped an imagined object/force, hold when they handled an imagined control or object, and embody when they used the body or space as a stand-in for a more abstract structure. When a gesture blended the modes, I coded them with compound labels, with the dominant mode appearing first, followed by the secondary mode. I used both layers because SQ1 is not only about which physical shapes appeared, but also about what participants seemed to be "doing" with their hands: acting out a move, drawing a path, pressing an imaginary button, or building a control metaphor. This is especially important in indirect mapping cases, where the same primitive can serve different meanings depending on intent.

Reliability and quality control

A second annotator independently coded a subset of the dataset (2 out of 10 participants; 20%) to check whether the coding scheme is applicable beyond a single rater. I assessed inter-rater reliability with Cohen's kappa for each tier, which measures agreement corrected for chance (Cohen 1960). Agreement was high for lower-level descriptive tiers (motion primitives; $\kappa = 0.71$, handshape; $\kappa = 0.83$, and palm orientation; $\kappa = 1.00$). For higher-level interpretive tiers, agreement was moderate (Müller modes $\kappa = 0.53$; McNeill types $\kappa = 0.56$). This pattern is expected: disagreements tend to cluster in conceptual labels rather than in visible movement features. Given the exploratory nature of the study and the interpretive load of type/mode coding, I treat these values as an adequacy check rather than a definitive reliability claim. When disagreements occurred, we discussed the case, clarified definitions, and updated the coding notes. I did not re-label the full dataset

based on discussion outcomes. Instead, I used the agreement check to make the coding protocol clearer and more stable. For e.g. the flick and extend primitives were discussed to be the same fundamental movement, hence the coding scheme was updated to merge the two labels under extend.

Analysis procedure for SQ1

I first summarize gesture form across all moves. I then report category-level profiles for Navigation, Single Attacks, Multi-hit Strings, and the Neutron Bomb (ambiguous Embodiment) stress test case. I finish with a cross-cutting synthesis of the mapping rationales that show up in think-aloud data.

Gesture appearance profile (quantitative summaries)

For each move category, I computed frequency distributions for motion primitives, spatial directions, handshapes, and palm orientations. I report the top components (e.g., top three primitives) with their percentage occurrence to make the dominant building blocks visible. I also report the distribution of complexity (simple vs. compound) per category. For timing, I summarize gesture durations using median and range. I use descriptive comparison rather than statistical testing. With N=10, tests would be underpowered and would encourage over-interpretation.

Representation strategy profile

I summarized McNeill types and Müller modes as category-level distributions. I also looked at typical pairings (e.g., enact + iconic) because they are often more informative than either label alone. This step supports SQ1's focus on what gestures express under indirect mapping pressure.

Thematic pass on think-aloud data

I conducted a thematic pass over the think-aloud notes to capture the reasoning participants used to justify gestures. I first labeled recurring reasoning patterns in the notes (open coding). I then merged closely related labels into a smaller set of themes that explain why gestures took particular forms. I treated repetition as a signal, but I also kept themes that appeared in only a few participants if they explained a distinctive design choice. I used short participant phrases, or close paraphrases from my notes, as anchors for each theme. This keeps the themes tied to the data. I treat the themes as descriptive mechanisms observed in this study, not as claims about how players in general reason. Finally, I used the themes to interpret the quantitative

profiles. For example, if Navigation shows high use of taps and palm-down stability, I connect that to think-aloud statements that frame the gesture as a keyboard tap or a joystick push.

How questionnaire and Paul move metadata were used

I used the pre-questionnaire in two ways. First, I use it to describe the sample's background and gesture readiness. Second, I use it for gentle pattern checks when interpreting think-aloud strategies. I do not compute correlations or claim causal links. When I mention a connection, I frame it as a plausible explanation supported by a participant's own phrasing and their background context. I also created a small metadata layer for Paul's moves. I annotated move timing and labeled each move's dominant reference frame as egocentric, allocentric, or both, drawing on prior discussion of player–avatar interaction and reference frames in fighting games (Mattiassi 2019). I use this metadata to support interpretation, especially for Navigation and Neutron Bomb, where reference-frame ambiguity is part of the indirect mapping pressure.

4.2 Results

Sample context and gesture readiness

This study included 10 right-handed participants (ages 19–31, mean ≈ 25). Most reported using PC (8/10) and console (7/10). Mobile (6/10) was also common. Motion-control exposure was limited. Six participants reported no motion-control experience. The others mainly referenced systems like Wii, Kinect, or VR.

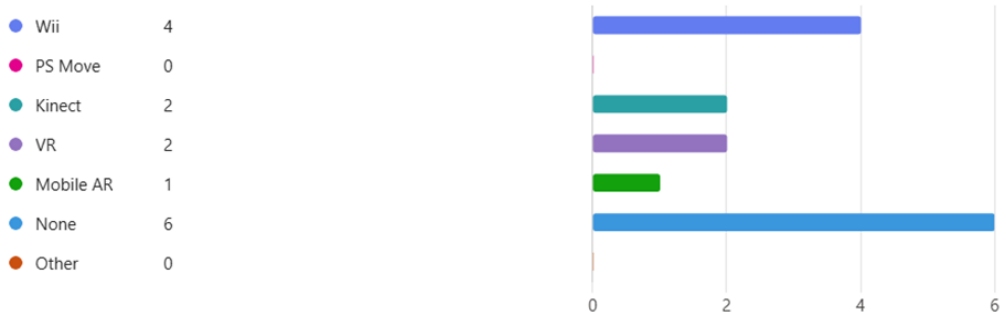


Figure 4.1: Motion Control Experience

General gaming familiarity varied. Weekly video game play ranged from 0–10 hours. Weekly fighting game play was low overall (many reported 0 hours; the highest was 3 hours/week). Tekken 8 familiarity was mostly "*played a few times*" (6/10). One participant played weekly, and one had never played. Participants also differed in how ready they felt to invent gestures. Half agreed they felt comfortable inventing gestures, while the rest were neutral. Embodiment attitudes were mixed. Only a minority agreed with "*I think of my on-screen character's limbs as my own*," and several disagreed. This matters for SQ1 because when body ownership is weak, participants may lean more on control metaphors (keyboard/joystick/touch logic) than on full-body simulation. A gentle pattern check fits that idea. P007, who framed gameplay through keyboard habits, said: "*I'm used to playing on a keyboard... feels like I'm tapping the left arrow key*" and "*feels like I'm using a normal keyboard.*" In contrast, P005, who reported minimal gaming background, described their early choices as "*kind of like mimicking the character so far.*" I treat these as suggestive examples, not as rules, given the sample size.

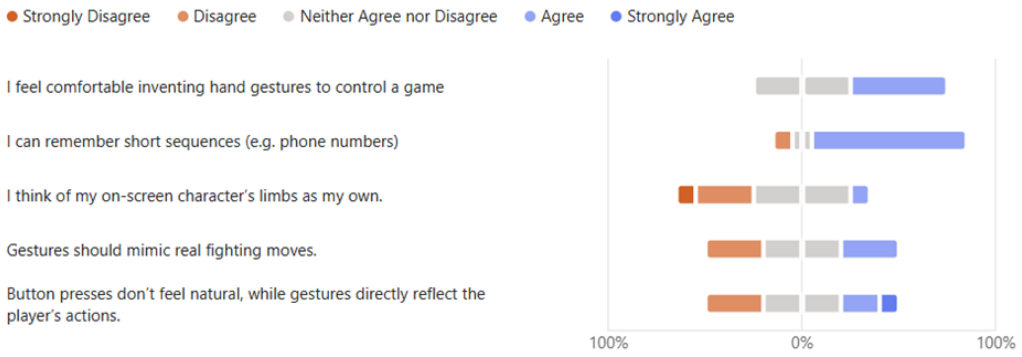


Figure 4.2: Participant Opinions

Overall gesture form across all moves (overview)

Across move categories, participants reused a small set of components rather than inventing completely new forms for every move. The most common building blocks were:

- simple motion primitives such as tap, extend, compress, rotate, tilt, drop
- stable hand configurations, especially fist/closed and flat hand

- a strong bias toward palm-down orientations (most visible in Navigation and Single Attacks)

Even when a move did not map naturally to a one-hand action, gestures still looked structured. Several participants tried to define a consistent "control grammar" early on. This included a neutral start position, opposite pairs (forward vs backward), and a small set of distinct actions. For example, P001 explicitly wanted "to decide on a starting position for the hand." P009 described a neutral reference using a number-line idea, where moving away from neutral counts as input, and returning does not.
At a high level, the category profiles in this dataset look like this:

- **Navigation**: mostly simple gestures; shortest durations
- **Single direct attacks**: still often simple, with stronger iconic/enact patterns
- **Multi-hit strings**: more often compound, with sequencing and combination logic
- **Neutron Bomb**: mixed framing (often "both" ego + allo), which fits its ambiguous mapping

Navigation gestures (primary indirect mapping case — the stress test)

Navigation is the clearest stress test in this dataset. Locomotion and sidesteps do not have a natural one-hand analogue. Here, "no direct body-to-avatar mapping" shows up as a design pressure. Gestures need to be easy, repeatable, and clear, even if they are not body-faithful.

Category summary

Navigation gestures were mostly simple (90% simple) and relatively fast (median duration 0.99s, range 0.52–1.75). They were most often built from tap and swipe-like primitives. Handshape was often fist/closed or point. Palm-down was dominant (≈61%). In representational terms, Navigation leaned egocentric (≈68%) and was often metaphoric (≈60%), with mold and hold appearing frequently. Overall, Navigation gestures in this study often read as control actions rather than locomotion reenactments.

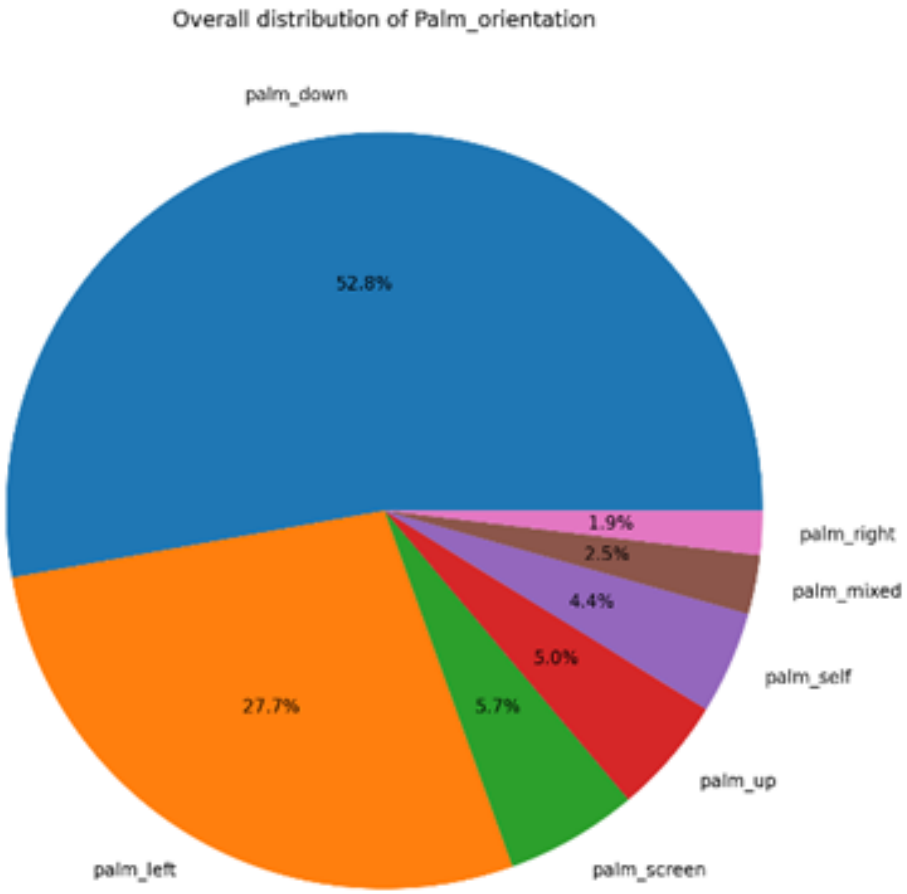


Figure 4.3: Overall distribution of palm orientation

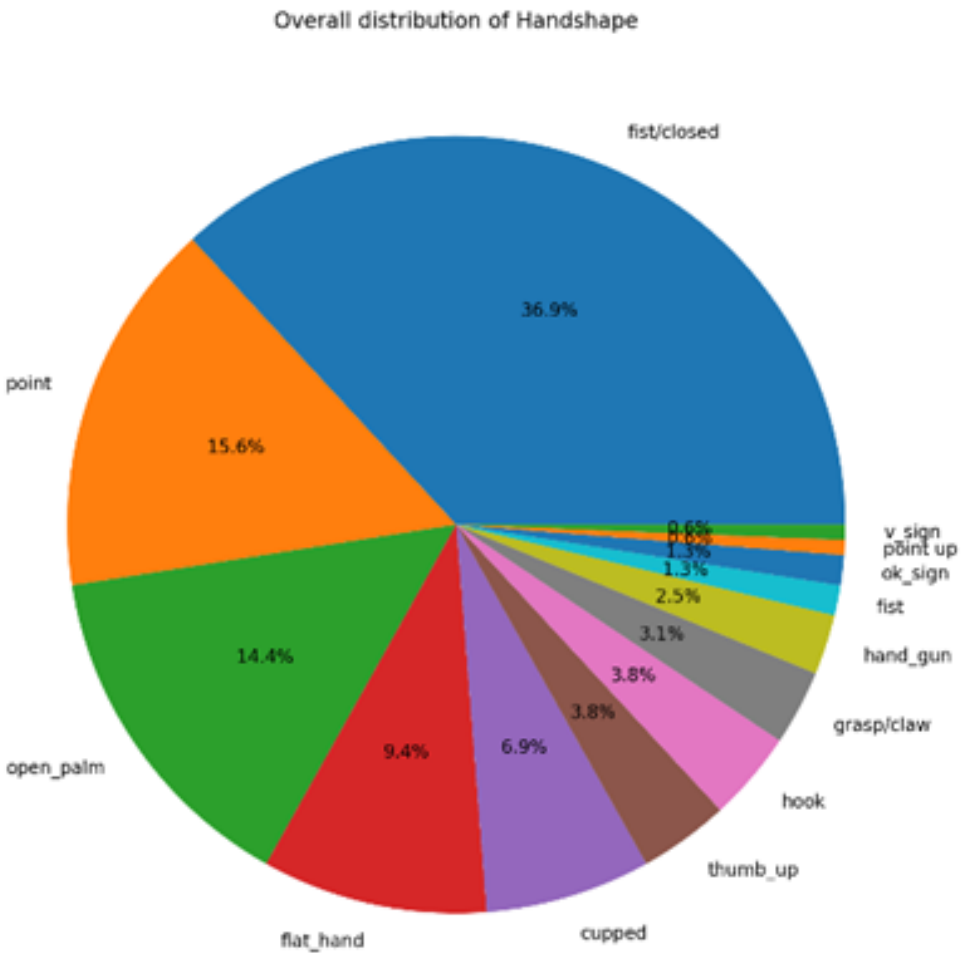


Figure 4.4: Overall distribution of handshapes

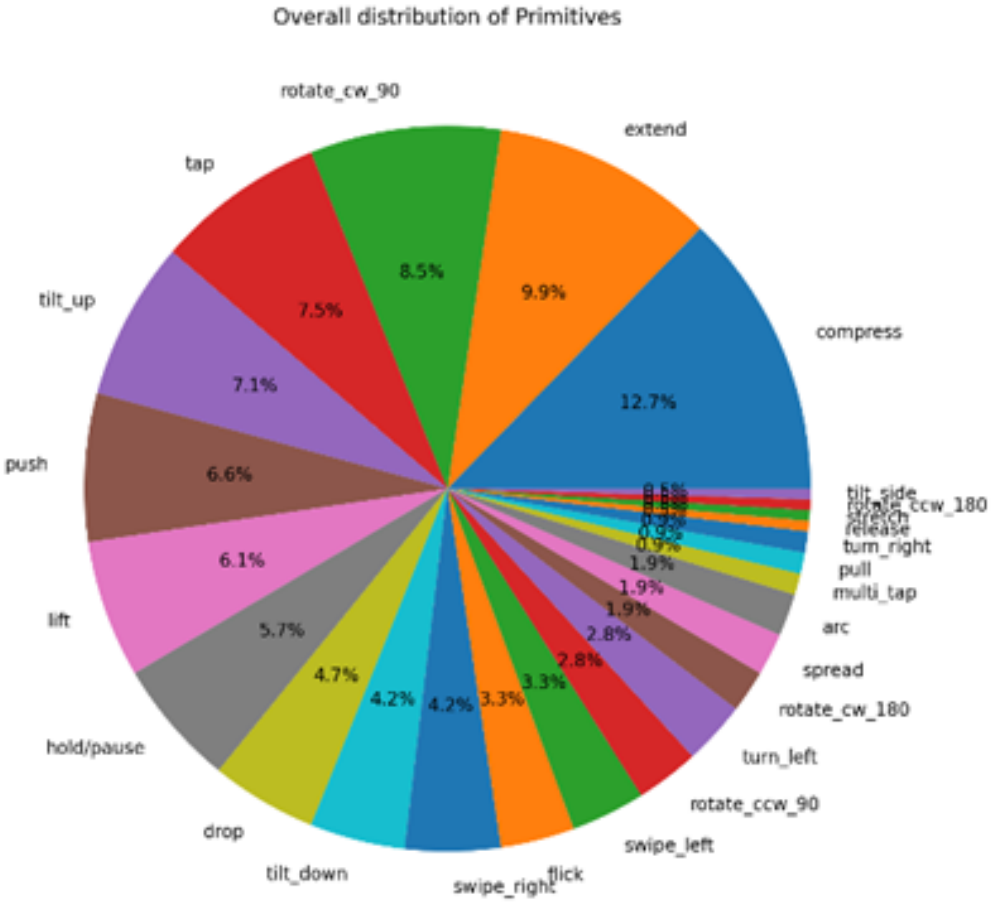


Figure 4.5: Overall distribution of primitives

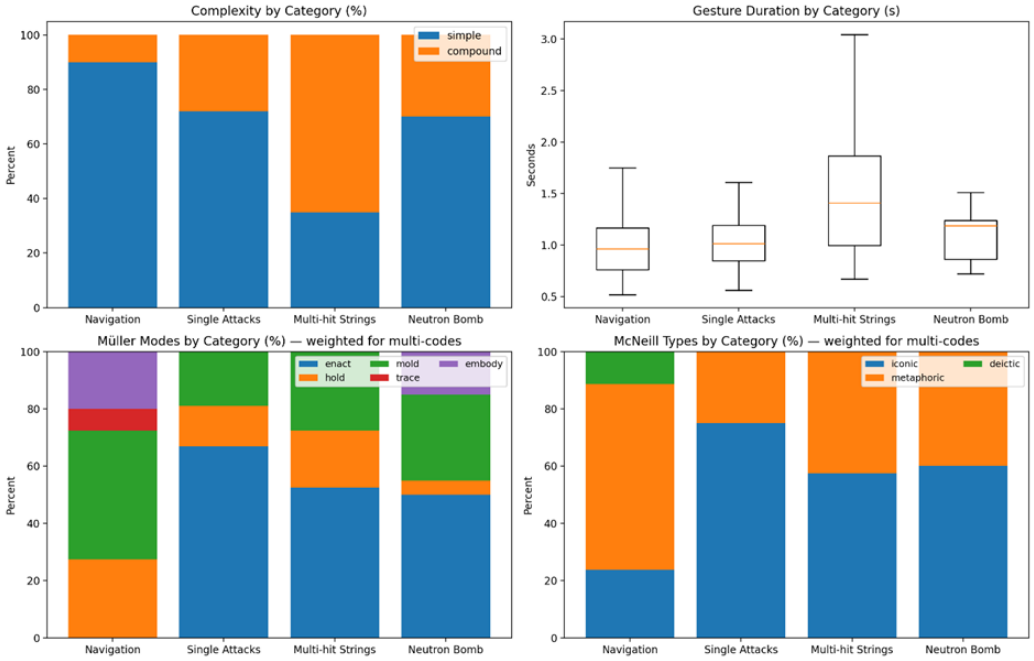


Figure 4.6: Overall distribution by category

Mental model link

In think-aloud, participants repeatedly described navigation as operating an input device.

- Keyboard framing. P007: "*feels like I'm tapping the left arrow key*" and for sidestep "*I will tap with my middle finger... feels like I'm using a normal keyboard.*"
- Joystick framing. P002 described a return-to-neutral rule: "*kind of like the joystick. if you push it out, it comes back to the normal position.*"
- Touch/drag framing. P004 suggested proportional control using "*a drag or a tap*" with two fingers.

This is also where repeatability showed up most clearly. P010 worried that "*going back three times might be difficult.*" Several participants also raised ambiguity concerns with mirrored gestures. P003 spent a long time trying to understand the mirrored-gesture issue, and P009 noted that a system could "*capture both gestures and can't differentiate.*"

Interpretation for SQ1: under indirect mapping pressure, Navigation gestures often looked like small, controller-like actions (tap/slide/drag) orga-

nized around a neutral home position and polarity rules, rather than "walking with the hand."

Single direct attacks (more obvious mapping, but still simplified)

Single attacks have a clearer body analogue (punch/kick). Still, the mapping is constrained by one-hand input limits, distinctiveness across many moves, and comfort.

Category summary

Gestures for Single Direct Attacks were mostly simple (72% simple, 28% compound), with a median duration of 1.07s (range 0.56–1.89). They were often built from extend and compress, commonly directed toward the screen (≈38%), and frequently performed as a fist/closed handshape (≈53%). Palm-down remained common (≈55%). Representation leaned allocentric (≈78%), with a strong iconic tendency (≈72%) and enact as the most common mode (≈66%). A frequent pairing was enact + iconic (≈55%). Overall, these attacks were often enacted, but in a reduced form.

Mental model link

Participants often kept one clear "essence cue" rather than acting out the full move.

- P001 explained uppercut using source-of-motion logic: "motion... comes from underneath... closest possible gesture..."
- P008 used fingers as limb stand-ins: "my thumb kinda looks like a leg extending to kick" and described it as "so much fun."
- P003 emphasized speed and repetition for punches: "can easily spam and hurt the other character."

Distinctiveness mattered here. P001 explicitly said punches should use finger differences because they need to be "*distinct*." That pressure likely contributed to smaller, more discrete triggers (finger actions, compact motions) instead of large arm-like movements.

Interpretation for SQ1: even when mapping is more direct, gestures still tended to be small and optimized. They kept attack meaning through iconic/enacted cues (direction, impact, trajectory) rather than full realism.

Multi-hit strings (temporal structure + chunking)

Multi-hit strings add a different pressure. Gestures now need to express sequence and timing. This is where gestures start to look like short phrases, not single signs.

Category summary

Multi-hit string gestures were more often compound (65% compound), with a median duration of 1.48s (range 0.67–3.04). They often combined compress with rotation or downstroke components. Movement directions were commonly toward the screen (≈33%) or down (≈28%). Handshape still often used fist/closed (≈37%), but variety increased (including occasional "hand gun"). Representation was mixed. Allocentric was most common (≈54%), but "both" framing also appeared frequently (≈39%). Iconic (≈56%) and metaphoric (≈44%) were closer here than in Single Attacks, with enact still most common (≈52%). This fits the idea that combos blend depiction with shortcut strategies.

Mental model link

Think-aloud points to three recurring approaches:

1. Compose: chain component gestures.
   P001 described having a method: "*I developed my own algorithm of developing the gesture.*" P004 wanted to "*do separate moves quickly*" because it would reinforce memory.

2. Compress: replace multiple hits with one shortcut.
   P003 wanted a single ring-finger tap for an entire combo later on, consistent with "one trigger = one function."

3. Re-time: adjust gesture duration to match the animation pace.
   P005 liked left kick because "*the lag was minimal... the sync was nice.*" Even when participants did not explicitly talk about timing, their concerns about practicality (too long, too hard to repeat) point to timing as an underlying constraint.

Interpretation for SQ1: under temporal pressure, gestures often became either structured sequences or compressed macros. They encoded order and timing more than body mechanics.

Neutron Bomb (ambiguous embodiment — the second stress test)

Neutron Bomb is not a clean punch/kick analogue, and it is not a simple navigation command either. It sits between "depict an action" and "trigger a special move." That ambiguity makes it a stress test in a different way than Navigation.

Category summary

The Neutron Bomb case showed mixed form. Gestures were 60% simple and 40% compound, with a median duration of 1.51s (range 0.67–2.42). Common primitives included compress and extend. Directions were split between down and toward the screen (each ≈25%), with up also present (≈20%). Handshape was divided: flat hand and fist/closed were tied as most common (each ≈33%). Palm orientations were more varied than in other categories (palm-down ≈38% was still most common but less dominant). Representation was often explicitly both egocentric + allocentric (≈80%). Iconic and metaphoric types were tied (≈50% each). Modes also split between enact (≈46%) and mold (≈39%). Overall, this category did not settle into one dominant strategy in this study.

Mental model link

Participants justified Neutron Bomb in different ways: motion shape, symbolism, and sometimes redesign due to collisions.

- P001 described it as "*like the emoji*" and started with "*turning clockwise.*" They then noticed overlap: "*gesture overlaps the uppercut,*" and tried to invent a clearer front-flip gesture.
- P007 used motion-shape matching: "*he is moving in a circle... it really represents what it does... my mind maps into that.*"
- P005 tied it directly to the avatar's path: "*gesture is easy to do and it is also how the character is moving.*"

Interpretation for SQ1: when a move is hard to embody cleanly, gestures often become motion sketches (circle/arc) or function cues, and participants may iterate when gestures collide with other moves.

4.3 Discussion

SQ1 asked: How do hand gestures look when they are performed to control fighting game actions, even when no direct body-to-avatar mapping exists? The results show that, in this study, participants rarely treated gesture control as "acting out Paul." Instead, they often approached it like designing a small input language that could plausibly survive repeated use in a fighting game. That shift matters for the main research question. Wizard-of-Oz prototyping did not only yield candidate gestures. It also surfaced the rules, constraints, and mental shortcuts participants relied on when the mapping was indirect. Those "design moves" are part of the potential of WoZ elicitation, because they point to what a real gesture system would need to support.

Indirect mapping pressure changes what "natural" means

One central pattern across categories is that indirect mapping pressure changed what participants treated as natural. For moves where a body-faithful gesture is hard to invent (navigation and Neutron Bomb), naturalness did not primarily mean realism. It more often meant that the gesture is quick, distinct, easy to repeat, and consistent with the rest of the set. In the quantitative profiles, this shows up as high simplicity in navigation, short median durations, and strategy profiles that lean metaphoric and "mold/hold" rather than enactment. In the think-aloud data, it shows up as participants explicitly talking about starting positions, opposites, mirrored-gesture problems, and repeatability. This is an important clarification for SQ1. The absence of direct mapping did not make gestures "random" or "arbitrary." It pushed gesture design toward constraints that resemble game input design: stability, low effort, and low confusion. In other words, indirect mapping pressure seems to move the goal from "depict the move" to "produce a reliable command that still feels motivated."

Navigation as the primary stress test: control metaphors and command grammar

Navigation is the clearest stress test in this dataset because locomotion and sidestepping do not have an obvious one-hand analogue. Participants seemed to resolve this by borrowing control metaphors and by building a command grammar. Several participants described navigation explicitly in terms of familiar input devices. P007 framed forward and backward as "tapping the left arrow key... WASD," and described sidestep in a way that matched "a

normal keyboard." P003 described "imagining holding a mouse and spamming a button." P002 used a joystick analogy to justify a neutral position that "comes back to the normal position." These are not small comments. They reveal that, for navigation, participants often treated the hand less like a proxy body and more like an input surface. For locomotion-style referents, the mapping problem is often less about depicting an avatar's body and more about finding a compact, discriminable command form; this aligns with gesture-set research showing that users frequently gravitate toward simple, directional, command-like structures when the referent is abstract (Wobbrock et al. 2009). This helps explain why navigation gestures were physically small in this study. Taps and short swipes are not just convenient, they match the action requirements of locomotion in a fighting game where you may need to trigger the same action repeatedly and quickly. This also connects to participant 10, who worried that "going back three times might be difficult." That concern is about repeated activation and effort, not realism. In a fighting game, the "best looking" gesture is not necessarily the best playable gesture. Another pattern in navigation is that several participants tried to establish a neutral "home" position and polarity rules. P001 explicitly wanted "to decide on a starting position for the hand." P009 described a center point "like a number line," where moving away from neutral counts as input. This kind of grammar building is a system-level behavior. Participants were not only designing single gestures, they were designing a set. That matters because a gesture vocabulary is not judged one gesture at a time. Every new gesture competes with existing ones for distinctiveness and recognition. The mirrored-gesture issue is a good example of how "set design" enters the picture. In this study, mirrored gestures were not only a representational choice, but also a recognition and ambiguity concern. 7 out of the 10 participants first choice was a mirrored gesture for the navigation moves. P003 "spent a long time" on the mirrored-gesture problem, and P009 noted that a mirrored gesture can lead the computer to "capture both gestures and can't differentiate." Even if no recognition algorithm was implemented in the WoZ setup, participants still reasoned about what a system could confuse. That suggests an opportunity for design: WoZ sessions can reveal where users expect recognition failure, and which gesture forms feel brittle under indirect mapping.

Neutron Bomb as the second stress test: ambiguity triggers hybrid strategies and redesign

Neutron Bomb is the second stress test as it is 'indirect' and ambiguous in what it demands from the hand. The move is a front flip with an arc and rotation. Participants can interpret it as a motion-shape problem (circle, spin, arc), an effect problem (a special move), or even a sequence problem (a forward step followed by a flip with sub-actions). As a result, Neutron Bomb showed the most mixed representational profile in this study, and it frequently triggered redesign. Several participants used global motion-shape matching. P007 said the character is "moving in a circle" and wanted to "mimic his motion." P005 described the gesture as "easy to do" and aligned it with how the character moves. At the same time, symbolic stand-ins appeared. P001 described it as "like the emoji," and initially used rotation. Importantly, P001 also noticed a collision with the gesture they invented earlier for the uppercut. They then started looking for a new gesture that better isolates the flip. That shows a realistic constraint: ambiguous moves can sometimes generate gestures that look like other gestures (rotations, arcs), which then forces users to either accept overlap or engineer distinctiveness. This is where indirect mapping pressure becomes visible as iteration pressure. For Neutron Bomb, participants were not only choosing a gesture. They were negotiating the space of acceptable gestures under three constraints at once: the gesture must feel motivated, must not collide with existing gestures, and must remain physically easy. In this study, that negotiation often produced a hybrid: gestures that are partially iconic (motion-shape cues) and partially metaphoric (special-move signaling). Work on silent gesture shows that people often combine different iconic strategies (e.g., acting + representing) depending on what kind of meaning they are trying to express, which supports interpreting the Neutron Bomb gestures in this study as a genuine ambiguity in what is being represented (action trajectory vs. outcome vs. salient features) rather than simple inconsistency (Ortega & Özyürek 2020)

From embodiment to minimum viable depiction: why single attacks still look compressed

Single direct attacks are the category where body depiction is most available. Punches have clear analogues in human movement. But because only hand and wrist are available for gestures, participants can't rely on full-body reenactment. Therefore, they have to either compress the move into a small set of hand-level kernels (direction, contact type, handshape), or abandon depiction and use symbolic/metaphor-based control gestures that function more like

input commands than miniature reenactments. In my study, participants rarely produced full reenactments. Instead, a useful interpretation is that participants often aimed for minimum viable depiction: one or two cues that preserve the move's essence while keeping the gesture compact and distinct. The think-aloud data shows this clearly. For uppercut, P001 justified the gesture by source-of-motion, saying the move "comes from underneath," and therefore turning the hand and closing into a fist is "closest." This is not a literal uppercut. It is a cue selection process. The participant keeps the "from below" idea and compresses the rest. Similar cue selection appears in the hammer punch, where participants often emphasized the downward strike, and in kicks where the thumb becomes a stand-in for the leg. Even when users can "act out" a move, mid-air interaction research highlights fatigue as a real constraint (the "gorilla arm" problem), which supports treating reduced-range, compact gesture forms as a sensible adaptation rather than as a failure of embodiment (Hincapié-Ramos et al. 2014). This also helps interpret why single attacks leaned strongly enact and iconic in the representation profile, while still using compact hand actions. Enactment does not have to mean full-body mimicry. In this study, enactment often meant enacting the core dynamic: an upward drive, a downward strike, a forward hit. These are "gesture kernels" rather than literal performances. Distinctiveness pressure is also stronger than it may look. Participants needed to separate five attacks that are all "hit the opponent," plus additional special moves. P001 explicitly said punches "need to be distinct" and therefore chose finger-based differentiation. To assign different physical forms to different moves, even when the moves are conceptually similar, is a system design decision. In this study, that pressure likely contributed to the high use of fist/closed handshape and palm-down orientation across attacks, combined with small differentiators such as specific finger extensions, small rotations, or direction cues. It also suggests that distinctiveness should not be treated as an emergent property in the final system. Rather it should be designed for.

Proxy-body strategies: embodiment without full-body ownership

A striking cross-category strategy is that participants often treated the hand as a proxy body, where fingers represent limbs. P002 explicitly wanted to "use fingers for arms and legs." P009 described "fingers as my two legs" for locomotion. P008 described the thumb as "a leg extending to kick someone." P010 mapped individual fingers to the avatar's hands. This strategy is worth discussing because it offers a middle ground between two extremes: full-

body embodiment versus pure controller metaphors. In this study, proxy-body mapping seems to allow embodiment even when participants do not fully identify the avatar's body as their own. That matters because the pre-questionnaire responses on "I think of my on-screen character's limbs as my own" were mostly neutral or negative. Yet the gesture data still contains strong iconic and enactment strategies for attacks. A plausible interpretation is that participants do not need to feel bodily ownership over the avatar to produce embodied gestures. They can externalize embodiment into a model: "my hand represents the character." That model can be literal (fingers as limbs) or schematic (fist as an impact unit). This also aligns with accessibility goals, because proxy-body mapping can remain small and comfortable while still feeling meaningfully connected to action.

