Understanding Freehand Cursorless Pointing Variability and Its Impact on Selection Performance

JAMES WHIFFING, University of Bath, UK
TOBIAS LANGLOTZ, Aarhus University, Denmark
CHRISTOF LUTTEROTH, University of Bath, UK
ADWAIT SHARMA, University of Bath, UK
CHRISTOPHER CLARKE, University of Bath, UK

ABSTRACT

Freehand pointing is a fundamental gesture commonly used for cursorless interactions. Prior work in HCI often elicits the same pointing behaviour – facing the target with an outstretched dominant arm and index finger. However, freehand pointing outside of HCI shows more variability across hand pose, usage, and coordination with gaze. To understand what variability exists and how it affects pointing performance, we collected data (N=23) using a hybrid motion capture system. To elicit a wide variety of pointing behaviours we included different levels of user effort and attention, as well as the widest range of target placements studied. We systematically characterised and described three distinct pointing behaviours, each with three different traits, ranging from accurate stereotypical pointing observed in prior works to more casual hip fire-style pointing. Our analysis demonstrates how different pointing behaviours affect pointing performance and highlights their importance when designing interactive systems for more naturalistic freehand pointing.

CCS Concepts: • **Human-centered computing** → **Pointing**; **Gestural input**; *Empirical studies in HCI*.

Additional Key Words and Phrases: Cursorless, Freehand, Pointing, Interaction Techniques, Gesture.

ACM Reference Format:
James Whiffing, Tobias Langlotz, Christof Lutteroth, Adwait Sharma, and Christopher Clarke. 2025. Understanding Freehand Cursorless Pointing Variability and Its Impact on Selection Performance. *ACM Trans. Comput.-Hum. Interact.* 1, 1, Article 1 (January 2025), 47 pages. https://doi.org/10.1145/3770583

1 INTRODUCTION

Freehand pointing is a fundamental gesture that is used for communicating spatial information. Pointing enables individuals to guide the attention of others effectively, and we learn how to point from a very early age to help facilitate human-to-human communication. The biomechanical motion of pointing with the hand has been used in human-computer interaction (HCI) design and studied across various contexts, mainly as an interaction technique to interact with distant objects and devices [44, 45, 54]. Pointing interactions commonly use a cursor or ray as a proxy for the user's attention, which is controlled and manipulated to provide precise input [48]. This visual feedback allows users to refine their pointing gesture, making very precise selections before selection takes place using a secondary input gesture such as dwell [45, 49, 54] or voice [5, 35, 59, 61]. However,

Authors' addresses: James Whiffing, jw2304@bath.ac.uk, University of Bath, Bath, UK; Tobias Langlotz, tobias.langlotz@cs.au.dk, Aarhus University, Aarhus, Denmark; Christof Lutteroth, cl2073@bath.ac.uk, University of Bath, Bath, UK; Adwait Sharma, as5339@bath.ac.uk, University of Bath, Bath, UK; Christopher Clarke, cjc234@bath.ac.uk, University of Bath, Bath, UK.

ACM 1073-0516/2025/1-ART1
https://doi.org/10.1145/3770583

the conventional way we point when communicating with other people does not require a cursor for visual feedback, and we are often surrounded by devices or in situations where visual feedback is not required (e.g. high-level of accuracy is not needed), possible (e.g. smart devices without a visual display), or desirable (e.g. shared public spaces).

Cursorless pointing offers a simple alternative that encodes both the location of the target (i.e. where they point) and the intent to select (i.e. that they point) in one physical movement without the need for direct visual feedback. While not as accurate as cursor-based pointing, cursorless pointing is compelling because it is a quick and efficient interaction technique. As a result, it has been used for interaction with smart-home devices [59], virtual agents in VR [61], robots [51, 56, 60], cross-device interactions [52], and for communicating spatial information in the real world to a system [25, 30]. It can also be used in combination with voice [5, 35] or gaze [40, 58] for multimodal interactions. As spatial computing and sensing technologies mature, we anticipate cursorless pointing to play a fundamental role when interacting naturalistically with devices and systems that combine real and virtual elements. For example, most AR-based HMDs feature hand tracking, but not all have eye tracking for techniques such as Gaze+Pinch [53], and we are seeing rapid advances in other form factors for spatial computing such as wearable AI-based pins (e.g. Humane Inc.'s AI pin [26]).

Interactions with technology should be as seamless as possible, and systems should embrace the individual differences between people. As a result, cursorless pointing should reflect the style of pointing we use in our everyday interactions and embrace individual differences. Research into human-to-human communication has demonstrated a wide variety of ways in which we point at objects in everyday contexts. This includes different arm poses (e.g. how raised or straight the arm is) [12, 13], the use of different hand shapes depending on proximity and intent of a target [10], and whether or not a user looks at the target when pointing [36]. In the context of interaction, cursorless pointing gestures lack an observer's perspective which would be present when pointing to an object for another person. Therefore, it is unknown whether similar variability is present for pointing abstractly for a spatial computer system because there is some evidence to suggest that pointing is performed differently based on the observer's relative position [23, 34, 43, 62]. Research into cursorless pointing gestures has mainly focused on accurately pointing at a target in front of the user in which a user is directly facing the target with their arm raised and outstretched and their index fingertip pointing towards the target [15, 44, 45, 54]. As a result, we lack empirical evidence regarding the extent to which this natural variability in pointing translates to interactions with abstract objects, and critically, how these variations impact performance metrics such as accuracy, duration, and user fatigue.

To move towards a vision where people can point as they choose, the goal of this paper was to gain a deeper understanding of what variability exists when performing freehand cursorless pointing, and how this variability affects pointing performance with existing pointing models. We addressed this by designing a study (n=23) that captures the broadest range of naturalistic pointing behaviours to date in the context of interaction. We spatially distributed targets around participants in contrast to prior work that constrains users to pointing at targets in front of them. We employed a minimally invasive motion-tracking system and elicited a range of naturalistic behaviours to see how pointing accurately at a target differs to pointing casually, as well as how concurrent cognitive tasks affect pointing performance.

Our results provide in-depth evidence that demonstrates that unconstrained cursorless pointing is a quick and efficient interaction technique and that users can accurately select a wide range of targets. In contrast to prior works, we show that different pointing behaviours exist and describe how they result in different levels of accuracy, effort and duration of the pointing gestures. More specifically, we contribute:

(1) The first set of empirically derived behavioural traits which can be used to characterise cursorless freehand pointing gestures. These include different variations of (i) *arm pose* which relates to how a person positions their arm and where this places the hand when pointing, (ii) *finger incorporation* which details how the finger is utilised within the pointing gesture, and (iii) *arm-torso alignment* which describes where the arm is placed with respect to the torso and sagittal axis.
(2) Analysis of a large, annotated freehand pointing gesture dataset (11,367 trials) which is publicly available[1], composed of full body and hand motion capture data which shows that:
    - Pointing accuracy, effort, and duration depend on what type of pointing behaviour someone adopts, in particular whether someone points with the intention of maximising accuracy or minimising effort and whether pointing is their primary task.
    - Pointing becomes much shorter (with the median time for a gesture to be held being half as long), less fatiguing (with the median NICER [38] value almost halving) but less accurate and with a lot more variability when a participant points casually or with a distraction.
    - Target locations affect how participants point at the targets, which in turn influences how accurately they can point. In particular, participants are more accurate when pointing at targets in the centre or to their dominant side.
    - Participants prefer to use both hands in order to point with the hand closest to the target, and can use both hands accurately. For example, pointing laterally with the non-dominant hand is more accurate than pointing medially with the dominant hand.
(3) Design recommendations to allow designers to incorporate individual differences in pointing behaviours and to better support the design of future freehand cursorless pointing interactions and research.

These contributions are of general relevance to the field of HCI as they provide data-driven insights into actual pointing behaviours that can be used to improve pointing-based interfaces and models. This is of specific relevance as unconstrained cursorless pointing is particularly well-suited for wearable and spatial computing systems where current and future devices will increasingly rely on freehand input. In addition, this paper provides compelling evidence about how designers and systems should accommodate and embrace individual user differences.

2 BACKGROUND AND RELATED WORK

This paper investigates how pointing behaviour varies in cursorless, mid-air interactions and how these variations impact pointing performance (time, accuracy, and fatigue). To contextualize our work, we review prior research in three key areas: (1) psychological studies examining pointing gestures in different contexts as well as interaction research that investigates (2) how these gestures are elicited and the behaviours typically observed and (3) how the pointing direction is inferred.

2.1 Variability in Pointing Gestures

The ability to comprehend and perform pointing gestures is understood to be a key component in the ability to learn language during infancy [63]. As such, there exists a large body of research exploring the different ways in which people perform pointing gestures in human-human communication. For example, while deictic pointing is often associated with manual gestures (i.e. through use of one's hand) [47], studies have shown that culture [14], occupation [37], and situational factors [37] can affect the choice of modality for pointing (e.g., hand or head gestures). Additionally, studies exploring manual pointing gestures find variability in the pose of the arm [2], the variety of hand shapes used for pointing [3, 10, 11], and the coordination of pointing gestures with gaze [29, 36].

[1]Full dataset is available from: https://doi.org/10.15125/BATH-01594 [68]

Prior work identifies two patterns in manual pointing gestures: (1) 'full' pointing, where the arm is fully extended and the fingertip occludes the target, and (2) 'partial' pointing, which involves limited arm movement with the hand indicating direction [2, 29, 31, 69]. Bangerter and Chevalley found that full pointing is more common when the target is visible, whereas partial pointing decreases as target distance reduces [2]. Aside from Bangerter and Chevalley [2], no prior work has explored arm pose in depth, and especially not in the context of cursorless pointing for interaction. In this paper, we show the importance of arm pose and how it affects the accuracy, fatigue, and time taken to point.

Pointing gestures can vary in hand shape, reflecting differences in intent and context. The most common form involves extending the index finger while keeping the rest of the hand closed. This pose is typically used for precise communication about an object or location [29], especially when the target appears small in the visual field—either due to distance or size—or when the gesture conveys an imperative intent, such as instructing someone to act upon the indicated location [10]. This is also the dominant pointing gesture among sign language users [3, 18]. In contrast, an open palm pointing gesture is often employed when the target is already known [29], making precise localization less critical. This gesture is commonly observed when pointing to oneself or indicating large or nearby objects [18]. Another variant involves pointing with a closed hand and extended thumb, which typically occurs when directing attention toward a location outside the speaker's field of view, such as to their side or behind them [29, 69].

Research also shows that pointing behaviour changes depending on whether an object is within the visual field and can be gazed at [29, 69]. When people point at targets outside their visual field, they rely on allocentric knowledge (using environmental cues) or egocentric knowledge (using their own body as a reference). Without these visual cues, pointing accuracy to the remembered targets decreases [36]. When communicating, the person pointing may realign to face towards the target when the location is previously unknown or where accuracy is required, otherwise choosing to remain unaligned and pointing with more casual gestures [29].

There is some evidence to show that people may adapt the way they point based on the position of an observer when pointing is used as a method of communication for another person. They may realign themselves to face the same direction as the observer when using pointing gestures to indicate directions, such as left and right, although they can also perform pointing gestures relative to their orientation with the observer (e.g. pointing to their right to indicate the observer's left) [31]. In addition, it is commonly assumed that the greatest accuracy when pointing is due to the pointer's eye gaze passing through their finger-tip, however observers may interpret the pointing gesture by estimating a line originating from the pointer's shoulder and through the finger-tip [23, 34, 43, 62]. This is believed to be a significant cause for misunderstanding between pointers and observers. Herbort and Kunde found that informing observers on how to interpret a pointing gesture (i.e. as a ray from the head and through the finger) can reduce miscommunication [23]. These findings have two important implications for interacting with spatial computing systems. First, it is unclear what type of variability is present with a spatial computing system when there is no clear representation of an observer in an abstract context. Second, people may have different mental models of how ray casting applies to pointing which we explore by applying different ray casting approaches posthoc.

2.2 Pointing Variability in Interaction Studies

Low-effort, casual pointing postures have been observed in human-human pointing literature [2, 10, 29, 32, 69], and have been applied to cursor supported pointing interactions [21, 39, 64]. However, we typically only see one form of pointing performed in cursorless interactions: high-effort, outstretched occlusion pointing. Table 1 provides an overview of cursorless interaction studies.

Table 1. Breakdown of previous freehand cursorless pointing research, including the interaction spaces and setups used. Target arrangements are listed as (Rows × Columns [× Depth]), with field of view (FoV) reported as (Horizontal × Vertical). The papers are ordered in chronological order.

Paper	Env.	Instruction	Selection	Arrangement	Distance	FoV	Technique
[17]	RW	Dominant hand, condition dependent	Click*	2D Plane (3 × 3)	1m	±8.53° × ±5.71°	EFRC-E, IFRC
[51]	RW	"Show the [Robot] an object by pointing at it"	N/A	3D (N/A)	-	N/A	HWRC, HRC-adj., FRC
[45]	RW	Dominant hand, naturally	Click* + Dwell (1s)	2D Plane (7 × 5)	2m – 3m	±46.40° × ±30.97°	EFRC, IFRC, FRC, HRC
[54]	RW	Ray Dependent†	Voice + Dwell (2s)	2D Plane (4 × 4)	3m	±18.44° × ±18.44°	EFRC, HRC-opt., IFRC
[44]	RW and VR	Dominant hand, naturally	Click* + Dwell (1s)‡	2D Plane (7 × 5)	2m	±33.96° × ±18.81°	EFRC, IFRC, FRC
[42]	RW	Straight arm, both eyes open, head fixed in place	N/A	2D Plane (9 × 9)	2.5m	±27.12° × ±19.80°	EFRC-E
[43]	VR	"point at the target with the index finger of the right hand"	Click*	Cylinder Interior (16 × 5)	4m	±180.00° × ±28.07°	EFRC, IFRC, FRC, HRC
[15]	VR	"[...] With the right index finger, [...] as they would to direct a friend's attention"	Click*	3D Cube (3 × 3 × 3)	1.5m – 3.5m	±33.69° × ±33.69°	ML Models
[50]	RW	Point with dominant hand, seated or standing	Click** + Voice	3D (N/A)	-	N/A	ML Models
[57]	VR	"as if they were pointing out the object to someone else"	Click*	2D (3 × 3) / 3D (N/A)	3m / 1.5m	±28.07° × ±28.07° / N/A	EFRC
[8]	VR	Point at an object for a nearby friend, naturally	N/A	3D Spherical (3 × 3)	0.9m – 3.3m	±30.00° × ±20.00°	ML Models
Ours	RW	Condition Dependent	N/A	2D Spherical (5 × 3)	2m	±70.00° × ±25.00°	EFRC, HFRC, IFRC, FRC

∗ Selection performed with non-dominant hand. † Ray-dependent instructions – HRC-opt.: point at targets, IFRC: Imagine ray extending one's finger, EFRC: Imagine ray originating from the camera in their equipped glasses, passing through their finger-tip. ‡ Dwell time was only imposed for initial data collection and omitted in system evaluation.