Reference frames and the fighting game context: why direction gets tricky

One detail that becomes more important in a fighting game than in many gesture-control examples is reference frame instability. In Tekken, "forward" and "backward" are relational. They depend on the opponent and on which side of the screen the character occupies. Paul Phoenix's allocentric and egocentric directions differ for some moves, which reflects this contextual dependence. In the gesture profiles, navigation directions often appear as left/right screen directions rather than world-level forward/back. That suggests participants may be implicitly designing for a single orientation (as seen during elicitation) rather than for side switching. Design implication wise, this suggests that a gesture vocabulary that uses screen-relative left/right for "forward" may feel natural in one stance, but it will break when the character changes sides, unless the system reinterprets it relative to the opponent. This provides a concrete implication: a gesture-based fighting game control system likely needs to manage reference frames explicitly. It can either (1) reinterpret gestures dynamically based on character facing, or (2) encourage relational mappings such as "toward opponent" versus "away," or (3) offer both and let users choose. Without that support, indirect mapping pressure will continue to push users toward mappings that feel locally correct but are globally fragile.

Multi-hit strings: gesture phrasing, chunking, and timing as part of form

For multi-hit strings, the problem changes from shape to phrasing. Participants must decide how to represent sequence and how to handle timing. In this study, participants seemed to rely on a small set of strategies, namely: compose, compress, and re-time. Composition appears when participants chain existing move gestures. P004 explicitly described doing "all 3 moves... no separate sign for it... one thing less to remember." This is an efficiency argument at the memory level: if the combo is already built from known parts, the gesture can reuse them. In contrast, compression appears when participants collapse a string into a single shorthand, often motivated by playability. P008 described choosing "only a single flick" and called it "underwhelming," but still kept it, which suggests a performance tradeoff: the gesture becomes less expressive but easier to execute repeatedly. Sequence-based commands are often learned as structured patterns rather than as literal depictions, and work on marking-menu learning supports the idea that users can become fluent with short, directional or chunked gesture patterns once the set is consistent (Kurtenbach & Buxton 1994). Re-timing is an interesting find because it connects gesture appearance to the on-screen animation. P005 commented that the gesture felt better when it aligned with the character's timescale, saying that earlier it felt like they "finished" the gesture while the character was still moving. This suggests that for combos and longer moves, the "look" of the gesture includes tempo. A gesture that is correct in shape but wrong in timing may feel disconnected. Since Paul's combo moves have longer durations in the move sheet, it is plausible that some participants implicitly matched gesture duration to animation length, even if they did not explicitly mention it. This is visible in the higher duration of invented gestures for multi-hit strings. This provides another design implication for the main research question. WoZ elicitation can reveal not only gesture shapes but also timing expectations. A future recognition system would need to decide whether it detects a discrete command quickly (allowing the animation to play out), or whether it treats the gesture as a continuous control that must last for the move's duration. Participants in this study seemed to assume both models at different times, which suggests that the final system might need to be explicit about timing semantics.

Engineering constraints: comfort, mobility, and "spammability" shape the vocabulary

The think-aloud data includes several comments about foreground constraints that can be easy to overlook when reading gesture descriptions. P002 avoided ring and pinky because they are "not very mobile." P008 changed backward to a finger flick because it was "cumbersome," and noted wrist discomfort for another move. P010 worried about repeating a gesture multiple times. These comments show that the gesture vocabulary is shaped by the body, but not only through representation. It is shaped through ergonomics, fatigue, and repetition demands. This is where the results connect directly to the main RQ's notion of potential. A gesture-based control system for fighting games will succeed or fail based on repeated use, not on one-off performance. WoZ elicitation can surface which gestures participants themselves consider sustainable. It can also highlight where the design space differs from typical gesture-control demos. Many gesture interfaces assume that a large gesture is acceptable because the user performs it occasionally. Fighting games are the opposite. Inputs are frequent, and the player cannot afford to "perform" the move each time. The data in this study suggests that participants already anticipate this and compress gestures accordingly.

Gaming familiarity as a gentle pattern check: control schemas versus enactment

The pre-questionnaire results do not allow strong claims about individual differences, but it provides context that helps interpret why participants reached for certain mappings. Prior work notes that elicited gestures can be shaped by "legacy" interaction habits (e.g., learned conventions from existing devices), so in this study I treat differences in participants' gaming background as a plausible influence on strategy choices rather than as a stable causal effect (Morris et al. 2014). In this study, fighting game exposure was generally low, while general gaming platform use (PC and console) was high. Under those conditions, it is plausible that participants lean on generic control schemas when facing abstract commands. The think-aloud examples fit that interpretation. P007, who framed navigation through keyboard logic, consistently referenced arrow keys and WASD. P003 referenced mouse spamming. P002 used a joystick return-to-neutral metaphor. In contrast, P005 described early design choices as "mimicking the character," and P001 talked about matching motion sources and developing an "algorithm" for building gestures move by move. These are not mutually exclusive strategies. But they suggest a continuum: some participants begin from controller metaphors

and only add depiction cues when needed, while others begin from depiction and only shift to metaphors when the move becomes hard to express (for example, "make a gun" for a high-impact special move). This pattern is also consistent with the mixed attitudes in the pre-questionnaire. Participants were not unanimous that gestures should mimic real fighting moves, and they were not unanimous that button presses feel unnatural. That mixed stance supports a design implication: the system should not assume a single "correct" gesture style. It should support both depiction-oriented and control-oriented gesture strategies.

What SQ1 suggests about the potential of WoZ elicitation for the main RQ

Taken together, the SQ1 findings suggest several ways WoZ elicitation has potential for designing a gesture-based fighting game controller. First, WoZ elicitation can reveal the gesture vocabulary that participants naturally compress toward when they consider repeatability. Even with only 10 participants, the patterns show that users often gravitate toward small, low-effort gestures and toward systematic rules for navigation. That is valuable input for an accessible design, because it highlights mappings that remain plausible under fatigue and repetition. Second, WoZ elicitation captures the logic participants use to make indirect mappings feel motivated. In this study, the key mechanisms were building a neutral home and polarity rules, selecting minimum viable depiction cues for attacks, using proxy-body mappings (fingers as limbs), composing or compressing multi-hit strings, and explicitly managing gesture collisions. These mechanisms can be turned into design scaffolds. A system could provide templates that align with these strategies rather than expecting users to invent everything from scratch. Third, WoZ elicitation exposes where ambiguity lives. Neutron Bomb is a good example. Participants did not simply disagree. They interpreted the move in different ways (motion-shape, icon/emoji, special-effect). That suggests the system should allow multiple mapping solutions for ambiguous actions and should provide a way to pick among them. In other words, WoZ can help identify which moves need pluralism rather than standardization. Finally, the data suggests that reference frame management is a core problem in fighting games. If the system does not support side switching and relational directions, "natural" gestures may fail in real gameplay. This is the kind of constraint that users may not articulate directly, but that becomes visible when you compare direction coding with the game's semantics. WoZ elicitation can therefore inform not only the gesture set but also the interpretation

layer: how the system should map a gesture to an action given the current game context.

Chapter 5

Sub-Question 2

5.1 Methodology

Purpose and operationalization

SQ2 asks:

> "Does player behavior reveal a convergent core set of control gestures, or a broad, divergent set?"

I answer this by treating convergence as something that can appear at more than one level. Participants can converge on a complete gesture solution, but they can also diverge in full form while still converging on shared components (for example, the same spatial intent or the same primitive action).
Therefore, I operationalize SQ2 using two constructs that I measure in parallel.
Whole-gesture convergence: for a given game move (referent), do participants converge to one or two dominant gesture families that represent the same overall solution strategy? I use this to judge whether a referent has a population-level "default" gesture solution.
Component-level convergence: even when full gestures differ, do participants still converge on stable components such as spatial direction (L1), motion primitives (L2), or handshapes (L3)? I use this to test whether shared intent exists even when surface form diverges.

Dataset used for SQ2

This methodology assumes that all gesture instances have already been annotated and consolidated into a master sheet. Each row corresponds to one

gesture instance for one referent (one participant per referent). The columns used in SQ2 include the coded layer values (L1, L2, L3) and the interpretive labels used later for mental model analysis.
I do not repeat the study procedure here. SQ2 begins at the point where gesture instances have been coded and normalized into consistent categories. Two practical constraints shaped the analysis design.
First, the gesture space in this study is high-variance. Participants map complex fighting-game actions to hand motion, which leads to many valid executions for the same intended control concept.
Second, I did not have additional researchers available to perform independent gesture-family clustering for inter-rater reliability. I therefore needed a protocol that reduces ad hoc grouping and makes my reasoning visible, so that the clustering can be audited.

Why I did not use plain Wobbrock or Ruiz agreement alone

My approach is inspired by Wobbrock et al. (2009) and Ruiz et al. (2011). In both cases, the authors group gestures into families and compute an agreement score per referent to estimate consensus and support design decisions. That logic is attractive because SQ2 is also about consensus. However, applying the same pipeline directly would hide too much structure in my data. There are two main issues.
Issue 1: gesture families are internally heterogeneous in my dataset. Wobbrock-style family agreement assumes that gestures inside a family are treated as equivalent for the purpose of agreement. In my study, two gestures can clearly reflect the same control strategy, but still differ in motion dynamics or handshape. If I treat these as fully identical, I lose the nuance that I later need to explain divergence patterns.
Issue 2: the original agreement score is based on binary equivalence. Wobbrock agreement effectively uses the rule "same family" versus "different family." This collapses partial agreement. Two participants might share spatial intent and a key primitive but differ in another component. A binary metric would score this as full disagreement, even though behaviorally it may indicate shared intent.
Because of these issues, I still compute Wobbrock agreement (it is a useful whole-gesture summary), but I complement it with a layered representation, component-level metrics, and a within-family similarity analysis. Together, these additions make convergence and divergence visible instead of compressing everything into a single number.

Coding scheme: three-layer gesture representation

I represent each gesture instance using three layers. The layers are ordered by importance for clustering and similarity, because they capture different levels of meaning and embodiment.

The three layers are spatial agreement (L1), movement primitive (L2), and handshape (L3). I do this because gesture form can be described as separable articulation units rather than one indivisible "whole gesture." This mirrors how sign-language phonology decomposes manual actions into core parameters (e.g., movement/location/hand configuration), which supports consistent coding and clearer comparisons across participants (Sandler 2012, Stokoe 2005). In elicitation research, agreement outcomes also depend strongly on what criteria are used to group proposals, so making these layers explicit helps keep the analysis interpretable and reproducible (Vatavu 2019).

L1 - Spatial intent (spatial agreement): This layer captures the intended direction or target of the gesture from the participant's perspective (for example, left as the participant's left). I code intended stroke direction of the active articulator, not the resulting displacement. Example: for tapping gestures on a table, I code spatial direction as downward, because the participant is performing a pressing motion even if contact constrains movement.

L2 - Motion primitives: This layer captures atomic motion units such as extend, compress, rotate, push, tilt, trace, and similar actions. A single gesture instance can contain multiple primitives. I treat the primitive list as a set of components that jointly describe the mechanics of the gesture.

L3 - Handshapes: This layer captures discrete hand configurations such as fist, flat hand, point, thumbs up, and similar forms. I code the active handshape used to perform the gesture, not the resting handshape. Example: if a participant rests in a fist but performs a punch by extending the index finger, I code the gesture handshape as point, not fist.

Within each layer, each category is treated as a single unit for counting and agreement. I also apply small "smoothing" rules where a difference is purely cosmetic and does not change the functional action. For example, loose fists are still fists, slightly cupped hands are still flat hands, and finger flicks that function as extension are grouped with extend.

Analysis pipeline overview

I answer SQ2 by running the same pipeline per referent. The pipeline has six steps. The early steps measure component-level structure, the middle steps construct and characterize gesture families, and the later steps quan-

tify whole-gesture convergence and translate distributions into interpretable dominance structures.
The steps are:

1. Layer-wise grouping (component distributions)
2. Component dominance per layer (component convergence)
3. Gesture family construction (3-layer clustering protocol)
4. Wobbrock agreement score over families (whole-gesture convergence)
5. Pairwise similarity inside popular families (within-family coherence)
6. Dispersion Metrics (within-family coherence)
7. Mental model coding for dominant solutions
8. Dominance structure classification and selection rules (core vs divergent outcome)
9. User-designed gesture set using dominance structure (how gestures are carried forward)

Layer-wise grouping (pre-clustering)

I start by grouping gestures independently by L1, L2, and L3 values. This produces frequency distributions per referent, such as:

- which spatial directions appear most often (L1)
- which primitives are most used (L2)
- which handshapes appear most often (L3)

This step deliberately ignores gesture families. I use it to detect component-level convergence that might be hidden if I only looked at complete gesture solutions.

Component-level dominance and agreement (layer-wise)

Whole-gesture agreement can be low even when participants share stable components. To measure this, I compute component dominance separately for L1, L2, and L3 using a Wobbrock-inspired concentration measure over component occurrences.

For a given referent and layer, let the layer contain component units u in U. Let c_u be the count of how many times unit u occurs in that layer for that referent. For L2 and L1, composite gestures can contribute multiple occurrences.

Let $C_{\text{total}} = \sum_{u \in U} c_u$.

Component dominance score is:

$$C = \sum_{u \in U} \left(\frac{c_u}{C_{\text{total}}} \right)^2$$

This behaves like agreement, but at the component level. It increases when one or two components dominate the distribution, and it decreases when usage is spread across many components.

I report this score per layer together with the dominant unit percentage. Example: "L1 dominant unit is left at 50%." This makes the component convergence readable and interpretable.

This is not the same as family agreement. Layer-wise dominance ignores families and asks a different question: how concentrated is the component usage across the full gesture space? Family analysis asks: how many complete solution clusters exist, and how large are they? I use both because they answer different parts of SQ2.

Creating gesture families (three-layer clustering protocol)

Next, I cluster gesture instances into gesture families. A family represents a full solution strategy for controlling the referent. I do not assume gestures in a family are identical. Instead, I group gestures that share the same core structure under a weighted similarity logic.

Weights are fixed and global across referents:

$$w_{L1} = 1/2$$

$$w_{L2} = 1/3$$

$$w_{L3} = 1/6$$

This encodes the assumption that spatial intent contributes the most to "sameness," followed by primitive mechanics, followed by handshape.
I weighted the layers as L1 > L2 > L3 because prior elicitation work suggests users often converge first on high-level motion/spatial parameters, while finer articulation details vary without necessarily changing the intended meaning (Ruiz et al. 2011, Wobbrock et al. 2009). This weighting also avoids over-fragmenting the dataset, since adding more (and stricter) criteria can sharply change measured agreement and make clusters look artificially "disagreeing" (Vatavu 2019).
I perform family construction manually, but I do not do it freely. I use the three-layer protocol as a guide, and I explicitly record notes describing the similarity rule that defines each family. These notes force consistency. They also make my grouping rationale visible when the same kind of pattern appears in a different referent.
I also treat participant intention as a legitimate tie-breaker when strict layer matching would be unfair. A simple example is the "thumbs up" strategy for a sidestep. One participant may extend the thumb into the thumbs-up (L1: up, L2: extend, L3: thumbs up) while another already starts in thumbs-up and instead tilts the wrist to aim the thumb (L1: self, L2: tilt, L3: thumbs up). The surface execution differs, but the control concept is the same. In such cases, intention prevents over-splitting families.
The output of this step is a set of mutually exclusive families per referent. For each family I store:

- Group ID
- Member participants
- Notes (the similarity rule)
- Frequency (family size)

Whole-gesture agreement using Wobbrock (family-based)

To quantify whole-gesture convergence, I compute the agreement score introduced by Wobbrock et al. over gesture families.
Let N be the number of participants for the referent. Let F be the set of gesture families for that referent. Let n_f be the number of participants whose gesture belongs to family f.
The agreement score A is:

$$A = \sum_{f \in F} \left(\frac{n_f}{N} \right)^2$$

What this means: A is the probability that two randomly selected participants independently produced gestures from the same family. It is not the same as "percent in the largest family." The score penalizes fragmentation because every additional family introduces disagreements across families. This is exactly what I want when I describe whole-gesture convergence.
In this study, I computed A over families rather than over layer values. This is a deliberate choice. Two gestures can share components but still reflect different overall control concepts. Family-based agreement captures convergence on complete solutions.

Pairwise similarity within popular families

Family construction is necessary for whole-gesture convergence, but it creates a risk: I could accidentally collapse gestures that are only loosely related. To make internal variation visible, I compute a pairwise similarity score inside the most prevalent family (and inside the top two families for tied cases). This gives me dispersion metrics that describe how coherent the "popular" family actually is. This also aligns with Tsandilas (2018) and Vatavu & Wobbrock (2015), who recommend complementing agreement score with additional evidence and being explicit about similarity assumptions.

Similarity function

For any two gestures g_a and g_b, I compute:

$$S(g_a, g_b) = w_{L1} \cdot L1(g_a, g_b) + w_{L2} \cdot L2(g_a, g_b) + w_{L3} \cdot L3(g_a, g_b)$$

where:
$L1(g_a, g_b)$ is spatial similarity
$L2(g_a, g_b)$ is primitive similarity
$L3(g_a, g_b)$ is handshape similarity
and the fixed weights are:

$$w_{L1} = 1/2 = 0.50$$

$$w_{L2} = 1/3 = 0.33$$

$$w_{L3} = 1/6 = 0.17$$

Computing L1, L2, L3 layer similarities

For each layer independently, I treat the coded values as unordered sets. Comma-separated values become a set.
Example:
Gesture A primitives: {rotate_cw_180, compress, extend}
Gesture B primitives: {rotate_cw_180, extend}
I then compute the symmetric difference between sets to count how many elements differ.

$$\text{symdiff} = A \oplus B$$

I map the number of differences to a similarity value using a fixed rule:
Number of differences $\rightarrow$ Layer similarity
$0 \rightarrow 1.00$
$1 \rightarrow 0.75$
$2 \rightarrow 0.50$
more than $2 \rightarrow 0.25$
I apply this mapping separately to L1, L2, and L3. This is important because it turns matching into graded similarity instead of a binary match.

Example calculation

If for a given pair:
L1 = 0.75
L2 = 1.00
L3 = 0.50
Then:

$$S = (0.50 \times 0.75) + (0.33 \times 1.00) + (0.17 \times 0.50)$$

$$= 0.375 + 0.33 + 0.085$$

$$= 0.79$$

Dispersion metrics inside popular families

Once I compute pairwise similarity scores for all gesture pairs in the popular family, I summarize internal coherence using mean similarity, variance, and standard deviation. These metrics let me say whether a "dominant family" is a tight default or a loose template.

Let a family contain n gestures. There are $k = n(n-1)/2$ unique pairs. Let $S_1 \dots S_k$ be the pairwise similarity scores.
Mean similarity:

$$\bar{S} = \frac{1}{k} \cdot \sum_{i=1}^{k} S_i$$

Variance:

$$Var = \frac{1}{k} \cdot \sum_{i=1}^{k} (S_i - \bar{S})^2$$

Standard deviation:

$$SD = \sqrt{(Variance)}$$

Interpretation:
High mean with low SD suggests a tight family with consistent execution. Lower mean or high SD suggests a looser family where participants share a strategy but vary in how they perform it.

Mental model coding for dominant solutions

To connect convergence patterns to participant reasoning, I summarize mental model labels inside the dominant family (or inside each top family for tied cases). I take the mode of the mental model labels within the family. If the family contains many different mental models with no clear mode, I code it as "all different."
This step helps me separate two cases that can look similar at the surface:

1. participants converge because they share a conceptual framing of the action

2. participants converge in form but for different reasons

I use mental model summaries as descriptive support, not as the primary convergence metric.

Dominance structure classification and selection rules

To translate family distributions into an interpretable "core versus divergent" outcome, I classify each referent into a dominance structure based on the top two family proportions.
For each referent, I compute:

- Largest family %

- Second-largest family %
- Dominance Gap = Largest% - Second largest%

Classification rules: Single-dominant: Largest $\geq 50\%$ and Gap $\geq 20\%$
Co-dominant: Largest and second are tied and both $\geq 40\%$
Bimodal: Largest and second are tied and both $< 40\%$
No dominant: otherwise

Why I use dominance structures:

I introduce dominance structures to avoid a common mistake in elicitation analysis: treating the "largest family" as consensus when it is only a plurality. In practice, several referents can have a largest-family value around 30–40%. In those cases, the most frequent family is better interpreted as the largest minority, not a stable default. This critique is also present in works from Viglialoro et al. (2020) and Tsandilas (2018). Dominance classification is therefore treated as the main evidence for whether a referent has a population-level default. Only Single-dominant and Co-dominant referents are interpreted as having defaults.
This also provides the motivation for using the Dominance Gap alongside the largest-family percentage. A simple 50% threshold can fail in two ways: a referent can be practically dominant even slightly below 50% if it is clearly separated from the second family, and a referent can reach 50% but still be unstable if the second family is close. Reporting the gap makes both magnitude and separation explicit.
Finally, dominance structure controls how I report dominant-family analyses. For Single-dominant referents, within-family similarity and mental model summaries describe the default. For No dominant referents, the same metrics are still informative, but I report them as properties of the modal family rather than "the consensus," to avoid over-claiming agreement. Ties are also handled carefully: two tied top families indicate polarized solutions, not dominance. Depending on their size, I treat them as Co-dominant (two stable defaults) or Bimodal (two popular minorities), and I report each tied family separately rather than merging them.

User-designed gesture set using dominance structure

Dominance structure also acts as an explicit selection policy for carrying gestures forward into the user-designed gesture set.

Single-dominant: I carry forward one gesture family as the default mapping for the referent. I describe it using its defining components and its within-family dispersion metrics.
Co-dominant: I carry forward two gesture families as equally valid defaults. I do not merge them. I report within-family metrics and mental model summaries separately for each family.
Bimodal: I do not claim a default. I carry forward the top two families as candidate mappings. This keeps the main patterns visible without implying population-level consensus.
No dominant: I do not select a fixed default. I treat the referent as requiring flexibility, either by allowing multiple mappings or by defining the mapping at the component level (for example, enforcing spatial intent while allowing variation in handshape).

How these metrics answer SQ2

I answer SQ2 by combining the outputs of the pipeline in a controlled way. For whole-gesture convergence, I rely on:

- Wobbrock agreement over families
- dominance structure labels (single-dominant, co-dominant, bimodal, no dominant)

For component-level convergence, I rely on:

- component dominance and dominant unit % per layer (L1, L2, L3)

To judge default precision (tight default versus template default), I rely on:

- within-family dispersion (mean similarity and SD) in the dominant family

To interpret why patterns appear, I use:

- mental model distributions inside dominant families as descriptive context

This mapping keeps the logic clean. I do not rely on a single number to claim convergence. I also avoid treating the "largest family" as consensus unless the dominance thresholds support it.

5.2 Results

This section reports the results for SQ2 by examining convergence at two levels: whole-gesture convergence, assessed through gesture families and agreement scores, and component-level convergence, assessed through layer-wise dominance. Results are presented descriptively; interpretation is reserved for the Discussion chapter.

Core default referents (single-dominant consensus)

These referents exhibit a single dominant gesture family with a clear dominance gap, indicating a population-level default gesture solution.

Forward

For Forward, whole-gesture agreement was moderate (A = 0.32), with five gesture families and a dominant family capturing 50% of participants (single-dominant; gap 30%). Component-level analysis shows convergence mainly at L1 (L1 = 0.36), with the dominant spatial unit "left" (50%) and dominant primitive "swipe" (40%). Handshape remains mixed (fist/flat/point = 30% each). Mental-model coding suggests the dominant family is metaphoric and uses mold, indicating that participants share an abstract mapping for locomotion rather than a strictly embodied action. Overall, Forward is classified as a core default, representing a usable default gesture with tolerance for handshape variation.

Left Punch

For Left Punch, whole-gesture agreement was high (A = 0.44), with three gesture families and a dominant family capturing 60% of participants (single-dominant; gap 40%). The dominant family is internally tight (mean similarity = 0.8867, SD = 0.0801). Component-level analysis shows convergence at L1 toward the screen (60%), with "extend" as the dominant primitive (50%) and "point" as the dominant handshape (50%). Mental-model coding indicates an iconic strategy using enact. Overall, Left Punch is classified as a strong core default with a coherent execution pattern.

Uppercut

For Uppercut, whole-gesture agreement was high (A = 0.42), with four gesture families and a dominant family capturing 60% of participants (single-dominant; gap 40%). Despite this, internal coherence within the dominant

family is low (mean similarity = 0.3607, SD = 0.2706), indicating substantial variation in execution. Component-level analysis shows strong convergence at L1, with an upward spatial unit dominant at 60%, while primitives remain distributed across compress, lift, rotate, and tilt combinations. Mental-model coding indicates an iconic, enact-based strategy. Overall, Uppercut is classified as a core default characterized by a shared action concept but diverse executions.

Hammer

For Hammer, whole-gesture agreement was moderate (A = 0.34), with four gesture families and a dominant family capturing 50% of participants (single-dominant; gap 30%). Internal dispersion is moderate (SD = 0.2484), indicating some variability in execution. Component-level analysis shows strong convergence at L1, with downward spatial intent dominant at 80%, followed by "compress" as the dominant primitive (40%) and "fist" as the dominant handshape (50%). Mental-model coding suggests an iconic mapping using enact. Overall, Hammer is classified as a core default with a clear embodied metaphor and moderate tolerance for execution differences.

Phoenix

For Phoenix, whole-gesture agreement was high (A = 0.42), with four gesture families and a dominant family capturing 60% of participants (single-dominant; gap 40%). Internal coherence within the dominant family is relatively high (mean similarity = 0.769, SD = 0.1267). Component-level analysis shows strong convergence at L1 toward the screen (80%), with dominant primitives "compress + rotate" (50%) and a dominant "fist" handshape (70%). Mental-model coding indicates an iconic, enact-based strategy. Overall, Phoenix is classified as a strong core default with consistent execution.

Polarized core referents (two stable defaults)

These referents show population-level convergence, but toward two equally strong gesture families rather than a single default.

Neutron Bomb

For Neutron Bomb, whole-gesture agreement was moderate (A = 0.34), with four gesture families and two co-dominant families capturing 40% of participants each (gap 0%). Both dominant families are internally coherent (Fam A: mean similarity = 0.8112, SD = 0.0724; Fam B: mean similarity =

0.8171, SD = 0.0897). Component-level convergence is moderate across layers, while mental-model coding differs between families: one family combines metaphoric and iconic enactment, while the other emphasizes iconic embodiment. Overall, Neutron Bomb is classified as a polarized core referent with two stable competing defaults.

Intent-convergent but form-divergent referents

These referents show low whole-gesture agreement, but clear convergence at the level of spatial intent, indicating shared action goals with diverse executions.

Backward

For Backward, whole-gesture agreement was low (A = 0.28), with four gesture families and no dominant family (largest 40%; gap 20%). Component-level analysis shows convergence primarily at L1, with the spatial unit "right" dominant at 60%. Primitives and handshapes remain more distributed. Mental-model coding of the modal family is iconic with mold. Overall, Backward is classified as intent-convergent but form-divergent.

Side Step In

For Side Step In, whole-gesture agreement was low (A = 0.20), with six gesture families and no dominant family (largest 30%; gap 10%). Component-level analysis shows strong L1 convergence toward the screen (60%), while L2 primitives and mental models are highly fragmented. Overall, Side Step In is classified as intent-convergent but form-divergent.

Side Step Out

For Side Step Out, whole-gesture agreement was low (A = 0.26), with five gesture families and no dominant family (largest 40%; gap 20%). Component-level analysis shows convergence at L1 toward the self (60%) and weak dominance of the "pull" primitive (30%). Mental models vary across families. Overall, Side Step Out is classified as intent-convergent but form-divergent.

Phoenix Smasher

For Phoenix Smasher, whole-gesture agreement was low (A = 0.28), with five gesture families and no dominant family (largest 40%; gap 10%). Component-level analysis shows strong convergence at L1 toward the screen (80%) and

moderate dominance of the "compress" primitive (50%). Mental-model coding in the modal family remains iconic and enact-based. Overall, Phoenix Smasher is classified as intent-convergent but form-divergent.

Component-convergent referents with weak spatial dominance

These referents do not converge strongly on spatial intent, but show clearer convergence in execution-related components.

Left Kick

For Left Kick, whole-gesture agreement was low (A = 0.20), with six gesture families and no dominant family (largest 30%; gap 10%). Spatial convergence is moderate, with the dominant L1 unit "left" at 50%. Stronger convergence appears at L2, with "extend" dominant at 50%, and at L3, with "flat hand" dominant at 50%. Mental-model coding indicates an iconic, enact-based strategy. Overall, Left Kick is classified as component-convergent but form-divergent.

Open-ended referents with weak structure

These referents exhibit weak structure across both gesture families and components, indicating limited convergence.

Hangover

For Hangover, whole-gesture agreement was low (A = 0.26), with four gesture families and weak bimodality (30% / 30%; gap 0%). Component-level convergence is weak across all layers (L1 = 0.215, L2 = 0.134, L3 = 0.266), and mental-model coding varies substantially, including families labeled as "all different." Overall, Hangover is classified as an open-ended, weakly structured referent.

Summary of SQ2 results

Across referents, the results reveal a small set of core defaults with whole-gesture consensus, one referent with two stable competing defaults, and several referents that lack whole-gesture convergence but still show strong

component-level agreement, especially in spatial intent. These patterns demonstrate that convergence in gesture-based control is multi-level and cannot be captured by whole-gesture agreement alone.

Table 5.1: Wobbrock Agreement Score per referent

Referent	**W Agreement Score**
Forward	0.32
Backward	0.28
Sidestep In	0.2
Sidestep Out	0.26
Left Punch	0.44
Left Kick	0.2
Uppercut	0.42
Neutron Bomb (Fam A)	0.34
Neutron Bomb (Fam B)	0.34
Hammer	0.34
Phoenix	0.42
Phoenix Smasher	0.28
Hangover (Fam A)	0.26
Hangover (Fam B)	0.26

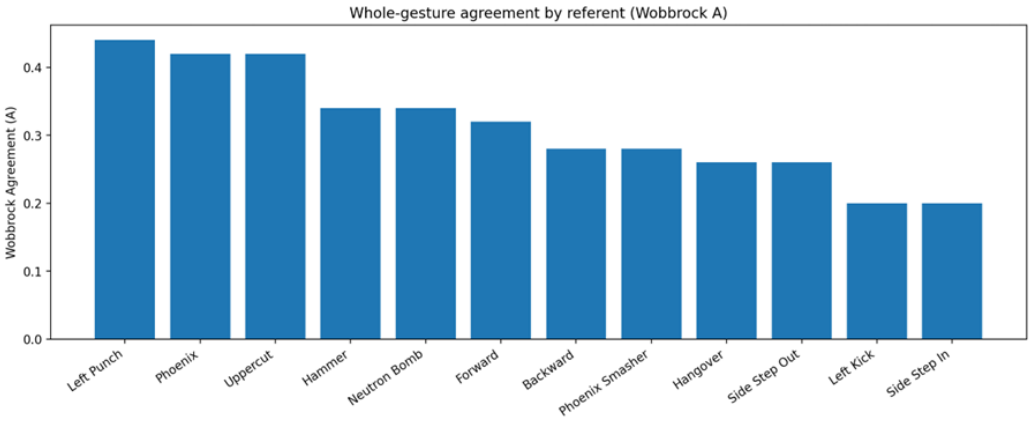


Figure 5.1: Whole gesture agreement by referent

Table 5.2: Dominance structure and consensus classification per referent

Referent	**No. of Families**	**Largest Family %**	**Second largest Family**	**Dominance Gap**	**Dominance Structure**	**Interpretation**
Forward	5	50%	20%	30%	Single-Dominant	Consensus
Backward	4	40%	20%	20%	No Dominant	Not consensus
Side Step In	6	30%	20%	10%	No Dominant	Not consensus
Side Step Out	5	40%	20%	20%	No Dominant	Not consensus
Left Punch	3	60%	20%	40%	Single-Dominant	Consensus
Left Kick	6	30%	20%	10%	No Dominant	Not consensus
Uppercut	4	60%	20%	40%	Single-Dominant	Consensus
Neutron Bomb (Fam A)	4	40%	40%	0%	Co-dominant	Consensus
Neutron Bomb (Fam B)	4	40%	40%	0%	Co-dominant	Consensus
Hammer	4	50%	20%	30%	Single-Dominant	Consensus
Phoenix	4	60%	20%	40%	Single-Dominant	Consensus
Phoenix Smasher	5	40%	30%	10%	No Dominant	Not consensus
Hangover (Fam A)	4	30%	30%	0%	Biomdal	Not consensus
Hangover (Fam B)	4	30%	30%	0%	Bimodal	Not consensus

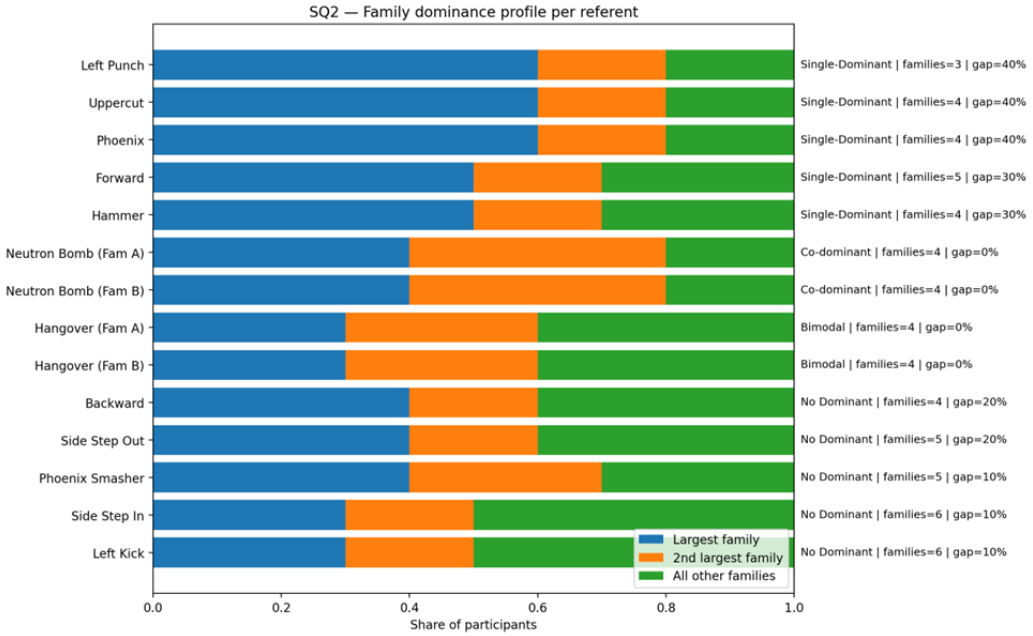


Figure 5.2: Family Dominance Profile per referent

Table 5.3: Within-family similarity statistics for the dominant gesture family

Referent	**Mean**	**Variance**	**StdDev**
Forward	0.715	0.0332	0.1823
Backward	0.8867	0.0064	0.0801
Side Step In	0.6667	0.0182	0.1347
Side Step Out	0.6933	0.0219	0.1479
Left Punch	0.8867	0.0064	0.0801
Left Kick	0.575	0.0975	0.3122
Uppercut	0.3607	0.0732	0.2706
Neutron Bomb (Fam A)	0.8112	0.0052	0.0724
Neutron Bomb (Fam B)	0.8171	0.0081	0.0897
Hammer	0.5597	0.0617	0.2484
Phoenix	0.769	0.0161	0.1267
Phoenix Smasher	0.5512	0.0054	0.0735
Hangover (Fam A)	0.4467	0.0373	0.1931
Hangover (Fam B)	0.6367	0.0073	0.0853

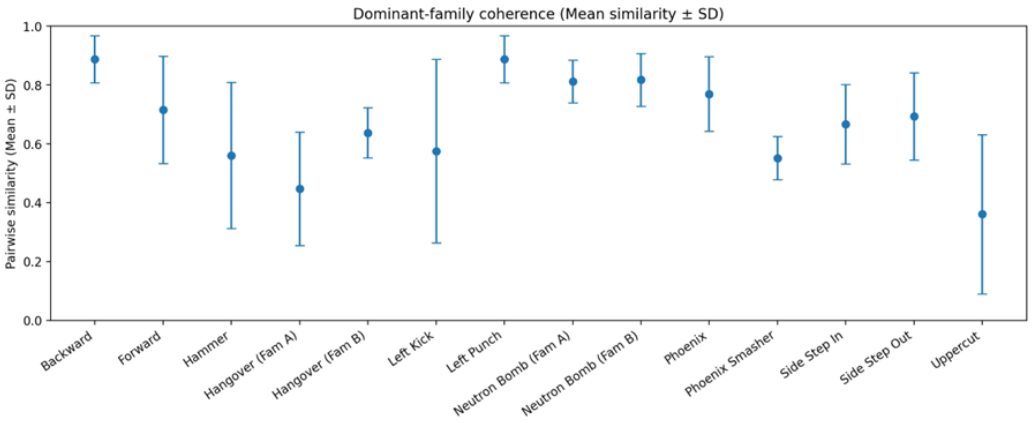


Figure 5.3: Dominant-Family Coherence

Table 5.4: Distribution of mental model types and modes (modal values)

Referent	**Types**	**Modes**
Forward	Metaphoric	Mold
Backward	Iconic	Mold
Side Step In	Metaphoric	All Different
Side Step Out	Metaphoric	All Different
Left Punch	Iconic	Enact
Left Kick	Iconic	Enact
Uppercut	Iconic	Enact
Neutron Bomb (Fam A)	Metaphoric + Iconic	Enact
Neutron Bomb (Fam B)	Iconic	Embody + Enact
Hammer	Iconic	Enact
Phoenix	Iconic	Enact
Phoenix Smasher	Iconic	Enact
Hangover (Fam A)	Iconic	Enact
Hangover (Fam B)	All Different	All Different

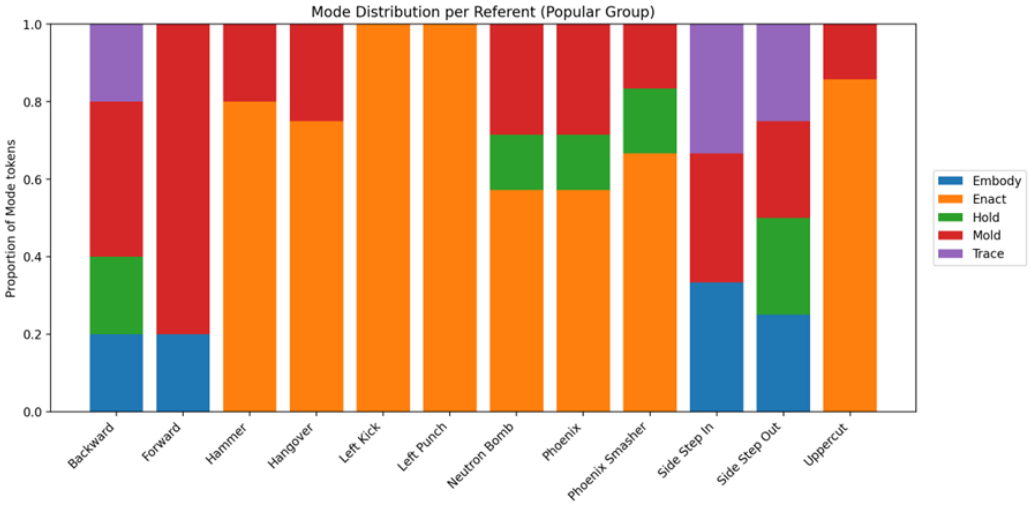


Figure 5.4: Mode Distribution per referent

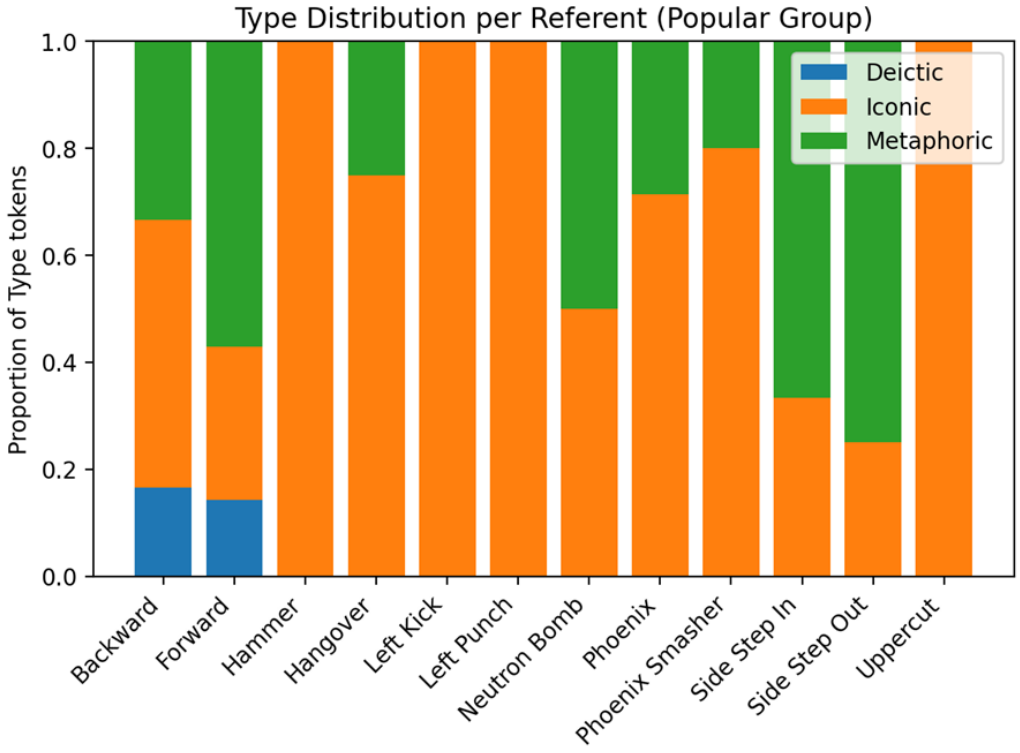


Figure 5.5: Type distribution per referent

Table 5.5: Component-level convergence scores and dominant units for spatial intent (L1), motion primitives (L2), and handshape (L3)

Referent	**L1**	**L2**	**L3**	**L1 Unit (%)**	**L2 Unit (%)**	**L3 Unit (%)**
Forward	0.36	0.28	0.28	left (50%)	swipe (40%)	fist, flat hand, point (30%)
Backward	0.44	0.3	0.3	right (60%)	swipe (40%)	fist (40%)
Side Step In	0.46	0.16	0.28	screen (60%)	extend, push, rotate (20%)	fist, flat hand, point (30%)
Side Step Out	0.46	0.18	0.22	self (60%)	pull (30%)	fist (30%)
Left Punch	0.3554	0.2727	0.38	screen (60%)	extend (50%)	point (50%)
Left Kick	0.36	0.2189	0.42	left (50%)	extend (50%)	flat hand (50%)
Uppercut	0.46	0.1822	0.46	up (60%)	compress, lift, rotate, tilt (30%)	fist (60%)
Neutron Bomb (Fam A)	0.2663	0.2899	0.3	left, up (40%)	rotate (50%)	flat hand (50%)
Neutron Bomb (Fam B)	0.2663	0.2899	0.3	left, up (40%)	rotate (50%)	flat hand (50%)
Hammer	0.3438	0.1468	0.38	down (80%)	compress (40%)	fist (50%)
Phoenix	0.66	0.2422	0.52	screen (80%)	compress, rotate (50%)	fist (70%)
Phoenix Smasher	0.313	0.136	0.3719	screen (80%)	compress (50%)	fist (60%)
Hangover (Fam A)	0.215	0.1338	0.2663	down, screen (50%)	compress, extend (40%)	fist (50%)
Hangover (Fam B)	0.215	0.1338	0.2663	down, screen (50%)	compress, extend (40%)	fist (50%)

5.3 Discussion

In this section, I discuss what the SQ2 results suggest about whether user-defined control gestures form a convergent core set or a broadly divergent set. My participant pool was small (N = 10). For that reason, I describe patterns as tendencies observed in this study rather than general claims about all players.

The main idea: convergence is multi-level

The results suggest that "agreement" does not appear in only one way. In my user study, participants sometimes converged on a single dominant gesture family for a referent. In other cases, whole gestures remained diverse, but participants still seemed to converge on what the gesture should express, especially at the spatial intent level (L1). This distinction matters for SQ2 because a low family-level agreement score does not automatically mean the mapping is unstructured.

The mix of higher and lower agreement across referents is consistent with prior elicitation findings where some commands naturally invite shared gestures, while others remain diverse even within the same study context (Wobbrock et al. 2009, Villarreal-Narvaez et al. 2020). On a component level, I found that there is stronger convergence on motion/spatial parameters than on detailed articulation. This matches earlier observations that consensus often appears in movement parameters even when the exact performed form varies (Ruiz et al. 2011, Vatavu 2019).

I therefore discuss convergence using four linked lenses that were measured in SQ2: (1) whole-gesture convergence (Wobbrock A and dominance structure), (2) coherence inside the most prevalent family (pairwise similarity and dispersion), (3) component convergence (layer dominance and unit dominance), and (4) mental model distributions (McNeill's types and Müller's modes of representation).

Whole-gesture convergence appears for a subset of referents

Across referents in my study, Wobbrock agreement scores ranged from roughly $A \approx 0.20$ to $A \approx 0.44$. As expected, A generally rose when a referent had fewer gesture families and a larger dominant family, and it dropped when gestures fragmented across more families. This alignment is useful because it suggests the family-level metric behaved consistently in this dataset.

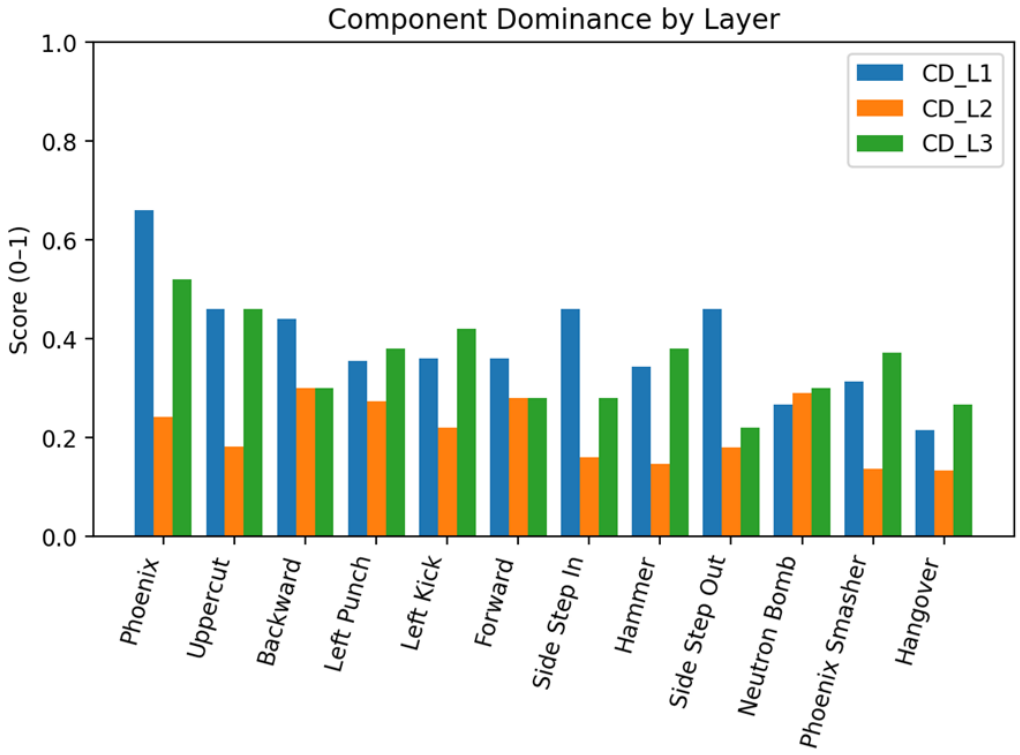


Figure 5.6: Component Dominance per layer

Dominance-structure classification makes the meaning of "agreement" clearer than A alone. In my user study, several referents met the single-dominant criteria (largest family $\geq 50\%$ and a clear gap to the second family). For these referents, participants in this study often produced a population-level default gesture family. In contrast, other referents did not meet dominance criteria even when they had a largest family around 30–40%. In those cases, the largest family is better treated as the largest minority rather than a default.

One referent in my study (Neutron Bomb) showed a different kind of convergence: two co-dominant gesture families of similar size. This pattern does not look like random divergence. Instead, it suggests that two stable solution strategies were attractive to different participants in this study.

A dominant family is not always a single standardized form

Dominance structure answers whether a default exists at the population level (within this study), but it does not tell us what kind of default it is. Pairwise similarity inside the dominant family adds that missing information.

In my study, some referents showed a "tight default" signature: the referent was single-dominant, and gestures inside the dominant family were also highly similar with low dispersion. This indicates that participants not only chose the same family, but also produced similar executions of that family.

Other referents showed a "template default" signature: the referent was still single-dominant, but within-family similarity was much lower and dispersion was higher. In these cases, participants in my study seemed to converge on the idea of the gesture family (often held together by spatial intent), while still varying widely in the mechanics of execution. Practically, this means the "default" is not one precise movement. It is closer to a template with multiple acceptable realizations.

This distinction matters for SQ2 because it prevents an over-simplified conclusion such as "the move converges." Instead, the results allow me to specify that in this study, some moves converged to a tight shared form, while others converged to a shared intent with flexible execution.

Divergence at the family level often hides component-level convergence

The strongest cross-referent pattern in my results is that component convergence often remained visible even when whole-gesture convergence was low.

Across referents in this study, L1 (spatial intent) tended to be the most stable layer, L2 (primitives) tended to be the most variable, and L3 (handshape) sat between them.
This pattern appears most clearly in the no-dominant referents. Although these referents did not produce a single dominant gesture family in my study, several still showed strong dominance of a single L1 unit (for example, clear targets such as toward screen, toward self, up, down, or right). This suggests that participants in this study often agreed on where the action should go in space, even when they disagreed on how to package that intent into a full gesture.
Put differently, whole-gesture divergence in these cases may reflect execution diversity layered on top of intent convergence. Participants shared an intended direction but implemented it using different primitives, different sequencing, or different hand configurations. This is exactly the scenario where a single agreement score would flatten meaningful structure.

Mental models help distinguish execution diversity from conceptual splits

The mental model labels provide context for why convergence or divergence might appear in this dataset. I treat these patterns cautiously due to the small sample (N = 10), but they still help separate two kinds of non-convergence observed in this study.
First, some referents showed low whole-gesture convergence while still maintaining relatively consistent mental model patterns. In those cases, the lack of a dominant family may be driven mainly by motor-level or packaging-level variation (different ways to implement the same idea), rather than by fundamentally different conceptual strategies.
Second, some referents showed fragmented mental model distributions, including cases where modes were effectively "all different." In my study, these referents tended to also show weaker family dominance. This pairing suggests that when participants in this study did not share a conceptual strategy for what the gesture should represent, convergence at the family level became less likely.
The co-dominant case is especially informative. In my study, the two largest families did not just differ in mechanics. They also differed in mental model profiles. This supports the idea that co-dominance can reflect a division in conceptual framing rather than minor motor variation.

Putting the metrics together: what patterns are most important to notice

Looking across the SQ2 columns together, three patterns stand out in this study.

Pattern 1: Whole-gesture convergence is selective. Only a subset of referents in this study produced a single dominant family. This suggests that a "core default" vocabulary may exist, but it may be bounded to moves that are easier to map in a consistent way.

Pattern 2: Spatial intent is a reliable anchor under diversity. Even when families fragment, L1 often remained stable. In my study, this was the most consistent form of convergence across divergent referents. This suggests that participants often share the same target in space even when their gestures look different.

Pattern 3: Similarity separates tight defaults from templates. Two referents can both be single-dominant, but their dominant-family similarity can be very different. This distinction prevents treating "dominant" as equivalent to "standardized," and it helps describe what kind of default is present.

Answering SQ2 carefully (based on N = 10)

Based on the results from my user study, player behavior appears to reveal a partially convergent gesture space rather than a purely divergent one. A small set of referents in this study formed population-level defaults at the whole-gesture family level, and one referent appeared to split into two stable competing families. At the same time, several referents in this study remained diverse at the whole-gesture level.

Importantly, this diversity was not always unstructured. For many non-dominant referents in this study, convergence shifted from whole-gesture form to component-level intent, most consistently at L1 spatial intent. In those cases, participants in my study often seemed to agree on what the gesture should express in space, while varying in primitives, handshape, or execution details.

With the above arguments, I answer the SQ2 as follows:

Within this study, convergence exists, but it is not uniform across moves and it does not always appear as a single standardized gesture per referent. Instead, convergence often appears as either (a) a dominant family with a tight shared form, (b) a dominant family that functions as a template with flexible realizations, (c) two stable competing families, or (d) component-level intent convergence without family-level dominance.

5.4 Research Artifacts and Reusable Outputs from SQ1 & 2

In addition to answering the analytical questions posed in SQ1 and SQ2, the study produced several tangible research artifacts. These artifacts emerged directly from the empirical analysis and extend the contribution beyond reported findings. They include a structured gesture taxonomy, a consolidated user-defined gesture set, and openly shared research infrastructure. Together, these outputs translate the analytical insights of SQ1 and SQ2 into reusable resources that can support replication, comparison, and further development in gesture-based interaction research.

Multi-Level Gesture Taxonomy

The analyses conducted in SQ1 and SQ2 revealed that elicited gestures exhibit structure at multiple interrelated levels. Participants did not only converge at the level of complete gesture families, but also at representational, primitive, morphological, and temporal layers. These recurring dimensions were sufficiently stable across referents to justify consolidation into a structured taxonomy.
Table 5.6 synthesizes the empirically observed dimensions into a multi-level classification framework.
This taxonomy is therefore not imposed a priori, but grounded in the empirical findings of this study. It reflects the representational, structural, and temporal dimensions that consistently shaped gesture production and convergence.

User Defined Gesture Set

Beyond structural analysis, SQ2 resulted in a consolidated user-defined gesture set. This set represents the dominant or convergent gesture families identified across participants for each referent. Rather than prescribing an optimal solution, it captures the most stable mappings that emerged empirically under elicitation. The set can serve as a baseline for future prototype implementations, comparative evaluations, or further refinement in gesture-controlled interactive systems.

Table 5.7: User-defined gesture set derived from SQ1–SQ2, structured by convergence type.

Referent	L1	L2; L3	Notes on flexibility
Core default referents (single-dominant consensus)			
Forward	Left	Swipe; point/fist/flat	Stable spatial intent; handshape varies.
Left Punch	Screen	extend; point	Tight mapping; consistent execution.
Uppercut	Up	lift/tilt; fist	Shared concept; L2 varies within the dominant family.
Hammer	Down	drop/tap/compress; fist	Clear downward intent; L2 variants acceptable.
Phoenix	Screen	compress/rotate/push; fist	Strong convergence across L1–L3.
Polarized core referent (two stable defaults)			
Neutron Bomb (Default 1)	Up	tilt; flat hand	Two competing defaults; treat both as valid mappings.
Neutron Bomb (Default 2)	Left	rotate; v_sign/flat_hand/fist	Handshape flexible within this default.
Intent-convergent but form-divergent referents			
Backward	Right	swipe/rotate/tilt; fist/point	Specify L1 intent; allow multiple L2 and L3 variants.
Side Step In	Screen	rotate/swipe/push/extend; flat/fist/point	Strong L1; fragmented execution components.
Side Step Out	Self	rotate/pull/tilt/arc; point/thumbsup/flat_hand/fist	Strong L1; broad L2/L3 variation expected.
Phoenix Smasher	Down+Screen	rotate/compress/push; fist	Compound L1 intent; consistent handshape.
Component-convergent referent			
Left Kick	Left/Screen	extend/rotate; flat/point	Weak whole-form; stronger convergence in L2–L3.
Open-ended referent (weak structure)			
Hangover	Up/Left/Down	compress/extend/rotate; flat/fist	No stable default; treat as exploratory/provisional.

Public Video Dataset

The study also produced a curated video dataset of isolated gesture instances. Each gesture is exported as a short clip (approximately 3–4 seconds) centered on the performed movement, so that individual executions can be inspected

without requiring access to full-session recordings. The clips contain little to no audio cue, and the only provided label is the name of the referent (i.e., the move the gesture was intended to represent). This format supports transparent inspection of gesture form and variability at the instance level, and it enables reuse for secondary analysis, comparative coding, or training and evaluation in future gesture research. The dataset is publicly archived on Zenodo. The DOI and the link can be found in the following link in my github repository: https://github.com/Wahaj6Ahmad/beyond-buttons-woz-gestures/blob/main/video-dataset/Zenodo-link

Replication Repository

To facilitate methodological transparency and reproducibility, a structured replication repository accompanies this work. The repository contains the scripts used for data cleaning, annotation processing, agreement computation, duration analysis, and table generation. It reflects the full analysis pipeline from raw ELAN exports to reported results. By making this infrastructure available, the study lowers the barrier for future gesture elicitation research and supports extension of the analytical framework beyond the present case study. The associated replication repository is available at my github at the following link: https://github.com/Wahaj6Ahmad/beyond-buttons-woz-gestures

Table 5.6: Multi-Level Taxonomy of Elicited Hand Gestures

Dimension	Classes	Description
Representational Meaning		
Müller's Modes of Representation	Enact, Mold, Hold, Trace	Describes how the gesture represents meaning (e.g., through enacted action, shaping, holding, or tracing).
McNeill's Gesture Types	Iconic, Metaphoric, Deictic, Beat	Linguistic-inspired classification capturing communicative form and relation to semantic content.
Form-Based Structure		
Primitives	Directional motion, rotation, thrust, tap, hold, etc.	Minimal motion components composing the gesture.
Handshape	Fist, open palm, flat hand, index extended, etc.	Static configuration of the hand and fingers during execution.
Palm Orientation	Up, Down, Left, Right	Orientation of the palm relative to the body.
Structural Composition		
Complexity	Simple, Compound	Whether the gesture consists of a single primitive or multiple sequential primitives.
Temporal Properties		
Duration	Continuous measure in seconds	Time from gesture onset to completion.

Chapter 6

Sub-Question 3

6.1 Methodology

Sub-question 3 asks:

> How usable and reliable are user-elicited gestures when performed at high speeds, as required by fast-paced fighting game mechanics?

I answer this question using Experiment 2, where participants attempted their own gesture set under progressively faster (speed ramp) prompting. My focus is on observable and felt outcomes under speed (errors, breakdowns, hesitation, simplification, fatigue, frustration, loss of control, flow interruption), rather than diagnostic constructs such as workload dimensions, as explained in Chapter 3.

Data sources

I used three data sources for each participant ($N = 10$):

1. **Experiment 2 video recordings.** These were used to verify what the participant actually did and to code behavioral indicators of breakdown under speed.

2. **Experiment 2 log data.** This log contains the prompt sequence, timing, and the Wizard-of-Oz (WoZ) command inputs.

3. **Post-experiment questionnaire transcripts.** These capture participants' reflections on errors, speed adaptation, fatigue/comfort, and coping strategies.

Step 1: Aligning logged outcomes using video

The log data reflects what the system registered, not only what the participant intended. In a WoZ setup, each action includes end-to-end latency: the participant perceives the prompt, recalls the gesture, executes it, the wizard interprets it, executes the command, and the logger records it. At high prompt speeds, this pipeline can become the limiting factor. When prompt intervals become shorter than the end-to-end latency, the log may mark a correct participant response as a "misinput" or "miss" simply because the next prompt has already appeared. Because sub-question 3 is about performance under speed, I first ensured that the log labels were aligned with what actually happened in the video.

Correction rules

I applied the following correction logic to ensure the log labels reflected participant intent as closely as possible:

- **On-time correct:** the wizard input matches the expected command for the active prompt window.
- **Late-correct:** the participant produces the intended gesture, but the wizard input lands after the next prompt onset (or is attributed to the next prompt in the logger).
- **Misinput (wrong):** the registered command does not match the intended prompt, and the video evidence supports that the participant executed the wrong gesture or mixed gestures.
- **No-response / incomplete:** no corresponding input is registered within a reasonable window, and the video indicates either no attempt, an incomplete attempt, or that the participant chose to skip.

I treated *late-correct* and repeated gesture behavior carefully. Some participants repeated a gesture (for fun or momentum) until the next prompt appeared. This behavior can create false negatives or apparent misinputs in the log, even when the gesture itself did not break down. In those cases, I relied on the video to determine whether the participant was still performing the prior gesture, had transitioned to the next prompt, or had lost track of the prompt sequence. It is interesting to note that an explicit instruction given to perform the gesture only once could have led to cleaner results, and less manual cleaning afterwards. I leave this afterthought for future researchers to improve upon.

Step 2: Video coding for speed-related breakdowns

After correcting log alignment, I coded the videos for speed-related usability and reliability outcomes. I used a small set of binary and categorical codes that directly map to sub-question 3.

Video codes

For each prompt, I coded:

- **Navigation confusion (0/1):** marked as 1 when the participant mixed navigation gestures (e.g., left vs. right, or using them interchangeably), or when the participant's actions were driven by prompt icon interpretation rather than their originally intended mapping.
- **Hesitation (0/1):** marked as 1 when the participant visibly paused before executing the gesture. This includes hesitation from recalling the gesture, or from interpreting the prompt.
- **Adapted / simplified (0/1):** marked as 1 when the participant simplified the gesture compared to their earlier form (e.g., reduced amplitude, removed a component, switched to a quicker variant).
- **Breakdown (0/1):** marked as 1 when the gesture failed (e.g., wrong finger, wrong gesture, partial execution).
- **Notes (free text):** a short field for contextual observations (e.g., what exactly was confused, what the participant said during the run, visible frustration, or coping strategies).

This coding step produces a behavioral account of speed effects that the log alone cannot provide. It also allows me to separate "gesture breakdown" from "system timing mismatch."

Rationale for the video codes (research basis)

I chose these codes because SQ3 targets the practical consequences of performing gestures under time pressure, not the underlying cognitive causes. This aligns with the ISO view of user experience as the user's *perceptions and responses* during and after interaction, including comfort, behaviours, and accomplishments (Standards 2022). In games user research, a standard way to study this is to combine behavioural observation during play with post-task verbal reports, because observation captures where interaction breaks

down while interviews capture how those moments felt (e.g., loss of control, frustration, disrupted flow) (Drachen et al. 2018). Under increasing time pressure, performance changes often manifest as visible delays and execution failures, consistent with modern speed–accuracy tradeoff findings in conflict and decision tasks (Mittelstädt et al. 2022). I therefore coded *hesitation* (visible pause before acting) and *breakdown* (wrong or partial execution) as direct behavioural markers of speed-related reliability loss. I coded *navigation confusion* because stimulus–response compatibility effects can strongly shape gesture responses, meaning directional cues (such as arrows) can pull responses toward the displayed direction and increase confusion when cues conflict with a learned mapping (Janczyk et al. 2019). Finally, I retained a short *notes* field to document context that the binary codes cannot capture (e.g., coping strategies, visible loss of control), which is commonly recommended when analyzing rich, time-based usability evidence such as video and verbal data (Fan et al. 2020).

Interviews as primary evidence for Adaptation

Although I initially included an *adapted/simplified* code, adaptation was rarely clear enough in the videos to support strong qualitative claims. Many changes were subtle (e.g., reduced amplitude or partial rotations) and could not be distinguished reliably from correct gesture execution. For that reason, I did not treat *adapted* as a primary qualitative signal in the video analysis. Instead, I relied mainly on post-experiment interviews to capture simplification and form drift, since participants could explicitly report what they changed and why, which complements behavioural logs and observation in games/user-experience evaluation (Drachen et al. 2018).

Step 3: Transcript extraction for felt consequences

I used the post-experiment transcript to extract only sub-question 3-relevant content. I focused on answers and remarks related to:

- **Errors and breakdowns:** what participants felt they got wrong, and why (as they described it).
- **Speed adaptation:** whether participants simplified, sped up, switched strategy, or chose to skip prompts.
- **Fatigue and comfort:** reported strain, discomfort, or sustained comfort under repetition.

- **Control and flow:** whether speed caused stress, loss of rhythm, or whether they stayed calm and continued.
- **Memory aids and coping:** strategies like saying prompts out loud, chunking moves, or accepting misses and moving on.

The rationale behind these is provided in Chapter 3.

Referent-to-factor linking

When participants mentioned a specific move (referent) in relation to a factor (fatigue, confusion, simplification, etc.), I recorded it as a linked entry rather than a general comment. For example: *"Phoenix was tiring"* becomes a link between `phoenix` and `fatigue`. This makes it possible to later summarize which referents repeatedly appear with which breakdown signals across the participant pool.

Step 4: Defining speed zones

To interpret performance under speed fairly, I treated prompt pace as the key independent variable. For each trial, I computed the time between prompt onsets (prompt interval), using the log timestamps.
I then grouped trials into three speed zones:

- **Feasible zone:** prompt interval comfortably above typical end-to-end latency, where on-time performance can be interpreted as gesture reliability.
- **Borderline zone:** prompt interval near the end-to-end latency, where outcomes are sensitive to small delays in execution or wizard input.
- **Overload zone:** prompt interval below end-to-end latency, where *late-correct* and *no-response* become expected outcomes even if gestures remain performable.

This zone framing is important for sub-question 3 because it avoids treating a timing ceiling as a gesture ceiling. It also matches the "game lens" reality: fighting games require on-time inputs, but gesture performability still matters for design.

Step 5: Comparing prompt pacing with gesture duration and game move duration

At high speed, failure can come from two different time bottlenecks. The participant may not have enough time to finish the gesture before the next prompt. Or the game may still be executing the previous move animation, which can disrupt feedback and timing.
To test this, I compared prompt intervals against two durations.

Prompt interval versus participant gesture duration

For each prompt, I checked whether the time to the next prompt was shorter than the participant's gesture duration for that referent. This identifies cases where the task schedule makes completion unlikely even with correct intent.

Prompt interval versus in game move duration

For each referent, I compared the prompt interval to the game's move duration. This identifies moves where the ramp demands a new action while the previous move still plays.
This comparison helps explain why some failures cluster during overload pacing even when misinputs do not increase much.

6.2 Results

This section presents the results from Experiment 2 and its questionnaire. The details on the data collected using automated logs in Experiment 2 can be found in the Chapter 3. As a brief reminder, reaction time (RT) is measured from prompt onset (when the prompt appears on screen) to the moment the wizard executes the corresponding command. This value represents the full end-to-end latency of the Wizard-of-Oz setup. It includes the participant perceiving the prompt, recalling and performing the gesture, and the wizard interpreting it and responding.
For this sub-question, usability and reliability is reported in two different ways. **On-time correct**-ness reflects whether the intended command was registered within the same prompt interval in which it was requested. This shows whether the gesture–to–command loop could keep up with the time pressure imposed by the experiment. **Intention-correct** additionally includes *late-correct* cases, where the participant produced the intended gesture but the command was registered after the task had already moved on to

the next prompt. Reporting both measures helps distinguish whether a gesture became difficult to perform, or whether it was performed correctly but registered too late due to the paced prompting and Wizard-of-Oz pipeline.

Quantitative performance by prompt pacing

There were a total of 1106 scored prompts. The overall **on-time correct** accuracy was 41.5%, while the overall **intention-correct** rate was 66.7%.

Speed zones

I compute prompt interval as the time from the current prompt onset to the next prompt onset. I then group prompts into three zones. **Feasible** is more than 4.0 seconds. **Borderline** is 2.0 to 4.0 seconds. **Overload** is 2.0 seconds or less.

Table 6.1: Prompt interval distribution by speed zone (Experiment 2)

Zone	**n_prompts**	**Median interval (s)**	**Min interval (s)**
Overload ($\leq$ 2.0s)	487	1.6	1.0
Borderline (2.0–4.0s)	378	2.6	2.0
Feasible ($>$ 4.0s)	231	5.5	4.1

Table 6.2: Outcome distribution across prompt speed zones (Experiment 2)

Zone	**On-time** (%)	**Late-correct** (%)	**Misinput** (%)	**No-response** (%)	**Other** (%)
Overload ($\leq$ 2.0s)	10.1	39.8	9.2	40.9	0.0
Borderline (2.0–4.0s)	59.3	19.8	9.8	10.6	0.5
Feasible ($>$ 4.0s)	80.5	2.6	10.8	5.6	0.4

Across zones, *misinput* stays in a narrow band. The overload drop is driven mainly by *late-correct* and *no-response*.

Participant level summary

Table 6.3 shows that across participants, strict on-time accuracy has substantial variation. Some participants maintain high on-time performance under speed (e.g., P002), while others show markedly lower values despite completing a similar number of scored prompts.
In contrast, intention accuracy is consistently higher than strict accuracy for all participants. This indicates that many trials marked as failures in

strict scoring correspond to cases where participants produced the intended gesture, but the command was registered too late relative to the task pacing.