Some works explicitly prescribe the pointing pose for participants to adopt [17, 42, 54]. Others may not give exact instructions on how to point, but rather describe the intent with which to point, e.g. "As you would point for a friend" [8], "as quickly and accurately as possible" [44], or describe how to visualise the pointing gesture, e.g. "imagine a ray originating from [between the eyes] and passing through the finger-tip" [54], which may implicitly influence the participants in adopting a higher-effort pose. Even when instructing participants to point 'naturally' [44, 45, 50], there may be an implicit level of accuracy expected by participants from simply partaking within a study. We look to explore how instruction may alter pointing behaviour, using language that describes different intents behind the pointing participants should perform, without prescribing specific pointing poses.

Possible pointing poses can also be constrained by the context in which the study was conducted. For example, prior research involving motion capture places markers on specific parts of the body that are known to contribute towards pointing (e.g. only on the index finger on the dominant hand [15, 42, 45]), which immediately introduces a bias such as the expectation of how to point (with the tracked index finger) and with which hand (the dominant hand that is tracked). Additionally, most pointing studies focus on target accuracy or task completion time rather than variability [8, 15, 42–45, 54]. More recently, Nakamura et al. elicited a wider variety of pointing behaviours, including what appears to be low-effort casual pointing, however they only focus on applying deep learning techniques to predict pointing direction and do not explore pointing performance in more depth [50]. In addition, the method in which participants confirm selection influences the way in which they point. By requiring participants to use their non-dominant hand to trigger selection [17, 44, 45, 50], only the dominant hand may be used for pointing. Similarly, enforcing a minimum dwell time [45, 54] can artificially extend natural pointing durations [49]. In contrast, we track both arms and allow participants to point naturalistically without enforcing any additional selection mechanism, instead labelling when pointing occurs post-hoc.

As pointing gestures become increasingly relevant for spatial computing, we can no longer assume targets will be placed only ahead of the user. The accuracy of cursorless pointing gestures has been explored across a wide range of targets. Several studies have focused on pointing for remote interaction with 'Ultrawalls', large 2D screens (sometimes curved) placed in front of the user [4, 44, 45, 54] and one has explored how target shape affects where on the target one points [57]. Several works have explored the accuracy of pointing at targets placed in any direction around the user in the real world [50, 51] and in virtual reality [43]. While these prior works have explored how target size, shape, and distance can impact the accuracy of pointing estimation techniques, they constrain how users should point and do not provide a detailed breakdown of how the target placement around the user affects pointing accuracy or alters pointing behaviour. In our work, we explore pointing behaviour across a wide range of target locations (140° horizontally and 50° vertically) and analyse how different target placements affect pointing

The trend towards spatial computing, usually in the form of head-mounted displays, will also introduce that we will interact with these devices in different contexts [20] including scenarios in which pointing may not be the primary task a user is focused on as it was in all existing studies. Simple pointing interactions could be used in parallel with another task (e.g. controlling a device while maintaining a conversation or walking the streets while selecting a new destination in a navigation application). Additionally, while research exists that shows pointing to targets is possible within and beyond one's field of view [29, 69], no work has explored how pointing beyond one's field of view affects selection performance. In our work, we explore pointing behaviour while participants are both focused on pointing as well as pointing when distracted.

Overall, most existing studies in HCI either implicitly or explicitly used or elicited freehand pointing techniques that involved a pointing pose where a user faces the target, raises and extends their arm, extends their index finger, and aligns the tip of the index finger with the target. This does not represent the larger observed variability reported in earlier work on natural pointing. In our work, we aim to explore more casual freehand pointing interaction by reducing prior constraints that have been introduced through the study context (e.g. such as focus on accuracy, limited tracking and pointing at 2D screen, no distraction). We argue that this research on casual pointing warrants more attention as pointing is increasingly used in ubiquitous and immersive technologies such as AR and VR head-mounted displays which are commonly used in casual contexts (e.g. while on the go, public transportation, or in public spaces). Common to these scenarios is that they require a spontaneous interaction where controllers are unlikely to be present, and where the general context is likely to require less precise interaction but with external distractors.

2.3 Inferring Pointing Direction for Cursorless Pointing

The most common approach to estimating the accuracy of a pointing gesture is to derive a vector, commonly referred to as a ray, from two points on the user's body. The intended target for selection is inferred through intersection with, or distance from, the ray. There are two common types of rays: head-rooted and hand-rooted. For head-rooted, eye finger ray casting (EFRC) is the most common, where the root of the ray originates from one of the eyes (EFRC-E) [17, 42] and passes through the tip of the index finger. However, for practical reasons it is more common to use the 'Cyclopean Eye' (mid-point between the eyes [1]), which we refer to as EFRC for simplicity. Other variants of head-rooted ray casting include optimising the origin of the ray, based on user ocular and hand dominance (HRC-opt.) [54], using head wrist ray casting (HWRC), or orientated by the difference between the head ray and the head-wrist ray (HRC-adj.) [51]. Other approaches have looked at only using gaze-based rays such as the head direction (HRC) [45]. Head-rooted ray casting has been shown to be more accurate than hand-rooted ray casting [17, 45], which includes forearm ray casting (FRC) and index finger ray casting (IFRC). Prior work has shown that systematic offsets exist when pointing and that these offsets can be corrected for using ordinary least squares models [44, 45]. These are especially important for the hand-rooted rays. More recent approaches have explored how machine learning (ML) techniques can be used to infer the pointing direction. For example, Dalsgaard et al. explored pointing in 3D space as both a classification and regression problem, trained against different feature sets (pose, motion, mobile) [15]. More recently, Nakamura et al. introduced *DeePoint*, a system which uses a transformer-based neural network in conjunction with a 2D pose estimator to extract the pointing vector from a sequence of image frames [50]. For interpretability, we focus on the common ray-casting techniques from the literature including EFRC, IFRC and FRC, as well as exploring head finger ray casting (HFRC).

2.4 Summary

Pointing can be performed in various ways. People point with a straight or bent arm [2, 29, 31, 69] and with either the index finger, whole hand, or thumb [3, 10, 18, 29, 69]. Additionally, a target does not need to be within one's visual field in order to point in its direction [29, 31, 36]. However, such variability is not observed in pointing interaction studies, where pointing may be constrained explicitly through instruction or implicitly through apparatus. In this paper we look to address this research gap, exploring how permitting pointing with either hand and by removing the need for an explicit selection mechanism that is more akin to human–human communication might elicit variability in pointing behaviour, and how instructions, distractions, and target position affect this variability, selection accuracy and user effort.

3 METHODOLOGY

We ran a data collection with participants to elicit different pointing behaviours and to understand and evaluate how variability affects pointing performance. In prior human–human communication research, casual pointing behaviours occurred in conversation. However, not all spatial computing systems support conversation with users (e.g. with voice commands) and some aim to only utilise the pointing modality. Therefore, to elicit natural and realistic pointing variability, we specifically designed our study to:

- **Elicit natural pointing gestures.** We included a condition that explicitly asked participants to point casually to promote natural pointing and a separate condition that encouraged accurate pointing. To avoid bias in the instructions, we purposefully did not provide any examples of pointing, nor did the experimenter suggest or perform any pointing gestures to not promote a specific pointing style.
- **Reveal changes in pointing style under different task loads.** With the advances in ubiquitous and spatial computing, we wanted to understand how task load affected pointing. Cursorless pointing as a selection technique involves a low-effort gesture that should not require much cognitive load. Therefore, one condition involved pointing only, and another condition required participants to complete a Stroop task at the same time to distract them and increase their task load.
- **Enable unconstrained pointing.** We used a marker-based motion capture system with markers on both hands to capture all five fingers and the wrist (rather than just the index finger on the dominant hand). Additionally, a markerless motion capture system was used for tracking the full body. Participants were not required to perform an explicit confirmation gesture (such as a hand-held button [15, 44, 50]) or self-label the pointing gesture. Instead, we identified pointing gestures post-hoc so that participants were not restricted to using one specific hand or finger and to avoid promoting certain hand gestures or shapes while pointing.
- **Explore a wide spatial arrangement of targets.** Targets were placed around the user, including the limits of peripheral vision, so that they could not be aligned to all target planes without re-orientating themselves. Targets were also presented at different heights (below shoulder level, at shoulder level, above shoulder level).

3.1 Conditions

We used a repeated measures within-subject design with Pointing Style (accurate or casual) and Focus (focused or distracted) as independent variables (IV). We asked participants to *"point as accurately and precisely as possible"* for the *accurate* pointing condition, emphasising that they should do so in a way that *"the system we are developing would recognise exactly where you are pointing at"*. In contrast, in the *casual* pointing condition participants were instructed to *"point as casually and relaxed as possible. [...] by this we mean that you are pointing in a more casual manner, and should be more physically relaxed compared to the accurate pointing you have done"*. In both conditions, we emphasised to participants to *"remember that we are developing a system that is going to be personalised to the way you point. In that respect, please do not do what you think we want to do, do whatever you find most comfortable and intuitive – there is no wrong way of pointing"*. The distracted condition involved participants completing a Stroop effect test [19]. This involved saying aloud the colour of a word displayed on the monitor, not the word itself. The words displayed represent both congruent (i.e. font colour matches the word) and incongruent (i.e. font colour does not match the word displayed) stimuli.

3.2 Study Apparatus

To support pointing at targets to which the participant is not directly aligned to, we looked to adopt a similar layout to Bihani et al. [4], having targets placed about the user, rather than directly ahead. This was to promote changes in arm pose due to target spread assuming that participants do not always turn to align with the target, similar to Dalsgaard et al. [15].

Our setup used 135 targets, which were grouped into 15 clusters of 9 (3×3), arranged into a 3×5 (rows × columns) array, see Figure 1b. For each target cluster, the distance between each target was 8.4 cm, the average error obtained using Mayer et al.'s error offset correction model for pointing ray-casting [45]. The middle row was located 1.4 m from the floor and 2 m away from the participant, and the top and bottom rows were pitched ±25° from the middle row (see Figure 1a). Each column was yawed ±35° relative to the adjacent column. To capture variability within participants and targets, we wanted to capture repetitions pointing toward the same target. Repetitions for all 135 targets would have been impractical due to the time required. Therefore, we used three targets from each cluster with three repetitions, retaining a total of 135 pointing gestures per condition. Figure 1b shows the subset of targets which were selected pseudo-randomly to ensure that within each cluster the targets did not share a row or column.

For our targets, we utilise 5 mm diameter RGB LEDs (grouped into clusters of 9, 3×3), although only the colour red was utilised within the study. The target clusters were tracked using unique rigid

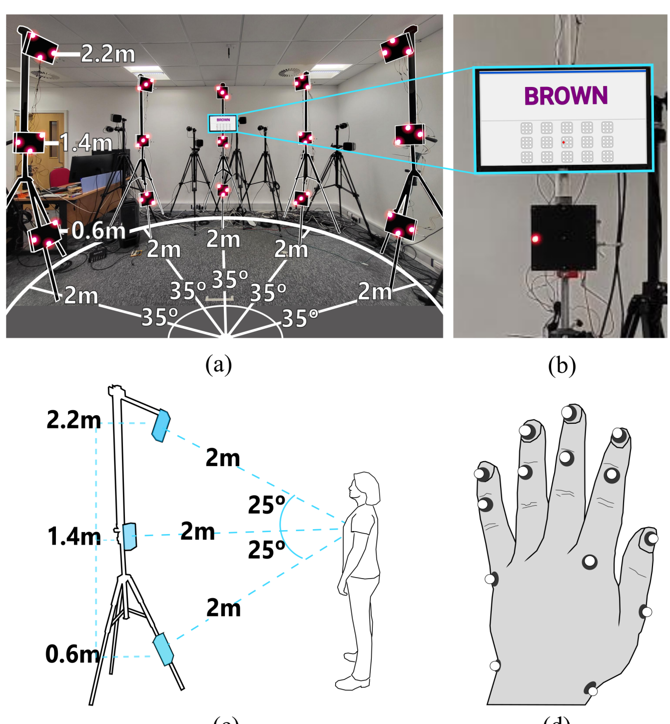

Fig. 1. a) Profile setup of a single column of targets relative to the participant. b) Arrangement of targets from the participant's perspective (image taken with fish-eye lens to capture full array of targets) showing the unique LEDs used for each target cluster. c) Top: Information screen displaying Stroop test and location of the target, Bottom: An active target. d) Placement of IR reflective markers for hand tracking.

bodies, with locations of the specific LED targets derived from the location and orientation of the tracked cluster. To record the activation of targets, an Adafruit ESP32 feather board micro-controller received target activation messages from a Java+Spring Restful application running the study, and sent a pulse over one of the trigger ports on the Qualisys Track Manager (QTM) sync unit, which created an event in QTM. Each column of targets was controlled by an Arduino which received instructions via Ethernet from a server managing the study logic. Each Arduino was additionally connected to a piezo buzzer located behind the middle target cluster of the respective column of targets in order to emit a tone when a target was activated. This was used to assist participants with locating the target (as we were not interested in search times) along with a visual map displayed on a 21" Dell P221H monitor that was placed between the top and middle centre targets (see Figure 1c).

3.3 Body Tracking

Similar to prior works, we utilised a motion capture system to track the participants' body movements [15, 42, 44, 45]. We used 12 Arqus infrared (IR) tracking cameras, and 10 Miqus cameras capturing RGB images at 1080p (4:3), with recording and marker tracking managed by QTM. The cameras provided coverage of a 4 m wide $\times$ 3 m deep $\times$ 2.5 m tall volume, within which the participant would be placed 2 m from the shorter edges, and $\sim$1 m from the long edge. The system was calibrated at the start of each day, with an average residual of 0.732 mm and standard deviation of 0.174 mm.

We used 28 6.5 mm IR reflective markers to track all fingers on both hands; two for each finger and four to capture the palm and wrist (see Figure 1d). To aid in the tracking of the hands, we employed QTM's AIM models and skeleton-assisted labelling, which can identify markers based on similarity to a pre-calibrated model and identify missing markers based on relative position to already labelled markers linked to a skeleton. We tracked the participants' body pose using the Miqus cameras and a markerless tracking system (Theia3D) which extracts the user's skeleton from videos using multiple perspectives. This removed the need to further instrument the participants.

The participants' eyes and gaze were tracked with Tobii Pro 3 glasses, which were themselves tracked using a rigid-body marker set (marker set 1) from Tobii. The Tobii glasses could not be worn over conventional glasses, but could be worn with corrective contact lenses. Due to technical issues with the eye tracking data, there is only valid gaze for a subset of participants, therefore we do not discuss eye tracking in this paper.