Table 6.3: Participant-level accuracy under speed (Experiment 2)

Participant	n_scored	Strict acc. (%)	Intention acc. (%)
P001	106	19.8	59.4
P002	77	80.5	88.3
P003	80	32.5	56.2
P004	90	41.1	87.8
P005	121	53.7	65.3
P006	111	37.8	59.5
P007	130	56.9	80.0
P008	96	47.9	69.8
P009	101	22.8	53.5
P010	194	32.5	58.2

Per referent performance snapshots

Table 6.4 below pools all participants and reports performance per referent group across all prompt intervals. **Intention-correct** accuracy percentage column includes the *on-time* plus *late-correct*. Same analysis on individual referents can found in Table B.2 in Appendix B

Table 6.4: Outcome distribution by referent group pooled across participants (Experiment 2)

Referent group	n	On-time (count)	Late-correct (count)	Mis-input (count)	No-response (count)	On-time accuracy (%)	Intention accuracy (%)
Navigation	364	141	88	40	93	38.7	62.9
Single direct attacks	457	202	115	37	103	44.2	69.4
Multi hit strings	183	70	47	21	44	38.3	63.9
Ambiguous Embodiment	92	46	25	9	12	50.0	77.2

Breakdown signals and coded flags

The log includes coded flags for *confused-navigation*, *hesitation*, and *breakdown* events. Table 6.5 and 6.6 report how often these flags appear by speed zone and by referent groups, respectively. Same data on individual referents can be found in Table B.1 in Appendix B.

Table 6.5: Breakdown signals by speed zone (Experiment 2)

Zone	**n**	**Confused navigation** (%)	**Hesitation** (%)	**Breakdown** (%)
Overload ($\leq$ 2.0s)	487	21 (4.3)	0 (0.0)	2 (0.4)
Borderline (2.0–4.0s)	378	29 (7.7)	15 (4.0)	7 (1.9)
Feasible ($>$ 4.0s)	231	16 (6.9)	14 (6.1)	9 (3.9)

Table 6.6: Breakdown signals by referent-groups (Experiment 2)

Referent group	**n**	**Confused navigation** (%)	**Hesitation** (%)	**Breakdown** (%)
Navigation	366	42 (11.5)	5 (1.4)	4 (1.1)
Single direct attacks	456	9 (2.0)	13 (2.9)	10 (2.2)
Multi hit strings	185	8 (4.3)	6 (3.2)	4 (2.2)
Ambiguous Embodiment	92	7 (7.6)	5 (5.4)	0 (0.0)

Time budget analysis: prompt pacing, gesture duration, and game move duration

This experiment creates a time budget on each prompt. At the fastest tail, the prompt interval approaches 1.0 to 1.3 seconds. Two additional constraints matter at this tail. First, some gestures take longer than the remaining time before the next prompt. Second, some in game moves take longer than the prompt interval, so the avatar animation can still be playing when the next prompt arrives. The table 6.7 compares the move durations with the gesture medians for all referent groups. The same comparison for each individual referent is present in Table B.3 in Appendix B.

Table 6.7: Median gesture execution time compared to Paul Phoenix move duration by referent group (Experiment 2)

Referent group	**Paul move duration (s)**	**Gesture median (s)**	**Gesture – Paul (s)**
Navigation	0.492	0.972	+0.481
Single direct attacks	0.864	1.033	+0.169
Multi hit strings	1.737	1.376	−0.361
Ambiguous Embodiment	1.393	1.187	−0.206

Positive values in the last column mean the gesture takes longer than the move animation. Negative values mean the move animation takes longer than the gesture. A side by side comparison can be observed in the Figure 6.1.

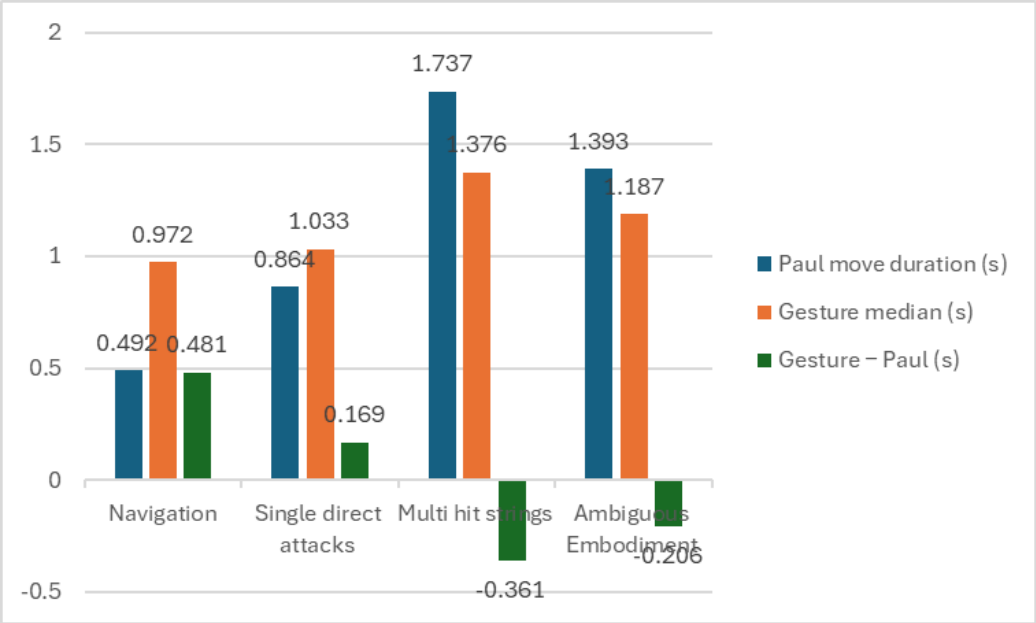


Figure 6.1: Gesture and Moves Duration Compared for all referent groups

Overload ceiling where the next prompt arrives before the gesture finishes

In the overload zone there are 487 prompts. In 115 prompts, the interval to the next prompt is shorter than the participant gesture duration. This is 23.6% of overload prompts. When prompt interval is shorter than the gesture duration, the outcome distribution shifts toward no response. The table below reports rates inside overload only.

Table 6.8: Overload outcomes under gesture time constraint (Experiment 2)

Condition	**n**	**No-response** (%)	**On-time** (%)	**Late-correct** (%)	**Mis-input** (%)
Interval $<$ gesture duration	115	56.5	3.5	34.8	5.2
Interval $\geq$ gesture duration	372	36.0	12.1	41.4	10.5

Overload ceiling where the game animation exceeds the prompt pacing

In overload, some long moves still animate for more than one second. For these moves, the prompt stream often requests a new action while the previ-

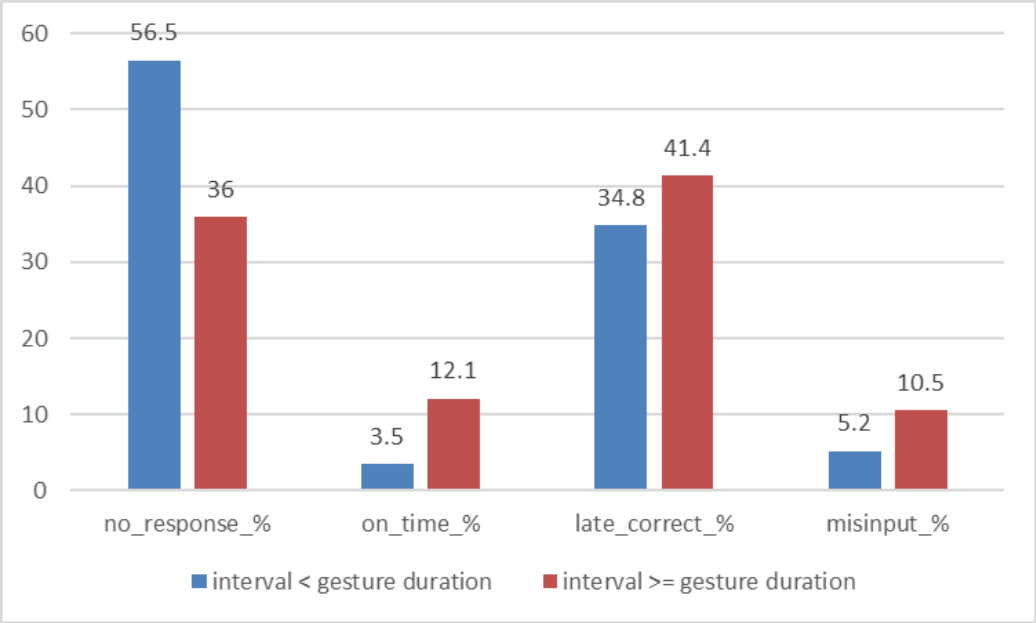


Figure 6.2: Overload outcomes under gesture time constrain

ous move would still be executing on screen. Figure 6.3 shows that *Hangover*, *Phoenix*, and *Phoenix Smasher* share most of the overload prompts. All 3 are among the longest moves by duration as can be seen in figure B.2 in Appendix B.

Granular speed analysis across referents and referent categories

This section goes beyond the three-zone summary (feasible, borderline, overload). Here, I treat *prompt interval* as the main definition of "speed." Prompt interval is the time available until the next prompt appears. This allows a more precise answer to the question: *at what speed does the game remain playable*, and *where is the tipping point where speed makes it unplayable*.

What changes over time for each referent category

Figure 6.4 shows **intention-correct** across prompt interval for each category. Figure 6.5 shows **strict on-time correctness** across the same range. Together, these figures show two key points.
First, strict correctness drops earlier than intention correctness. In other

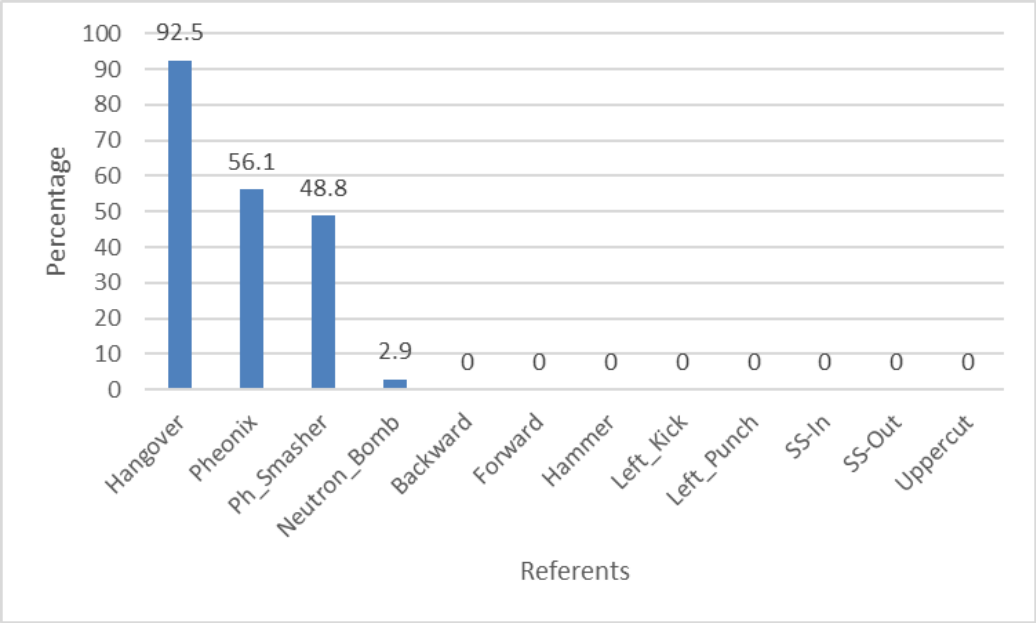


Figure 6.3: Share of overload prompts where the next prompt arrives before Paul Phoenix finishes the move (Experiment 2)

words, the mapping can remain correct at the intention level even while the system stops registering it on time. Second, the categories do not degrade equally. Navigation and multi-hit moves show earlier instability than single-hit moves.

Sweet spot, tipping point and unplayable region

To find a practical "sweet spot" for playability in this setup, I looked for a narrow prompt-interval range where strict performance stays acceptable while intention-correct stays high and omissions stay low. Table 6.9 reports performance in small prompt-interval bands.

Sweet spot: The clearest playable compromise appears around **2.2s to 2.6s** between prompts. If I pick a single tight answer, it is around **2.3s to 2.4s**. In this range, strict correctness stays around the 50% level while intention-correct stays around 80% or above, and no-response remains close to 10%.
Across all participants:

- **2.2–2.4s:** strict 57.7%, intention 84.6%, late-correct 26.9%, no-response 11.5%, misinput 3.8% ($n = 52$)

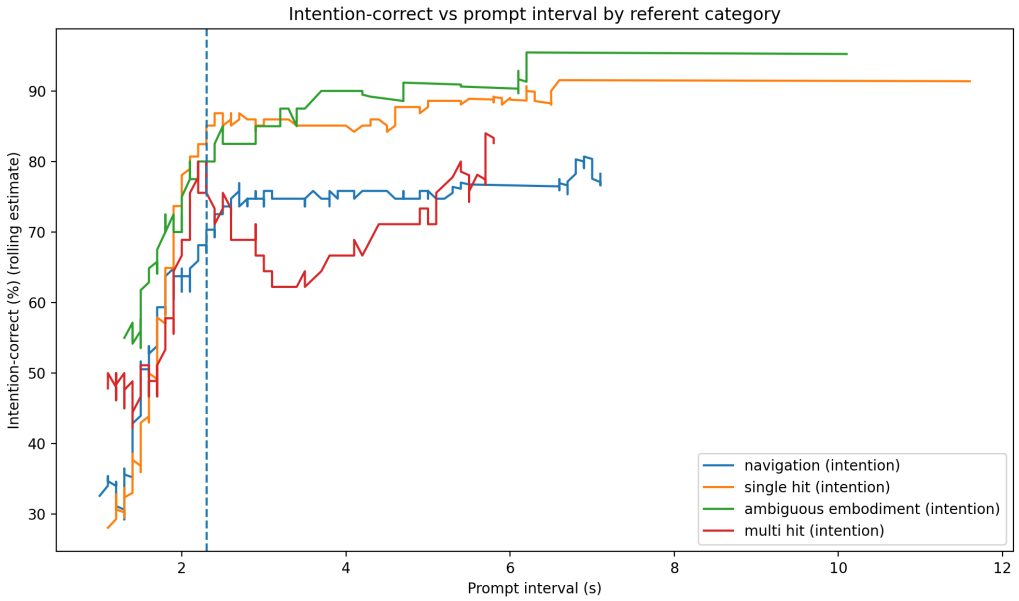


Figure 6.4: Intention correct vs prompt interval by referent category

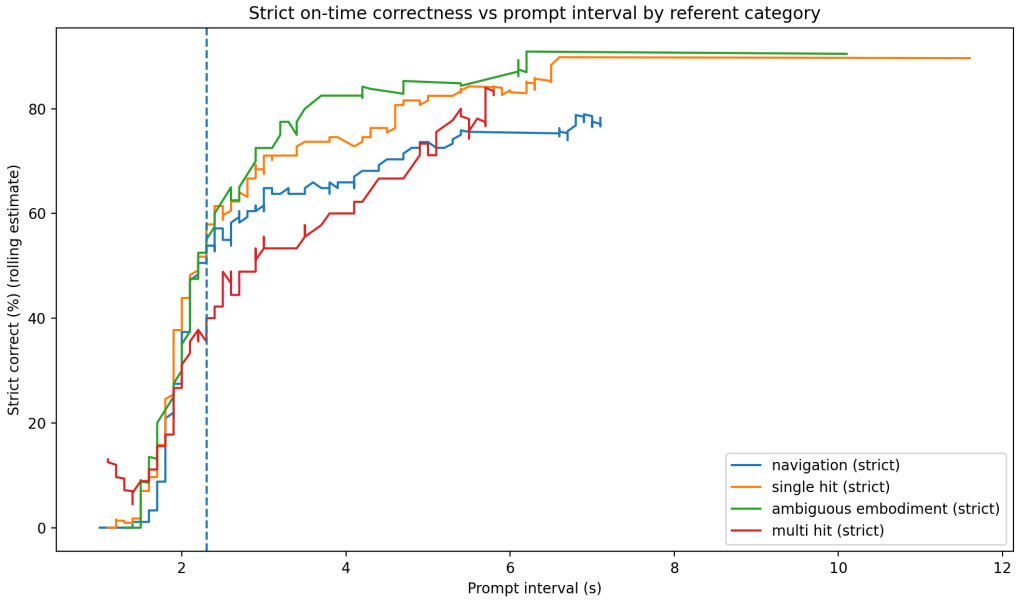


Figure 6.5: Strinct on-time correctness vs prompt interval by referent category

Table 6.9: Outcome rates by prompt interval band across all trials (Experiment 2)

Prompt interval band	n	Strict correct (%)	Intention correct (%)	Late-correct (%)	No response (%)	Misinput (%)
2.2–2.4s	52	57.7	84.6	26.9	11.5	3.8
2.4–2.6s	85	57.6	80.0	22.4	10.6	9.4
2.0–2.2s	106	42.5	73.6	31.1	20.8	5.7
1.8–2.0s	99	19.2	62.6	43.4	22.2	15.2
1.0–1.8s	340	3.2	42.4	39.1	49.1	7.9

- **2.4–2.6s:** strict 57.6%, intention 80.0%, late-correct 22.4%, no-response 10.6%, misinput 9.4% ($n = 85$)

Tipping point: The sharp breakdown begins around **2.0s**. Below this, strict correctness drops quickly and late-correct plus no-response become the dominant outcomes. This is easiest to see in Table 6.9:

- **2.0–2.2s:** strict 42.5%, intention 73.6%, no-response 20.8% ($n = 106$)
- **1.8–2.0s:** strict 19.2%, intention 62.6%, late-correct 43.4%, no-response 22.2% ($n = 99$)

At this point, the prompt stream begins to outrun what the participant and wizard can complete within a single prompt window.

Unplayable region: Below **1.8s**, strict correctness collapses and the task becomes dominated by omissions and timing overflow. In the **1.0–1.8s** band, strict correctness is 3.2% and no-response is 49.1% ($n = 340$). Even intention-correct drops below 50% (42.4%). This is not a small performance degradation. It is a collapse.
Figure 6.6 visualizes this as continuous curves across prompt interval, and Figure 6.7 shows the same collapse across time within the ramp.

Change across the ramp timeline

Figure 6.7 summarizes performance over time using progress deciles within each run. This view confirms the same tipping point using a timeline rather than seconds.
Around the middle of the run, the ramp crosses into the region where strict correctness begins to fall rapidly. For example:

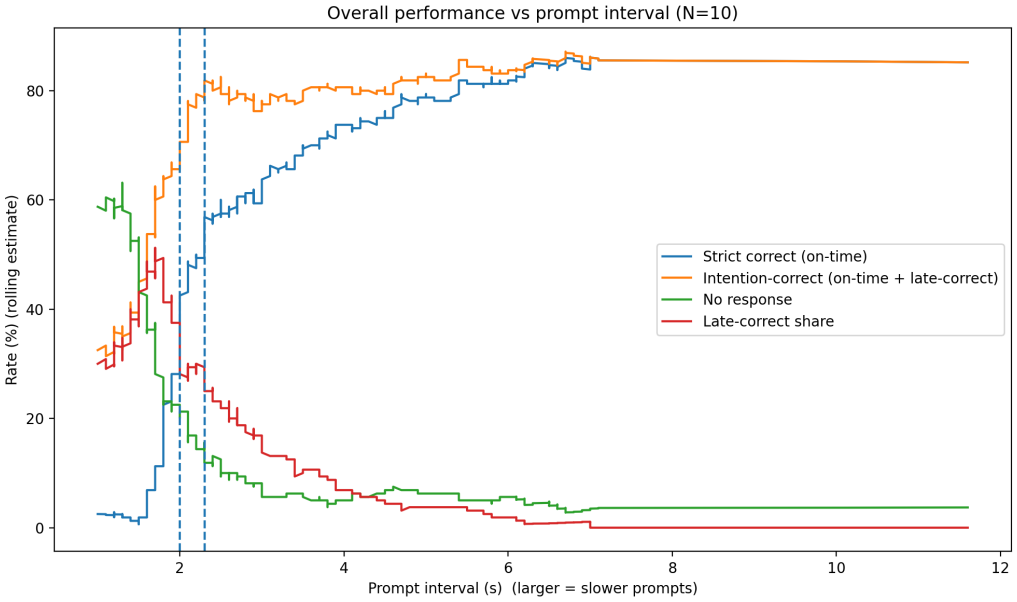


Figure 6.6: Overall performance vs prompt interval (N=10)

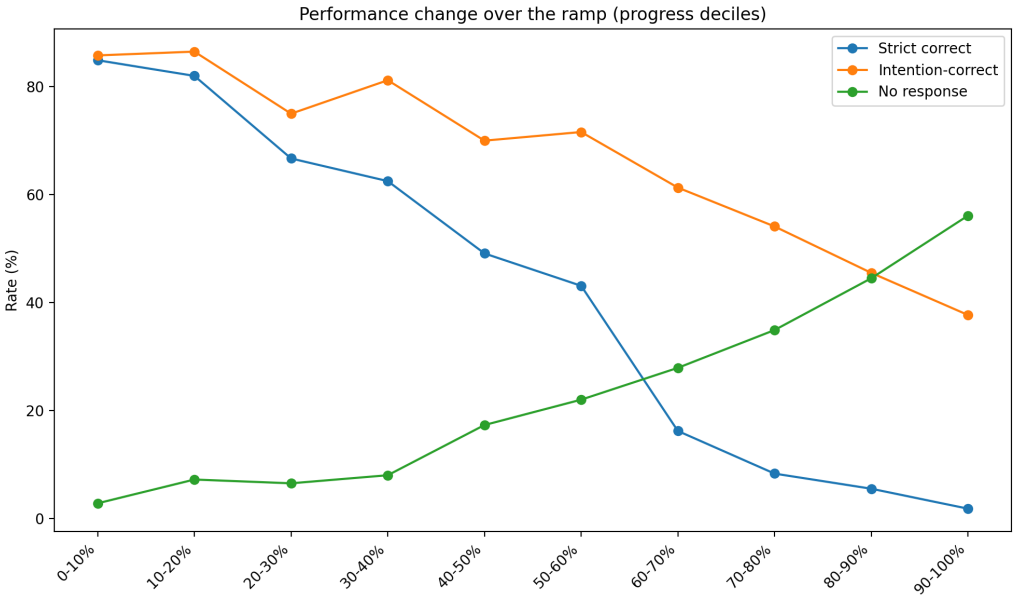


Figure 6.7: Performance change over the ramp (progress deciles)

- At about 40–50% progress, the median prompt interval is 2.4s and strict correctness drops to 49.1%.
- After about 60–70% progress, the median prompt interval reaches 1.9s and strict correctness falls to 16.2%.

So the breakdown does not only happen at the very end. It starts once prompt intervals approach the 2-second range.
In the Discussion section, I use these points to argue about what "reliable under speed" means in a fighting game context, and how to design gesture sets that remain playable as pacing tightens.

Qualitative notes from post experiment interviews

This subsection summarizes what participants reported after Experiment 2. I report outcomes and felt consequences. I do not claim root causes beyond what participants stated.

Navigation and prompt interpretation

Several participants described navigation as the most sensitive part of the task. Under time pressure, they sometimes reoriented their gestures to follow the arrow direction shown on screen, or mixed up forward, backward, and sidestep movements. This pattern is also visible in the log data, where navigation referents show the highest rate of *confused-navigation* flags.

Confusable gestures at speed

Some participants reported confusion between attacks that were assigned visually similar gestures, such as *Left Punch* and *Left Kick*, or *Phoenix* and *Phoenix Smasher*. In the log, these cases appear as misinputs. However, misinput rates do not increase sharply in the overload phase, suggesting that most high-speed failures are due to timing constraints rather than a sudden increase in gesture confusion.

Small practical adaptations

A few participants mentioned making their gestures smaller or dropping minor components, such as a wrist rotation, as the speed increased. These changes were generally subtle. The *adapted* flag appears only rarely in the log, indicating that most gestures remained structurally similar even at higher speeds.

Control and flow at the extreme tail

At the fastest part of the ramp, several participants reported a shift in focus. Instead of attending to the avatar, they began tracking the prompt icons directly. They described moments where the prompt stream moved faster than both the avatar animation and the WoZ loop, leading to brief interruptions in their sense of control and flow.

Fatigue and comfort

Overall, participants did not report strong fatigue during this short speed ramp. When discomfort was mentioned, it tended to be localized, such as wrist strain from resting posture or repeated rotational movements, rather than general exhaustion.

Referent to Factor index

This descriptive table summarizes which referents were explicitly linked to factors extracted from the interviews; however, it does not cover every single mention of the referents. This table can be found in the Appendix A.

6.3 Discussion

I answer sub-question 3 using Experiment 2's speed ramp, where the same personal gesture set gets pushed from comfortable pacing into time pressure. In the Results, I separated two "game lens" realities: (1) strict real time reliability, and (2) whether the participant still managed to perform the intended gesture even if the timing slipped (*late correct*). I keep using that split here because it explains most of what happens as speed increases.
The granular analysis adds something important. The three-zone summary (feasible, borderline, overload) tells me *that* performance breaks. The prompt interval curves tell me *when* it breaks, and what "playable" means in seconds, in this specific setup.

What the speed ramp really tested

The ramp did not only test the gestures. It tested the whole loop: seeing the prompt, deciding, moving the hand, and the system logging the move. In this study, reaction time includes the Wizard of Oz step, so it captures an end to end pipeline, not just hand speed. Tight deadlines amplify the cost of any delay, even if the action itself stays "correct." Claypool & Claypool

(2006) describe this using a deadline framing in games, where small delays push otherwise correct actions past the point where the game can still use them.
That is basically what the fast tail of the ramp became in this experiment.

Prompt interval is the real definition of speed in this setup

The granular results treat *prompt interval* as the definition of speed. This helps because it matches how the task feels. Each prompt interval is a time budget. The player must perceive the prompt, recall the mapping, execute the gesture, and the wizard must respond before the next prompt appears.
Figures 6.6 and 6.7 show that performance does not fade slowly and evenly. It bends, then it collapses. This pattern matches the idea of a deadline limit. When the task tightens the window, performance can look stable until it crosses a threshold, then it breaks fast (Claypool & Claypool 2006, Heitz 2014).
Table 6.9 lets me describe three practical regions that sit underneath the zone summary:

- A **sweet spot** where strict on time outcomes remain acceptable, intention stays high, and omissions stay relatively low.
- A **tipping point region** around the two second range where strict on time correctness drops quickly, and late correct plus no response become the dominant outcomes.
- An **unplayable region** where strict on time correctness collapses and omissions dominate, and even intention correctness drops sharply.

This matters for the meaning of "reliable under speed." It tells me that reliability does not only depend on whether a gesture makes sense or stays distinct. Reliability depends on whether the whole loop can finish inside the current prompt interval.

Reliability under speed looked like a deadline problem, not a confusion problem

As pacing increased, strict **on time** correctness dropped sharply, but the drop did not come mainly from people doing the wrong gesture. Table 6.2 shows that the overload drop comes from people being **late** or giving **no**

response. The granular table (Table 6.9) shows the same story in finer steps. As the interval approaches the two second range, the system starts missing the window more often, even before it reaches the fastest tail.
This pattern suggests that gestures did not suddenly become confusing for participants. Instead, the time budget stopped matching what a person can complete reliably, especially when the next prompt arrives while the previous action still sits in progress.
Motor control and decision research shows that when you force faster responding, people trade accuracy for speed, or they miss the deadline entirely. Heitz (2014) describes this as the speed accuracy tradeoff. In my data, this tradeoff appears more as **deadline misses** (late correct and no response) than as **misinputs** (wrong commands). This distinction matters. It means strict reliability fails mainly because the clock wins, not because the mapping collapses.

What *late correct* means once we look at sweet spot and tipping point

In this experiment, **late correct** means that the participant performed the intended gesture, but the system logged the mapped command after the next prompt already appeared. So late correct is a timing failure.
The granular results make late correct even clearer. Late correct grows strongly once the task enters the tipping point region (Table 6.9). This shows that late correct is not a rare accident. It becomes a predictable outcome when the prompt interval gets close to the loop's limit.
The time budget analysis supports this explanation. In overload, prompt intervals can be shorter than measured gesture durations. When that happens, strict on time correctness becomes physically unrealistic. Figure 6.2 and Table 6.8 show that when the interval is shorter than the gesture duration, no response becomes much more common. This gives a concrete meaning to late correct and no response together:

- **Late correct** often means "I did the right thing, but I finished after the window."
- **No response** often means "the window moved on so fast that I could not even finish something usable."

So in this context, late correct is evidence of gesture performability under speed, but it is also evidence that the control loop cannot meet real time deadlines once the interval enters the two second range.

Not all referent categories break at the same speed

The category curves add an important nuance. Figures 6.4 and 6.5 show two consistent facts.

First, strict correctness drops earlier than intention correctness across categories. This supports the main interpretation that timing fails before meaning fails.

Second, the categories do not degrade equally. Navigation and multi hit moves show earlier instability than single hit moves. This matches what I also saw in the pooled referent group snapshot (Table 6.4) and in the qualitative notes where participants described navigation as sensitive under time pressure.

For navigation, the prompt design likely adds extra pressure. The arrows and direction cues act like strong spatial signals. Under speed, people often follow the strongest and fastest cue, which can pull them away from their own mapping. Stimulus response compatibility research supports this. When a stimulus strongly suggests a spatial response, people respond faster and more reliably when the mapping matches that cue, and struggle more when it conflicts (Miles & Proctor 2009, Ivanoff et al. 2014). This gives a reasonable explanation for why navigation shows earlier instability and why the confused navigation flags concentrate there (Tables 6.5 and 6.6).

For multi hit strings, the earlier instability also makes sense in a timing view. Multi hit actions ask for faster chaining and tighter rhythm. Even if the gesture stays conceptually clear, the task gives less room to complete each part and still land inside the window. So multi hit performance becomes speed fragile earlier than single hits.

The breakdown starts in the middle of the run, not only at the end

Figure 6.7 shows that strict performance starts falling rapidly around the middle of each run, not only at the very end. This matters for usability because it means participants spend a meaningful part of the run in a mixed state. They still understand the mappings and often still perform the correct gestures, but the system increasingly fails to register them on time.

This is the stage where trust and rhythm take damage. In a game context, once correct input does not reliably produce timely output, players stop planning and start reacting to the most immediate cues. Claypool & Claypool (2006) discuss how latency affects player actions and responsiveness in games, and the same logic applies here because my reaction time captures the full loop delay. This also matches what participants reported in the interviews,

where they described shifting attention from the avatar to the prompt icons at the extreme tail.

Why the move animations also mattered

Sometimes the next prompt appears while the previous in game move still plays. Figure 6.3 shows that long moves account for a large share of these cases. This creates a second timing conflict. Even if the player finishes the gesture, the animation and the visual feedback can lag behind the prompt stream.
This matters for usability because people need fast and consistent feedback to feel in control. Nielsen (1993) describes how delays break the feeling of direct manipulation and interrupt the user's flow of thought. Flow research makes a similar point in experience terms. People stay engaged when they can act, get feedback, and feel they control the outcome Csikszentmihalyi et al. (2014). In this study, the prompt stream can move faster than both the avatar animation and the WoZ loop. That makes feedback harder to use. It helps explain why participants described losing rhythm and tracking prompts directly in the fastest tail.

Why "breakdown signals" dropped at the fastest speeds

One interesting pattern in Table 6.5 is that hesitation and visible breakdown markers do not spike in overload. They appear more in the slower zones.
The granular results help interpret this. Once the task crosses the tipping point region, the window becomes too tight for hesitation to appear. Participants either execute immediately, accept lateness, or skip. This fits deadline based accounts of performance. As the deadline tightens, behavior can look cleaner on the surface, while missed actions increase underneath (Claypool & Claypool 2006, Heitz 2014).
So I do not read low hesitation in overload as comfort. I read it as the task removing the time needed for hesitation to show up.

Answering SQ3

Reliability

In my study, gestures stayed reliable under speed only while the prompt interval still gave enough time to complete an action and get it logged. The zone comparison (Table 6.2) shows the same pattern at a high level. The granular bands (Table 6.9) make the timing limit more concrete. A narrow region in

the mid two second range still supports acceptable strict performance while intention remains high and omissions remain relatively low. Then the sharp breakdown begins around the two second range. Below that, strict on time reliability drops quickly and late correct plus no response become the dominant outcomes. Below the lower bound shown in the table, strict reliability collapses.
So, in this setup, strict real time reliability hits a **throughput and timing limit**. It does not fail mainly because gestures become confusing. It fails because the system cannot complete the loop inside the time window once the interval crosses the tipping point (Claypool & Claypool 2006, Heitz 2014).

Usability

Usability held up better than strict reliability. Even when strict on time performance dropped, intention correctness stayed higher for longer. Figures 6.6 and 6.4 show that people can still produce the intended gesture across smaller intervals, even when the system stops registering it on time. This means the gestures remain performable beyond the point where the strict system becomes reliable.
However, usability still degrades once the task enters the tipping point region, because rhythm and feedback start to break. Prompts arrive before gestures finish (Figure 6.2), and sometimes before the avatar finishes animating the previous move (Figure 6.3). This creates a loop where the participant can do the right thing but still feel out of sync. That kind of delayed or misaligned feedback reduces perceived control and interrupts engagement (Nielsen 1993, Csikszentmihalyi et al. 2014, Claypool & Claypool 2006).
These results are limited to the participants in my experiment and to this Wizard of Oz pipeline. With the low participant count, I cannot claim the exact thresholds generalize outside this setup. Still, the overall pattern is strong and consistent. Strict reliability fails first, intention lasts longer, and the tipping point appears once the prompt interval approaches the two second range.

Chapter 7

Sub-Question 4

7.1 Methodology

The sub-question 4 asked: To what extent do user elicited hand gestures show qualities associated with potential immersion in fighting games? I did not attempt to measure immersion as a full gameplay state because my study uses a Wizard of Oz setup with predefined tasks. Instead, I evaluated *immersive potential* by analyzing participants' self report indicators across two experiments:

- Experiment 1, where participants invented gestures,
- and Experiment 2, where they performed those gestures under increasing speed.

I used a theory guided qualitative analysis approach and I kept the analysis traceable to the raw transcripts at every step.

Data sources

The dataset for SQ4 consists of post experiment interviews from ten participants. I conducted the interviews immediately after each experiment using a semi structured questionnaire. The questionnaire targeted three immersion related constructs: intuitive physical interaction, mental imagery and embodiment, and sense of control. Participants sometimes referenced specific moves while answering, but the interview did not systematically cover every move for every participant. For that reason, I treated named moves as examples and I analyzed immersion related qualities at the level of the overall experience.

Step 1: Transcribing interview recordings

I started the analysis by transcribing the audio recordings of the interviews. I used an intelligent verbatim approach. I kept the meaning, phrasing, and hesitation markers when they affected interpretation, but I removed filler words that did not change meaning. This choice improves readability while still preserving the participant's intent. After the first transcription pass, I did a second pass while listening again. I corrected obvious mishearing, clarified unclear pronouns when context made the referent unambiguous, and marked any remaining uncertainty.

Step 2: Choosing a theory guided coding approach

I used a deductive coding strategy because SQ4 is explicitly framed through concepts drawn from immersion and flow literature. This approach is often described as directed content analysis. It begins with a set of theory based categories and uses them to guide coding (Hsieh & Shannon 2005). I selected three constructs as the organizing framework for immersive potential. First, intuitive physical interaction (IPI) reflects the bodily and sensorimotor quality of control and aligns with the sensory dimension of the game play experience (Ermi & Mäyrä 2005) and with work that links body movement to engagement in games (Bianchi-Berthouze et al. 2007). Second, mental imagery and embodiment (MIE) capture imaginative involvement and the felt link between gesture and avatar action. This aligns with the imaginative component of immersion (Ermi & Mäyrä 2005) and with staged accounts of immersion where deeper involvement depends on the player sustaining attention and investment (Brown & Cairns 2004). Third, sense of control (SC) captures agency and mapping predictability. Flow theory emphasises the importance of perceived control and immediate feedback for deep engagement (Csikszentmihalyi et al. 2014). Related HCI work also shows that controller type can shape enjoyment and the experience of the game self (Birk & Mandryk 2013).

Step 3: Coding the transcripts

I coded the transcripts using a fixed codebook with three top level codes: IPI, MIE, and SC. Based on the participant's emphasis, I assigned a primary code to their answers. If a segment touched multiple constructs, I coded it under the construct that was most central to the participant's framing. I used analytic memos to record overlaps when needed. This decision keeps the dataset structured while still preserving nuance, which aligns with prac-

tical coding guidance for qualitative analysis (Saldana 2021). In addition to the construct code, I tagged each segment with a valence label that captures whether the segment supports or constrains immersive potential. I used three valence labels. Positive evidence indicates that the participant described naturalness, character linked imagery, or strong agency. Negative evidence indicates friction, confusion, loss of control, physical strain, or being pulled out of the experience. Mixed evidence indicates that the participant reported both supportive and limiting aspects, or that the outcome depended on speed, gesture type, or pacing. I applied the same coding procedure separately for Experiment 1 and Experiment 2. This separation matters because Experiment 1 captures how participants made sense of gestures while inventing them, while Experiment 2 stress tests those gestures under time pressure and reveals breakdowns, adaptations, and attention shifts.

Step 4: Summarizing each participant into three deliverables

After coding each transcript, I produced three participant level deliverables. These deliverables are descriptive summaries that stay close to the participant's own wording. I did not make cross participant claims at this stage.

Deliverable A is an immersive potential matrix for the participant. It reports IPI, MIE, and SC for Experiment 1 and Experiment 2. For each construct and phase, I summarized the balance of evidence as High, Mixed, or Low. These labels are not numerical scores. They act as a compact description of whether the participant's comments were mostly supportive, conditional, or mostly constraining.

Deliverable B contrasts Experiment 1 and Experiment 2 for the same participant. I wrote this contrast by looking for changes in emphasis and valence across the two phases. For example, a participant may describe strong control during gesture invention but report a loss of control when the prompt stream becomes fast.

Deliverable C lists disruption moments as negative evidence. I treated disruptions as moments where the interaction became noticeable as a control problem. These moments include explicit reports of being pulled out, losing control, confusion, or physical discomfort. They also include implied disruptions such as attention shifting away from the avatar toward prompts. When participants mentioned specific moves, I recorded them as referent examples inside these deliverables. I did not treat referents as analysis units because the interview coverage was not standardized across participants.

Step 5: Building the cross participant dataset in a case by construct matrix

Once I completed the participant level deliverables for all ten participants, I compiled them into a single analysis workbook. I used a case by construct matrix because it supports cross case comparison without flattening qualitative evidence into a single metric. Matrix based displays are a common approach in qualitative analysis because they help researchers keep evidence organized and traceable (Saldana 2021). The master sheet contains one row per participant and six rating fields: IPI, MIE, and SC for Experiment 1 and Experiment 2. For each rating, I also stored a short evidence summary and any referent examples that the participant named. This structure allowed me to compare patterns across participants while retaining links to the underlying text.

Step 6: Converting ratings into counts for descriptive summary

To describe the distribution of evidence across the sample, I converted the High, Mixed, and Low labels into counts per construct per experiment. I report these as frequencies rather than scores. This choice avoids implying measurement precision and fits the exploratory nature of the study. The counts provide a compact overview of how many participants expressed mostly supportive versus conditional experiences for each construct.

Step 7: Cross participant experiment contrast

I then produced a cross participant experiment contrast table. For each construct, I summarized what participants tended to report in Experiment 1 and how that pattern changed in Experiment 2. I used this table to capture the stress test effect of speed. In practice, this table emerged from comparing the evidence summaries in the master matrix and identifying recurring shifts, such as attention moving from the avatar to the icon prompts, or gesture execution becoming sensitive to switching overhead.

Step 8: Standardizing disruption moments into a taxonomy

To make negative evidence comparable across participants, I standardized disruption moments into a taxonomy. I recorded each disruption as a separate instance with a short title, its trigger, its effect, and any referent example

that the participant named. I then grouped instances into higher level categories such as speed overload, prompt design issues, switching overhead, gesture interference, animation pacing mismatch, and physical strain. This taxonomy serves two purposes. First, it makes it possible to report which failure modes appeared most often in this dataset. Second, it supports later discussion by separating problems caused by the gesture concept from problems caused by pacing, prompting, or physical load.

Step 9: Trustworthiness and traceability

I used several practices to support trustworthiness. I kept the analysis traceable by storing evidence summaries alongside ratings and by preserving a full list of disruption instances. I also made memos during coding to capture ambiguity, overlaps between constructs, and decisions about valence. These practices support transparency and allow later readers to see how I moved from raw text to summary patterns (Nowell et al. 2017). I also interpret all patterns cautiously because the participant pool is small and the study context differs from a real fighting game match. The goal is not to claim population level effects. The goal is to describe what participants in this study reported, and to identify conditions that appear to support or disrupt immersive potential during early stage gesture based control design.

SQ4 Methodology summarized

The SQ4 results section reports findings in four layers. First, I present an overview using the construct rating counts. Second, I report results construct by construct, separating Experiment 1 from Experiment 2. Third, I summarize the experiment contrast to highlight what changed under speed pressure. Fourth, I report disruption categories as negative evidence and I use referents only as illustrative examples. This structure keeps the results readable while staying grounded in the interview data.

7.2 Results

For each construct within each experiment, I summarized the balance of evidence as High, Mixed, or Low. These labels are qualitative summaries, not numerical scores. Where participants mentioned specific moves, I treat those referents as examples, not as a standardized per-move evaluation.

Overview of construct ratings across participants

In Experiment 1, IPI was mostly High across participants, while MIE was mostly Mixed. SC was often High, but some participants framed control as conditional or still emerging. In Experiment 2, ratings shifted toward Mixed across all three constructs, reflecting speed pressure and system constraints.

Construct	E1 High	E1 Mixed	E1 Low	E2 High	E2 Mixed	E2 Low
IPI	9	1	0	2	8	0
MIE	1	9	0	0	7	3
SC	7	3	0	2	7	1

Table 7.1: Distribution of High, Mixed, and Low ratings across constructs and experiments.

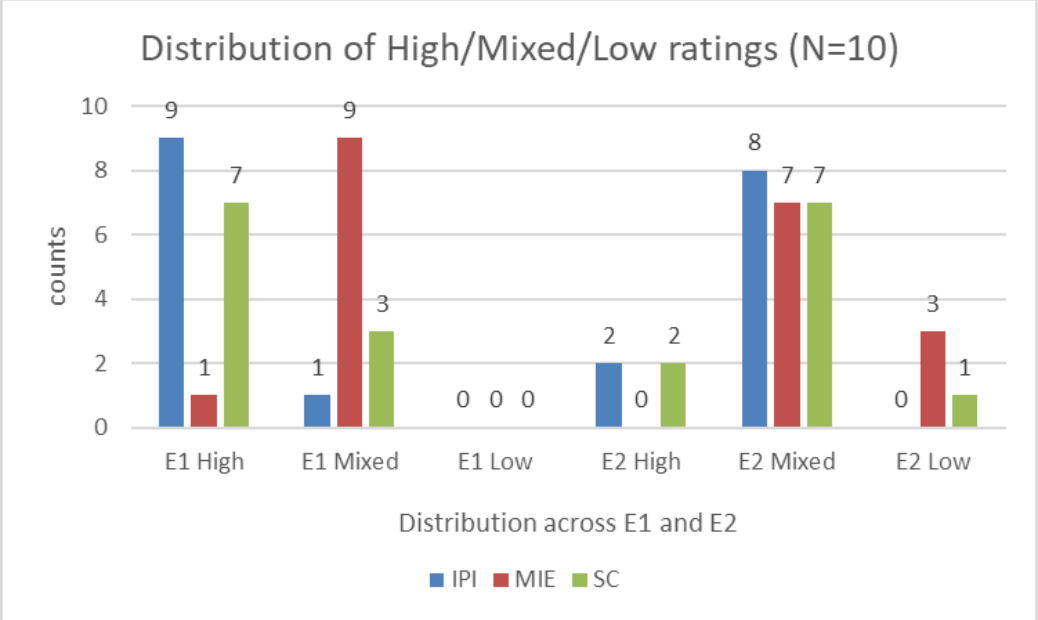


Figure 7.1: Distribution of High/Mixed/Low ratings (N=10)

Intuitive Physical Interaction (IPI)

Experiment 1: Gesture Invention

In Experiment 1, 9 participants were rated High and 1 was rated Mixed for IPI. Participants often described their invented gestures as natural because

Participant	E1_IPI	E1_MIE	E1_SC	E2_IPI	E2_MIE	E2_SC
P001	High	Mixed	High	Mixed	Mixed	Mixed
P002	High	Mixed	Mixed	High	Mixed	High
P003	High	Mixed	High	Mixed	Low	Mixed
P004	High	Mixed	High	High	Mixed	High
P005	High	Mixed	Mixed	Mixed	Low	Low
P006	High	Mixed	Mixed	Mixed	Mixed	Mixed
P007	Mixed	Mixed	High	Mixed	Low	Mixed
P008	High	Mixed	High	Mixed	Mixed	Mixed
P009	High	High	High	Mixed	Mixed	Mixed
P010	High	Mixed	High	Mixed	Mixed	Mixed

Table 7.2: Case-by-construct rating grid for all participants across Experiments 1 and 2.

the gestures resembled how they would explain or act out the move. Several participants also described intentionally simplifying gestures to keep them repeatable and comfortable.
Across the interviews, IPI evidence in Experiment 1 commonly included (a) movement similarity (doing a punch-like or kick-like action), (b) everyday metaphors for navigation (pointing or moving "something" forward/back), and (c) comfort-first gesture choices that still felt meaningful. When participants raised IPI concerns in this phase, they typically focused on remembering direction mappings, or on combination/sequence moves that felt less natural to compress into a single gesture.

Participant-level IPI evidence (Experiment 1):

- P001: Reported gestures felt natural because they tried to mimic the game action. Named uppercut as "exactly how I did" and "felt like you were doing it in real." Kick also "worked well."
- P002: Described jab as the most intuitive because it is "quick and simple," so the gesture naturally becomes quick/simple too. Reported quick moves (punching/kicking/hammer) as easier to mimic. Noted that long succession moves are harder to interpret intuitively.
- P003: Reported the interaction felt "really natural," partly because they kept a mouse-like resting posture. Described gestures as quick/responsive and "fun." Most moves felt easy/smooth (e.g., index tapping, forward/back). Minor difficulty: using the finger next to the middle finger.

- P004: Reported most gestures felt natural and intuitive. They actively optimized for repeatability and comfort (avoiding motions that would become uncomfortable over repeated play). They also noted some gestures felt more immediately intuitive (finger motions, thumb poke).

- P005: Said the gesture creation felt natural and intuitive, and they intentionally kept gestures simple and "easy for people to do." Movement gestures (e.g., directional movement) were singled out as especially natural because they match everyday pointing/indicating.

- P006: Described the experience as "interactive and indulging" and said they tried to make gestures as natural as possible. Their rationale for naturalness relied on a body-logic mapping (left/right dynamics translated into single-hand motion).

- P007: Said gestures were "as intuitive as they could be," but also framed the experience as control-mapping (like assigning buttons). Unnaturalness came from how much motion a single move required and using finger flicks instead of simple taps.

- P008: Reported gestures felt intuitive "for the most part," especially directional movement (pointing as a natural everyday action). Also described flicking for hits as highly satisfying and fun, with stronger felt fit between gesture and on-screen action.

- P009: Reported gestures felt "very natural," mainly because they could map hand space to character space (placing the character on a 2D plane and moving accordingly) and map limbs to fingers. Main drawback was remembering what they invented.

- P010: Reported gestures felt naturally intuitive because they were similar to the movements. Strong example: left kick felt like an obvious everyday way to "explain a kick" to someone.

Experiment 2: Increasing Speed

In Experiment 2, 2 participants were rated High and 8 were rated Mixed for IPI. Many participants still described the gestures as broadly natural, but they also described conditions where that naturalness weakened. These conditions typically appeared when the pace increased, when gestures required larger travel or orientation changes, or when rapid switching made execution less stable.

A recurring pattern in Experiment 2 was that IPI became more sensitive to interaction overhead. Participants often described difficulty in switching quickly, returning to a neutral posture, or differentiating direction-heavy navigation cues. Some participants reported that the gesture itself remained intuitive, but the prompt stream, icon recognition, or timing pressure made the overall interaction feel less smooth.

Participant-level IPI evidence (Experiment 2):

- P001: Said gestures still felt natural overall as speed increased, but certain gesture qualities became problematic: (1) directional navigation remained confusing under time pressure, (2) gestures with larger wrist travel / up–down motion became slow or hard.
- P002: Said gestures stayed "somewhat" natural/intuitive as speed increased, and overall, everything got easier over time as they got used to the mappings.
- P003: Some gestures stayed easy (punching, finishers, index-tap punch), but movement-related gestures (rotations, forward/back) became difficult at speed. The main issue was switching between signals and returning to a natural hand position quickly.
- P004: Reported gestures stayed natural/intuitive even as speed increased and felt easier than keyboard/joystick. Main constraint was transition overhead: gestures that move far from a neutral position became harder at speed because returning to neutral takes time.
- P005: Directional movement still felt intuitive, but at higher speed they began mixing up icons/gestures and reported that some gestures were no longer performed exactly as intended because their hand was "flying in air" (not resting).
- P006: Said gestures still felt natural as speed increased. Main difficulty was confusion between arrow prompts (color-driven perception made it harder to distinguish 4 directions), not the other icons. Reported several gestures became easier over time (uppercut, hammer, neutron bomb).
- P007: Said the experience overall was intuitive enough, but they would change several gestures. Main issue was inconsistency within the punch family (sometimes fist-based, sometimes tapping), causing confusion when seeing a fist cue.

- P008: Said gestures remained intuitive "for the most part," but speed reduced the earlier satisfaction and made performance feel more like keeping up. Noted that gestures requiring large wrist/orientation changes become awkward and time-costly at speed.

- P009: Reported some gestures stayed intuitive (especially kicking and punching), but overall experience at speed became less "intuitive" because they disliked following icon prompts and struggled to switch gestures quickly in fast gameplay.

- P010: Said gestures largely stayed intuitive, but speed revealed difficulty with (a) keeping up with signs/icons, and (b) specific gestures/moves—especially Phoenix and the Hammer+Phoenix combination. Participant also began simplifying gestures during play.

Mental Imagery and Embodiment (MIE)

Experiment 1: Gesture Invention

In Experiment 1, 1 participant was rated High and 9 were rated Mixed for MIE. Most participants described some connection between their gestures and the character's action, but the form of that connection varied. Some participants described imagining the character or the force/impact of the move, while others described primarily thinking about the hand and the control mapping.
A common feature across the dataset was that participants did not usually describe a stable experience of "being the character." Instead, MIE appeared in several different ways: (a) character-focused planning (designing gestures based on what the character should do), (b) partial bodily simulation (e.g., focusing on force or impact rather than full-body mimicry), and (c) a command-like stance (gestures as inputs that resemble the move without self-identification).

Participant-level MIE evidence (Experiment 1):

- P001: Did not describe "picturing the character" visually, but described embodied simulation through force (thinking about "the force with which he hits") and feeling like their body was "mimicking" the character/force.

- P002: Reported a mix of sometimes imagining the character/move (e.g., replicating forward movement then punch for a wave-dash + smash), but also actively prioritizing hand comfort. For quick moves

(kicking/punching/hammer) they reported moments of "mimicking the character."

- P003: Said they were more hand-focused than character-focused, choosing gestures for ease and memorability (mapping moves to fingers). Still indicated it was "easy" to picture the character over time, but their core reasoning emphasized hand mapping and combinations.
- P004: Described balancing two things: (1) aligning with how the character moves and (2) what feels comfortable for the hand. They explicitly noted that when hand motion becomes less aligned, they may feel disconnected.
- P005: Reported thinking primarily about what the character should do first, then secondarily about what is easy for the hand. However, they explicitly said they did not feel like their body was mimicking the avatar (only wrist moves).
- P006: Reported imagining themself as the user playing, not imagining being the character. Still described a partial embodiment link: their arm felt like it was doing what the character was doing (but not whole-body mimicry).
- P007: Reported thinking much more about the hand than the character ("what can I do easily quickly?"). Said they could picture the character if asked, but it was not their default strategy. Embodiment was strongest only in the literal gestures.
- P008: Reported thinking "100% character" in terms of what the character should do, but also clearly stated they do not experience the avatar's limbs as their own and did not describe vivid mental imagery.
- P009: Reported thinking about both their hand movement and the character doing the move. Found it satisfying to see the avatar hit based on their gestures and described expressing aggression physically through the hand/fingers.
- P010: Reported thinking in "character-move terms" (e.g., "left kick is how I imagine it"), but explicitly said they were not thinking 'I am the one fighting'. Their gesture strategy was "closest familiar movement that explains what the fighter is doing."

Experiment 2: Increasing Speed

In Experiment 2, 0 participant was rated High, 7 were rated Mixed, and 3 were rated Low for MIE. Participants often reported that mental imagery and embodiment were more fragile under speed pressure than IPI. Several participants described shifting attention away from the avatar and toward icons or their own hands as the pace increased.
When participants described reduced MIE in Experiment 2, they typically linked it to one of two conditions: (a) the prompt stream becoming too fast to maintain a coherent sense of what the character is doing, or (b) a mismatch between gesture timing and avatar animation timing. In these moments, participants described the experience as more command-following than character-linked.

Participant-level MIE evidence (Experiment 2):

- P001: Reported that at higher speeds they lost the connection because everything happened too fast ("you don't know what exactly is going on between you and the character"). Explicitly said it felt immersive when slow, then diminished when pace became too fast.

- P002: Initially reported being more focused on "performing the right action" than feeling "one with the character." With familiarity and especially when seeing the character perform the move in response, they reported that it started to click and felt better.

- P003: As speed increased, their comments focused on prompt-following challenges (switching, symbol confusion, delay) rather than maintaining a character-focused mental image. They did not describe sustained "being the character" at high speed.

- P004: Reported they could picture the character doing the move, but repeatedly noted misalignment: their gesture finishes faster than the avatar animation, especially for combination moves. They said it would be more enjoyable if character motion stayed aligned.

- P005: When pace increased, they stopped looking at the character and focused on the icons; they described the experience as "following commands," similar to icon-driven dance/step games. They also noted lag-like effects where they were still doing the previous gesture.

- P006: Continued to imagine themself as the user with the character replicating actions; explicitly stated they would not want to visual-

ize themself as the character more. They described the interaction as becoming immersive especially at medium pacing.

- P007: Explicitly said they were not paying much attention to the character; interaction became "that symbol means do this." Visual attention shifted to anticipating what comes next rather than watching execution.

- P008: Reiterated that they don't "imagine" being the character; they experience gestures as commands whose nature resembles the move. At higher speed, less intuitive/awkward gestures and heavy wrist/orientation moves were reported to pull them out of the experience.

- P009: Reported they could still imagine the character doing what they wanted, and that picturing an aggressive punch helped them perform it. However, they also described themselves as a "button smasher" who typically does not plan moves.

- P010: Reported being able to think of the character performing moves ∼70% of the time and said they could imagine the character even with eyes closed if not driven by prompts. However, at higher speeds they stopped looking at the character.

Sense of Control (SC)

Experiment 1: Gesture Invention

In Experiment 1, 7 participants were rated High and 3 were rated Mixed for SC. Many participants described a strong cause–effect link between their gestures and the character's actions. Participants often compared this feeling favorably to traditional controllers, describing gesture input as more direct or more connected to the move concept.
Mixed SC ratings in Experiment 1 usually reflected one of two stances: (a) the participant felt in control in principle but emphasized that the mapping was still new and would require learning, or (b) the participant framed control as intentionally bounded, preferring a balance where gestures support engagement without becoming fully realistic.

Participant-level SC evidence (Experiment 1):

- P001: Compared to a normal controller, reported more control and stronger connection because gestures felt like doing the action "in real," which made the character feel more relatable.

- P002: “Yes and no”: not fully in control yet because it was fresh/new and self-invented, but “yes” because it felt like actually moving with the character rather than pressing buttons. Reported strong control for the Phoenix Smash.
- P003: Reported strong control: “no delay,” actions happened at the right time. Named the fist/punch as strongly matching the character (Phoenix Smash).
- P004: Reported a stronger sense of control and freedom than joystick/keyboard, partly because gestures are less “monotonous” than keys and may take less time to learn.
- P005: Reported a clear sense of control because the character followed their gestures “exactly,” and said gestures were a good match. At the same time, they distinguished this from feeling like they were fighting; it felt like directing the character.
- P006: Reported “some sense of control,” but also argued that full control is not ideal because it would reduce the “game” feel and push toward realism. They framed the desired experience as a boundary between immersion and realism.
- P007: Reported a clear sense of control (“seeing a thing I came up with control the character”). Strongest mapping example was the literal punch gesture.
- P008: Reported a “robust sense of control” for the limited moves because the character reliably did what the gesture commanded; emphasized that lack of lag is critical to this feeling.
- P009: Reported feeling in control because gestures reduce the “intermediate layer” of a console and feel closer to doing what the character does. Still, they flagged a control concern for longer/sequence moves (Phoenix Smasher / Hangover): they couldn’t move fingers reliably and in time.
- P010: Reported strong control and directness (“fighter was my hand”). Most mappings felt reasonable/doable, and they felt the gestures made sense enough that an observer could infer the move from the motion.

Experiment 2: Increasing Speed

In Experiment 2, 2 participants were rated High, 7 were rated Mixed, and 1 was rated Low for SC. Under speed pressure, participants frequently described control as conditional on timing. Some participants maintained a strong sense of control when the system feedback felt responsive and when the pace was manageable. Others described losing control beyond a speed threshold or when switching demands and animation pacing interfered with predictable execution.
Reports of reduced SC in Experiment 2 often involved either (a) prompt pace outstripping the participant's ability to switch gestures cleanly, (b) the character still finishing a move while the next prompt arrives, or (c) misrecognition and drift when gestures became less precise at speed.

Participant-level SC evidence (Experiment 2):

- P001: Reported that once speed increased past a point, they lost control ("not really... not that much"), tied to difficulty keeping up with icons, directional confusion, and cognitive load. Also suggested design changes to improve control at speed.

- P002: Reported that even when fast, it still felt like controlling the character because feedback matched their movements (no delay; wrong action only once, barely noticed). Did not want to simplify gestures; felt they were already quick/efficient.

- P003: Some gestures still felt in control (finisher/fist, index finger punch, kick), but movement gestures were "mostly getting out of control." They attributed loss of control to fast switching + delay and confusing navigation.

- P004: Reported feeling completely in control and able to keep up with prompts but identified a system pacing ceiling: prompts could change faster than the character can execute moves, causing missed or duplicated executions.

- P005: Reported that at high speed they felt mainly "in control of following the icons" rather than controlling the character. Control decreased as memory demands increased and gestures became mixed/confused, especially when their hand position differed slightly from what they had invented earlier.

- P006: Described losing the character "a bit" at higher speed mainly because the character needs time to finish the prior move before executing the next ("ritual"/recovery), even when the participant could perform the gesture quickly.
- P007: Still reported control ("hands do something and character does something"), but felt less because attention focused on reacting to the next prompt. Suggested control/engagement could improve with stronger audio feedback and social play context (playing a friend, reactions).
- P008: Said they still felt in control in terms of character responsiveness, but could not keep up with prompt speed later on. Also noted character reset/animation time affects chaining, and repeated gesturing during reset feels more "committed."
- P009: Reported they still felt directly controlling the character when fast and linked this to "immersion" (thinking of the player as themself and acting to "inflict damage"). At the same time, control was constrained by gesture-switching speed.
- P010: Reported strong control for many basics (kick, jab, uppercut, hammer, forward/back) and felt recognition was immediate for those. Control degraded mainly around (a) Phoenix (difficult + tiring + not intuitive in their design), (b) Hammer+Phoenix (higher cognitive load).

Experiment-level contrast summary (E1 vs E2)

To summarize the shift from Experiment 1 to Experiment 2, Experiment 1 results emphasize meaning-making during gesture invention, while Experiment 2 results emphasize stability under time pressure. Across participants, IPI often remained broadly positive but became conditional on switching and gesture travel. MIE was more variable in Experiment 1 and more likely to weaken in Experiment 2 when attention shifted to prompts. SC was strong in Experiment 1 for many participants, and in Experiment 2 it was often described as preserved only when timing and feedback supported predictable action.
Table 7.3 summarizes the cross-participant contrast extracted from the coded dataset.

Construct	Experiment 1 patterns	Experiment 2 outcome
Intuitive Physical Interaction	Most participants described inventing gestures as natural when they could mimic or simulate the move; comfort-driven simplifications appeared early for some actions.	Naturalness often remained for individual/simple actions, but time pressure surfaced issues: directional ambiguity, gesture switching overhead, and combo complexity.
Mental Imagery & Embodiment	Imagery varied: some participants described character- or force-based simulation; others stayed hand-focused and treated gestures as control mappings rather than acting/inhabiting the avatar.	At higher speed, several participants reported a shift toward prompt-following and reduced capacity to "picture" the character; connection became pace- and feedback-dependent.
Sense of Control	Many participants described strong agency from direct gesture-to-action mapping and "doing the move" themselves; perceived control was often compared favorably to a traditional controller.	Perceived control frequently became conditional on system pacing, cue clarity, and gesture distinctiveness; several described a threshold where control degraded when prompts or animations outpaced them.

Table 7.3: Cross-participant contrasts across constructs and experiments.

Disruption moments (negative evidence)

I treated "disruption moments" as negative evidence for immersive potential. These moments include explicit reports of being pulled out of the experience, losing control, confusion, or physical strain, as well as implied disruptions such as attention shifting away from the character toward prompts. Across the ten participants, I coded 73 disruption instances into a standardized taxonomy of ten categories. Table 7.4 lists disruption categories and how often they appeared across participants and instances.
The most frequently observed category was physical discomfort/fatigue/strain, followed by timing-related constraints (animation pacing mismatch) and cognitive load around combos/sequences. The categories below describe typical triggers and effects as reported in the interviews.

Category	Participants_n	Instances_n
Physical discomfort / fatigue / strain	9	17
Animation pacing mismatch / lag / recovery time	6	9
Cognitive load / memory / planning (combos/sequences)	6	9
Gesture similarity & interference (family inconsistency)	6	8
Transition overhead & switching (return-to-neutral)	6	7
Speed pressure & pace threshold	5	7
Attention shift away from avatar (icon-/prompt-/hand-focused)	4	4
Prompt / symbol design confusion	4	4
System interpretation / misrecognition	3	5
Directional mapping ambiguity (navigation)	3	3

Table 7.4: Summary of challenge categories with participant counts and total coded instances.

Physical discomfort / fatigue / strain

Appeared in 9 of 10 participants, across 17 coded instances. Biomechanical difficulty, discomfort, or fatigue (e.g., sustained mid-air posture, large wrist travel, fine finger isolation).
Example instances:

- P001: Large wrist-travel / up–down gestures at high speed (including "hammer" and other up–down patterns)
- P002: Localized physical discomfort / finger limitations

Animation pacing mismatch / lag / recovery time

Appeared in 6 of 10 participants, across 9 coded instances. Mismatch between gesture timing and avatar/system timing (e.g., recovery frames, prompts advancing while avatar still animating), breaking cause–effect.
Example instances:

- P002: System pacing conflict: next prompt appears while character still performing

- P003: Loss of control beyond a pace threshold (gesture switching + delay)

Cognitive load / memory / planning (combos/sequences)

Appeared in 6 of 10 participants, across 9 coded instances. Hesitation/errors from interpreting multi-step sequences or remembering/combining cues, especially for combos.
Example instances:

- P001: Combo prompts as cognitive disruption
- P002: Successive/long move gestures as a weak point for "intuitive interpretation"

Gesture similarity & interference (family inconsistency)

Appeared in 6 of 10 participants, across 8 coded instances. Different moves mapped to too-similar gestures (or inconsistent variants), causing confusion between actions under pressure.
Example instances:

- P002: Hand-monitoring / needing visual confirmation
- P003: Gesture confusion between similar attacks (uppercut vs hammer)

Transition overhead & switching (return-to-neutral)

Appeared in 6 of 10 participants, across 7 coded instances. Extra movement needed to switch between gestures or return to a start state, interrupting chaining and continuity.
Example instances:

- P001: Neutral reset requirement disrupts continuity
- P003: Movement gestures degrade at speed (forward/back/rotation)

Speed pressure & pace threshold

Appeared in 5 of 10 participants, across 7 coded instances. Reported "too fast" moments where response capacity was exceeded and the interaction felt disruptive or disconnected.
Example instances:

- P001: Speed threshold where “connection” breaks
- P007: Motion cost vs keyboard expectation (gesture effort disrupts intuitiveness)

Attention shift away from avatar (icon-/prompt-/hand-focused)

Appeared in 4 of 10 participants, across 4 coded instances. Attention pulled toward prompts or correct execution rather than the character/experience (command-following mode).
Example instances:

- P004: Attention shift toward “doing the gesture correctly” and checking execution
- P005: Attention pulled from character to icons (command-following mode)

Prompt / symbol design confusion

Appeared in 4 of 10 participants, across 4 coded instances. Ambiguity or inefficiency in cue design (icons, labels, colors, layouts) that slowed parsing or misled interpretation.
Example instances:

- P002: Prompt text labels caused early confusion (icon–word mismatch)
- P005: High-speed “crash”: participant + system confusion as pace escalated

System interpretation / misrecognition

Appeared in 3 of 10 participants, across 5 coded instances. Wrong system classification due to gesture drift, posture changes, or under-exaggeration; participants compensated by exaggerating/stabilizing.
Example instances:

- P005: Hand posture shift at speed (not resting hand) caused gesture drift
- P008: Less exaggerated gestures distort more under stress

Directional mapping ambiguity (navigation)

Appeared in 3 of 10 participants, across 3 coded instances. Confusion distinguishing direction-based movement cues/gestures (e.g., left/right/back/forward/U-turn), often leading to mixed-up navigation responses.
Example instances:

- P001: Directional navigation confusion (persistent across both experiments)
- P003: Symbol/prompt confusion for navigation (especially "U-turn" cue)

7.3 Discussion

Throughout this chapter, I interpret patterns cautiously because my participant pool is small ($N = 10$) and the setting differs from a real fighting game match. When I say that participants "tended to" report something, I mean within this study and within this dataset.

What the results suggest at a glance

The pattern observed in 7.1 for Experiment 1 suggests that, during gesture invention, most participants found it easy to create gestures that felt physically natural and controllable, while the character imagery component varied more across people.
The pattern observed in 7.1 for Experiment 2 shows that speed pressure did not fully remove intuitive interaction or control for most participants, but it made both more conditional. MIE was the most fragile construct under speed.
Negative evidence reinforces this shift. Disruption moments clustered around physical strain, timing constraints, switching overhead, gesture interference, and cognitive load in combos. These disruptions matter because immersion models treat breakdowns as barriers. Once players notice the interface as an interface, they often exit the effortless mode that supports deeper involvement (Brown & Cairns 2004).

Intuitive Physical Interaction as a basis for sensory involvement

The strongest evidence for immersive potential in my dataset sits in Intuitive Physical Interaction. In Experiment 1, participants usually described

their gestures as natural because they resembled the move concept. They often chose gestures that feel like a punch, a kick, or an impact. They also used simple metaphors for navigation, such as pushing, pulling, or indicating direction. This aligns with Ermi & Mäyrä (2005)'s sensory immersion component, where audiovisual and bodily engagement can pull attention into the game world. It also aligns with work on body movement and engagement, which argues that movement imposed or afforded by a controller can increase engagement and affective involvement (Bianchi-Berthouze et al. 2007).
The results also show that intuitive interaction is not the same as doing a realistic full body reenactment. Several participants intentionally kept gestures compact to make them repeatable and comfortable. That choice still produced high IPI in Experiment 1. This matters for fighting games. Players need speed, precision, and repetition. If a gesture system demands large, tiring motions, it can increase friction rather than immersion.
Experiment 2 clarifies what "intuitive" means under time pressure. Participants often still called the gestures natural, but they also described when the interaction stopped feeling smooth. A gesture can remain meaningful while becoming hard to execute fast, especially when it requires travel, rotation, or an explicit return to a neutral posture. This matches an important point from (Brown & Cairns 2004). Engagement has barriers, and controls are one of them. If the input system demands too much attention or physical effort, it becomes a barrier rather than a path to deeper involvement.
In my dataset, this barrier appears most clearly in switching and transition overhead. Several participants described that the hard part was not the gesture itself but the change between gestures at speed. This also helps interpret why IPI shifted from mostly High to mostly Mixed in Experiment 2. Participants did not suddenly lose an intuitive understanding of the moves. Instead, timing turned the interaction into a coordination problem.