All sensing apparatus sampled data at $100Hz$ as to provide the greatest sample rate common to each system, for which we could retain 1080p resolution and reduce the impact of motion blur for the markerless video capture. Tracking rates for points derived from the marker motion capture can be found in Appendix A. These are calculated as the percentage of frames that a labelled point is tracked during a pointing gesture. We omit points for the non-pointing hand, as these are not used when analysing the pointing gestures, along with points derived from the Theia3D, for which some trials were excluded (see Appendix B) due to processing failures, or found to have 100% tracking. The non-index finger tip markers had poor and inconsistent tracking between trials, due to occlusion of markers throughout the gesture, as such we do not process hand shapes as we originally intended.

3.4 Data Labelling

We developed a semi-automatic labelling tool to identify three pointing phases [51]:

- *Start phase* involves the initial ballistic movement of the arm raising the hand into position.
- *Hold phase* is where the user is aiming at the target and the pointing pose is held steady, possibly with small refinements. The accuracy of the pointing gesture can be calculated in this phase.
- *End phase* brings the pointing gesture to an end and involves the arm returning to a neutral position.

To identify these phases we assume each pointing gesture starts from rest as they were instructed to return to a neutral position between trials. The experimenter observed participants during data collection to ensure these instructions were followed, and all pointing gestures were manually inspected posthoc to validate this was the case. We therefore expect to see two ballistic phases (start and end phases) using the velocity of the hand. We identified these phases using peak detection on both hands after smoothing the velocities with a Butterworth filter to retain only velocities above $0.2ms^{-1}$. If exactly two peaks were found on the same hand, and two peaks were not found for the other hand, the trial would be automatically labelled with the hand used and the two ballistic phases. The hold phase can then be inferred as the period between the end of the first ballistic movement and the start of the second ballistic movement.

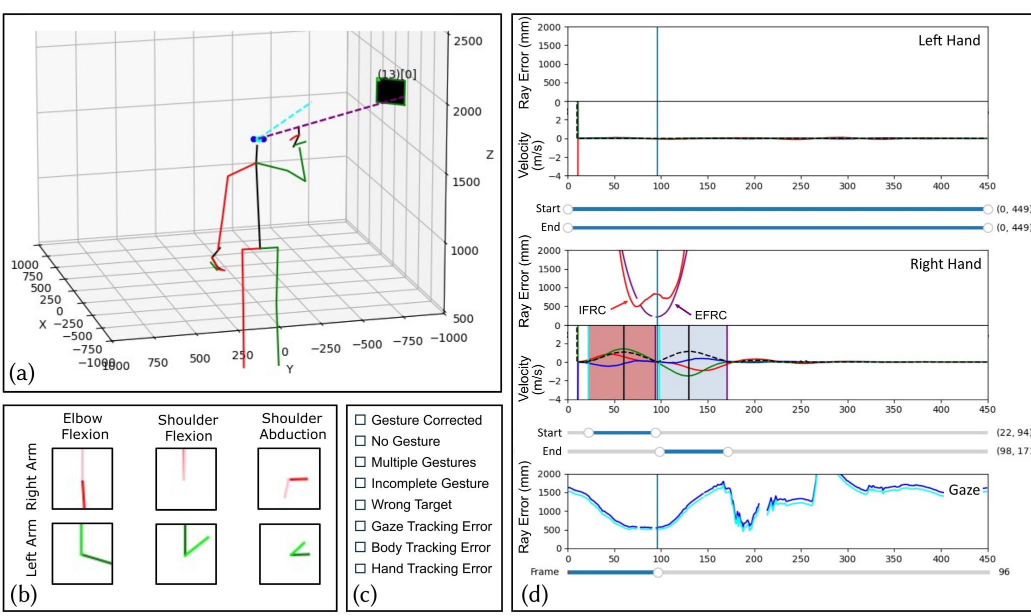

Fig. 2. An indicative screenshot highlighting the key components of the visualisation and data annotation tool used to verify or manually label pointing gesture phases. This contains: a) a render of the motion capture data, including the gaze ray (dashed cyan), the EFRC-Cyclops ray (dashed purple), and the active target (green circle on black box). b) Key joint angles for each arm, projected onto a plane around the joint. c) Exclusion flags used to drop invalid trials. d) The EFRC and IFRC ray errors for each hand (red and purple lines respectively), the labelled 'Start' and 'End' pointing phases (red and blue regions respectively) on the pointing hand plotted over the hand velocity (dashed black line), and the gaze error (cyan line to LED, blue line to target centre).

Due to the variability in the pointing gestures, we were unable to tune the peak detection to confidently capture every pointing gesture automatically. Therefore, we developed a tool [2] to visualise and manually verify each trial (see Figure 2). Figure 2a shows a render of the motion capture, gaze, and some computed rays for a given trial, which we can use to verify the pointing gesture performed was valid (i.e. started from rest, pointed to the correct target, only performed one pointing gesture). In the event the automatically detected phase labels are incorrect, or phase labels could not be derived from detected peaks, we can manually assign the pointing phases for the trial using the sliders below the hand velocity plots seen in Figure 2d.

3.4.1 Excluded Trials. In total, our dataset consists of 11,367 trials out of a possible 12,420 (91.5%). Each of the 8.5% of dropped trials were excluded trials from the dataset for one of the three following reasons: 1) data collection failure (4%), 2) data processing failure (2.33%), or 3) invalid gestures (2.15%). Data collection failures occurred during the study and were a result of data captured not being saved or being unable to be collected (e.g. recordings not being saved, hardware not collecting data). Processing failures resulted from data being collected but the required data could not be extracted (e.g. marker dropout, technical failure when extracting skeleton from video). Invalid gestures included pointing to the wrong target, performing multiple gestures, or performing a delayed correction. We do not believe this had a noticeable effect on data bias, resulting in a discrepancy of 2.8% between the conditions with the largest and smallest contributions to the dataset. A full breakdown of the reasons and number of excluded trials can be found in Appendix B.

3.5 Outcome Measures

To better understand pointing variability, we collected the following outcome measures for each participant.

3.5.1 Subjective Questionnaires. To validate that participants adopted appropriate pointing styles, and to capture their subjective experience, we used the following validated questionnaires:

- **Borg Rating of Perceived Exertion (BRPE)** is a scale from 6 to 20 which represent a subjective measure of the amount of effort and exertion required to perform a physical task [6].
- **NASA Task Load Index (NASA-TLX)** is a set of subscales used to evaluate different aspects of workload: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration [22].
- **Intrinsic Motivation Inventory (IMI)** is a validated scale measuring intrinsic motivation in the context of an activity [9, 46, 55]. We chose four of the IMI subscales: Interest/Enjoyment, Perceived Competence, Effort/Importance, and Pressure/Tension.

3.5.2 Temporal Characteristics. To determine how pointing durations changed within our study, we calculated four temporal metrics for each pointing trial:

- **Overall gesture time** is defined as the period from the start of the gesture to the end of the gesture, and represents how long it would take for a user to make a pointing gesture and reset to the neutral position, ready for another interaction (e.g. another pointing gesture).
- **Start phase duration** is the duration of the initial ballistic movement which occurs from rest until the hand "locks on" to the target.
- **Hold phase duration** is the duration in which the arm is held in place pointing at the target. In previous studies, participants were required to hold the pointing gesture for a minimum

[2]The data labelling and visualisation tool is available from: https://doi.org/10.15125/BATH-01594 [68]

time for experimental purposes (e.g. 1 second [45]). In this study, there are no restrictions on hold time so that we can understand more naturalistic behaviour.
- **Selection time** is the sum of the start and hold phase, and represents the shortest time in which a selection would have been made. This assumes that for selection, the position of the ray is calculated using all of the values during the hold phase (e.g. taking the median).

3.5.3 Biomechanical Characteristics. For each trial we compute three metrics related to the biomechanics of the gesture performed. These are:

- **Torque about the shoulder** is derived from the forces acting on the shoulder from the movement and position of the arm's centre of mass (see Figure 3). We use the equations provided by Hincapié-Ramos et al., which describe the resultant torque derived from the weight of the arm and any movement of the arm, on its centre of mass [24]:

$$Torque = (\overrightarrow{r} \times \overrightarrow{force}) - ((\overrightarrow{r} \times m\overrightarrow{g}) + I\overrightarrow{\alpha}) \quad (1)$$

where r is the distance to the centre of mass of the arm, $force$ is the force tangential to r from acceleration of the arm, m is the mass of the arm, I is the moment of inertia, and $\overrightarrow{\alpha}$ is the angular acceleration. We do not weigh participants or measure the volume of participant arms, therefore we use the average weights for each part of the arm reported by Hincapié-Ramos et al. [24].
- **NICER** is a model proposed by Li et al. [38] which estimates a user's level of exertion as a percentage of how much effort has been expended in maintaining their arm raised over a given period of time, compared to the duration of time one could maintain that effort. We calculate this for the duration of each pointing gesture. This is an extension on Hincapié-Ramos et al. 'Consumed Endurance' [24], wherein Li et al. adjust the model to work when exertion is below 15% and better represent the torque at the shoulder when the shoulder is abducted more than 90°[38].
- **Normalised hand movement** is the distance travelled by the pointing hand from the start to the end of the pointing gesture. Since we cannot assume that each participant will have the same arm span, we normalise the elapsed distance by the average length of the forearm, used when pointing, captured during the pointing gesture.
- **Handedness** represents which hand a participant chose to use for the pointing gesture. In this study there was no requirement to use a specific hand.

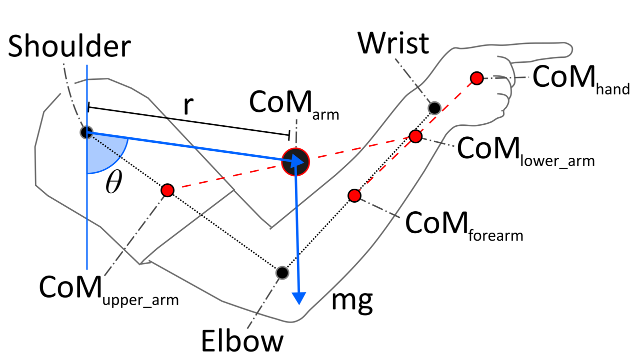

Fig. 3. Kinematic diagram of the arm, used to derive the force vector required to calculate the torque about the shoulder, using the Centre of Mass (CoM) of components of the arm.

To understand how accurately participants pointed we posthoc apply four ray casting approaches from the literature that can be categorised into two types based on the different sources for the ray: head-rooted and arm-rooted.

Head-Rooted Rays. Ray casting models originating from a point on or in the user's head. These assume pointing is performed such that the finger-tip occludes the target, and is visualised in Figure 4a:

- **Eye Finger Ray Casting (EFRC):** The ray originates within one of the user's eyes and intersect the user's index finger [28]. However, the choice of eye is difficult because eye dominance is dynamic and changes [67]. Therefore, we used the 'cyclops' eye which is the mid-point between the eyes [33].
- **Head Finger Ray Cast (HFRC):** The ray originates from the centre of the head, defined as the mid-point between the ears, and passes through the index finger-tip [8]. We explore this as an alternative approach for where the exact eye positions may not be known.

Arm-Rooted Rays. Ray casting models originating from a point on the user's arm or hand, which presume pointing is performed such that the arm/hand is orientated towards the target, visualised in Figure 4b:

- **Index Finger Ray Cast (IFRC):** Sets the ray origin to the base of the index finger (i.e. base knuckle) and determines orientation of the ray by having the ray intersect the tip of the index finger [65].
- **Forearm Ray Cast (FRC):** An elbow rooted technique, casting a ray which extends the user's forearm [51]. The origin of the ray is the elbow and passes through the wrist.

3.5.4 Pointing Accuracy. For every frame within each trial, we computed the four rays for the pointing hand as a yaw and pitch normalised by the normal of the plane defined by the torso. The angular error is then calculated using the pythagorean theorem using the median error for the pitch and yaw values during the hold phase of the pointing gesture. The errors are calculated based on the angular offset to the point at which the rays intersect with the plane for that specific set of LED targets.

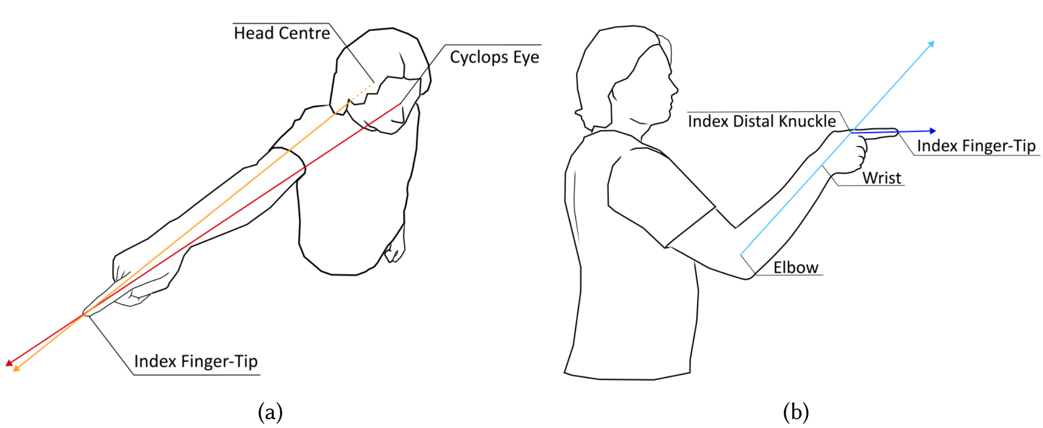

Fig. 4. Visualisations of different ray cast models. a) visualisation of different head-rooted ray casting approaches: HFRC (orange), EFRC-Cyclops (red), and b) visualisation of different arm-rooted ray casting approaches: IFRC (blue), FRC (cyan).

Systematic Offset Compensation Models. Prior work has shown that systematic offsets exist when pointing, and that these offsets can be corrected for using ordinary least squares models [44, 45]. We therefore calculate corrected rays using Mayer et al.'s approach to systematically compensate for pointing error [45], which has been shown to have good external validity [44].

For each ray-casting approach, two error compensation models were created, one to correct yaw and the other to correct pitch, both of which use the polynomial model from Mayer et al. [45]:

$$f(x,y) = ax^4 + by^4 + cx^3y + dxy^3 + ex^3 + fy^3 + gx^2y^2 + hx^2y + ixy^2 + jx^2 + ky^2 + lxy + mx + ny + o \quad (2)$$

Where $f(x,y)$ is the estimated error, x is the median yaw of the pointing gesture, and y is the median pitch. Due to the differences in pointing variability elicited during this study and the wider range of target positions, we trained the offset correction models on our dataset rather than using the previously reported models, with the model coefficients available in Appendix C.