Mental imagery and embodiment as imaginative immersion, and why it weakened at speed

Mental Imagery and Embodiment showed the most variation across participants. In Experiment 1, most participants were rated Mixed for MIE. Many participants did connect their gesture to what the character is doing, but they described that connection in different ways. Some participants explicitly pictured the avatar or imagined impact and force. Others framed gestures as commands that resemble the move but did not report a strong "I am the character" feeling. In my opinion, this is not a failure, rather it reflects how imaginative involvement works in games. Ermi & Mäyrä (2005)

describe imaginative immersion as a component that can vary with content and player mindset. Brown & Cairns (2004) also note that deeper immersion stages involve emotional and cognitive investment. A short task based Wizard of Oz study does not easily create that kind of investment.
Experiment 2 provides clearer evidence about the conditions that support or break imagery. Under increasing speed, several participants described shifting attention away from the character and toward the prompts or their hands. This is consistent with the idea that imaginative involvement competes for attention (Brown & Cairns 2004, Ermi & Mäyrä 2005). When the player uses most of their attention budget to decode icons and plan the next gesture, there is less room to maintain a stable mental image of the avatar's action. This helps explain why MIE dropped to Mixed or Low for many participants in Experiment 2. This pattern also highlights an important methodological point. My Experiment 2 prompt stream is a useful stress test for reliability, but it can also push participants into a command following mode. In that mode, the player treats the avatar as output confirmation rather than as a focus of imagination. In a real fighting game, players also process cues quickly, but they do so while reading an opponent, reacting to spacing, and anticipating outcomes. Those elements can support imagery because they provide context and stakes. My setup reduced that context, so it may have reduced the opportunity for imaginative immersion even when the gesture mapping was meaningful.
Even with these constraints, some participants still described moments of embodiment, especially when the gesture concept matched the move and the feedback felt immediate. This matches the idea that embodiment can be supported by congruent sensorimotor cues (Francese et al. 2012). However, the results suggest that gesture meaning alone is not enough. Timing and attention demands can mute the imaginative layer.

Sense of control as the gatekeeper for immersion and flow

Sense of Control sits at the center of the immersive potential story. In Experiment 1, most participants described strong control because they could predict what would happen from their gesture. They often framed this as directness. They said the character felt like an extension of their hand, or they compared it favorably to a button press. This aligns with both immersion and flow work. Brown & Cairns (2004) treat control as a barrier. Players must learn and accept the control scheme before they can move toward engrossment. Flow theory from Csikszentmihalyi et al. (2014) also emphasizes

clear goals, immediate feedback, and a sense of control as conditions that support deep engagement.

Experiment 2 shows that control becomes conditional when time pressure rises. The dominant pattern is not total loss of agency. Instead, participants described a threshold where control depends on timing, gesture switching, and system responsiveness. This is consistent with work that links controller naturalness and mapping to immersion but also warns that naturalness alone does not guarantee immersion (Rogers et al. 2015). When the system pacing or feedback breaks the cause effect loop, players start managing the interface instead of acting through it.

Two disruption categories support this interpretation strongly. First, several participants reported animation pacing mismatch or recovery time as a reason for losing control. If the character still finishes a move while the next prompt arrives, the player can feel late even when they execute correctly. Second, misrecognition and gesture interference break predictability. Once the system interprets a gesture as the wrong move, players stop trusting the mapping. They then slow down, watch their hands more, and simplify gestures. That adaptation can help performance, but it can also reduce the feeling of effortless control that supports immersion.

This is where gesture based control has a clear design tradeoff. Rich, expressive gestures can feel satisfying in isolation, but they can undermine control when used repeatedly at speed. In my data, many participants implicitly solved this by simplifying gestures during Experiment 2. This suggests that the most immersive gesture set for a fighting game may not be the most realistic or the most expressive. It may be the one that preserves agency under pressure.

Putting the three constructs together: what "immersive potential" looks like in this study

Taken together, the results suggest a layered form of immersive potential. First, most participants can reach a strong baseline of intuitive physical interaction during gesture invention. This baseline supports sensory involvement because the player is physically doing something that fits the action concept. Second, many participants report a strong sense of control during invention. This supports engagement because the mapping feels understandable and predictable. Third, mental imagery and embodiment appear as an extra layer that some participants access more easily than others, and that weakens under speed when attention shifts to prompts.

This layered view matches the multi component view of immersion from

(Ermi & Mäyrä 2005). Gesture control can strengthen sensory involvement, and speed trials can create challenge based intensity. But imaginative immersion depends on whether the player can keep the character in mind and whether the interaction stays transparent. Brown & Cairns (2004)' staged model helps interpret why many participants do not report deep identification. A short controlled session can support engagement and moments of engrossment, but total immersion usually requires sustained focus, emotional investment, and the absence of noticeable barriers.
In my dataset, the clearest barriers were not conceptual misunderstanding of gestures. They were physical and temporal. Physical strain appeared in 9 of 10 participants, and timing mismatch appeared in 6 of 10. This echoes the body movement literature. Movement can increase engagement, but it also changes effort and fatigue (). So immersive potential in gesture based fighting games depends on designing movements that feel meaningful while staying light enough for repetition.

What the disruption taxonomy reveals about failure modes

Disruption moments provide negative evidence for immersive potential. They show where the interface becomes noticeable and where attention shifts from the fight to the control problem. The most common disruption categories in this dataset were: Physical discomfort / fatigue / strain (9/10 participants), Animation pacing mismatch / lag / recovery time (6/10 participants), Cognitive load / memory / planning (combos/sequences) (6/10 participants), Gesture similarity & interference (family inconsistency) (6/10 participants), Transition overhead & switching (return-to-neutral) (6/10 participants), Speed pressure & pace threshold (5/10 participants).
These categories cluster into three broader barrier types. The first is physical cost. Many participants described fatigue or discomfort, often in the wrist or forearm, and often linked to rotations and repeated travel. The second is time coordination. Participants described losing rhythm when prompts arrived too fast, when the avatar needed recovery time, or when switching required a neutral reset. The third is mapping ambiguity. Participants described interference between similar gestures, confusion in symbol prompts, and occasional misrecognition. All three barrier types can interrupt engagement. Brown & Cairns (2004) describe barriers as obstacles to deeper immersion. In my study, these barriers appeared most clearly during the speed stress test.
Importantly, the disruption data does not imply that gesture control is un-

suitable for fighting games. It shows where the design must be careful. If a gesture set keeps the mapping clear, reduces interference, and stays physically light, then the system can keep the player in a state where control feels effortless. That state is the condition under which sensory and imaginative involvement can build rather than collapse.

Limitations and how they shape interpretation

Several limitations constrain how far I can generalize from these results. First, the participant pool is small (N = 10) and the study only sampled a limited action set. Second, Wizard of Oz control and the use of prompts reduce ecological validity. Participants did not play a full competitive match, and they did not need to make tactical decisions against an opponent. Third, the speed condition may have increased cognitive load in an artificial way, since participants focused on decoding icons rather than on reading the game state. Finally, the findings rely on self report. Participants described their experiences clearly, but self report does not fully capture moment to moment engagement.

Even with these constraints, the dataset still supports a useful interpretation. It shows which immersion related qualities are easy for participants to access in a gesture design process, and which qualities become fragile under pressure. This is precisely what "immersive potential" means in a prototype stage.

Answering SQ4

Within this study, user elicited hand gestures showed strong evidence of intuitive physical interaction and a generally strong sense of control during gesture invention. These two qualities support a baseline of engagement because participants can understand the mapping and can act through it with confidence. Mental imagery and embodiment appeared more unevenly across participants. It was present as character focused planning or as moments of acting out the move, but it often weakened under speed when attention shifted to prompts and switching. As a result, the immersive potential of the gesture set in this study looks strongest when the interaction stays physically light, the mapping stays clear, and the system timing supports predictable control. When those conditions fail, disruptions appear and the interaction becomes a control management task rather than an embodied fight.

Chapter 8

Design Considerations Derived from the Study Findings

Scope and intended use

This chapter summarizes *design considerations* suggested by the patterns observed in my two Wizard-of-Oz experiments. I avoid framing these points as general design implications, because the study is small (N=10), tied to a specific testbed (Tekken 8, Paul Phoenix), and evaluated under an experiment-specific speed-ramp protocol. Instead, I treat the following as **design considerations** that emerged from the data and may be useful when translating the results into prototypes or follow-up studies.
To reduce reliance on reader memory, each consideration is grounded in a short recap of the relevant findings from the Results chapters. In brief:

- in Sub-question 1, participants reused a small set of components (simple primitives, stable handshapes, palm-down bias) and often constructed a "control grammar" with neutral start positions and opposite pairs (Figures 4.3–4.6);
- in sub-question 2, convergence was multi-level: some referents showed population-level defaults, while others remained form-diverse but intent-convergent (Tables 5.1–5.5; Figures 5.1–5.6);
- in sub-question 3, performance under the speed ramp dropped mainly due to time pressure (late-correct and no-response) rather than a surge in misinputs, with a clear tipping point around ≈2.0 s prompt intervals (Tables 6.1–6.9; Figures 6.4–6.7);

- and in sub-question 4, perceived intuitive physical interaction and sense of control were often positive during invention (Experiment 1) but became more conditional under speed and system constraints (Tables 7.1-7.4).

1. Considerations for gesture–command mapping

Consideration 1.1: Treat each referent as its own mapping problem (policy per referent)

Sub-question 2 shows that different referents exhibit different convergence regimes. Some have a clear single dominant family (e.g., Left Punch, Uppercut, Phoenix; Tables 5.1–5.3), while others lack whole-gesture consensus but still show strong spatial-intent convergence (e.g., Side Step In/Out; Table 5.5), and Neutron Bomb shows two stable co-dominant families (Table 5.2).

Possible design consideration: Rather than assuming a single global rule ("one move, one exact gesture"), a gesture system could label each move with a *mapping policy* that matches its observed convergence regime:

- **Tight default:** one recommended gesture form, evaluated with stricter tolerance (fits single-dominant + coherent cases; Tables 5.2–5.3).
- **Template default:** one recommended gesture idea with allowed variation (fits cases where a dominant family exists but internal similarity is low; e.g., Uppercut; Table 5.3).
- **Two-default split:** two equally valid official options (fits co-dominant Neutron Bomb; Table 5.2).
- **Flexible intent:** accept multiple realizations as long as they preserve the stable intent cue (fits intent-convergent but form-divergent referents; Table 5.5).

This framing aligns with elicitation literature showing that some commands naturally elicit shared solutions while others remain diverse (Wobbrock et al. 2009, Ruiz et al. 2011).

Consideration 1.2: Use dominance structure as a cautious heuristic for "how many defaults"

Sub-question 2 explicitly classifies referents by dominance structure (Table 5.2) and shows that agreement scores alone do not capture execution diversity (Tables 5.1–5.3).

Possible Design Consideration: Dominance structure can be used as a *heuristic* (not a rule) to decide whether it is reasonable to propose one default, two defaults, or no default: single-dominant → one default; co-dominant → two defaults; no-dominant/bimodal → avoid claiming a population-level default and instead rely on templates or intent-based acceptance. The benefit is that the design stance is anchored in the observed distribution (Table 5.2), instead of designer preference.

Consideration 1.3: When form stays diverse, anchor the mapping on spatial intent (L1)

Several referents show low whole-gesture agreement (Table 5.1; Figure 5.1), yet strong L1 dominance (Table 5.5; Figure 5.6). For instance, Side Step In/Out are form-diverse but show clear toward-screen / toward-self intent dominance (Table 5.5).

Possible Design Consideration: In such cases, a practical design stance is to treat **L1 spatial intent** as the primary anchor and allow variation in L2 motion primitives and articulation, as long as the intent cue remains stable. This matches prior observations that consensus can appear in movement parameters even when detailed gesture form differs (Ruiz et al. 2011, Vatavu 2019).

Consideration 1.4: Make the reference frame explicit because navigation is an indirect-mapping stress test

In Sub-question 1, navigation gestures were predominantly small, controller-like actions organized around neutral/home references and opposite pairs, not locomotion reenactments (Figure 4.6). Participants explicitly raised mirrored ambiguity concerns during navigation mapping.

Possible Design Consideration: If a gesture set relies on "left / right / forward / back," the design may need to explicitly define the reference frame (screen-relative vs. body-relative) and keep it consistent. This is especially

relevant for navigation, where the mapping is already indirect, and where confusion flags were most frequent in the speed task (Table 6.6).

Consideration 1.5: Provide multiple templates for referents that invite different mental models

Sub-question 1 shows that some moves (notably Neutron Bomb) invite mixed representational strategies and gesture collisions, with participants iterating when overlap occurs. Sub-question 2 confirms that Neutron Bomb supports two coherent but competing families (Tables 5.2–5.3).

Possible Design Consideration: For such referents, instead of forcing one metaphor, the system could offer multiple gesture templates (e.g., motion-shape sketch vs. symbolic cue) and let the player select the one that matches their reasoning. This follows the user-defined gesture principle of designing around participant mental models (Wobbrock et al. 2009).

2. Considerations for gesture–command mapping

Consideration 2.1: Under time pressure, throughput constraints dominated breakdowns

In Sub-question 3, overall strict on-time correctness was 41.5% while intention-correct was 66.7%, indicating many "failures" were correct gestures that arrived too late in the prompt window (Sub-question 3 overview). Across speed zones, the overload drop is driven mainly by late-correct and no-response, while misinput stays in a narrower band (Table 6.2). The tipping point begins around ≈2.0 s prompt intervals (Table 6.9; Figures 6.6–6.7).

Possible Design Consideration: In this dataset, the dominant constraint under speed was not gesture confusability but **time-to-execute and time-to-switch**. A cautious design hypothesis is therefore: for time-critical actions, prefer gestures with short travel and compact structure, and treat "fast completion" as a first-order constraint before optimizing expressiveness or distinctiveness. The broader speed–accuracy tradeoff literature supports the general idea that performance degrades under deadlines (Heitz 2014).

Consideration 2.2: "Late-correct" reflects a timing mismatch in the protocol, but points to game-relevant window sensitivity

Sub-question 3 defines late-correct as "intended gesture produced, but registered after the prompt interval," which is specific to the paced prompting and WoZ loop (Sub-question 3 definitions). Late-correct becomes dominant in overload (Table 6.2), and the gap between strict and intention-correct widens as intervals shrink (Figures 6.4–6.6).

Possible Design Consideration: The exact late-correct phenomenon is protocol-specific (prompt pacing). However, the underlying pattern still suggests that gesture control becomes fragile when the interaction demands **tight timing windows**. In fighting game terms, this corresponds more closely to "missing the actionable window" (e.g., punish timing) rather than choosing the wrong move. A prototype informed by this result could explore timing supports that are already common in fighting games, such as input buffering or early commitment, while keeping the design claim cautious: this study motivates these mechanisms as candidates to test, not as confirmed requirements (Claypool & Claypool 2006).

Consideration 2.3: Feedback should confirm recognition quickly to support chaining and planning

In Sub-question 4, sense of control in Experiment 2 became conditional on timing, switching, and feedback responsiveness (SC results; Table 7.3). In Sub-question 3 interviews, participants described loss of flow when the pace outstripped what they could execute and interpret, especially at the extreme tail (Sub-question 3 qualitative notes).

Possible Design Consideration: A reasonable hypothesis from these results is that gesture systems for fast play may benefit from immediate, unambiguous confirmation that an input was registered (or queued), independent of long move animations. This aligns with classic response time guidance in HCI about keeping interactions responsive (Nielsen 1993), but the main justification here is the observed control fragility under speed in Experiment 2 (Sub-question 4).

Consideration 2.4: Navigation deserves extra protection because it was both frequent and error-sensitive

Navigation was the clearest indirect-mapping stress test in Sub-question 1 (navigation results), and in Sub-question 3 it showed the highest rate of

confused-navigation flags (Table 6.6). Navigation also degraded earlier than single-hit moves in the speed curves (Figures 6.4–6.5).

Possible Design Consideration: In this dataset, navigation appears to be a weak link under time pressure. One cautious design direction is to make navigation mappings maximally spatially compatible and easy to parse, reducing translation effort. Compatibility research provides general support for faster responses when mappings align with spatial cues (Ambrosecchia et al. 2015).

3. Experience-related observations (control, embodiment, comfort)

Consideration 3.1: Embodiment appeared as "partial fit," not full-body role-play

In Sub-question 1, even direct attacks were commonly enacted in reduced form, keeping "essence cues" rather than full realism (single attacks results). In Sub-question 4, Mental Imagery & Embodiment (MIE) was mostly Mixed in Experiment 1 and shifted toward Mixed/Low in Experiment 2 (Table 7.1), often becoming fragile under speed and attention shifts.

Possible Design Consideration: A cautious takeaway is to design for embodiment as a spectrum: allow gestures to communicate meaning through partial cues (direction, impact, trajectory) rather than requiring full physical simulation. This fits immersion work treating engagement as multi-component rather than a single "immersed/not immersed" state (Brown & Cairns 2004, Cairns et al. 2019).

Consideration 3.2: Sense of control depended on stable, predictable response timing

In Sub-question 4, Sense of Control (SC) was often High in Experiment 1 (7 High, 3 Mixed) but shifted toward Mixed in Experiment 2 (Table 7.1). Participant reports of reduced SC in Experiment 2 frequently cite timing thresholds, switching overhead, and pacing mismatch (SC results; Table 7.3). Timing-related disruptions also appear in the disruption taxonomy (animation pacing mismatch; Table 7.4).

Possible Design Consideration: These results suggest that perceived control is supported when the system responds quickly and consistently, and becomes fragile when timing drifts or outcomes feel unpredictable. A prototype informed by this study should therefore prioritize low end-to-end latency and consistent recognition behavior, so that failures (when they happen) are repeatable and learnable rather than "random." This is consistent with accounts linking control/involvement to deeper engagement (Brown & Cairns 2004, Birk & Mandryk 2013).

Consideration 3.3: Comfort and gesture compression looked like natural adaptation under speed

In Sub-question 3 interviews, some participants reported making gestures smaller or dropping components as speed increased (Sub-question 3 qualitative notes). In Sub-question 4, physical discomfort/strain is the most frequent disruption category (9 participants; 17 instances; Table 7.4), and IPI in Experiment 2 becomes more conditional on travel, switching, and awkward orientations (IPI results).

Possible Design Consideration: A cautious design hypothesis is that gesture systems should *expect* compression and micro-adaptations under time pressure and repetition. Supporting compact variants as legitimate (rather than as "incorrect") may help preserve both performance and comfort, particularly for high-frequency actions. Prior work on body movement and engagement supports treating physical design as a first-order factor for experience (Bianchi-Berthouze et al. 2007, Nadia Bianchi-Berthouze & Bianchi-Berthouze 2012).

Consideration 3.4: Prompts were an evaluation scaffold; the transferable insight is about attention and overhead under speed

The prompt stream is specific to Experiment 2's speed-ramp method (Sub-question 3 methods/definitions). In Sub-question 3 qualitative notes and Sub-question 4 MIE results, several participants reported attention shifting away from the avatar toward prompts/icons or "doing the correct gesture," especially at the extreme tail (Table 7.4 includes "attention shift away from avatar").

Possible Design Consideration: Because prompts are not part of normal fighting game play, the prompt-following effect should be interpreted as a property of the evaluation scaffold, not as a claim about real matches.

The transferable design insight is narrower: under high time pressure, **extra attention overhead** (whether from prompts, complex cues, or uncertain recognition) can disrupt imagery and perceived control. If prompts are used at all, they fit best as a training or evaluation tool (e.g., drills, speed ramps), and should be designed to fade out once the player can self-initiate inputs.

4. Prototype-level organization considerations

Consideration 4.1: A two-layer architecture follows directly from the convergence patterns

Sub-question 2 separates referents into (a) core-default candidates with consensus, (b) polarized co-dominant cases, and (c) intent-convergent but form-divergent cases (Sub-question 2 summary; Tables 5.1–5.5).

Possible Design Consideration: A prototype could mirror this structure by separating a **core backbone** (moves with stronger consensus) from a **flexible layer** (moves where the study suggests multiple stable solutions or weaker structure). This is a direct way to translate the Sub-question 2 typology into implementable design choices, while keeping the claim cautious.

Consideration 4.2: Onboarding effort can be proportional to the convergence regime

Sub-question 1 indicates that some participants developed a control grammar early (neutral positions, polarity rules), while Sub-question 2 shows that not all moves share the same convergence profile (Tables 5.2–5.5).

Possible Design Consideration: A practical hypothesis is to tailor onboarding by policy: tight defaults can be taught as "the standard form," template defaults can be taught as "the idea + allowed variants," co-dominant cases can present two first-class options, and flexible-intent cases can emphasize the spatial goal rather than exact articulation. This keeps learning demands aligned with how structured the mapping actually appeared in this dataset.

Consideration 4.3: Success criteria may differ for defaults versus intent-based acceptance

Sub-question 2 shows cases where shared L1 intent exists without shared full-form consensus (Table 5.5), which implies that judging success only by

exact form would misrepresent shared structure.

Possible Design Consideration: For "tight default" moves, success can be framed as form reliability; for intent-based moves, success can be framed as meeting intent constraints (especially direction), even if articulation varies. This is compatible with agreement analysis work arguing that consensus can be captured beyond strict form matching (Vatavu 2019).

Consideration 4.4: Combos were represented as either sequencing or compression

In Sub-question 1, multi-hit strings were predominantly compound (65%), with participants describing three recurring strategies: composing sequences, compressing into a shortcut, or re-timing to match animation (multi-hit results). Sub-question 3 and Sub-question 4 also highlight cognitive load for combos/sequences as a disruption category (Table 7.4).

Possible Design Consideration: A prototype may benefit from supporting both: (1) **sequencing** (gesture tokens with timing rules) and (2) **macros** (single shortcut gestures), because both strategies appear in participant reasoning and relate differently to cognitive load and time pressure.

5. Confusion and interference signals in the results

Consideration 5.1: Keep gesture families separable at the levels that survive speed

In Sub-question 3, misinput does occur (Table 6.2) and interviews mention confusions between similar gesture assignments (Sub-question 3 qualitative notes), though misinput does not spike as sharply as timing failures in overload. Sub-question 4 also includes "gesture similarity & interference" as a disruption category (Table 7.4).

Possible Design Consideration: A cautious takeaway is that family design should preserve distinct cues even under compression and switching. Practically, that means separability at robust layers (especially spatial intent and gross primitives), not only in fine articulation that may disappear under speed.

Consideration 5.2: Neutral home positions and explicit polarity rules can reduce directional ambiguity

Sub-question 1 describes repeated participant use of neutral reference points and opposite pair logic for navigation (navigation results), and Sub-question 3 shows navigation as the highest "confused-navigation" group (Table 6.6).

Possible Design Consideration: A reasonable prototype heuristic is to formalize a neutral home posture and define opposite pairs explicitly (forward/back, in/out), especially for navigation. This directly mirrors what participants constructed informally in Sub-question 1 and targets the most frequent confusion signal seen in Sub-question 3.

Consideration 5.3: Avoid treating one canonical execution as "the gesture" when within-family variation is high

Sub-question 2 shows that some dominant families are internally coherent (e.g., Left Punch; Table 5.3), while others have substantial within-family dispersion (e.g., Uppercut; Table 5.3). Sub-question 3 and Sub-question 4 also show that speed and posture changes can cause drift (Table 7.4 includes misrecognition/system interpretation and switching overhead).

Possible Design Consideration: Where within-family dispersion is high, it may be more realistic to document "acceptable variation" (template stance) rather than enforcing a single canonical execution. This reduces the risk of penalizing executions that still match the shared intent and family concept (Wobbrock et al. 2009, Vatavu 2019).

Synthesis: Summary of testable design hypotheses

Across themes, the results support a three-layer lens for designing and evaluating prototypes:

- **Physical layer:** gesture travel, switching overhead, and comfort (Sub-question 3 time constraints; Table 6.8; Sub-question 4 disruption taxonomy; Table 7.4).
- **Cognitive layer:** mapping structure, intent cues, and reference frames (Sub-question 1 navigation mental models; Sub-question 2 L1 dominance; Tables 5.4–5.5).

- **System layer:** timing behavior, feedback, and consistency (Sub-question 3 tipping point around $\approx$2.0 s; Table 6.9; Sub-question 4 conditional control under speed; Table 7.3).

Importantly, these considerations are not meant as universal prescriptions. They are concrete hypotheses derived from the observed convergence patterns (Sub-question 2) and the specific failure modes surfaced by the speed-ramp method (Sub-question 3), combined with the experience evidence about control, imagery, and disruption (Sub-question 4). Their main value is to provide a structured starting point for building and testing a recognition-backed prototype in future work.

Chapter 9

Answering Main Research Question

My main research question asked:

> What is the potential of hand gestures collected through Wizard of Oz prototyping in the design of a gesture based control system for fighting games?

To answer this, I treat *potential* as something practical. It is not about whether gesture control looks cool in a demo. It is about whether Wizard of Oz elicitation produces design input that is actually useful for building a fighting game control system. Across the four sub questions, my results show that WoZ does more than generate a list of gestures. It reveals

1. how players build meaning when the mapping is indirect,
2. where a shared default gesture is likely to exist and where it is not,
3. what breaks first under speed and time pressure, and
4. which parts of the experience support immersion and which parts become fragile.

Together, these points describe what WoZ gestures can offer to a fighting game controller design, and also what conditions must hold for that offer to remain realistic in play.

WoZ elicitation captures "input design thinking," not just gesture shapes

Sub-question 1 showed that in this study, participants rarely approached gesture control as literal acting. They did not mostly try to "be Paul." Instead, many participants behaved like they were designing a small input language that they could repeat many times in a fight. That shift matters for the main question because a fighting game does not reward one-off expressive movement. It rewards repeatable commands that stay distinct, fast, and consistent under stress. In my data, this indirect mapping pressure changed what participants treated as "natural." Natural often meant quick, easy to repeat, low effort, and easy to tell apart within the set, more than "realistic."
Navigation made this especially clear. Locomotion and side steps do not have an obvious one-hand analogue. Participants in this study often solved that by borrowing control metaphors from familiar devices, like keyboard arrows, WASD, joystick return-to-neutral, or mouse clicking. They also tried to create grammar rules for the set, like a neutral home position and opposites, and they worried about mirrored gestures causing recognition confusion. Even though the WoZ setup did not implement recognition, participants still reasoned about recognition limits. That is valuable for design because it shows what users expect to be brittle, and what they expect to be robust, even before an algorithm exists.
For attacks, participants had more room for depiction, but they still compressed. They often aimed for what I described as minimum viable depiction. They kept one or two cues that carry the essence of the move, like "from below" for uppercut, "downward strike" for hammer, or thumb as a stand-in for a leg in kicks. This shows how players can stay motivated and "move-like" without requiring full body reenactment. It also fits the reality that a fighting game asks for frequent repetition, and that large mid-air gestures can cause fatigue.
Neutron Bomb acted as a second stress test. It is ambiguous. Participants in this study interpreted it in different ways, like motion shape, special move effect, icon style shortcut, or sequence. That ambiguity often triggered redesign and collision management with other gestures. This reveals an important part of the "potential" story. WoZ elicitation can expose not only stable mappings, but also where the move concept itself invites multiple valid interpretations. In those cases, the design problem is not "find the one correct gesture." It is "support more than one reasonable mapping strategy."
So the first part of the main answer is this: **WoZ elicitation captures how players turn fighting game actions into a usable command space.**

It shows the constraints players care about, the rules they invent to stay consistent, and the cue selection logic they use to make indirect mappings feel motivated. That kind of data is exactly what a designer needs when they want to build a real vocabulary, not just collect gesture examples.

WoZ data can reveal both shared defaults and structured diversity

Sub-question 2 addressed whether player behavior reveals a convergent core set or a broad divergent set. In this study, the answer was not "one or the other." Convergence appeared in multiple forms. Sometimes participants converged on a dominant gesture family for a move. Other times, whole gestures stayed diverse, but participants still tended to converge on what the gesture should express, especially at the spatial intent level. That matters because it means low agreement at the full-gesture level does not automatically mean the space is random.

Two specific ideas from sub-question 2 matter for the main research question. First, a "default" can exist in different strengths. In this study, some moves showed a tight default, where participants not only chose the same family but also produced very similar executions. Other moves showed a template default, where participants seemed to converge on the idea of the gesture family, often held together by shared spatial intent, while varying widely in the exact mechanics. This distinction is important because it changes what "core set" means. A core set is not always a single standardized movement. Sometimes it is a stable concept with flexible realizations.

Second, divergence at the family level often hid component-level structure. In this study, L1 spatial intent tended to be the most stable layer across many moves, even when gestures fragmented into many families. This suggests that even when people disagree on packaging, they may still agree on the spatial target of the action. For a fighting game controller, that is a useful anchor because spatial intent often aligns with what a move means in play, like toward opponent, away, up, down, or toward screen.

This gives the second part of the main answer: **WoZ elicitation can identify where standardization is plausible and where flexibility is more realistic, and it can do that without reducing everything to one agreement number.** It shows whether a move supports one shared default, two stable competing solutions, a shared template, or mainly shared intent under varied execution. That is a strong kind of "potential," because it turns gesture collection into a structured map of the design space.

Speed pressure turns gesture control into a deadline problem

Sub-question 3 tested usability and reliability under speed using a ramp that pushed each participant's own gesture set from comfortable pacing into time pressure. The key result is that strict on-time reliability dropped mostly because of timing misses, not because participants forgot their mappings or mixed gestures up. In other words, the limit looked like a deadline problem. As the prompt interval shrank, the system stopped giving enough time for the human loop to complete. In this setup, reaction time also included the Wizard step, so it captured an end-to-end loop, not only hand motion. Tight deadlines amplify every delay in that loop.

This is where "late-correct" matters. In this experiment, late-correct meant that the participant did the intended gesture, but the system logged it after the next prompt had already appeared. So late-correct is not a meaning failure. It is a timing failure. When prompts arrive before a gesture can realistically finish, strict on-time correctness becomes physically unrealistic. At that point, late-correct becomes evidence of performance ability under speed, but it also signals that the pacing no longer matches human execution time. That also explains why late-correct and no-response rise together. Under overload, participants often either finish and become late, or they drop it and skip.

This speed result connects back to sub-question 1 in a useful way. In sub-question 1, participants already compressed gestures toward small and repeatable forms because they anticipated repetition demands. Sub-question 3 shows why that instinct makes sense. Fighting game inputs live inside small timing windows. A gesture can be conceptually right and still fail as a fighting game input if the time budget does not support it, or if the feedback loop becomes messy during chaining.

So the third part of the main answer is this: **WoZ gestures have potential, but speed becomes the filter that decides which ideas can survive in a fighting game context.** The data suggests that reliability limits come mainly from throughput and timing, not from gesture meaning confusion. That means the path from elicitation to implementation must treat timing as a first-class design constraint, not an afterthought.

Immersive potential appears, but it depends on control staying transparent under pressure

Sub-question 4 looked at immersion-related qualities, using the constructs in my framework. In this study, the strongest evidence sat in intuitive physical interaction and sense of control during gesture invention. Many participants described gestures as physically natural and controllable. They often felt the mapping made sense, and they could predict outcomes from their movement. That kind of predictability supports engagement because it reduces the feeling of fighting the interface.

Mental imagery and embodiment showed more variation across participants, and it weakened under speed. In Experiment 2, several participants shifted attention away from the character and toward prompts or their own hands. That shift makes sense because time pressure forces planning and decoding. When the task consumes attention, there is less room for imaginative involvement. This does not mean gestures cannot support imagination. It suggests that imagery is more fragile and more dependent on the surrounding play context, and on whether the interaction stays smooth enough to remain in the background.

The disruption taxonomy helps connect sub-question 4 back to sub-question 3 and sub-question 1. The most common disruptions clustered around physical cost, time coordination, and mapping interference. In other words, immersion did not mainly break because participants suddenly stopped understanding their gesture meanings. It broke when the interface became noticeable due to strain, missed timing, switching overhead, or inconsistent outcomes. This matches a simple idea that runs through the whole thesis. In a fighting game, players need to act quickly, get feedback quickly, and keep their focus on the fight. When the control scheme demands too much attention or effort, it stops feeling like "I am doing the move" and starts feeling like "I am managing a system."

So the fourth part of the main answer is this: **WoZ elicitation can surface gesture mappings that support intuitive action and a strong sense of control, which are key ingredients for immersive potential, but that potential becomes conditional under speed and fatigue.** The gestures can support engagement when the mapping feels direct and the system loop stays stable. When the loop becomes inconsistent or demanding, the immersive layer weakens.

Answer to the main research question

Based on the patterns observed in my study, **hand gestures collected through Wizard of Oz prototyping show strong potential as design material for building a gesture-based control system for fighting games**, because the method reveals more than a list of gesture ideas. It reveals the mapping logic players rely on when the avatar mapping is indirect. It shows where a shared default gesture concept is likely to form and where diversity remains structured. It also exposes timing and speed limits that affect what can stay reliable at fighting game pace. And it highlights which parts of the experience support intuitive action and a sense of control, and which parts become fragile when the interaction adds effort, confusion, or delay.

Chapter 10

Limitations & Future Work

Limitations of my work

I interpret these results with care, mainly because of the study scale and the way the Wizard of Oz setup shapes what the data can show.

Sample size and action scope

The participant pool was small (N = 10) and the studies covered a limited set of fighting game actions. Because of this, I do not treat the exact rates or dominance patterns as something that will hold for all players. I treat them as signals about what can happen, and what design pressures show up, within the scope of this study.

Wizard of Oz and task structure

Both experiments used Wizard of Oz control, and parts of the work relied on prompt driven tasks. This reduces ecological validity. Participants did not play a full competitive match, and they did not have to make real tactical decisions against an opponent. This matters most for immersion findings, since the interaction context was more controlled and more "task-like" than real fighting game play.

Timing and the extreme speed conditions

In sub-question 3, reaction time reflects the full Wizard of Oz loop, not just participant movement. At the fastest prompt intervals, the time budget sometimes became unrealistic, where prompts could arrive before a gesture

or even the on-screen move could reasonably finish. For that reason, I interpret overload-zone outcomes as a stress test that highlights pacing limits and control-loop fragility, not as a precise measure of human maximum performance.

Gesture coding as an abstraction

Some findings depend on how gestures are described and grouped into families and layers. That abstraction is necessary to make sense of diverse gestures, but it also means another researcher could draw boundaries slightly differently. I therefore rely most on patterns that show up across multiple lenses, such as stable intent even when surface form varies.

Immersion evidence relies on self report

Finally, immersion-related conclusions rely mainly on participant self report. Participants explained their experiences clearly, but self report does not fully capture moment-to-moment engagement. I therefore interpret sub-question 4 as evidence about immersive potential during prototype-stage interaction, not as a final verdict on immersion in competitive play.

Future Work & Recommendations

This thesis shows what Wizard of Oz gesture elicitation can tell us in a fighting game setting. The next steps are about checking how well these ideas hold up when the system is real, the timing rules are strict, and the player is under real match pressure. I also think other researchers can reuse the method and the analysis approach in several nearby domains.

Implement and test a real system

A direct next step is to build an actual gesture recognizer and repeat the same core tasks. Wizard of Oz helps collect good mapping ideas, but it avoids key implementation choices. A real system forces decisions about recognition thresholds, latency, feedback timing, and what the system does when input is unclear. It also lets future work separate human execution limits from system delay, which makes reliability results easier to interpret.

Re-test speed with fighting game timing rules

My speed results mainly point to timing pressure as the bottleneck. Future work should repeat the speed ramp without a wizard and test timing rules that match fighting games, such as buffering and early commitment. This would help answer a practical question: when players perform the intended gesture, what timing window makes the system feel responsive and fair, without accepting random noise.

Test gestures in real play, not only prompt streams

Another step is to move beyond prompt-following tasks and test the gestures in game-like play. This can start with training drills and then move to short sparring and full matches. The goal is to see whether the gesture set still feels manageable when the player must also watch the opponent, make tactical decisions, and deal with pressure. This is also the context where immersion claims become stronger, because the player is responding to the game state, not just to prompts.

Study learning and long-term stability

My work captures first-time invention and short-term performance. Future work should test what happens with practice. A longer study can show whether gestures become faster, more consistent, and more compact over time. It can also show whether early diversity settles into stable personal habits. This matters because a gesture controller only becomes valuable if it supports fluency, not just novelty.

Expand the move set and stress reference-frame issues

Future work should include more move types, especially defensive actions, stance-related actions, and multi-step sequences. It should also stress side switching more directly, since direction and left-right logic can become unstable when the character changes sides. A mapping can look good in elicitation but still fail in match reality if it depends on a fixed viewpoint.

Chapter 11

Conclusion

This thesis started from a simple tension. Traditional controllers work well for speed and precision, but they force players to learn an arbitrary mapping between small physical actions and dramatic on-screen moves. Gesture control promises a closer link between intention, body movement, and game action, but real systems face limits around recognition, segmentation, fatigue, and time-critical reliability. Because of that, gesture controllers often end up small, pre-defined, and based on designer intuition rather than evidence.

To work through that tension, I used Wizard-of-Oz prototyping to study gesture-based control as a design space before committing to automatic recognition. I also used fighting games as a high-pressure testbed, not because the contribution is only about Tekken, but because this genre makes timing, mistakes, and control feel impossible to hide. If a gesture idea cannot survive here, it is unlikely to survive in other fast, command-heavy interactive contexts.

The main takeaway is that WoZ-elicited hand gestures have practical potential for designing a fighting game gesture controller, but not in the simplistic sense of "people will naturally act out moves." The potential sits in what the method reveals about control design. WoZ exposes how players turn indirect game actions into a usable command space, what kinds of defaults are likely to emerge, where diversity remains but still has structure, and where performance pressure becomes the hard filter that decides what ideas can realistically work in play. In other words, WoZ does not just collect gesture shapes. It produces design evidence about meaning, consistency, speed limits, and the experience conditions that keep the controller feeling transparent rather than demanding.

At the same time, these conclusions should be read in the scope of this study. I worked with a small participant pool, so the results do not claim population-level truths about how all players will design or sustain gestures.

I therefore treat the findings as patterns observed under my specific study setup, task framing, and gesture set. Where I draw broader implications, I do so cautiously and with the understanding that they remain grounded in this dataset. The value of the work is not that it "proves" one universal gesture vocabulary, but that it shows how WoZ can surface structured tendencies and practical constraints that can guide future design and validation.

This matters because it reframes gesture control away from novelty and toward viability. A gesture set can be expressive and still fail if it cannot meet timing windows, if it causes strain under repetition, or if it pulls attention away from the actual game. So the path forward is not to chase "naturalness" as literal reenactment, but to treat gesture control like any other serious input method: it must be learnable, repeatable, distinct within a set, and stable under time pressure.

In the end, this thesis shows that Wizard-of-Oz prototyping is a strong early-stage method for gesture controller design in fast interactive systems. It helps a designer separate what is promising from what is fragile, and it does that before investing in complex sensing and recognition pipelines. That is the core contribution of this work: a grounded way to explore gesture control as a real controller option, with clear constraints, clear opportunities, and clear design direction.

Appendix A

Referent to Factor Index

Table A.1: Referent–factor links derived from video and interviews

Factor	Referent	# Participants	Participants
Control and flow interruption	Backward	1	P003
	Forward	1	P003
	Hammer	1	P003
	Left_Kick	1	P003
	Left_Punch	1	P003
	SS–In	1	P003
	SS–Out	1	P003
	Uppercut	1	P003
Fatigue and comfort	Backward	2	P003, P010
	Neutron_Bomb	2	P003, P010
	Forward	1	P003
	Hammer	1	P001
	Left_Punch	1	P006
	Phoenix	1	P010
	SS–In	1	P003
	SS–Out	1	P003
Hesitation and recall pauses	Neutron_Bomb	2	P001, P003
	Uppercut	2	P003, P006
	Hammer	1	P010
	Hangover	1	P001
	Left_Punch	1	P003
	Phoenix	1	P010
	SS–In	1	P010
Navigation prompt confusion	Backward	2	P003, P006
	Forward	2	P003, P006
	SS–In	2	P003, P006
	SS–Out	2	P003, P006
	Hammer	1	P003
	Left_Kick	1	P003
	Neutron_Bomb	1	P003
	Ph_Smasher	1	P003
	Phoenix	1	P003
	Uppercut	1	P003
Simplification and form drift	Forward	2	P003, P010
	Hammer	2	P001, P006
	Backward	1	P010
	Neutron_Bomb	1	P006
	Phoenix	1	P010
	SS–In	1	P003
	SS–Out	1	P003
	Uppercut	1	P006
Other (contextual or mixed factors)	Phoenix	4	P001, P003, P006, P010
	Uppercut	3	P001, P003, P006
	Backward	2	P001, P006
	Forward	2	P001, P006
	Left_Kick	2	P001, P010
	Ph_Smasher	2	P001, P003
	SS–In	2	P006, P010
	SS–Out	2	P003, P006
	Hammer	1	P010
	Hangover	1	P001
	Left_Punch	1	P003
	Neutron_Bomb	1	P006

Appendix B

Breakdown Signals and coded flags - Sub-question 3

Table B.1: Referent-linked breakdown signals across all speed conditions (Experiment 2)

Referent	**n**	**Confused navigation** (%)	**Hesitation** (%)	**Breakdown** (%)
Backward	90	9 (9.9)	1 (1.1)	2 (2.2)
Forward	90	12 (13.3)	1 (1.1)	0 (0.0)
Hammer	91	1 (1.1)	4 (4.3)	4 (4.3)
Hangover	93	3 (3.3)	4 (4.4)	2 (2.2)
Left_Kick	91	1 (1.1)	2 (2.2)	2 (2.2)
Left_Punch	91	2 (2.2)	2 (2.2)	2 (2.2)
Neutron_Bomb	92	7 (7.6)	5 (5.4)	0 (0.0)
Phoenix_Smasher	92	5 (5.4)	2 (2.2)	2 (2.2)
Phoenix	92	2 (2.2)	2 (2.2)	2 (2.2)
Sidestep-In	93	8 (8.6)	1 (1.1)	1 (1.1)
Sidestep-Out	93	13 (14.4)	2 (2.2)	1 (1.1)
Uppercut	91	3 (3.3)	3 (3.3)	0 (0.0)

Table B.2: Outcome distribution by referent pooled across participants (Experiment 2)

Referent	n	On-time (count)	Late-correct (count)	Mis-input (count)	No-response (count)	On-time accuracy (%)	Intention accuracy (%)
Sidestep-In	93	37	30	7	19	39.8	72.0
Hammer	92	41	23	5	23	44.6	69.6
Neutron Bomb	92	46	25	9	12	50.0	77.2
Phoenix Smasher	92	32	22	13	25	34.8	58.7
Phoenix	92	35	30	7	20	38.0	70.7
Backward	91	34	20	11	25	37.4	59.3
Hangover	91	38	25	8	19	41.8	69.2
Left Kick	91	41	23	9	18	45.1	70.3
Left Punch	91	47	21	7	16	51.6	74.7
Uppercut	91	38	18	9	26	41.8	61.5
Forward	90	36	17	8	28	40.0	58.9
Sidestep-Out	90	34	21	14	21	37.8	61.1

Outcomes by referent (pooled across participants)

Time budget analysis: prompt pacing, gesture duration, and game move duration

Positive values in the last column mean the gesture takes longer than the move animation. Negative values mean the move animation takes longer than the gesture. A side by side comparison can be observed in the Figure B.2.

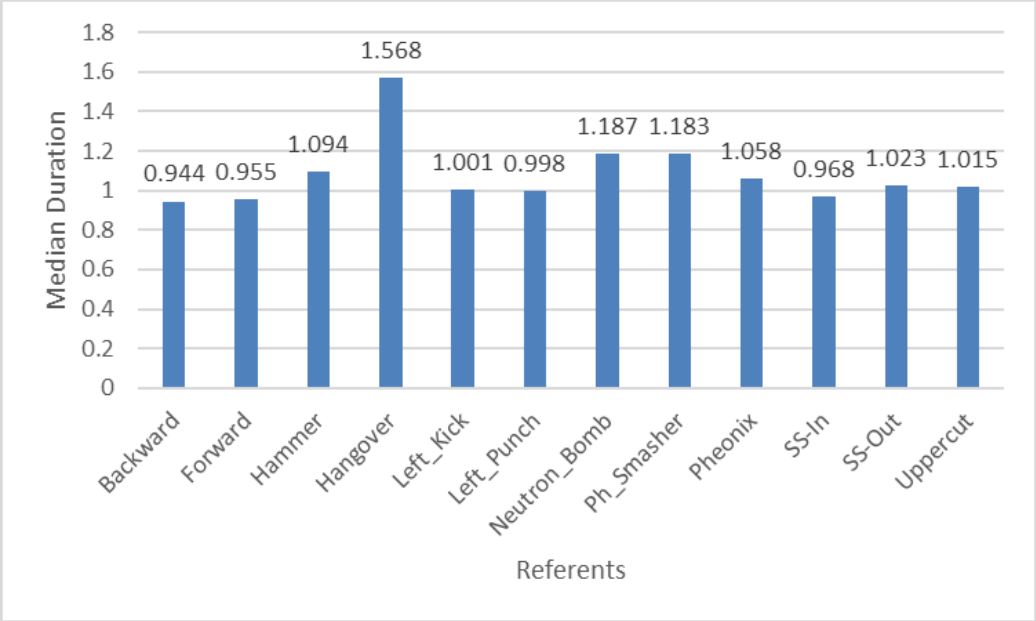


Figure B.1: Participants' Mean Gesture Duration

Table B.3: Median gesture execution time compared to Paul Phoenix move duration (Experiment 2)

Referent	**Paul move duration (s)**	**Gesture median (s)**	**Gesture – Paul (s)**
Backward	0.286	0.944	+0.659
Forward	0.320	0.955	+0.635
Hammer	0.800	1.094	+0.294
Hangover	1.907	1.568	−0.339
Left Kick	0.746	1.001	+0.255
Left Punch	0.420	0.998	+0.578
Neutron Bomb	1.393	1.187	−0.206
Phoenix Smasher	1.567	1.183	−0.384
Phoenix	1.633	1.058	−0.575
Sidestep-In	0.694	0.968	+0.274
Sidestep-Out	0.667	1.023	+0.356
Uppercut	0.720	1.015	+0.295

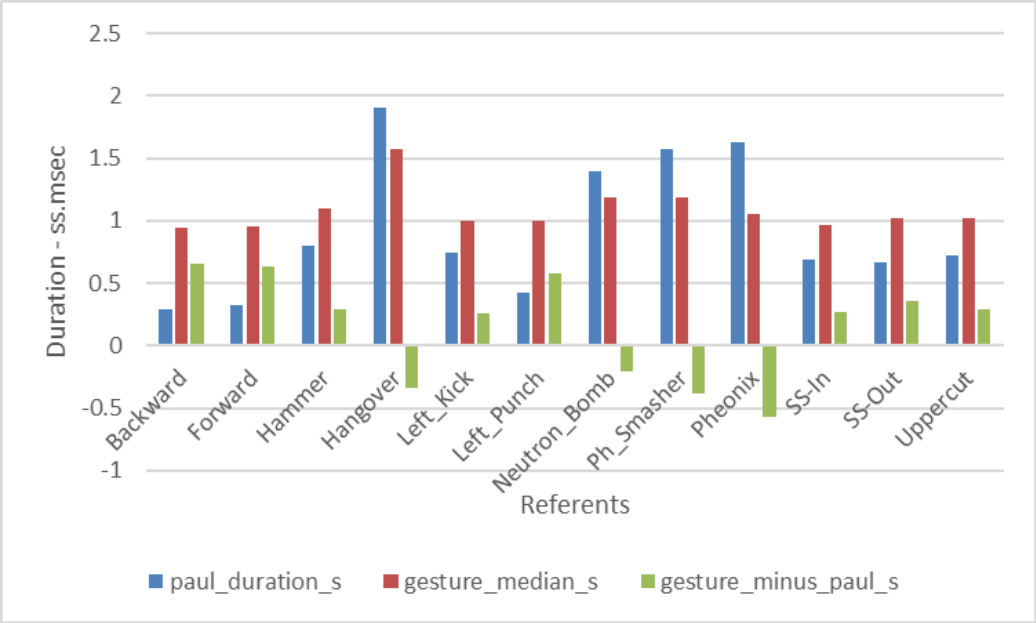


Figure B.2: Gesture and Moves Duration Compared

Appendix C

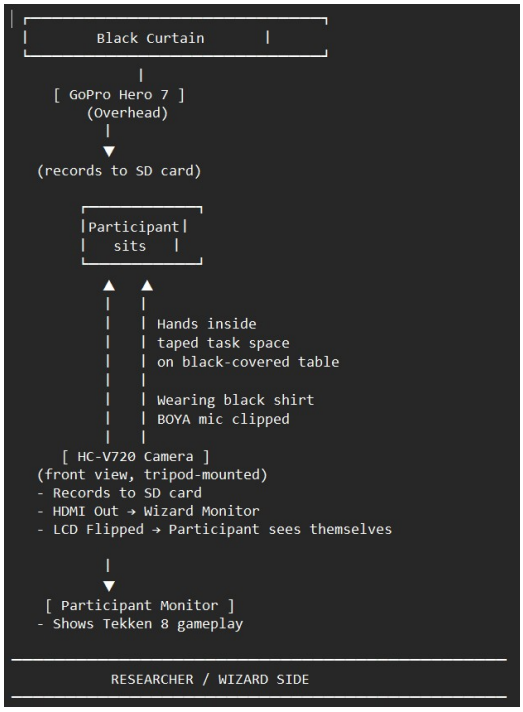
User Study Setup

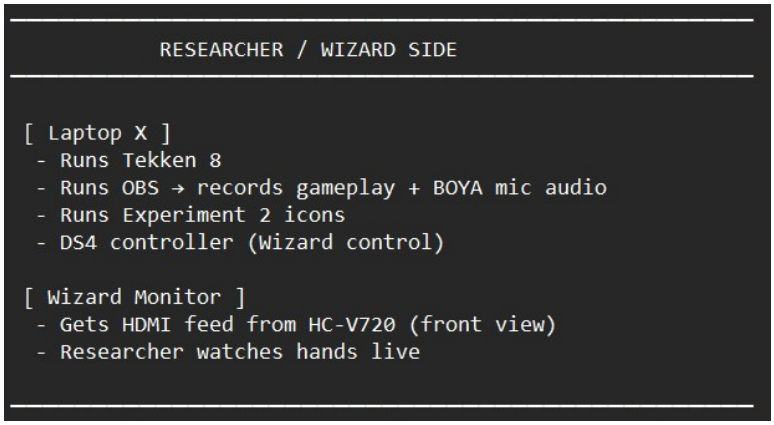
The following document contains the details on the setup of the experiment, both physical as well as the software.

User Study Setup

Wahaj Ahmad

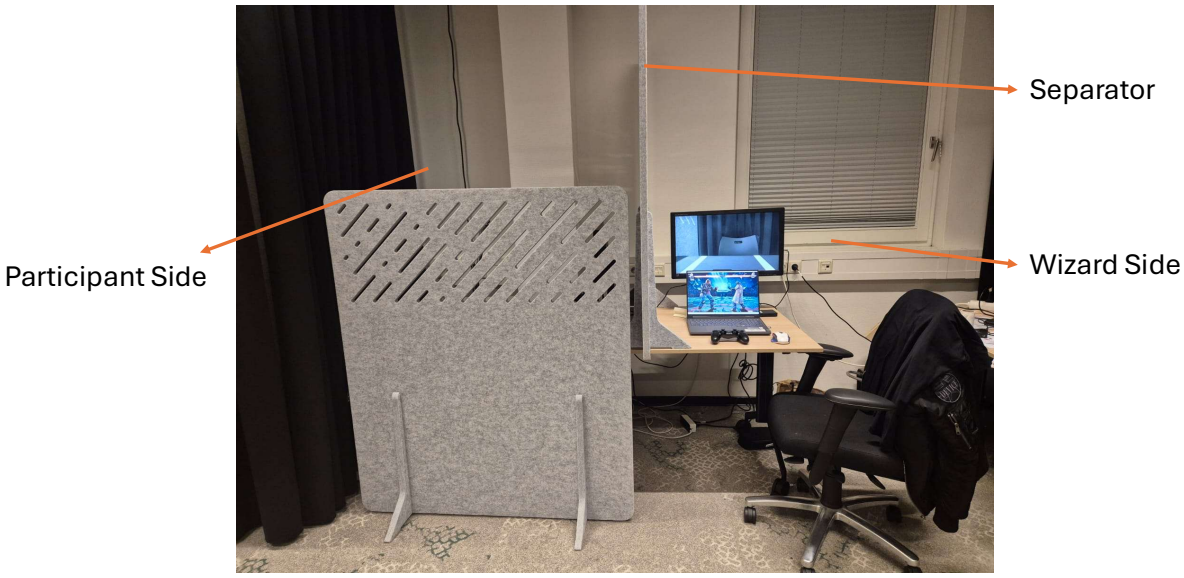
Setup



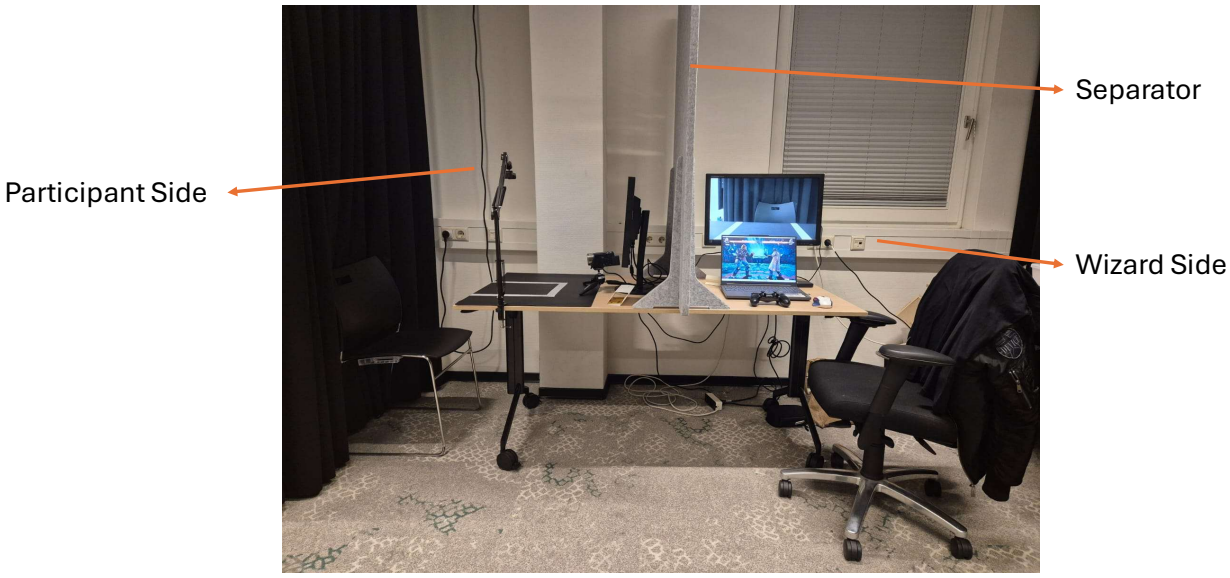


Note: The wizard and researcher are the same person

Complete setup

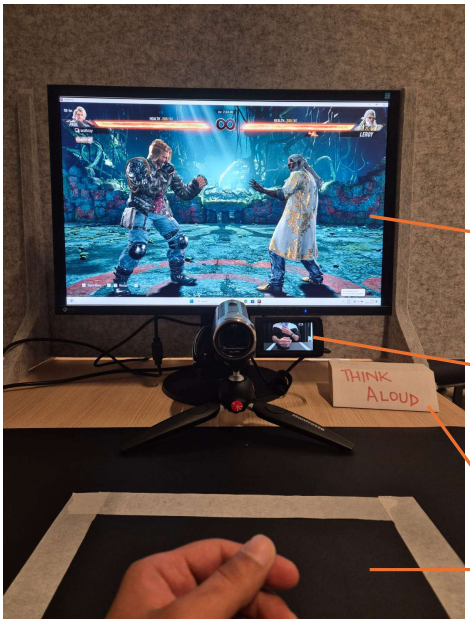


Complete setup (without participant blind)



Participants side in focus

Participant's POV



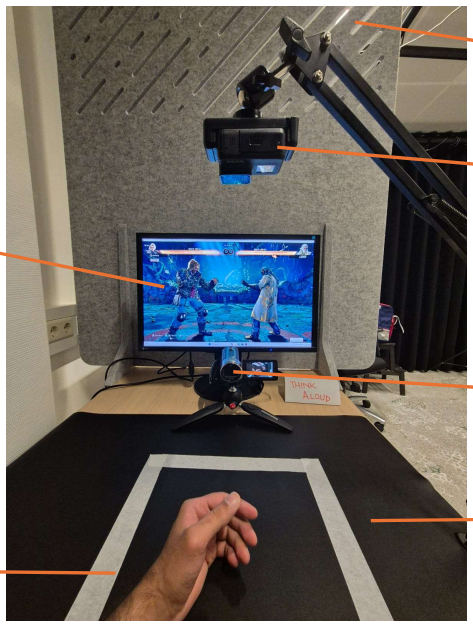
Display

Participant Screen on camera

Soft reminder to think aloud

Task Space

Without zoom or wide lens



Blind/ separator

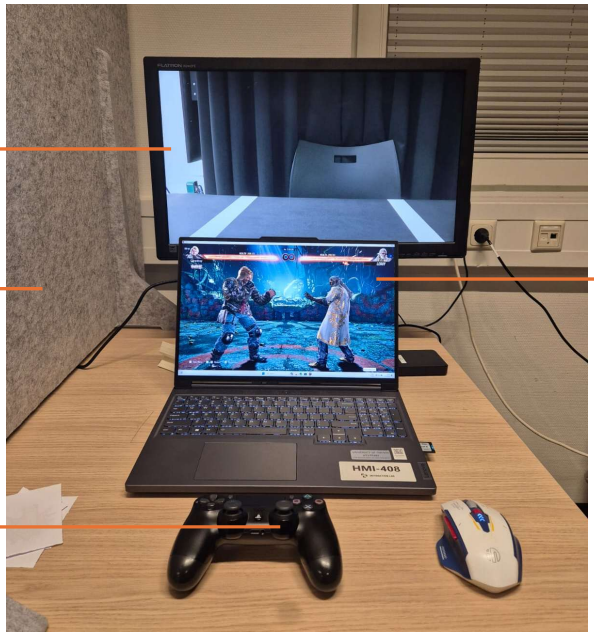
Camera From top

Camera From front

Black mat to increase contrast

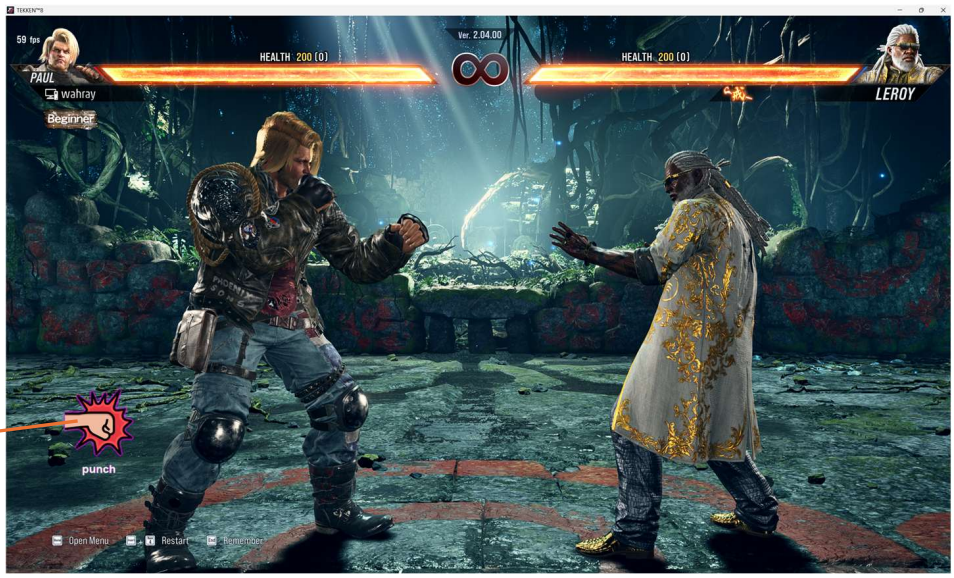
With 0.6x wide zoom

Wizard/Researcher’s POV



Gameplay screen in Experiment 2

Move Prompt
for participant



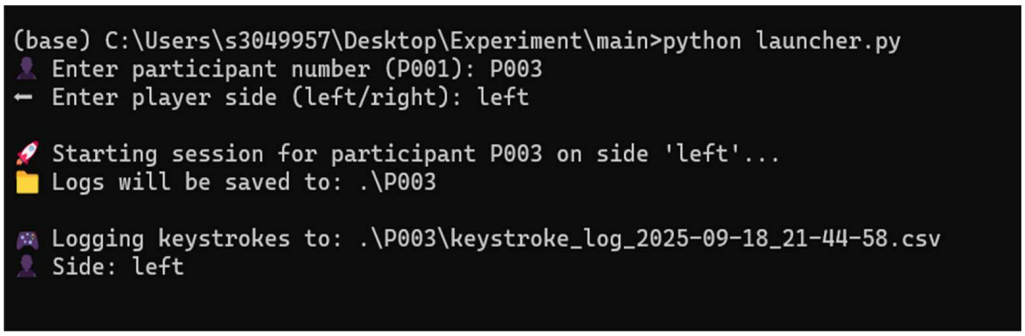
Software Overview

- **Launcher.py** – starts overlay, logger and matcher
  | \
  | \
- **Image_overlay.py**
  - shows move cues
  - saves *session_log.txt*

  Keystroke_logger.py
  - logs wizard inputs
  - saves *keystrokes_log.csv*

- **Match_logs.py**
  - parses both logs
  - creates *matched_log.txt* (also excel)

Launcher.py



- Takes participant number for naming conventions
- Takes player side (to show the right icons)

Session_log.txt

```
Session started at 2025-09-18_17-31-51

[1] Time: 10.8s | Move: left | Display: 5.00s | Wait: 2.00s
[2] Time: 17.8s | Move: right | Display: 4.90s | Wait: 2.00s
[3] Time: 24.7s | Move: ss-b | Display: 4.80s | Wait: 2.00s
[4] Time: 31.5s | Move: ss-f | Display: 4.70s | Wait: 1.90s
[5] Time: 38.1s | Move: punch | Display: 4.60s | Wait: 1.90s
[6] Time: 44.6s | Move: kick | Display: 4.50s | Wait: 1.90s
[7] Time: 50.9s | Move: uppercut | Display: 4.40s | Wait: 1.80s
[8] Time: 57.2s | Move: neutron | Display: 4.30s | Wait: 1.80s
[9] Time: 63.3s | Move: hammer | Display: 4.20s | Wait: 1.80s
[10] Time: 69.2s | Move: phoenix | Display: 4.10s | Wait: 1.70s
```

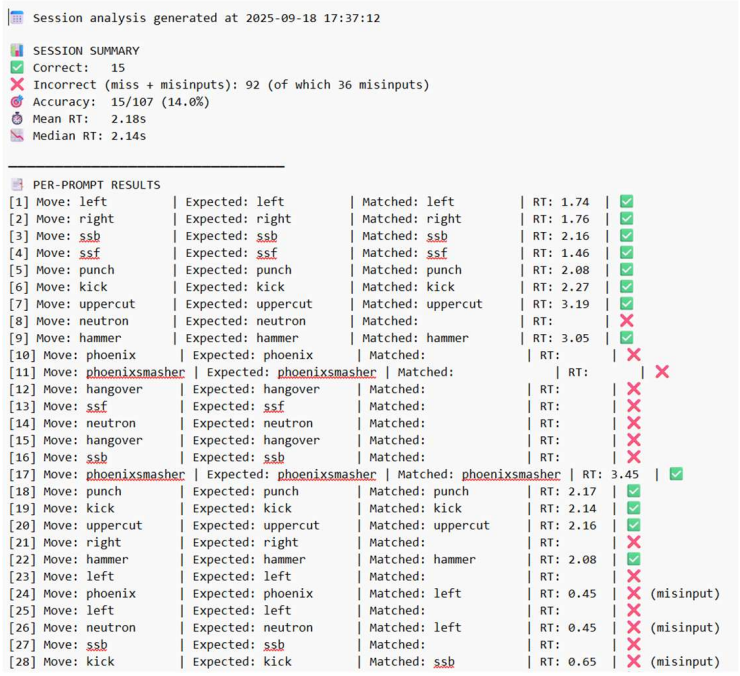
Details:
Time - the time at which the icon is displayed
Move - name of the move for which the icon is displayed
Display – the duration for which the icon is displayed on screen
Wait – the time duration between appearance of the next icon

Keystroke_log.csv

```
Time (s),Key,Gesture
1.977,6,kick
12.545,1,left
19.561,2,right
26.858,3,ssb
32.958,4,ssf
40.178,5,punch
46.867,6,kick
49.762,6,kick
54.094,8,uppercut
63.246,9,neutron
66.350,0,hammer
74.661,8,uppercut
80.506,=,phoenixsmasher
96.461,9,neutron
101.024,+,hangover
106.065,3,ssb
110.751,=,phoenixsmasher
114.371,5,punch
119.041,6,kick
```

Details:
Time - the time at which the input is detected
Key – the identity of the key pressed
Gesture – the move ties to the key press

matched_log.txt (also excel)



```
Session analysis generated at 2025-09-18 17:37:12

SESSION SUMMARY
Correct:   15
Incorrect (miss + misinputs): 92 (of which 36 misinputs)
Accuracy:  15/107 (14.0%)
Mean RT:   2.18s
Median RT: 2.14s

--------------------------------------------------

PER-PROMPT RESULTS
[1] Move: left         | Expected: left         | Matched: left         | RT: 1.74  | ✅
[2] Move: right        | Expected: right        | Matched: right        | RT: 1.76  | ✅
[3] Move: ssb          | Expected: ssb          | Matched: ssb          | RT: 2.16  | ✅
[4] Move: ssf          | Expected: ssf          | Matched: ssf          | RT: 1.46  | ✅
[5] Move: punch        | Expected: punch        | Matched: punch        | RT: 2.08  | ✅
[6] Move: kick         | Expected: kick         | Matched: kick         | RT: 2.27  | ✅
[7] Move: uppercut     | Expected: uppercut     | Matched: uppercut     | RT: 3.19  | ✅
[8] Move: neutron      | Expected: neutron      | Matched:              | RT:       | ❌
[9] Move: hammer       | Expected: hammer       | Matched: hammer       | RT: 3.05  | ✅
[10] Move: phoenix     | Expected: phoenix      | Matched:              | RT:       | ❌
[11] Move: phoenixsmasher | Expected: phoenixsmasher | Matched:           | RT:       | ❌
[12] Move: hangover    | Expected: hangover     | Matched:              | RT:       | ❌
[13] Move: ssf         | Expected: ssf          | Matched:              | RT:       | ❌
[14] Move: neutron     | Expected: neutron      | Matched:              | RT:       | ❌
[15] Move: hangover    | Expected: hangover     | Matched:              | RT:       | ❌
[16] Move: ssb         | Expected: ssb          | Matched:              | RT:       | ❌
[17] Move: phoenixsmasher | Expected: phoenixsmasher | Matched: phoenixsmasher | RT: 3.45 | ✅
[18] Move: punch       | Expected: punch        | Matched: punch        | RT: 2.17  | ✅
[19] Move: kick        | Expected: kick         | Matched: kick         | RT: 2.14  | ✅
[20] Move: uppercut    | Expected: uppercut     | Matched: uppercut     | RT: 2.16  | ✅
[21] Move: right       | Expected: right        | Matched:              | RT:       | ❌
[22] Move: hammer      | Expected: hammer       | Matched: hammer       | RT: 2.08  | ✅
[23] Move: left        | Expected: left         | Matched:              | RT:       | ❌
[24] Move: phoenix     | Expected: phoenix      | Matched: left         | RT: 0.45  | ❌ (misinput)
[25] Move: left        | Expected: left         | Matched:              | RT:       | ❌
[26] Move: neutron     | Expected: neutron      | Matched: left         | RT: 0.45  | ❌ (misinput)
[27] Move: ssb         | Expected: ssb          | Matched:              | RT:       | ❌
[28] Move: kick        | Expected: kick         | Matched: ssb          | RT: 0.65  | ❌ (misinput)
```

Details:
Some details about the session.
RT = Reaction time
RT = (time at which input is detected) –
(time at which the icon appears)

Move – the move of the icon displayed
Expected – the gesture expected
Matched – the move matched by the wizard input
RT – reaction time on the input
Check/Cross – represent if the expected input matches the user input

This final output is not perfect. The matching is not very accurate especially at higher speeds. I will further refine the algorithm for this.

Appendix D

Post-Experiment Questionnaire

The following document contains the interview questions asked form the participants after each experiment.

Post Experiment Questionnaire

Experiment 1 – Inventing Gestures:

Intuitive Physical Interaction – (SQ4 – Immersion)

Question 1.1:

“Did the gestures you created feel natural and intuitive—as if you were directly acting out the character’s moves?”
(Or: “Did it feel like your body knew what to do?”)