3.6 Procedure

Figure 5 provides an overview of the procedure. Upon arrival, we confirmed that participants were able to partake in the study by verifying that they can, without the use of glasses, identify the LED targets, and read the Stroop test words and distinguish the colours (purple, brown, red, green, and blue). If not, they were excluded from the study. In the event they only had an issue with two colours in the Stroop test, we omitted one of the colours. We then provided them with a participant information sheet and they signed a physical consent form. We then tested the participant's ocular dominance and collected participant demographics (e.g. hand dominance, age, familiarity with pointing interaction). Participants began with a warm-up session to familiarise them with the study setup. This involved a sequence of five targets which they were asked to point at in an accurate manner. Once participants were comfortable pointing accurately, they familiarised themselves with pointing casually. This allowed participants to understand the differences between conditions. Once participants were comfortable with the setup we attached the hand tracking markers and fitted the eye tracking glasses. We captured a 10 second recording of the participant flexing their

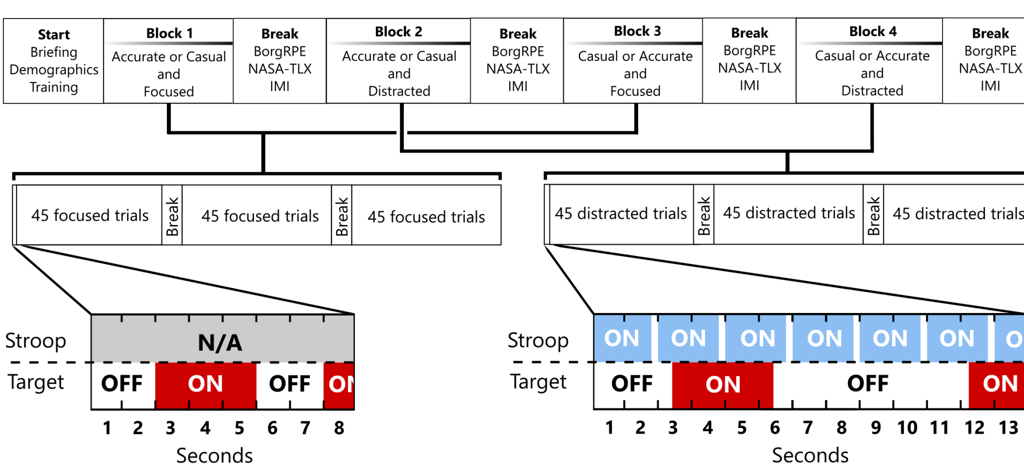

Fig. 5. A visual representation of the study procedure showing how (Top row:) each participant completed four blocks of pointing, where (Middle row:) each block consists of 135 trials split up into three recordings with an optional break in between, and (Bottom row:) each trial consists of a 3-second LED with 2-second break in the focused condition, while multiple Stroop tasks were incorporated in the distracted condition.

fingers and rotating their hands while keeping them in-place, which we manually labelled and used as reference for the QTM AIM models.

With the briefing, setup, and warm-up concluded, we began the data collection. We counterbalanced Pointing Style using a balanced Latin square design because prolonged pointing with an extended arm can cause arm fatigue (also referred to as Gorilla Arm [7]). Within Pointing Style, participants always performed the non-distracted pointing condition before the distracted condition so that they were familiar with the pointing task. As well as counterbalancing for Pointing Style we allowed for breaks between sets of pointing trials to reduce any effect of fatigue. Each condition was split into three recordings of 45 pointing trials each – 3-4 minutes in the focused conditions, 4-5 minutes in the distracted conditions. One recording covered all 15 target clusters, and three targets per target cluster. Each recording used the same targets that were pre-selected, but presented in a randomised order. Participants were explicitly instructed to return to a neutral, rest position in between pointing gestures.

In the focused condition the target LEDs lit up for 3 seconds with a 2 second pause in-between to allow participants to return to a neutral, rest position. In the distracted condition the targets lit up for 3 seconds but were designed to light up 0.5 seconds before or after a subsequent Stroop test, with some randomness either side to reduce predictability. This was designed so that pointing disrupted, or was disrupted by, the Stroop test. This resulted in an average delay between target deactivation and activation of 2.49 seconds, with there always being at least two Stroop tests provided while a target was active. The Stroop test was administered with a 2 second cadence (displayed for 1.9 seconds, with a 0.1 seconds with the screen cleared). As two subsequent Stroop tests could be the same, the display of a new Stroop was accompanied by an audible notification. There was always a sequence of three Stroop tests prior to the activation of the first LED target.

Between conditions, participants completed the BRPE, IMI, and NASA-TLX questionnaires and the Tobii glasses were re-calibrated. At the mid-point and end of the study, participants were interviewed regarding the accurate and pointing styles. Before the first Stroop task, participants were introduced to the task and given a short familiarisation session so that they were familiar with how to answer the Stroop and point at the same time. Once all conditions were complete participants conducted an exit interview. The study took approximately 2 hours (~20 minutes briefing and setup, ~25 minutes for each condition).

3.7 Participants

The final dataset consists of data collected from 23 participants (8 identified as male, 15 as female), aged 20 - 56 (M = 26.95, SD = 8.28) who were recruited through mailing lists, social media, and posters. 28 participants took part in the study but due to technical issues resulting in missing data only 23 were included for analysis. This resulted in 12 participants in the Accurate–Casual split, and 11 participants in the Casual–Accurate split. All participants were screened prior to taking part to ensure they were aged 18 or over, had normal or corrected to normal hearing and vision, could distinguish between the colours used within the Stroop test, and did not have any movement-related conditions. 22 participants reported themselves as right-handed and one reported to be ambidextrous. 11 participants reported that they were right-eye dominant, and 11 reported that they were left-eye dominant, one participant had ambiguous ocular dominance and was not assigned an eye dominance. The participant who was ambidextrous reported left-eye dominance. We also measured participants' experience with spatial interaction and pointing. 14 stated that they had no prior experience with a pointing interaction system, seven reported having some exposure, and two reported multiple times. The data collection was given a favourable opinion by the Biomedical Sciences Research Ethics Committee at the University of Bath, which is one of the

Table 2. Angular error (°) for each ray with and without the adjustment from offset correction models [45]. We report these as the Median (and IQR).

	Error (°)		
Ray	Uncorrected	Corrected	p
IFRC	14.2 (8.2)	7.78 (3.22)	$< .001^{***}$
FRC	19.2 (8.00)	5.61 (5.12)	$< .001^{***}$
HFRC	3.11 (1.54)	2.54 (1.26)	$< .001^{***}$
EFRC (Cyclops)	3.03 (2.65)	2.92 (1.52)	.151

university's central institutional ethics review boards. Participants were compensated £20 for their time.

4 CONDITION ANALYSIS

We analysed the temporal, biomechanical and accuracy outcome measures by averaging them over each participant for the four conditions (Pointing Style × Focus) so that we could perform two-way repeated measures ANOVAs (RM-ANOVA). We checked for normality using Shapiro-Wilks test, and for presence of extreme outliers defined as three times the interquartile range (IQR) above the third quartile (Q3) or three times the IQR below the first quartile (Q1). We performed an aligned rank transformation prior to running the RM-ANOVA when one of the cells in our 2×2 design was not normally distributed or if there was an extreme outlier present [70]. When the data is normally distributed we report the mean (M) and standard deviation (SD), if not we report median (Mdn) and interquartile range (IQR) which are better representations of central tendency and variability for skewed distributions.

For brevity, we report full details of the statistical tests including partial eta-squared (η_p^2) with 5% and 95% confidence intervals in brackets in Appendix D and Appendix E. The interpretation values are $\eta_p^2 = 0.01$ (small), $\eta_p^2 = 0.06$ (medium), and $\eta_p^2 > 0.14$ (large). Wilcoxon tests are reported with effect size, r, which varies from 0 to (close to) 1, where 0.10 - 0.3 (small effect), 0.30 - 0.5 (moderate effect) and ≥ 0.5 (large effect). The analysis was conducted in R 4.5.1, with code and data available in supplementary material.

4.1 Application of Systematic Offset Compensation Models

Prior work has shown that systematic offsets can be corrected for during pointing tasks [44, 45], discussed in subsubsection 3.5.4. We begin by applying these models to our dataset to understand their effect. Appendix C details the model coefficients derived through fitting the model polynomials to our dataset via ordinary least squares.

Figure 6 shows the distributions for the different rays, with and without correction (which are agglomerated per participant for RM-ANOVA analysis). There was a significant interaction between the ray casting approaches and corrective model on the angular error, $F(3, 154) = 70.07$, $p < .001$, $\eta_p^2 = 0.577$ [0.494, 1.000]. We checked whether the corrective models reduced the error for each ray using a Wilcoxon test per ray and posthoc correcting the four tests with Holm-Bonferroni corrections. Table 2 shows the median errors for each ray across the whole dataset. All rays are significantly improved by the systematic compensation model ($p < .001$, $r \geq 0.786$) apart from EFRC Cyclops ($p = .151$, $r = 0.304$). As a result, we use the corrected models for the rest of the analyses because they improve the accuracy for three out of the four rays, especially in the case of the arm-rooted rays. Interestingly, FRC is significantly worse than IFRC when not corrected ($p < 0.001$), however it becomes more accurate once the systematic compensation is applied ($p = 0.016$). This

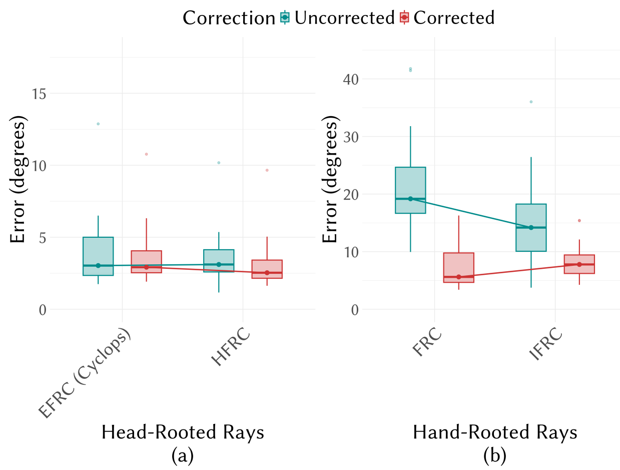

Fig. 6. Box plots showing the distributions for the corrected and uncorrected ray-casting approaches, split up into: (a) head-rooted rays – EFRC (Cyclops) and HFRC, and (b) arm-rooted rays – FRC and IFRC

Table 3. Table showing the median and interquartile range for participants across all subjective measures with the exception of IMI Interest Enjoyment which is mean and standard deviation because the data is normally distributed. Significant main effects are highlighted using *** for p < .001, ** for p < .01, and * for p < .05. Significant interactions between Pointing Style and Focus are highlighted using †. Full statistical details can be found in Appendix D.

		Pointing Style		Focus	
		Accurate	Casual	Focused	Distracted
	BRPE	13.00 (3.75)***	9.5 (5.00)***	9.00 (3.00)***	13.0 (4.00)***
NASA-TLX	Effort	62.50 (48.75)*	55.00 (53.75)*	32.50 (38.75)***	77.50 (22.50)***
	Frustration	20.00 (38.75)	17.50 (40.00)	12.50 (15.00)***	37.50 (38.75)***
	Mental Demand	57.50 (48.75)**	27.50 (53.75)**	20.00 (17.50)***	70.00 (25.00)***
	Physical Demand	50.00 (30.00)***	27.50 (40.00)***	30.00 (33.75)***	52.50 (42.50)***
	Temporal Demand	50.00 (48.75)	45.00 (48.75)	22.50 (30.00)***	65.0 (20.00)***
	Performance	85.00 (23.75)	82.50 (28.75)	85.00 (15.00)	80.0 (35.00)
IMI	Effort Importance	5.70 (1.90)**	4.70 (2.60)**	4.20 (2.60)***	5.70 (1.75)***
	Interest Enjoyment	4.93 (1.39)	4.76 (1.31)	4.50 (1.32)**	5.20 (1.29)**
	Perceived Comp.	5.67 (1.29)	5.42 (1.12)	5.50 (0.96)*	5.92 (1.46)*
	Pressure Tension	3.10 (1.90)	2.80 (2.35)	2.40 (1.60)***	3.60 (1.80)***

mimics results in previous research that showed the same phenomena [45]. Similarly, while HFRC is significantly better than EFRC with the systematic offset correction applied ($p < 0.001$), there is no significant difference without ($p = 0.345$).

4.2 Manipulation Check

The manipulation checks validate that the participants self-reported responses aligned with the respective conditions and that the study instructions elicited different types of pointing. Table 3 shows the results of the subjective questionnaires, and Appendix D provides details about the

Table 4. Table showing the median and interquartile range for participants across all outcome measures based on condition. Significant main effects are highlighted using *** for p < .001, ** for p < .01, and * for p < .05. Significant interactions between Pointing Style and Focus are highlighted using †. Full statistical details can be found in Appendix E.

	Pointing Style		Focus	
	Accurate	Casual	Focused	Distracted
Overall Gesture (ms)†	2305.06 (700.28)***	1787.32 (322.32)***	2210.66 (772.05)***	1864.50 (402.00)***
Start Phase (ms)	801.01 (94.10)***	728.61 (105.63)***	787.52 (139.58)***	747.68 (116.96)***
Hold Phase (ms)†	587.33 (558.41)***	257.59 (166.78)***	535.38 (689.70)***	276.60 (288.77)***
Selection Time (ms)†	1453.07 (548.29)***	962.61 (248.51)***	1275.24 (832.64)***	1028.35 (289.41)***
Shoulder Torque (Nm)	14.42 (4.05)*	14.17 (3.01)*	14.27 (3.69)	14.23 (3.10)
NICER (%)†	1.70 (1.11)***	1.07 (0.61)***	1.41 (1.10)***	1.06 (0.84)***
Norm. Hand Movement	6.93 (1.03)***	6.45 (1.36)***	6.86 (1.02)*	6.79 (1.27)*
Right Hand Usage (%)	61.05 (40.42)***	59.77 (16.57)***	60.15 (45.08)	60.63 (42.60)
EFRC (Cyclops) (°)†	2.58 (0.95)***	3.09 (2.27)***	2.73 (0.99)**	2.96 (2.25)**
HFRC (°)†	2.06 (0.92)***	2.84 (2.06)***	2.08 (1.09)***	2.72 (2.16)***
IFRC (°)	6.75 (2.43)***	8.03 (4.47)***	6.85 (3.16)**	7.74 (4.31)**
FRC (°)	5.17 (2.12)**	6.27 (5.36)**	5.31 (2.81)	5.93 (5.35)

statistical tests including effect sizes. Participants reported that they exerted themselves less when pointing casually, and that this was significantly less mentally and physically demanding than pointing accurately. This demonstrates that we correctly elicited different styles of pointing from participants, aligning with the instructions provided to point in a less physically demanding manner. We also see how pointing during the distracted condition was significantly different across all the subjective measures apart from NASA-TLX performance, showing that participants thought they performed equally well in all conditions. Interestingly, despite being more demanding and frustrating, participants found the distracted pointing condition to be more enjoyable – likely due to the abstract nature of the focused pointing condition.

4.3 Temporal Analysis

Table 4 provides descriptive statistics for the different temporal characteristics which includes details about significant main effects, full statistical details can be found in Appendix E. Figure 7 visualises the distributions for the temporal characteristics across conditions.

We observed how casual pointing is significantly quicker than accurate pointing at each individual phase of the pointing gesture, as well as for the overall gesture time. In addition, the temporal characteristics were much less variable (based on IQR) when pointing casually compared to the accurate condition. This means that participants were significantly quicker in the start phase, which consists of the initial ballistic arm movement, when pointing casually, implying that either their hand movements are quicker or that they are not moving the arm as far. Once the target was acquired, participants held the pointing gesture for over twice as long in the accurate condition, compared to pointing casually. Similarly, participants pointed significantly more quickly during the distracted condition compared with the focused condition. We also observed significant interactions between Pointing Style and Focus on the overall gesture time, the hold phase, and the selection time. This can be seen in Figure 7, which shows that being distracted affects accurate pointing much more, resulting in bigger differences between the focused and distracted conditions.

4.4 Biomechanical Analysis

Table 4 provides descriptive statistics for the biomechanical characteristics which includes details about significant main effects, full statistical details can be found in Appendix E. Figure 8 shows the distributions across conditions for shoulder torque, NICER, normalised movement of the pointing hand, and hand usage. These outcome measures represent the physical effort, strain, and mechanical constraints involved in the pointing interactions and add further context to the temporal differences we see across conditions.

There is no significant difference in shoulder torque between the distracted and focused task, which could imply that participants moved quicker in order to achieve lower gesture times. However,

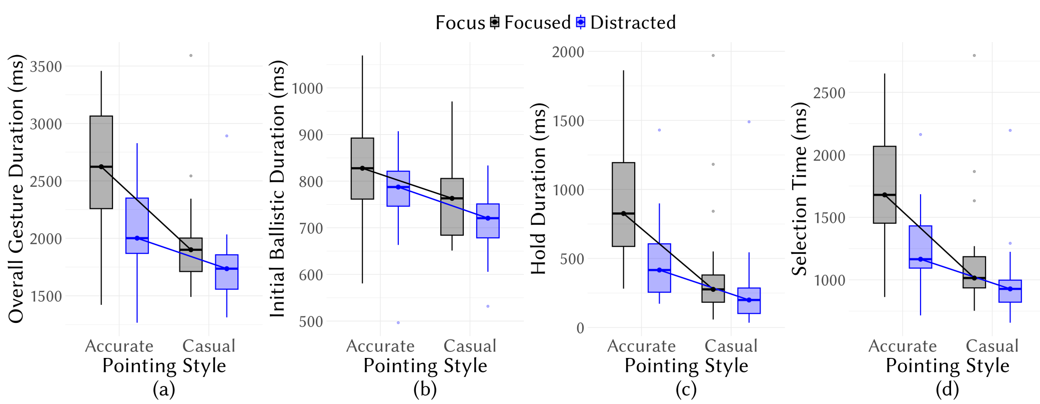

Fig. 7. Box plots showing the distributions and effect of Pointing Style and Focus on the temporal characteristics of the pointing gestures: (a) overall gesture time, (b) start phase duration, (c) hold phase duration, and (d) selection time. Lines connect the medians to better demonstrate the relationship between distracted and focused conditions.

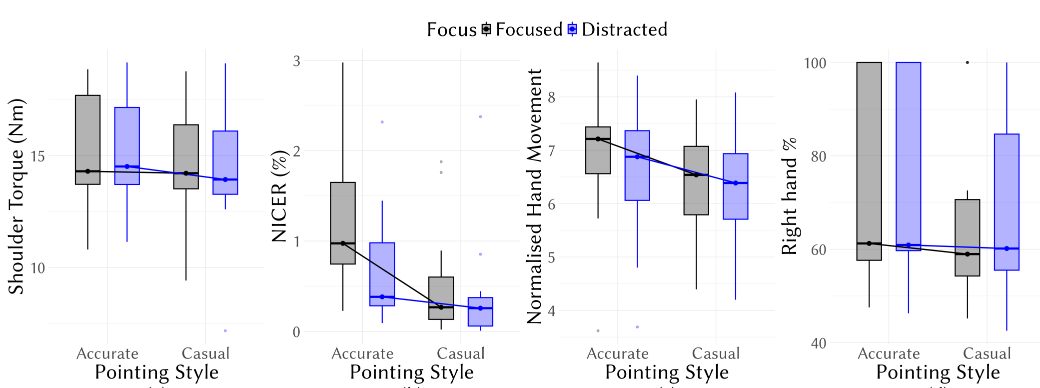

Fig. 8. Box plots showing the distributions and effect of Pointing Style and Focus on the biomechanical characteristics of the pointing gestures: (a) shoulder torque, (b) NICER, (c) normalised hand movement, (d) handedness. Lines connect the medians to better demonstrate the relationship between distracted and focused conditions.

we observe a significant increase in shoulder torque for accurate pointing which suggests not all casual pointing was as physically straining.

The NICER metrics demonstrate that cursorless pointing is low-effort across all conditions, with values between 0–3.5%. They also reveal how participants expended significantly less effort in the casual pointing condition, which is what we expected. Interestingly, they also expended less effort during the distracted pointing condition. There was also a significant interaction, and Figure 8b shows that the difference in NICER due to Focus is larger in the accurate condition.

Participants were found to move their hand significantly more in the accurate condition compared to the casual condition, with a small, but significant, increase in hand movement in the focused condition compared to the distracted condition. Figure 8c shows the decrease in hand movement (normalised by the participant's forearm length) between the accurate and casual conditions.

All participants recorded themselves as right handed except one who identified as ambidextrous. For this analysis, we look at the percentage of trials in which participants used their right hand. Figure 8d shows the difference in right (i.e. dominant in this case) hand usage across conditions, demonstrating how participants were more likely to use their right hand when pointing accurately.

4.5 Accuracy Analysis

We have seen how casual pointing, and pointing when distracted, affects temporal and biomechanical features for the better with reduced gesture times and effort. However, we would expect these to be a trade-off with the accuracy of the pointing gestures. In this section, we explore how Pointing Style and Focus affect pointing accuracy for four ray casting approaches: EFRC (Cyclops), HFRC, IFRC, and FRC. Distributions for the error with each ray in each condition are shown in Figure 9. Table 4 provides descriptive statistics for the ray casting accuracies which includes details about significant main effects, full statistical details can be found in Appendix E.

We observed similar patterns for the head-rooted rays: EFRC (Cyclops) and HFRC. A significant interaction for both showed that pointing during the casual–distracted condition was significantly

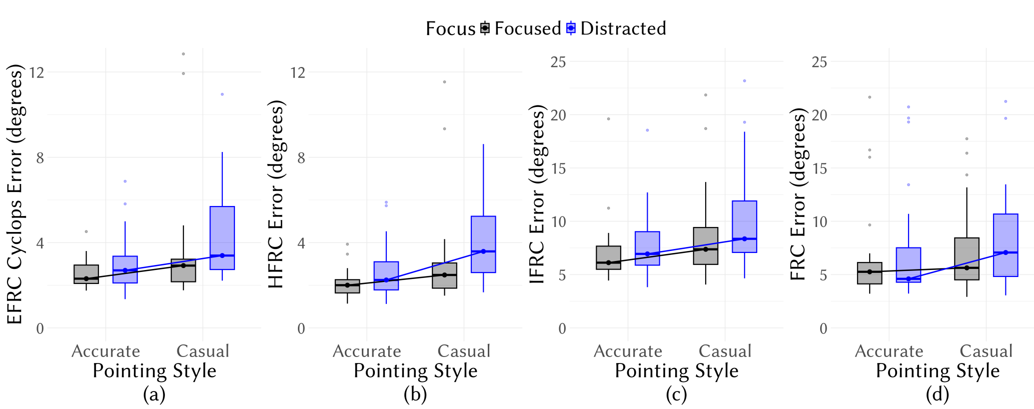

Fig. 9. Box plots showing the distributions and effect of Pointing Style and Focus on the accuracy of the pointing gestures using four different ray-casting approaches: (a) EFRC (Cyclops), (b) HFRC, (c) IFRC, and (d) FRC. Lines connect the medians to better demonstrate the relationship between distracted and focused conditions.

less accurate than all other conditions. Focus did not affect pointing accuracy in the accurate condition – i.e. when asked to point accurately, participants were just as accurate in the focused condition as the distracted. For the EFRC (Cyclops) ray, the casual–focused condition was not significantly different to either of the the accurate conditions, however for HFRC the accurate–focused condition was significantly more accurate than than the casual–focused condition.

There were no significant interactions between Pointing Style and Focus for the hand-rooted ray casting approaches: IFRC and FRC. For IFRC, accurately pointing produces significantly less error than casual pointing, and pointing when distracted resulted in a significantly higher error compared with focused pointing. For FRC, there is only a significant effect for Pointing Style, where casual pointing is significantly less accurate.

4.6 Condition Summary

The overall summary statistics for the conditions in Table 3 and Table 4 demonstrate that our study design successfully elicited different pointing techniques that significantly affect the temporal, biomechanical, and accuracy characteristics of the pointing gestures. Our results show how, when given the choice, participants performed different styles of pointing both objectively and subjectively depending on the condition. In particular, when pointing casually participants pointed faster in the ballistic phase, held the pose for less time, and did so in a manner that is less fatiguing. However, casual pointing is significantly less accurate. When using EFRC the median participant error for casual pointing is worse by 0.51° (equivalent to 17.8 mm at 2 m and 26.8 mm at 3 m). This is more pronounced for IFRC where the median participant error is 1.28° less accurate (equivalent to 44 mm at 2 m and 67.2 mm at 3 m). However, pointing casually is much more variable than pointing accurately, especially when participants were distracted. For example, the 90th percentile participant error for the casual distracted pointing condition is worse by 4.64° (equivalent to 162.41 mm at 2 m and 243.62 mm at 3 m) than pointing accurately and focused. Similarly, this is more pronounced for IFRC where the median participant error is 9.18° less accurate (equivalent to 325.02 mm at 2 m and 487.48 mm at 3 m) for casually pointing when distracted compared with accurately pointing when focused. The figures and box plots also show how there is a lot of variability across participants. In the next section, we will explore the underlying causes of these differences in more depth.

5 BEHAVIOURAL ANALYSIS

In this section, we explore the underlying variability observed in the data in section 4. We begin by providing a description of the approach taken to extract and characterise different pointing behaviours. In this section, we do not perform null hypothesis testing because the identified pointing behaviours have different numbers of trials associated with them. Instead, we use descriptive statistics and visualisations to describe and analyse the data. Unlike section 4, all visualisations are plotted over the whole dataset without agglomerating per participant or condition, unless stated.

5.1 Methodology

We analysed all the pointing gestures to extract distinct but common pointing behaviours across participants. We first visualised the data using histograms and trial renders (e.g. Figure 2), and used knowledge gained from observations during the unsuccessful clustering approaches to identify relevant features. We then explored several unsupervised clustering techniques, with and without PCA dimensionality reduction, including K-Means, Agglomerative, and DBScan clustering. Some of these techniques highlighted key features in pointing gestures, such as elbow flexion and shoulder abduction, due to the emergence of clusters around specific values for these features, e.g. a high elbow flexion irrespective of shoulder abduction, or a low shoulder abduction irrespective of elbow flexion. However, due to overlapping expression of these features, clusters were also found from interactions between these features, such as clusters for high elbow flexion and low shoulder abduction, along with many other features. Additionally, clusters produced using these techniques had poorly defined boundaries, in part due to the aforementioned overlapping of feature expression, but also due to the high and continuous variation in the data between participants and targets. Attempts to perform clustering on the data subsets, e.g. by participant or target, did not improve the clustering process. Therefore, we developed a semi-supervised method to classify and label the pointing gestures based on hierarchical Gaussian Mixture Model clustering.

5.1.1 Hierarchical Gaussian Mixture Model Clustering. At a high-level, we performed a form of hierarchical clustering, where instead of slowly merging subsets of data based on a distance metric (starting from each data-point as it's own subset), we segment the dataset into smaller and smaller subsets via the use of Gaussian Mixture Models (GMMs). Figure 10 illustrates the process which involves performing GMM fitting with 2–10 components to identify prominent components within the features of the gestures (Figure 10a), selecting the number of components that has the best BIC score for each feature (Figure 10b), and selecting the feature with the greatest variance once normalised by the min-max range for the feature (Figure 10c). Once we have selected a feature and fitted Gaussian components for that feature, we split the current data by the intersections of the Gaussians (Figure 10d). This is visualised in Figure 10d for the shoulder abduction feature, in which

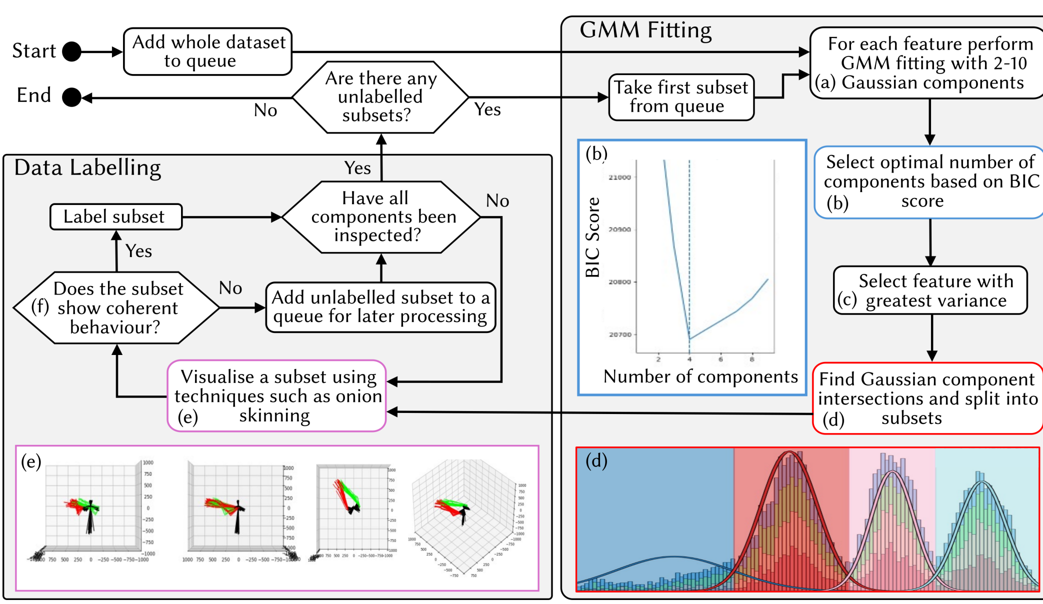

Fig. 10. A flow chart that illustrates the semi-supervised approach we used that uses Gaussian Mixture Modelling to identify unique pointing behaviours. Important steps are highlighted with (a)–(f), with additional figures used for (b), (d), and (e). (b) highlights the BIC score for GMM fittings across a range of components, upto ten, where four is the optimal number of components. (d) shows the calculated intersections of the four Gaussian components for the shoulder abduction feature, despite only having three rows of targets in the study. (e) shows a visualisation of multiple trials within a subset of data overlaid atop of each-other (onion skinning) from four perspectives.

there are four components even though there are only three rows of targets. For each of these new subsets, we visualise them via onion skinning from multiple perspectives (Figure 10e) and via trial renders (e.g. Figure 2) sampled within the subset. We inspect these subsets to check if a given subset contains a common set of poses that could be identified as a single, coherent behaviour (Figure 10f). If this is the case we do not perform any further segmentation on the subset and save it with a label describing the behaviour. Otherwise we add the subsets to the back of the processing queue and begin the process with the next unlabelled subset. This was repeated until all trials were labelled, resulting in a fully annotated dataset of different but unique pointing behaviours.