- If YES:

“What made them feel intuitive or natural to you?”
“Can you give an example of a gesture you thought worked especially well?”

- If NO:

“What felt awkward or unnatural about them?”
“Was there something about the move or feedback that didn’t match your expectations?”

Imaginative and Mental Imagery – (SQ4 – Immersion)

Question 1.2:

“When you were coming up with gestures, were you imagining the character doing the move, or were you just thinking about your hand movements?”

- If CHARACTER-focused:

“Did that mental image help you come up with the gesture?”
“Did it feel like your body was mimicking the avatar?”

- If HAND-focused or neutral:

“Did you find it hard to picture the avatar while doing the gestures?”
“Would seeing the full character animation before inventing gestures help?”

Sense of Control (Agency / Natural Mapping) – (SQ4 – Immersion)

Question 1.3:

“Do you feel like your gestures would give you a strong sense of control over the character’s moves in a full game?”

- If YES:

“What made you feel in control?”
“Did any gestures feel especially well matched to what you wanted the character to do?”

- If NO or uncertain:

“Was there a disconnect between the gesture and what the character would do?”
“What would help improve that sense of control?”

Gesture Preference

Question 1.4:

“Were there any gestures you liked/disliked more than others?”
→ *[Follow-up]*: “What made them better/worse for you?”

Physical Comfort

Question 1.5:

“Did any gesture feel physically awkward or uncomfortable to perform?”
→ [Follow-up]: “Would you change that gesture if you could?”

Experiment 2 – Gestures at Increasing Speed:

Intuitive Physical Interaction – (SQ4 – Immersion)

Question 2.1:

“As the game speed increased, did the gestures still feel natural and intuitive to perform?”

- If YES:

“What helped you keep the gestures smooth or natural even at higher speed?”
“Were there any gestures that became easier over time?”

- If NO:

“Which gestures started to feel awkward or harder to do?”
“Was it due to the speed, the gesture design, or something else?”

Imaginative and Mental Imagery – (SQ4 – Immersion)

Question 2.2:

“At higher speeds, were you still able to picture your character performing the moves as you gestured?”

- If YES:

“Did that mental image help you keep up with the pace?”
“Was there a moment where it stopped working?”

- If NO:

“What got in the way of visualizing the avatar—speed, stress, gesture confusion?”
“Would a better prompt or animation help you stay in sync?”

Sense of Control (Agency / Natural Mapping) – (SQ4 – Immersion)

Question 2.3:

“Did it still feel like your gestures were directly controlling the character when things got fast?”

- If YES:

“What helped you maintain that feeling of control?”
“Was it consistent across all gestures?”

- If NO or partial:

“What broke that connection for you—lag, recognition, gesture complexity?”
“How could that feeling of control be improved?”

Extra Question (measures presence of disruption)

Question 2.4:

“Did you experience any moments during fast gameplay where the gesture felt disruptive, awkward, or pulled you out of the experience?”

- If YES:

“Which gesture and when?”
“What exactly broke the flow for you?”

- If NO:

“What helped the interaction stay smooth and immersive for you?”

Speed Adaptation – (SQ3 – Reliability)

Question 2.5:

Did you have to change or simplify any gestures as the speed increased?
→ *[Follow-up]*: Which gestures changed the most? Why?
→ *[Follow-up]*: Did simplifying them help, or did it feel like a compromise?

Question 2.6:

Were there any gestures that became hard or impossible to do at high speed?
→ *[Follow-up]*: What did you do in those moments?

Errors and Breakdowns – (SQ3 – Reliability)

Question 2.7:

Did you make more mistakes as the pace increased?
→ *[Follow-up]*: Were the mistakes due to physical difficulty, remembering the gestures, or something else?

Question 2.8:

Did any gestures start to feel confusing or inconsistent at faster speeds?
→ *[Follow-up]*: Do you think they’d still work well in a real-time game?

Fatigue and Comfort – (SQ3 – Usability)

Question 2.9:

Did you feel any physical tiredness while performing the gestures at high speed?
→ *[Follow-up]*: In which body parts?
→ *[Follow-up]*: Were certain gestures more tiring than others?

Question 2.10:

If you had to play a 5-minute match using just these gestures, how do you think your body would feel afterward?
→ *[Follow-up]*: Which gestures would you want to avoid repeating?

Overall Impression

Question 2.11:

Overall, would you enjoy playing a fighting game using these gestures at high speed?
→ *[Follow-up if no]*: What would need to change to make it more enjoyable?
→ *[Follow-up if yes]*: What makes it enjoyable for you?

Extra Questions

- Which gestures felt most natural or fun to perform
- Which gestures were the hardest to do or remember?
- Would you use these gestures in a real game? Why or why not?

Appendix E

Participant Information Sheet

The following document is the information sheet provided the participants of the user study.

Participant Information Sheet

Study title
Beyond Buttons: Evaluating Neuroscientifically Refined Hand Gestures in Fast-Paced Fighting Games

Researcher
Wahaj Ahmad, M.Sc. Interaction Technology
University of Twente, Faculty EEMCS, Human-Media Interaction
w.ahmad-1@student.utwente.nl +31 6 84431110

Supervisors
Dr. Dennis Reidsma Dr. G.W.J. Bruinsma D.P. Davison PhD

---

1. What is the purpose of this study?

You are invited to help develop and evaluate hand-gesture controls for a popular fighting game (Tekken 8). In this first phase, we collect the natural gestures you spontaneously use to execute game tasks. Later phases will refine these gestures using neuroscientific principles and test them in real gameplay (*not part of the current study*).

2. Why have you been invited?

You are aged 18–35 with normal or corrected vision. We ask you to nominate yourself.

Exclusion Criteria:

> No upper-limb impairments - You must have full, pain-free range of motion in both wrists, hands and fingers, and no history of recent injury or surgery to these areas.

Screening process:

> Before the session begins, the researcher will ask you a few simple questions about your arm, wrist, hand, or finger health (*found in Participant Information Questionnaire*). You will also be asked to perform a brief range-of-motion check (e.g., opening and closing your hand, bending your wrist up/down). If you report any relevant impairment or cannot comfortably complete these movements, you will be excluded for your own safety and data validity.

3. Do you have to take part?

No – participation is entirely voluntary. You may decline to answer any question and withdraw at any time without giving a reason. If you choose to leave, we will destroy any data we have recorded from you.

4. What will taking part involve?

- **Duration:** 60–90 minutes total.
- **Setting:** Laboratory room at the University of Twente.
- **Equipment:** Hand-camera, face-camera, screen capture, and audio recorders.

- **Procedure:**

    1. Briefing and signing consent.
    2. Participant Information Questionnaire
    3. Experiment 1:
       You will need to invent and perform a hand gesture (*input*) to execute an in-game action (*command)*. You will be shown a specific move (SM) of an in-game character along with a distinct icon for that command and a verbal prompt by the researcher. You perform hand gestures for these commands five times each, thinking aloud.
    4. Short interview.
    5. Experiment 2:
       This experiment will test the frequency of the gestures invented in Experiment 1. The icons shown in Experiment 1 will now appear on the game-play screen. You will perform the gestures for the commands that the icons represent. The speed of the icons will increase gradually. You will be required to keep up to the best of your ability. If you miss a gesture, do not worry and simply move on.
    6. Short feedback session.

5. Possible risks or discomforts

You may feel mild frustration if your gesture does not yield the expected in-game result. We may limit the retries to a specific number or time.

6. How will your data be recorded and used?

- **Recordings:** Face, hand-cam, audio, and screen-capture video.
- **Use:** Your hand-cam footage will form a de-identified gesture corpus for ML model training; face and audio recordings support analysis of think-aloud data but will not be reused outside this study.
- **Storage & confidentiality:** The data will be stored securely with the researcher, accessible only to the research team, and anonymized before publication. No one outside the project will receive identifying information.
- **Retention:** Facial data will be kept until the end of the research, then securely deleted. The hand gesture data may be kept forming a gesture corpus for future studies.

You have the right to request access, correction, or erasure of your personal data.

7. What will happen to the results?

We will use the data to refine gesture sets and validate them in later user tests, and publish findings in academic journals and conferences. You may request a plain-language summary once the thesis is complete.

8. Contact for further information

For any further information you can contact the researcher.

- **Researcher:** Wahaj Ahmad (w.ahmad-1@student.utwente.nl, +31 6 84431110)

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the supervisor or the ethics committee at the University of Twente.

- **Supervisor:** Dr. Dennis Reidsma, d.reidsma@utwente.nl
- **Ethics questions:** Secretary, Ethics Committee EEMCS, ethicscommittee-cis@utwente.nl

Thank you.

Appendix F

Informed Consent Form

The following document is the consent form each participant signed prior to participation in the user study.

Consent Form for Beyond Buttons: Evaluating Neuroscientifically Refined Hand Gestures in Fast-Paced Fighting Games.

Researcher: Wahaj Ahmad, M.Sc., University of Twente

YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated ____ /____ /2025, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	○	○
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	○	○
I understand that taking part in the study involves being video-recorded (face and hand), audio-recorded, and that my gameplay inputs will be logged.	○	○
Risks associated with participating in the study		
I understand that taking part in the study involves the following risks: You may feel mild frustration if your gesture does not yield the expected in-game result.	○	○
Use of the information in the study		
I understand that information I provide will be used. My hand-cam footage will be used in anonymized form in future machine-learning research.	○	○
I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team.	○	○
I agree that my information can be quoted in research outputs	○	○
I agree to be audio and video recorded.	○	○
Future use and reuse of the information by others		
I give permission for the hand-cam footage that I provide to be archived so it can be used for future research and learning.	○	○

Signatures

_____	_____	_____
Name of participant	Signature	Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____Wahaj Ahmad_____	_____	_____
Researcher name [printed]	Signature	Date

Study contact details for further information:

Wahaj Ahmad (w.ahmad-1@student.utwente.nl, +31 6 84431110)

Contact Information for Questions about Your Rights as a Research Participant

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee Information & Computer Science: ethicscommittee-CIS@utwente.nl

Appendix G

Debrief Letter

The following document is the debrief letter each participant was given after the study ended.

Debriefing & Disclaimer Form

Study title
Beyond Buttons: Evaluating Neuroscientifically Refined Hand Gestures in Fast-Paced Fighting Games

Researcher
Wahaj Ahmad, M.Sc. Interaction Technology
University of Twente, Faculty EEMCS, Human-Media Interaction
w.ahmad-1@student.utwente.nl +31 6 84431110

Supervisors
Dr. Dennis Reidsma Dr. G.W.J. Bruinsma D.P. Davison PhD

Thank you for participating!

This debrief explains aspects of the study we did not fully reveal beforehand:

1. **Wizard of Oz setup:** During gameplay, you were told the game responded directly to your gestures. In fact, a researcher (the "wizard") was hidden behind a screen and controlling the character via a gamepad to collect your natural gesture repertoire.
2. **Why deception?** We withheld the wizard's role to ensure you provided genuine, intuitive gestures rather than imitating on-screen prompts.
3. **Your rights:** You may now withdraw your data if you wish, up to two weeks from today, by contacting w.ahmad-1@student.utwente.nl.
4. **Follow-up:** If you'd like a summary of results or wish to ask questions, please provide your email here: _____.
5. **Contact ethics committee:** For concerns about your rights, contact the Secretary, Ethics Committee EEMCS, University of Twente at ethicscommittee-cis@utwente.nl.

Participant signature (to confirm you've been fully debriefed): _____

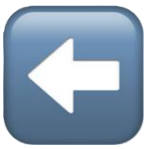
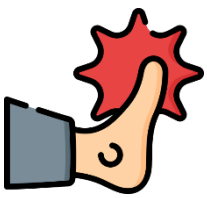
Date: _____

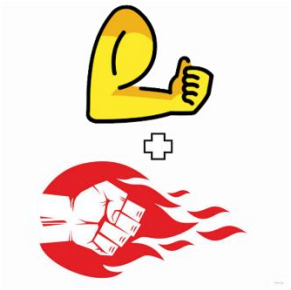
Appendix H

Wizard Cheat Sheet / Prompt Icons

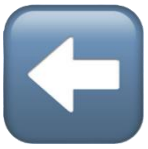
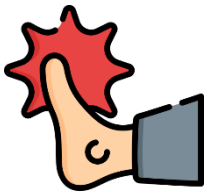
The following document contains the prompt icons for each move that was used in the study. This was also used as a cheat sheet by the wizard to help execute the commands.

Wizard Cheat Sheet – Left Side

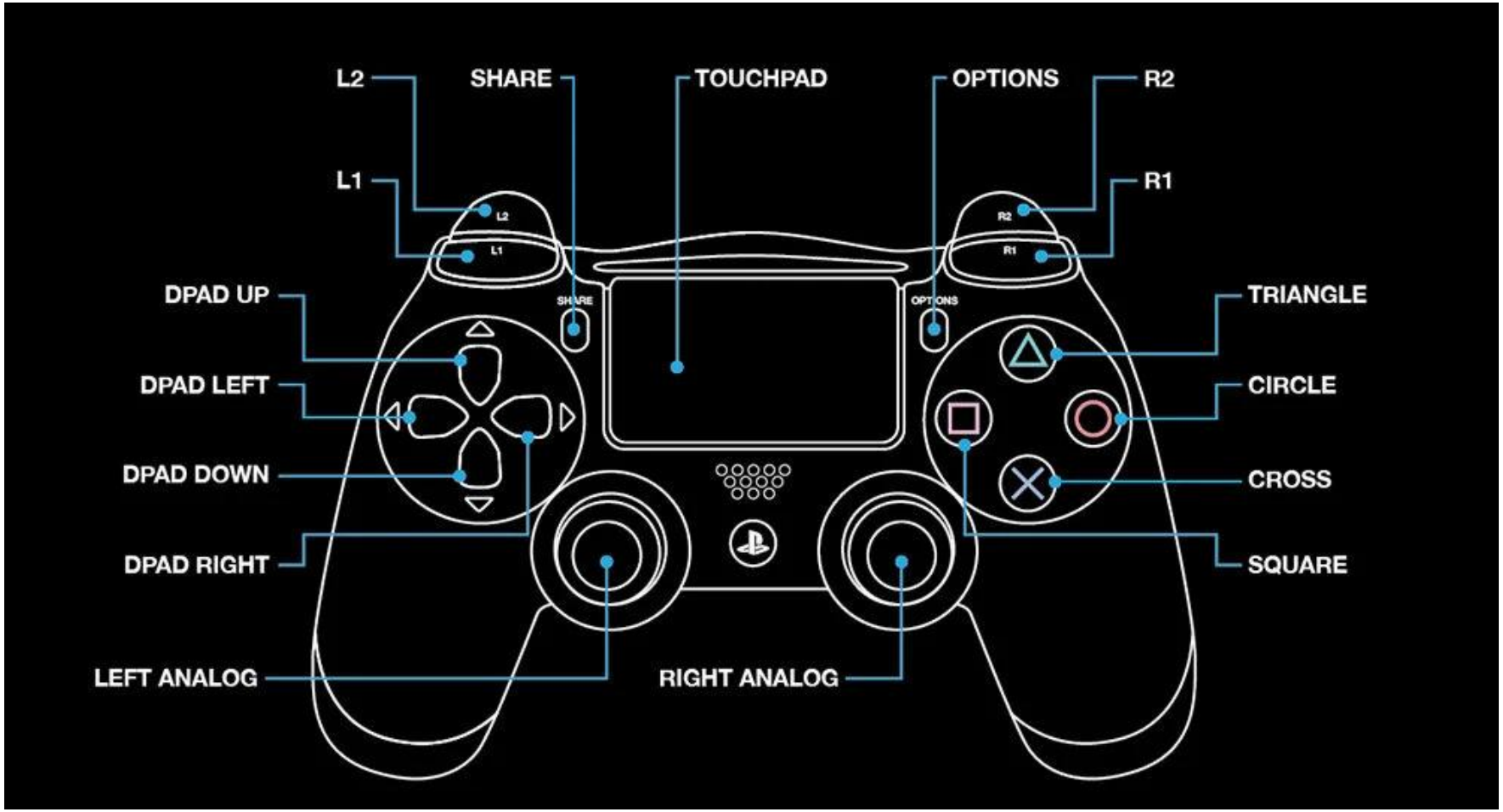
Move Name	Move Icon	Macro Input
Forward		D-Pad = Right OR Left Analog = Right
Backward		D-Pad = Left OR Left Analog = Left
Sidestep Forward		D-Pad = Down Quick press OR Left Analog = Down Flick
Sidestep Backward		D-Pad = Up Quick press OR Left Analog = Up Flick
Left Punch		Square
Right Punch		Triangle
Left Kick		Cross
Right Kick		Circle
Grab		Right Analog = Right

Uppercut – Kurenai		Right Analog = Up
Neutron Bomb		Right Analog = Down
Hammer Punch		R1
Phoenix Smasher		L1
Hammer Phoenix		R2
Hangover		L2

Wizard Cheat Sheet – Right Side

Move Name	Move Icon	Macro Input
Forward		D-Pad = Left OR Left Analog = Left
Backward		D-Pad = Right OR Left Analog = Right
Sidestep Forward		D-Pad = Down Quick press OR Left Analog = Down Flick
Sidestep Backward		D-Pad = Up Quick press OR Left Analog = Up Flick
Left Punch		Square
Right Punch		Triangle
Left Kick		Cross
Right Kick		Circle
Grab		Right Analog = Left

Uppercut – Kurenai		Right Analog = Up
Neutron Bomb		Right Analog = Down
Hammer Punch		R1
Phoenix Smasher		L1
Hammer Phoenix		R2
Hangover		L2


L2
SHARE
TOUCHPAD
OPTIONS
R2
L1
R1
DPAD UP
TRIANGLE
DPAD LEFT
CIRCLE
DPAD DOWN
CROSS
DPAD RIGHT
SQUARE
LEFT ANALOG
RIGHT ANALOG

Bibliography

Al-Shamayleh, A., Ahmad, R., Abushariah, M., Alam, K. & Jomhari, N. (2018), 'A systematic literature review on vision based gesture recognition techniques', *Multimedia Tools and Applications* **77**(21), 28121–28184.

Alyami, S., Luqman, H. & Hammoudeh, M. (2024), 'Reviewing 25 years of continuous sign language recognition research: Advances, challenges, and prospects', *Information Processing and Management* **61**(5).

Ambrosecchia, M., Marino, B. F. M., Gawryszewski, L. G. & Riggio, L. (2015), 'Spatial stimulus-response compatibility and affordance effects are not ruled by the same mechanisms', *Frontiers in Human Neuroscience* **9**, 283.

Baddeley, A. (2000), 'The episodic buffer: a new component of working memory?', *Trends in Cognitive Sciences* **4**(11), 417–423.

Bandura, A. (2010), Self-Efficacy, *in* 'The Corsini Encyclopedia of Psychology', John Wiley & Sons, Ltd, pp. 1–3. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470479216.corpsy0836.

Bianchi-Berthouze, N., Kim, W. W. & Patel, D. (2007), Does Body Movement Engage You More in Digital Game Play? and Why?, *in* A. C. R. Paiva, R. Prada & R. W. Picard, eds, 'Affective Computing and Intelligent Interaction', Springer, Berlin, Heidelberg, pp. 102–113.

Birk, M. & Mandryk, R. L. (2013), Control your game-self: effects of controller type on enjoyment, motivation, and personality in game, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', CHI '13, Association for Computing Machinery, New York, NY, USA, pp. 685–694.

Brown, E. & Cairns, P. (2004), A grounded investigation of game immersion, *in* 'CHI '04 Extended Abstracts on Human Factors in Computing Systems', CHI EA '04, Association for Computing Machinery, New York, NY, USA, pp. 1297–1300.

Cairns, P., Power, C., Barlet, M. & Haynes, G. (2019), 'Future design of accessibility in games: A design vocabulary', *International Journal of Human-Computer Studies* **131**, 64–71.

Cheng, H., Yang, L. & Liu, Z. (2016), 'Survey on 3D Hand Gesture Recognition', *IEEE Transactions on Circuits and Systems for Video Technology* **26**(9), 1659–1673.

Claypool, M. & Claypool, K. (2006), 'Latency and player actions in online games', *Commun. ACM* **49**, 40–45.

Cohen, J. (1960), 'A coefficient of agreement for nominal scales', *Educational and Psychological Measurement* **20**, 37–46. Place: US.

Cowley, B., Charles, D., Black, M. & Hickey, R. (2008), 'Toward an understanding of flow in video games', *Comput. Entertain.* **6**(2), 20:1–20:27.

Csikszentmihalyi, M., Abuhamdeh, S. & Nakamura, J. (2014), Flow, *in* M. Csikszentmihalyi, ed., 'Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi', Springer Netherlands, Dordrecht, pp. 227–238.

Dahlbäck, N., Jönsson, A. & Ahrenberg, L. (1993), 'Wizard of Oz studies — why and how', *Knowledge-Based Systems* **6**(4), 258–266.

DashFight (2024), 'Tekken 8 Control Scheme | DashFight'.

Drachen, A., Mirza-Babaei, P., Nacke, L., Drachen, A., Mirza-Babaei, P. & Nacke, L., eds (2018), *Games User Research*, Oxford University Press, Oxford, New York.

Dye, M. W., Green, C. S. & Bavelier, D. (2009), 'Increasing Speed of Processing With Action Video Games', *Current Directions in Psychological*

Science **18**(6), 321–326.

Ermi, L. & Mäyrä, F. (2005), Fundamental Components of the Gameplay Experience: Analysing Immersion, *in* 'Proceedings of DiGRA 2005 Conference: Changing Views: Worlds in Play'.

Fan, M., Shi, S. & Truong, K. N. (2020), 'Practices and challenges of using think-aloud protocols in industry: an international survey', *J. Usability Studies* **15**(2), 85–102.

Foottit, J., Brown, D., Marks, S. & Connor, A. (2014), An intuitive tangible game controller, Vol. 02-03-December-2014.

Francese, R., Passero, I. & Tortora, G. (2012), Wiimote and kinect: Gestural user interfaces add a natural third dimension to HCI, pp. 116–123.

Han, J., Shao, L., Xu, D. & Shotton, J. (2013), 'Enhanced computer vision with Microsoft Kinect sensor: A review', *IEEE Transactions on Cybernetics* **43**(5), 1318–1334.

Heitz, R. P. (2014), 'The speed-accuracy tradeoff: history, physiology, methodology, and behavior', *Frontiers in Neuroscience* **8**, 150.

Hincapié-Ramos, J. D., Guo, X., Moghadasian, P. & Irani, P. (2014), Consumed endurance: a metric to quantify arm fatigue of mid-air interactions, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', CHI '14, Association for Computing Machinery, New York, NY, USA, pp. 1063–1072.

Hsieh, H.-F. & Shannon, S. E. (2005), 'Three approaches to qualitative content analysis', *Qualitative Health Research* **15**(9), 1277–1288.

Höysniemi, J., Hämäläinen, P. & Turkki, L. (2004), Wizard of Oz prototyping of computer vision based action games for children, *in* 'Proceedings of the 2004 conference on Interaction design and children: building a community', ACM, Maryland, pp. 27–34.

Höysniemi, J., Hämäläinen, P., Turkki, L. & Rouvi, T. (2005), 'Children's intuitive gestures in vision-based action games', *Communications of the*

ACM **48**(1), 44–50.

Ionescu, D., Ionescu, B., Gadea, C. & Islam, S. (2011), A Multimodal Interaction Method that Combines Gestures and Physical Game Controllers, *in* '2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN)', pp. 1–6.

Isbister, K. & Schaffer, N. (2008), *Game usability: Advancing the player experience*, CRC Press, Boca Raton, FL.

Ivanoff, J., Blagdon, R., Feener, S., McNeil, M. & Muir, P. H. (2014), 'On the temporal dynamics of spatial stimulus-response transfer between spatial incompatibility and Simon tasks', *Frontiers in Neuroscience* **8**, 243.

Janczyk, M., Xiong, A. & Proctor, R. W. (2019), 'Stimulus-Response and Response-Effect Compatibility With Touchless Gestures and Moving Action Effects', *Human Factors* **61**(8), 1297–1314.

Janke, V. & Marshall, C. R. (2017), 'Using the Hands to Represent Objects in Space: Gesture as a Substrate for Signed Language Acquisition', *Frontiers in Psychology* **8**.

Juul, J. (2012), *A Casual Revolution: Reinventing Video Games and Their Players*, MIT Press. Google-Books-ID: r4vRa7MGbYIC.

Kelley, J. F. (1984), 'An iterative design methodology for user-friendly natural language office information applications', *ACM Trans. Inf. Syst.* **2**(1), 26–41.

Kendon, A. (2004), *Gesture: Visible Action as Utterance*, Cambridge University Press. Google-Books-ID: hDXnnzmDkOkC.

Kurtenbach, G. & Buxton, W. (1994), User learning and performance with marking menus, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', CHI '94, Association for Computing Machinery, New York, NY, USA, pp. 258–264.

Laura Ermi, Ermi, L., Frans Mäyrä & Mäyrä, F. (2005), 'Fundamental Components of the Gameplay Experience: Analysing Immersion', **3**, 15–27. MAG ID: 2160367221.

Leganchuk, A., Zhai, S. & Buxton, W. (1998), 'Manual and cognitive benefits of two-handed input: an experimental study', *ACM Trans. Comput.-Hum. Interact.* **5**(4), 326–359.

Madapana, N., Gonzalez, G., Zhang, L., Rodgers, R. & Wachs, J. (2020), 'Agreement Study Using Gesture Description Analysis', *IEEE Transactions on Human-Machine Systems* **50**(5), 434–443.

Mai, N. T. T., Hai, T. T. T. & Son, N. V. (2011), Wizard of Oz for Designing Hand Gesture Vocabulary in Human-Robot Interaction, *in* '2011 Third International Conference on Knowledge and Systems Engineering', pp. 232–238.

Mattiassi, A. D. A. (2019), 'Fighting the game. Command systems and player-avatar interaction in fighting games in a social cognitive neuroscience framework', *Multimedia Tools and Applications* **78**(10), 13565–13591.

McGloin, R., Farrar, K. M. & Krcmar, M. (2011), 'The Impact of Controller Naturalness on Spatial Presence, Gamer Enjoyment, and Perceived Realism in a Tennis Simulation Video Game', *Presence: Teleoperators and Virtual Environments* **20**(4), 309–324.

McNeill, D. (1992), *Hand and mind: What gestures reveal about thought*, Hand and mind: What gestures reveal about thought, University of Chicago Press, Chicago, IL, US. Pages: xi, 416.

Michailidis, L., Balaguer-Ballester, E. & He, X. (2018), 'Flow and Immersion in Video Games: The Aftermath of a Conceptual Challenge', *Frontiers in Psychology* **9**.

Miles, J. D. & Proctor, R. W. (2009), 'Reducing and restoring stimulus-response compatibility effects by decreasing the discriminability of location words', *Acta Psychologica* **130**(1), 95–102.

Mittelstädt, V., Miller, J., Leuthold, H., Mackenzie, I. G. & Ulrich, R. (2022), 'The time-course of distractor-based activation modulates effects of speed-accuracy tradeoffs in conflict tasks', *Psychonomic Bulletin & Review* **29**(3), 837–854.

Mohamed, N., Mustafa, M. B. & Jomhari, N. (2021), 'A Review of the Hand Gesture Recognition System: Current Progress and Future Directions', *IEEE Access* **9**, 157422–157436. Conference Name: IEEE Access.

Morris, M. R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., Schraefel, M. C. & Wobbrock, J. O. (2014), 'Reducing Legacy Bias in Gesture Elicitation Studies', *ACM Interactions Magazine* .

Müller, C. (2014), Gestural Modes of Representation as techniques of depiction, pp. 1687–1701.

Nadia Bianchi-Berthouze & Bianchi-Berthouze, N. (2012), 'Understanding the role of body movement in player engagement', *Human-Computer Interaction* **28**(1), 40–75. MAG ID: 1515243648.

Nielsen, J. (1993), 'Response Time Limits: Article by Jakob Nielsen'.

Nijhar, J., Bianchi-Berthouze, N. & Boguslawski, G. (2012), Does Movement Recognition Precision Affect the Player Experience in Exertion Games?, *in* A. Camurri & C. Costa, eds, 'Intelligent Technologies for Interactive Entertainment', Springer, Berlin, Heidelberg, pp. 73–82.

Nowell, L., Norris, J., White, D. & Moules, N. (2017), 'Thematic Analysis: Striving to Meet the Trustworthiness Criteria', *International Journal of Qualitative* **16**.

Ortega, G. & Özyürek, A. (2020), 'Types of iconicity and combinatorial strategies distinguish semantic categories in silent gesture across cultures', *Language and Cognition* **12**(1), 84–113.

Paliyawan, P., Sookhanaphibarn, K., Choensawat, W. & Thawonmas, R. (2015), Towards universal kinect interface for fighting games, *in* '2015 IEEE 4th Global Conference on Consumer Electronics (GCCE)',

pp. 332–333.

Pasch, M., Bianchi-Berthouze, N., van Dijk, B. & Nijholt, A. (2009*a*), Immersion in Movement-Based Interaction, *in* A. Nijholt, D. Reidsma & H. Hondorp, eds, 'Intelligent Technologies for Interactive Entertainment', Springer, Berlin, Heidelberg, pp. 169–180.

Pasch, M., Bianchi-Berthouze, N., van Dijk, B. & Nijholt, A. (2009*b*), 'Movement-based sports video games: Investigating motivation and gaming experience', *Entertainment Computing* **1**(2), 49–61.

Pietschmann, D., Valtin, G. & Ohler, P. (2012), The Effect of Authentic Input Devices on Computer Game Immersion, *in* J. Fromme & A. Unger, eds, 'Computer Games and New Media Cultures: A Handbook of Digital Games Studies', Springer Netherlands, Dordrecht, pp. 279–292.

Pisharady, P. & Saerbeck, M. (2015), 'Recent methods and databases in vision-based hand gesture recognition: A review', *Computer Vision and Image Understanding* **141**, 152–165.

Preece, J., Sharp, H. & Rogers, Y. (2015), *Interaction Design: Beyond Human-Computer Interaction*, John Wiley & Sons. Google-Books-ID: n0h9CAAAQBAJ.

Qualitative Data Analysis (n.d.).

Riek, L. D. (2012), 'Wizard of Oz studies in HRI: a systematic review and new reporting guidelines', *J. Hum.-Robot Interact.* **1**(1), 119–136.

Rogers, R., Bowman, N. D. & Oliver, M. B. (2015), 'It's not the model that doesn't fit, it's the controller! The role of cognitive skills in understanding the links between natural mapping, performance, and enjoyment of console video games', *Computers in Human Behavior* **49**, 588–596.

Ruiz, J., Li, Y. & Lank, E. (2011), User-defined motion gestures for mobile interaction, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', ACM, Vancouver BC Canada, pp. 197–206.

Sagayam, K. M. & Hemanth, D. J. (2017), 'Hand posture and gesture recognition techniques for virtual reality applications: a survey', *Virtual Reality* **21**(2), 91–107.

Saldana, J. (2021), *The Coding Manual for Qualitative Researchers*, SAGE. Google-Books-ID: RwcVEAAAQBAJ.

Sandler, W. (2012), 'THE PHONOLOGICAL ORGANIZATION OF SIGN LANGUAGES', *Language and linguistics compass* **6**(3), 162–182.

Simor, F. W., Brum, M. R., Schmidt, J. D. E., Rieder, R. & Marchi, A. C. B. D. (2016), 'Usability Evaluation Methods for Gesture-Based Games: A Systematic Review', *JMIR Serious Games* **4**(2), e5860. Company: JMIR Serious Games Distributor: JMIR Serious Games Institution: JMIR Serious Games Label: JMIR Serious Games.

Sloetjes, H. & Wittenburg, P. (2008), Annotation by Category: ELAN and ISO DCR, *in* N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis & D. Tapias, eds, 'Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)', European Language Resources Association (ELRA), Marrakech, Morocco.

Standards, E. (2022), 'BS EN ISO 9241-940:2022 Ergonomics of human-system interaction Evaluation of tactile and haptic interactions'.

Stokoe, W. C. (2005), 'Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf', *Journal of Deaf Studies and Deaf Education* **10**(1), 3–37.

Sweetser, P. & Wyeth, P. (2005), 'GameFlow: a model for evaluating player enjoyment in games', *Computers in Entertainment* **3**(3), 3–3.

Teixeira, J. M., Farias, T., Moura, G., Lima, J. P., Pessoa, S. & Teichrieb, V. (2006), 'GeFighters: an Experiment for Gesture-based Interaction Analysis in a Fighting Game'.

Tennant, R. A. & Brown, M. G. (1998), *The American Sign Language Handshape Dictionary*, Gallaudet University Press. Google-Books-ID: 27WtFCWcEucC.

Tsandilas, T. (2018), 'Fallacies of Agreement: A Critical Review of Consensus Assessment Methods for Gesture Elicitation', *ACM Trans. Comput.-Hum. Interact.* **25**(3), 18:1–18:49.

Vatavu, R.-D. (2019), The Dissimilarity-Consensus Approach to Agreement Analysis in Gesture Elicitation Studies, *in* 'Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems', ACM, Glasgow Scotland Uk, pp. 1–13.

Vatavu, R.-D. & Wobbrock, J. O. (2015), Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit, *in* 'Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems', CHI '15, Association for Computing Machinery, New York, NY, USA, pp. 1325–1334.

Viglialoro, R., Condino, S., Turini, G., Mamone, V., Carbone, M., Ferrari, V., Ghelarducci, G., Ferrari, M. & Gesi, M. (2020), 'Interactive serious game for shoulder rehabilitation based on real-time hand tracking', *Technology and Health Care* **28**(4), 403–414.

Villarreal-Narvaez, S., Vanderdonckt, J., Vatavu, R.-D. & Wobbrock, J. O. (2020), A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?, *in* 'Proceedings of the 2020 ACM Designing Interactive Systems Conference', DIS '20, Association for Computing Machinery, New York, NY, USA, pp. 855–872.

Williams, K. D. (2014), 'The effects of dissociation, game controllers, and 3D versus 2D on presence and enjoyment', *Computers in Human Behavior* **38**, 142–150.

Wobbrock, J. O., Morris, M. R. & Wilson, A. D. (2009), User-defined gestures for surface computing, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', CHI '09, Association for

Computing Machinery, New York, NY, USA, pp. 1083–1092.

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J. & Reed, B. (2015), 'Factors influencing the latency of simple reaction time', *Frontiers in Human Neuroscience* **9**, 131.

Yasen, M. & Jusoh, S. (2019), 'A systematic review on hand gesture recognition techniques, challenges and applications', *PeerJ Computer Science* **2019**(9).