Once the above process was completed, we used Decision Trees to fit to the data to extract the ranges within relevant features which describe the pointing behaviour. This allows us to describe the behaviours beyond rendered examples and written descriptions of the body pose. Based on preliminary testing, we set the maximum depth to six, and limited the trees to a maximum of five features. We performed 5-fold cross-validation, from which we took the model with the greatest accuracy.

5.1.2 Gesture Features. To identify pointing behaviours we computed target invariant features to describe each pointing gesture, derived from the motion capture within each of the pointing phases. We used target invariant features so that future pointing gestures can be classified into a behaviour without the system requiring *a priori* knowledge of the target the user is pointing at. We started with 348 features, however only a subset were used in our analysis. We reduced the set of features by dropping correlated features (where $p > 0.8$), retaining the features with the greatest variance, and based on prominent features we observed during the initial unsupervised clustering attempts. This resulted in nine features which are detailed below.

Joint Distances. Due to the different body proportions between participants, we normalised the distances by the median distance found between the participant's elbow and hand for the active pointing arm and used the following joint features in the behaviour labelling process. We used (1) *Hand elevation* which is the distance of the hand from the shoulders, relative to the transverse plane, and (2) *Hand extension* defined as the distance of the hand away from the near shoulder.

Joint Angles. An extension on joint distances proposed in [41], where the angle between two joints is calculated. We used (1) *Elbow flexion* computed as the angle between the near shoulder and hand, about the elbow. (2) *Shoulder abduction* defined as the pitch between the elbow and far shoulder, about the near shoulder, relative to the transverse plane. (3) *Finger flexion* which is the angle between the index finger-tip and the elbow about the wrist. (4) *Hand yaw* about the near shoulder, relative to the sagittal plane.

Ray Alignments. We normalised the rays to the sagittal axis and observed unexplained variance in the alignments between different rays that did not appear to be correlated with target direction. The rays also represent conceptually important features for pointing (e.g. where the finger is directed). We used (1) the cosine between the EFRC and IFRC rays, (2) the change in yaw between the EFRC and FRC rays, and (3) the cosine between the EFRC and FRC rays.

5.2 Pointing Behaviours

Based on our analysis, we found three unique pointing behaviour traits: (1) *Arm Pose*, (2) *Finger Incorporation*, and (3) *Hand Torso Alignment*, each of which is composed of three different behaviours. We first provide an overview, before exploring each trait in more depth and how it affects pointing performance. Features reported to distinguish between different pointing traits

Table 5. The number of trials for each combination of arm pose and finger incorporation, with percentages representing how many trials made up the arm pose. This demonstrates how outstretched ADS is the most common type of pointing, but that occluded pointing is more common for the proximal arm pose.

Arm Pose	Finger Incorporation	Number of trials
Outstretched	ADS	7749 (78.8%)
	Occluded	1577 (16.0%)
	Unoccluded	512 (5.2%)
Proximal	ADS	304 (26.6%)
	Occluded	581 (50.8%)
	Unoccluded	258 (22.6%)
Hip Fire	Finger	386 (100.0%)

are derived from the Decision Trees which had accuracies of 92.1% (arm pose), 82.5% (finger incorporation), and 91.7% (hand torso alignment) – full confusion matrices can be found in Appendix F.

Arm Pose describes how a user holds their arm during the pointing gesture, and can be classified into either *outstretched*, *proximal*, or *hip fire*:

- *Outstretched* involves the hand being held at arm's-length, with an elbow flexion greater than 132°. Outstretched pointing is the 'full' form pointing described by Bangerter and Chevalley [2] and is the most common type of free hand-based pointing in the HCI literature.
- *Proximal* pointing involves raising the hand to eye-level, but with the hand kept close to the head with an elbow flexion below 138°. Proximal pointing is a form of 'partial' pointing [2].
- *Hip Fire* is performed with minimal elevation of the arm and abduction of the shoulder, less than −45° from the sagittal axis, with much of the hand placement being controlled by movement about the elbow. Hip fire pointing is another form of 'partial' pointing [2].

Finger Incorporation describes how one utilises their finger while pointing and is closely linked with arm pose (Table 5 shows the combinations of arm pose and finger incorporation). Finger incorporation can be classified into either *aim down sights (ADS)*, *occluded*, or *unoccluded*:

- *Aim Down Sights (ADS)* involves looking down the index finger at the target, such that the IFRC and EFRC rays are aligned within 20°. This was the most common usage of the finger during outstretched pointing.
- *Occluded* involves the finger-tip occluding the target while the IFRC ray is not pointing towards the target which was the most common finger usage in the proximal pointing style. This is similar to the Gaze&Finger interaction explored in [40], wherein participants occlude the target with their finger-tip, as a selection mechanism for gaze interaction.
- *Unoccluded* is where the finger vector is being used to primarily indicate target direction while the IFRC ray does not closely align with the EFRC ray and the finger-tip does not occlude the target. This is the only usage of the finger for hip fire pointing.

Arm Torso Alignment describes how someone places their arm with respect to their torso and the sagittal axis and can be classified into either *medial*, *aligned*, or *lateral*:

- *Medial:* The hand is placed such that the yaw about the shoulder is less than −35°. We would expect to see this alignment for the left hand targets if pointing with the right hand.
- *Aligned:* The hand is placed ahead of the body, which results in a yaw about the shoulder between −35° and 2°.

Table 6. Gesture performance metrics for the arm pose trait. Temporal, biomechanical, and accuracy metrics are reported as median (and IQR). Behaviour distributions are reported as number of trials and the percentage.

	Measure	Outstretched	Proximal	Hip Fire
Temporal (ms)	Overall Gesture Duration	2050 (770)	1750 (500)	1505 (420)
	Start Phase Duration	760 (180)	690 (200)	590 (150)
	Hold Phase Duration	410 (600)	190 (290)	170 (260)
	Selection Time	1200 (690)	940 (340)	810 (310)
Biomechanical	Shoulder Torque (Nm)	14.97 (3.31)	12.64 (2.71)	8.65 (3.84)
	NICER (%)	1.38 (1.03)	0.60 (0.66)	0.28 (0.21)
	Norm. Hand Movement	6.82 (2.80)	5.91 (2.00)	4.26 (1.04)
	Right Handed	6699 (68.09%)	965 (84.43%)	288 (74.61%)
Accuracy (°)	IFRC	6.25 (5.34)	11.69 (11.84)	17.28 (13.92)
	FRC	4.51 (4.43)	17.91 (13.63)	19.33 (14.26)
	EFRC-Cyclops	2.24 (1.99)	3.91 (4.33)	8.40 (21.84)
	HFRC	2.00 (1.76)	3.44 (3.59)	8.65 (18.67)
Conditions	Number of participants	23 (100.00%)	13 (56.52%)	9 (39.13%)
	Number of Trials	9838 (86.55%)	1143 (10.06%)	386 (3.40%)
	Accurate-Focused	2661 (27.05%)	250 (21.87%)	0 (0.00%)
	Accurate-Distracted	2439 (24.79%)	246 (21.52%)	48 (12.44%)
	Casual-Focused	2515 (25.56%)	371 (32.46%)	135 (34.97%)
	Casual-Distracted	2223 (22.60%)	276 (24.15%)	203 (52.59%)

- *Lateral:* The hand is placed away from the body, such that the yaw about the shoulder is $> 2^\circ$. Pointing laterally aligns the hand, arm, and gaze with the target in a straight line compared with aligned pointing. Interestingly, the lateral stance is commonly seen in Olympic shooting competitions due to better alignment between the eyes and hand when the arm is held laterally.

5.2.1 Arm Pose. Table 6 shows the breakdown of the outcome measures for the different arm poses and Figure 11 shows a visual representation of the different arm poses. In section 2, we can see how the vast majority of interaction studies focus on, or observe, 'full' pointing gestures [2]. Full pointing is where the hand is brought up to point while maintaining a straight arm. This was the most common type of pointing observed in our study, which we denote as outstretched pointing.

How participants used different arm poses varied a lot within our dataset. Ten participants showed no change in arm pose, adopting only the outstretched arm-pose (P1, P3, P7, P8, P10, P11, P14, P15, P19) and instead only changed the temporal characteristics across conditions, for example spending longer in the hold phase during the accurate condition. A further seven used outstretched for over 90% of their pointing gestures. However, six participants performed 'partial' pointing for at least 30% of their gestures. In 'partial' pointing, the arm has a bend about the elbow that is typically less than 140°. We identified two sub-sets of partial pointing: 1) proximal and 2) hip fire.

Proximal involves elevating the hand as would be done with outstretched pointing, however the hand is kept much closer to the face. It is similar to gaze-hand alignment techniques that have been recently introduced for selection [66]. Proximal pointing was observed across all conditions, although it was most commonly used in the Casual-Focused condition. Proximal pointing require less effort to perform and is quicker than outstretched pointing, however it is nearly twice as inaccurate. This could be due to the shorter distance between the ray origin and intersection, which

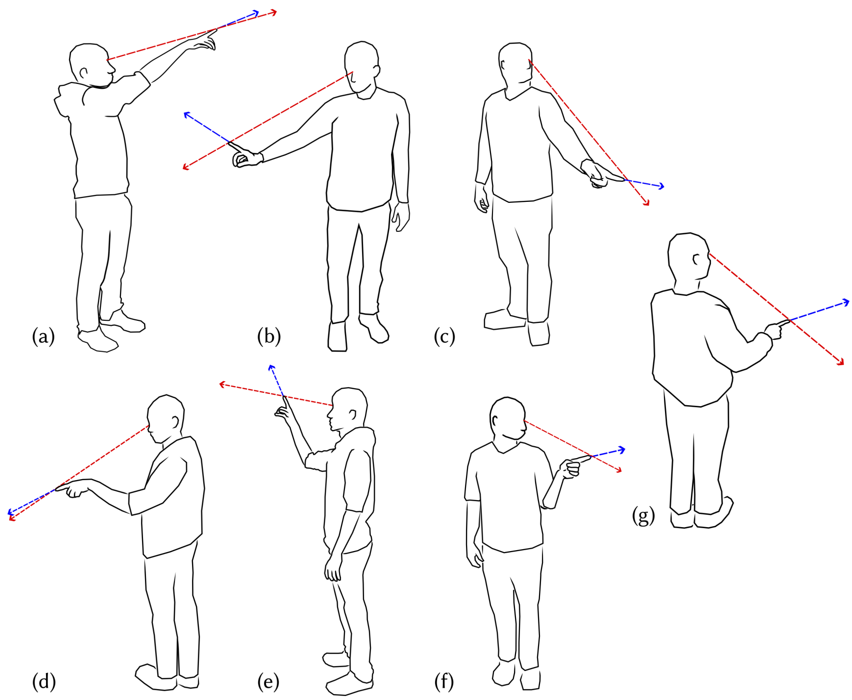

Fig. 11. We identified three unique pointing traits including different arm poses and finger incorporations. The outstretched arm pose (a)–(c) was the most common type of arm pose with the finger (a) ADS, (b) occluded, and (c) unoccluded. The proximal arm pose (d)–(f) was the second most common with all three types of finger (d) ADS, (e) occluded, and (f) unoccluded. Finally, (g) shows the hip fire style pointing technique. The red dashed rays represent the EFRC ray, and the blue dashed rays represent the IFRC rays.

amplifies small errors in the alignment of the ray. Despite this, proximal pointing achieves a median accuracy for corrected head-rooted rays of below 4° which equates to about 14 cm at 2 m. Six participants performed proximal pointing for more than 10% of their trials (P4, P5, P13, P20, P22, P23), some choosing to point proximally to be accurate, while others chose proximal for casual. Only three participants performed proximal pointing other than outstretched pointing for the majority (P4, P5, P13). Two of these participants also performed proximal pointing when pointing medially (P4, P13).

Hip fire differs from proximal because the majority of the positioning of the hand occurs due to flexion of the elbow and rotation of the upper arm, with minimal abduction of the shoulder ($\leq -45°$ compared to the transverse plane). Hip fire is the quickest and least physically demanding arm pose, requiring far less shoulder torque and movement of the hand. It was only performed by a minority of participants, none of whom used it for the Accurate-Focused condition, although it was used in 48 of the Accurate-Distracted trials. Hip fire pointing is more than twice as inaccurate as proximal pointing using the head-rooted rays ($> 8°$). As hip fire does not involve raising the finger to the head, one may expect the arm-rooted ray casting approaches might perform better than the head-rooted rays. However, looking across all targets we see that the arm-rooted ray casting approaches appear to be much worse. Only five participants performed hip fire for more than 8% of their trials (P4, P5, P13, P20, P23).

Table 7. Gesture performance metrics for the finger incorporation trait. Temporal, biomechanical, and accuracy metrics are reported as median (and IQR). Behaviour distributions are reported as number of trials and the percentage.

	Measure	ADS	Occluded	Unoccluded
Temporal (ms)	Overall Gesture Duration	2070 (800)	1890 (640)	1720 (510)
	Start Phase Duration	760 (180)	730 (180)	690 (190)
	Hold Phase Duration	430 (610)	280 (470)	190 (280)
	Selection Time	1220 (720)	1040 (540)	920 (360)
Biomechanical	Shoulder Torque (Nm)	15.07 (3.45)	14.24 (2.28)	12.78 (5.30)
	NICER (%)	1.38 (1.05)	1.11 (1.02)	0.59 (0.89)
	Norm. Hand Movement	6.94 (2.86)	6.08 (2.23)	5.50 (2.90)
	Right Handed	5638 (70.01%)	1501 (69.56%)	813 (70.33%)
Accuracy (°)	IFRC	5.86 (4.82)	9.25 (9.96)	13.69 (13.54)
	FRC	4.50 (4.34)	6.10 (9.16)	13.12 (14.26)
	EFRC-Cyclops	2.13 (1.82)	2.74 (2.36)	8.13 (10.82)
	HFRC	1.89 (1.62)	2.45 (2.18)	7.10 (9.14)
Conditions	Number of participants	23 (100.00%)	23 (100.00%)	23 (100.00%)
	Number of Trials	8053 (70.85%)	2158 (18.98%)	1156 (10.17%)
	Accurate-Focused	2381 (29.57%)	473 (21.92%)	57 (4.93%)
	Accurate-Distracted	2080 (25.83%)	450 (20.85%)	203 (17.56%)
	Casual-Focused	2054 (25.51%)	637 (29.52%)	330 (28.55%)
	Casual-Distracted	1538 (19.10%)	598 (27.71%)	566 (48.96%)

5.2.2 Finger Incorporation. Table 7 shows the breakdown of the outcome measures for the different ways in which the finger is used during pointing. ADS was the most typical way in which the finger was utilised while pointing, wherein the user lines up the target with their finger such that their finger is also pointing toward the target. This can be observed where the EFRC and IFRC rays are aligned within 20° of each other. Exploring how the finger was incorporated into the pointing gesture was more complex than the arm pose. Table 5 shows how the use of the finger is strongly linked to the arm pose used which is visually shown in Figure 11, and Figure 12 shows how the different combinations affect the pointing accuracy for the different arm and finger combinations.

Exploring how the finger is incorporated into the pointing gestures helps us to better understand the difference in accuracy between IFRC and FRC rays. FRC has lower errors for both ADS (Mdn: 4.35, IQR: 4.13) and occluded (Mdn: 4.49, IQR: 4.74) finger incorporation when pointing with an outstretched arm pose, compared with IFRC (Outstretched ADS - Mdn: 5.80, IQR: 4.74, Outstretched occluded - Mdn: 8.20, IQR: 7.54). These pointing behaviours make up the majority of the dataset. However, IFRC is much more accurate for proximal ADS pointing compared with FRC because of the misalignment between forearm and the index finger, the latter of which is directed at the target (see Figure 11 (d) for an example of proximal ADS).

For proximal pointing, occluded was the most common type of finger incorporation, and Figure 12 shows how both arm-rooted ray casting approaches perform poorly for this type of pointing. This is because the forearm does not align with the target due to the elbow flexion (as in Figure 11 (d–f), and the finger is also perpendicular to the target which is especially problematic. However, the head-rooted ray casting approaches (HFRC and EFRC (Cyclops)) are much more accurate as one would expect. In fact, these show comparable accuracy compared to the more conventional outstretched ADS style of pointing.

Unoccluded finger-based pointing is the least accurate pointing style, especially in the case of the hip fire technique. It features heavily in the Casual-Distracted condition which explains some of the variability in section 4, and participants held the gesture for less than half the time compared with ADS, suggesting little effort. Interestingly, despite the unoccluded finger incorporation relying

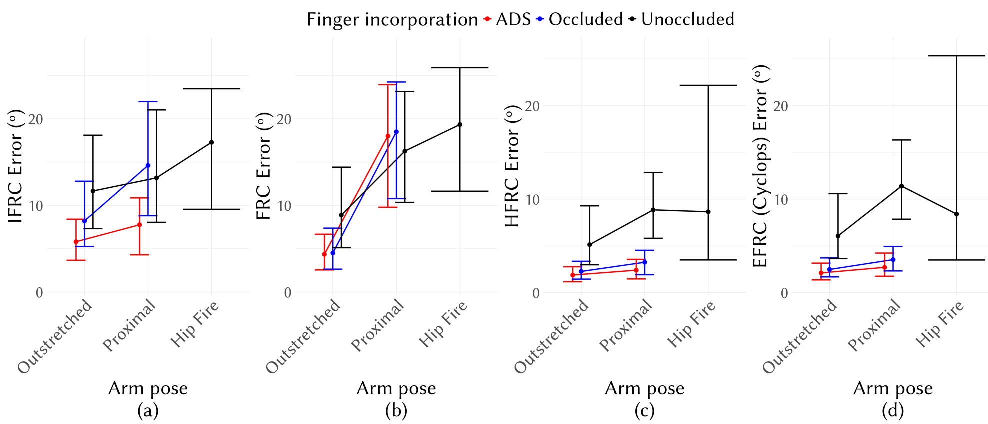

Fig. 12. Pointing accuracy showing the median and interquartile ranges across the whole dataset broken down by arm pose and finger incorporation traits for the different ray casting approaches: (a) IFRC, (b) FRC, (c) HFRC, and (d) EFRC (Cyclops).

Table 8. Gesture performance metrics for the hand torso alignment trait. Temporal, biomechanical, and accuracy metrics are reported as median (and IQR). Behaviour distributions are reported as number of trials as well as the percentage.

	Measure	Medial	Aligned	Lateral
Temporal (ms)	Overall Gesture Duration	2090 (830)	1930 (710)	1990 (730)
	Start Phase Duration	850 (220)	750 (160)	740 (170)
	Hold Phase Duration	280 (500)	300 (530)	390 (570)
	Selection Time	1150 (730)	1060 (650)	1160 (650)
Biomechanical	Shoulder Torque (Nm)	14.96 (3.51)	15.13 (3.57)	14.6 (3.36)
	NICER (%)	1.12 (1.19)	1.18 (1.11)	1.31 (1.03)
	Norm. Hand Movement	8.00 (2.46)	6.87 (2.41)	6.37 (2.80)
	Right Handed	1161 (99.40%)	1625 (88.08%)	5166 (61.84%)
Accuracy (°)	IFRC	9.70 (7.72)	7.40 (6.07)	6.25 (5.65)
	FRC	10.94 (9.84)	5.07 (5.67)	4.71 (4.73)
	EFRC-Cyclops	3.63 (3.39)	2.34 (2.27)	2.30 (2.10)
	HFRC	3.09 (3.18)	2.14 (1.92)	2.03 (1.89)
Conditions	Number of participants	19 (82.61%)	23 (100.00%)	23 (100.00%)
	Number of Trials	1168 (10.28%)	1845 (16.23%)	8354 (73.49%)
	Accurate-Focused	344 (29.45%)	511 (27.70%)	2056 (24.61%)
	Accurate-Distracted	302 (25.86%)	406 (22.01%)	2025 (24.24%)
	Casual-Focused	270 (23.12%)	498 (26.99%)	2253 (26.97%)
	Casual-Distracted	252 (21.58%)	430 (23.31%)	2020 (24.18%)

much more heavily on the direction of the finger, the head-rooted ray casting approaches are still far better than the arm-rooted approaches. This suggests that conceptually, participants were still pointing in a way that they are occluding the target without raising their hands. We explore this in more depth later in this section.

5.2.3 Hand Torso Alignment. Table 8 shows a breakdown of the outcome measures based on the three different arm torso alignments which are shown in Figure 13. One would expect at least 20% of trials to be performed with the arm and torso aligned, due to interactions with the centre column, however we observed that this was not the case. Lateral pointing gestures comprised the majority of all pointing gestures (73.49%), indicating that participants preferred to twist their torso away from the target when pointing or utilise the hand closest to target when pointing (61% of trials used the right hand, 39% used the left). This is also the case for the accurate conditions suggesting that participants felt comfortable pointing accurately with their non-dominant hand. We observed

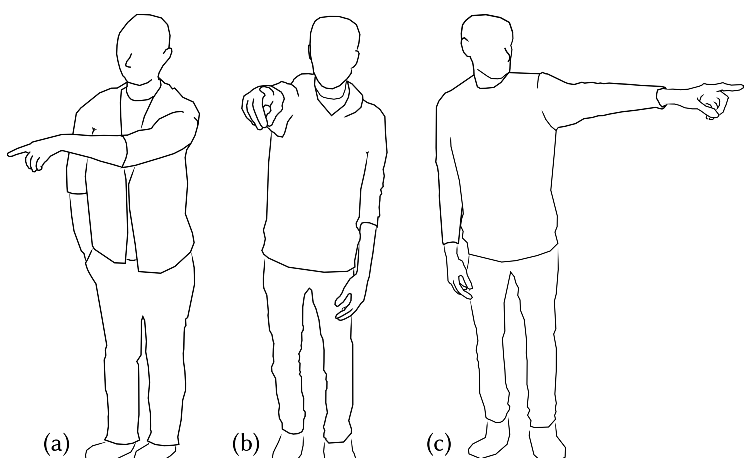

Fig. 13. Hand torso alignment describes how the hand is placed about the body relative to the torso, by taking the angle between the hand an the transverse plane, about the shoulder of the pointing arm. We identified three types of alignment: (a) medial, (b) aligned, and (c) lateral

Table 9. The number of trials for each combination of arm pose and hand torso alignment, with percentages representing how many trials made up the arm pose. This demonstrates how outstretched ADS is the most common type of pointing, but that occluded pointing is more common for the proximal arm pose.

Arm Pose	Alignment	Number of trials
Outstretched	Lateral	7472 (76.0%)
	Aligned	1567 (15.9%)
	Medial	799 (8.1%)
Proximal	Lateral	645 (56.4%)
	Aligned	217 (19.0%)
	Medial	281 (24.6%)
Hip Fire	Lateral	237 (61.4%)
	Aligned	61 (15.8%)
	Medial	88 (22.8%)

how some pointing gestures towards the centre target are lateral because participants positioned their shoulder forward while pointing. The majority of pointing gestures were performed with the dominant hand, however medial gestures were performed almost exclusively with the dominant hand (99.4%), and at higher rates in the accurate condition compared to the casual.

Not all participants exhibited all types of pointing. Four participants did not point medially in any of the conditions, and two others exclusively used their dominant hand in the accurate conditions by adopting medial pointing for targets on their non-dominant side or by turning to face the target. These two participants compose the majority of the medial pointing gestures. Interestingly though, medial pointing appears to be less accurate than both aligned and lateral pointing.

Table 9 shows how each type of arm pose was performed in each of the different types of alignments, and Figure 14 illustrates how the different types of alignment affected the accuracy for the different arm poses. For head-rooted ray casting, lateral is the most accurate for proximal pointing, with little difference between aligned and lateral for outstretched pointing. For hip fire, the aligned pointing gestures are much less accurate than both medial and lateral pointing. This difference may lie in that lateral pointing better aligns the hand, arm, and user's head with the gaze which may contribute to better accuracy.

For the arm-rooted rays, medial pointing is more accurate for the hip fire arm pose, especially for the FRC ray where outstretched and hip fire have similar accuracies. This implies that the biomechanical constraints of pointing across the body may turn some outstretched gestures into something that appears to be more like hip fire in some cases. Lateral pointing has similar accuracy to aligned when using the hand based rays for hip fire. Both outstretched and proximal pointing show a small general trend towards lateral being the most accurate with the arm-rooted rays.

5.3 Target Position

Figure 15 shows how the different target columns affect pointing accuracy for the arm poses. This demonstrates how participants can accurately point across a wide range of target positions, with little difference between the columns. For outstretched and proximal pointing, we see the same general trends across all rays which shows that accuracy is slightly better when pointing towards the right-hand targets compared to the left. Figure 16a shows how this could be due to the dominant

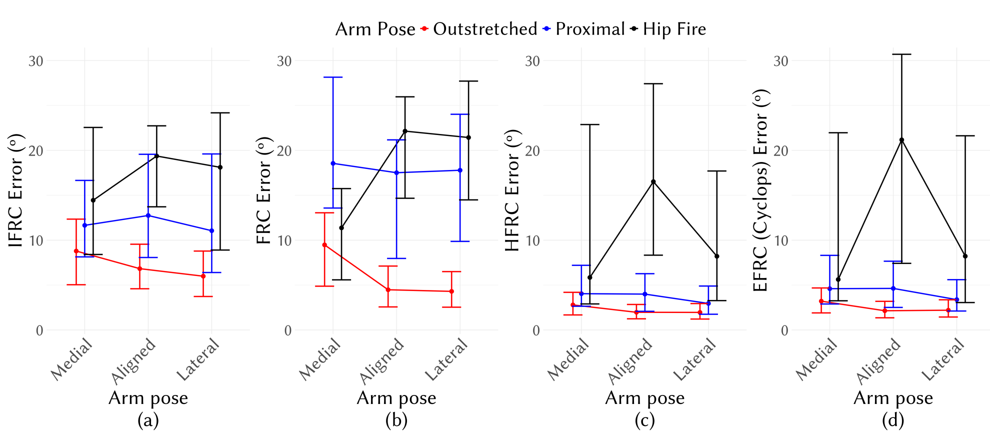

Fig. 14. Pointing accuracy showing the median and interquartile ranges across the whole dataset broken down by hand torso alignment and arm pose traits for the different ray casting approaches: (a) IFRC, (b) FRC, (c) HFRC, (d) EFRC (Cyclops).

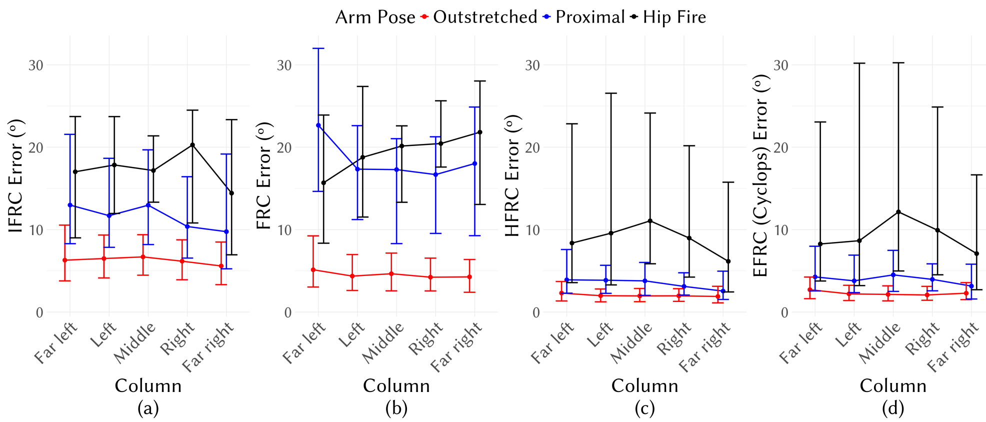

Fig. 15. Pointing accuracy showing the median and interquartile ranges across the whole dataset broken down by arm pose and target column for the different ray casting approaches: (a) IFRC, (b) FRC, (c) HFRC, (d) EFRC (Cyclops).

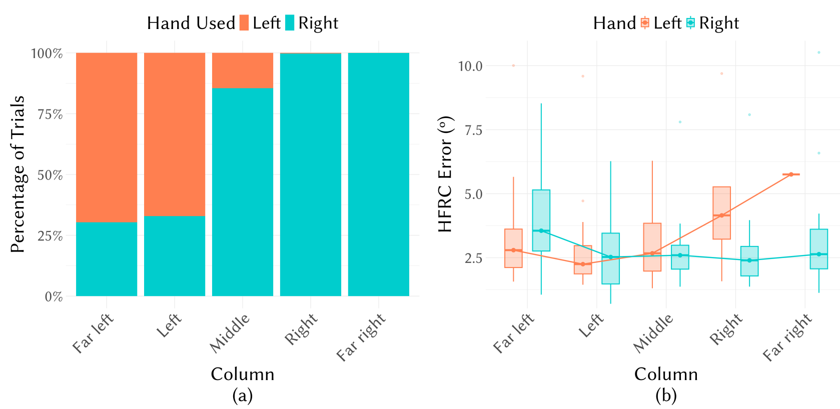

Fig. 16. (a) Stacked chart showing how the left hand is primarily used for left-hand targets, where it used more often than the right-hand and (b) a breakdown of HFRC error based on hand used across the target columns which shows left-hand pointing on the left-most targets is more accurate than right-hand targets, reinforcing that lateral pointing is more accurate than medial.

right hand being used more often for the centre, and almost exclusively for the right-hand targets. The majority of participants did not restrict themselves to pointing with the dominant hand, even when trying to point as accurately as possible. Figure 16b shows how usage of the hand on the same side of the body as the target yields a smaller median error than attempting to point medially, and Table 8 shows how participants were much less likely to turn and face the target to point, choosing instead to point laterally (or indeed medially) to targets at their sides. In particular, for hip fire we observe how the right hand columns exhibits much less variability which aligns with our previous observations of hand torso alignment.

Figure 17 shows how the different target heights have a much bigger effect on pointing accuracy for the different arm poses. For arm-rooted rays, the accuracy is most consistent for the top targets across the different pointing behaviours. Both proximal and hip fire are much less accurate when

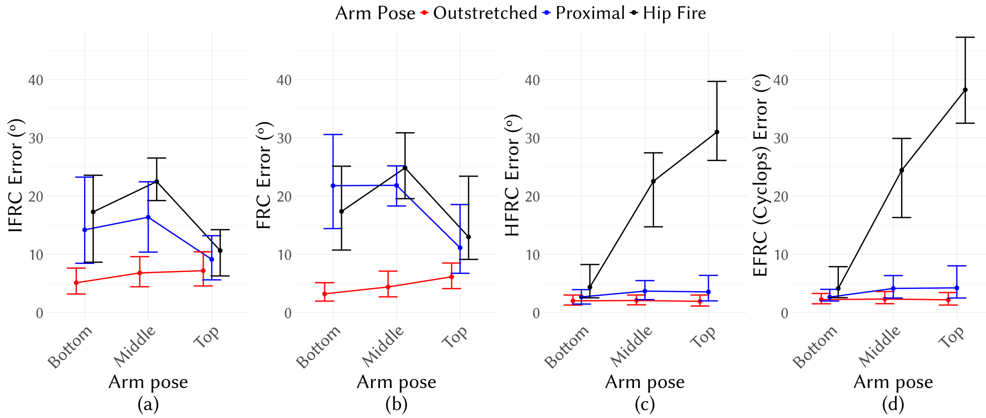

Fig. 17. Pointing accuracy showing the median and interquartile ranges across the whole dataset broken down by arm pose and target row for the different ray casting approaches: (a) IFRC, (b) FRC, (c) HFRC, (d) EFRC (Cyclops).

pointing at the bottom and middle targets compared to the top. In the case of proximal, this will likely be caused by the misalignment of the finger (for non-ADS proximal pointing) or forearm (by the very nature of proximal pointing). However, in contrast to arm-rooted rays, the accuracy when pointing at the bottom target is the most consistent when using head-rooted rays, with hip fire becoming much less accurate when pointing at the middle and top targets.

Figure 17 reveals important insights about the nature of hip fire pointing. For arm-rooted rays, hip fire is more accurate when pointing at the top targets compared to the middle and bottom rows, however the opposite is true for the head-based rays. In fact, both arm-rooted rays are much more accurate for the top targets compared with the eye-rooted rays which contrasts all the results seen so far. However, the arm-rooted rays are still much less accurate for the bottom targets compared to the head-rooted rays. To explore this further, we created a hybrid ray using the yaw from the HFRC ray, with the pitch from the IFRC ray. The error and variability of this hybrid ray is much lower (Mdn = 7.40, IQR= 5.25) compared to both HFRC (Mdn = 30.99, IQR= 13.57) and IFRC (Mdn = 10.64, IQR= 7.94) rays. Conceptually, this demonstrates how the yaw of the hip fire gesture is being performed by occluding the finger with the target, however the pitch is being directed by the finger. When pointing at the bottom targets using hip fire, the accuracy approaches that of outstretched and proximal pointing for both HFRC (Mdn = 4.36, IQR= 5.72) and EFRC (Cyclops) (Mdn = 4.13, IQR= 5.31). Interestingly, pointing at the middle targets has similar accuracy for hip fire across both arm-rooted (e.g. IFRC: Mdn = 22.49, IQR= 7.30) and head-rooted (e.g. HFRC: Mdn = 22.54, IQR= 12.71) rays at approximately 20°, and using the hybrid approach only marginally improves upon either (Mdn = 20.60, IQR= 8.87).

5.4 Behaviour Summary

Allowing participants to express their own individual pointing styles reveals different pointing behaviours. Not only do people change the temporal properties of their pointing gestures when pointing casually, they also change the fundamental way in which they point. We have identified and characterised three unique pointing traits which help to explain the variability in the data found in section 4. These findings reveal several important insights for the practical design of freehand cursorless pointing systems:

(1) People can point accurately, quickly and efficiently at a wide range of targets.
(2) Do not restrict people to pointing with their dominant hand because participants are likely to use, and can point accurately with, both hands.
(3) To embrace a wider variety of pointing behaviours while having consistent pointing accuracy, targets used for selection should be placed:
  (a) At the top when using hand-rooted rays;
  (b) At the bottom when using head-rooted rays.
(4) Place targets in the centre or on the user's dominant side for maximum accuracy.
(5) Use head-rooted rays unless users are likely to use the hip fire style of pointing.
(6) Systems designed for hip fire pointing should account for target height and switch between head- and arm-rooted rays accordingly.
(7) Use HFRC instead of EFRC when pointing at a wide range of targets because it is less variable and after offset correction becomes significantly more accurate.
(8) In line with previous research, systematic offset compensation should be used for hand-rooted rays as this significantly increases the accuracy.

6 DISCUSSION

Cursorless pointing is an intuitive gesture that encodes both semantic ('I want to select') and spatial ('this') elements in one seamless gesture that has been explored for interaction in both, distant 2D displays [42, 44, 45, 54], and in 3D environments [8, 15, 51]. We know from prior research that people perform a variety of pointing gestures when pointing in non-interactive contexts, including variations in their arm pose [2, 3, 10–12, 18, 29, 69]. However, cursorless pointing interaction in HCI has primarily involved or elicited participants performing the same pointing gesture: an outstretched straightened arm with the index finger aimed towards the target. This is because pointing has been explicitly prescribed to participants [17, 42], constrained to the dominant arm [15, 17, 44, 45, 50, 54], or implicitly imposed through the tracking apparatus (e.g. only tracking the index finger-tip [15, 44, 45]) or instructions provided (e.g. to be as accurate as possible [8, 15, 42, 44, 45, 54]). No prior work has reported or explored any differences in pointing gestures or the implications for how these differences affect interaction. The key contribution of this work is addressing this knowledge gap through the exploration of variability within cursorless pointing by unconstraining the way in which participants point, introducing distractions while pointing, and asking participants to point in a more casual and relaxed manner.

Through analysis of the collected pointing gestures we demonstrate how people point differently depending on whether they are trying to point accurately (as in many prior studies) or if they are encouraged to point casually and more naturally. We also saw changes in their pointing depending on whether pointing is their primary task, and depending on where the target is located. The observed behaviours mirror the 'full' and 'partial' pointing behaviours previously observed in human–human communication literature [2, 29, 69] which re-confirms our elicitation study and the general study environment. More importantly, we show that the observed different behaviours affect the speed, accuracy, and effort required for the interaction. With the ability to choose which hand to point with, participants were more likely to utilise the hand on the same side of the body as the target, even when pointing accurately, rather than preferring the dominant hand. The practical implications of this are that pointing gestures should not be restricted to a specific hand.

Table 10. Accuracy and interaction times reported for previous freehand cursorless pointing in prior research, compared with those reported in this paper. The papers are ordered in chronological order.

Paper	Target Dist. (m)	Condition	Accuracy EFRC	HFRC	IFRC	FRC	Other	Hold / Dwell Time (s)
[17]	1	Gaze-Shot	$< 2^\circ$	-	-	-	-	-
		Snipe-Shot	-	-	$< 5^\circ$	-	-	-
		Snap-Shot	-	-	$< 7^\circ$	-	-	-
[51]	-	-	-	$25^{\circ *}$	-	39°	$22^{\circ **}$	$0.1 - 2.5$
$[45]^{\|}$	2	-	10.48° ($36.7cm$)	-	10° ($35cm$)	12.75° ($44.7cm$)	-	1
	3	-	7.7° ($40.4cm$)	-	6.87° ($36cm$)	8.6° ($45.1cm$)	-	1
[54]	$2.2 - 3.4$	-	3.9°	-	-	25.7°	$3.7^{\circ \dagger}$	-
$[44]^{\|}$	2	VR	2.41° ($8.4cm$)	-	3.69° ($12.9cm$)	8.64° ($30.2cm$)	-	-
	2	RW	2.29° ($8cm$)	-	4.35° ($15.2cm$)	7.72° ($27cm$)	-	-
$[42]^{\S \|}$	2.5	$E_R H_R^{\$}$ $L_{eye}^{\therefore}$	5.04° ($22cm$)	-	-	-	-	-
		$E_R H_R^{\$}$ $R_{eye}^{\therefore}$	2.75° ($12cm$)	-	-	-	-	-
		$E_R H_L^{\$}$ $L_{eye}^{\therefore}$	4.58° ($20cm$)	-	-	-	-	-
		$E_R H_L^{\$}$ $R_{eye}^{\therefore}$	2.75° ($12cm$)	-	-	-	-	-
		$E_L H_R^{\$}$ $L_{eye}^{\therefore}$	1.83° ($8cm$)	-	-	-	-	-
		$E_L H_R^{\$}$ $R_{eye}^{\therefore}$	4.35° ($19cm$)	-	-	-	-	-
		$E_L H_L^{\$}$ $L_{eye}^{\therefore}$	2.75° ($12cm$)	-	-	-	-	-
		$E_L H_L^{\$}$ $R_{eye}^{\therefore}$	4.12° ($18cm$)	-	-	-	-	-
[43]	4m	-	5.73°	-	63.86°	39.10°	-	-
[50]	-	-	-	-	-	-	$13.58^{\circ \ddagger}$	-
[8]	$0.9 - 3.3$	-	-	-	-	-	$1.48^{\circ \#}$	$0.8 - 3^{\S}$
Ours	2	Accurate	2.58°	2.06°	6.75°	5.17°	-	0.59
	2	Casual	3.09°	2.84°	8.03°	6.27°	-	0.26
	2	Focused	2.92°	2.08°	6.85°	5.31°	-	0.54
	2	Distracted	3.39°	2.72°	7.74°	5.93°	-	0.28

$*$ Ray cast from the centre of the head through the wrist. $**$ Ray cast from the centre of the head, orientated by the difference between the head ray and head-wrist ray. † Ray cast from one of four points on the head based on the user's ocular and hand dominance. ‡ Arm pose and ray derived from an ML model using 2 seconds of video (30 frames). ∴ EFRC rays using the Left eye (L_{eye}) and Right eye (R_{eye}) as the origin. \$ Participant eye dominance (right: E_R, left: E_L) and hand dominance (right: H_R, left: H_L). # Ray derived from an ML model taking using 0.67 seconds (20 frames) of input data. § Values derived from provided figures. ‖ Error originally reported in cm and converted into an angular error.

Additionally, if choosing to use recognition systems, these should be trained on a wider variety of poses as to ensure all pointing behaviours can be recognised. Most importantly, we argue that cursorless pointing systems should embrace user individuality, supporting users in pointing in the way that is preferable to them rather than requiring users to adapt to the way in which the system supports pointing.

Despite the different observed pointing behaviours, our study shows in general that cursorless pointing is a quick and efficient selection technique for touchless interactions. Consistent with prior works [8, 49, 51], we observed that pointing hold phase durations (akin to a dwell time) can range from under 1 second to spanning multiple seconds. In this paper we further explore this spread and find the hold phase significantly varies depending on whether someone is pointing accurately or casually, and whether or not they are distracted while pointing. In contrast to prior work, Table 10 shows how the median hold phase durations in the accurate and focused cases ($0.54 - 0.59s$) are much lower when participants are able to point in a more unconstrained manner. In addition, casual and distracted pointing is performed much faster and with a much shorter hold phase: $0.26 - 0.28s$. Overall, we demonstrate how selection time for a cursorless pointing gesture is around 1.5 seconds for accurate pointing and less than 1 second when pointing casually – far quicker than other touchless interaction techniques involving cursor-based pointing techniques, which require explicit selection or activation mechanisms.

While trying to point accurately and with pointing as the primary task, we find pointing accuracy across systematically compensated ray casting techniques to be consistent with prior works, where head-rooted techniques are significantly more accurate than hand-rooted ones [43–45], across the spread of targets. However, when faced with distractions or while encouraged to point casually and more naturally, pointing accuracy is significantly reduced across all ray casting techniques. Specific pointing behaviours can also affect the accuracy of different rays, most notably for the hand-rooted rays. For instance, we find that proximal pointing, and hip-fire to elevated targets, result in worse performance for FRC rays, while occluded pointing negatively impacts the IFRC ray. Despite the variability in the pointing accuracy, we observe that use of the systematically compensated HFRC ray is generally the best choice, regardless of the pointing context or behaviour, except for Hip-Fire gestures to targets at or above shoulder level.

Ideally, cursorless pointing systems should not enforce a minimum pointing duration as this will artificially prolong interactions, making them less natural. To avoid the Midas Touch problem [27] that may arise without an enforced dwell time, especially for casual and distracted pointing, future research should explore advanced sensing techniques and alternative selection disambiguation techniques, such as models which use hand motion [71] or gaze-behaviour [16], to identify whether the gesture was intended for interaction. Additionally, while we can recommend the use of HFRC ray casting with systematic offset compensation generally, the way in which pointing occurs affects how accurately someone will interact with a system. This variability should be taken into account when designing future interactive systems using cursorless pointing. This could be through providing adaptations to the choice of ray casting technique to optimise accuracy, or through awareness of when pointing may be inaccurate and may require disambiguation.

6.1 Limitations and Future Work

We observed large variability both across and within participants even though our participants were a relatively homogenous sample, all workers or students of the same university and all without mobility impairments. Pointing is known to vary between cultures [14, 37] which future work should explore, as well as how people with upper mobility impairments may need to adapt their behaviours. Furthermore, during the study design, we ran into the challenge of how to elicit casual pointing. For example, in our discussions, we often observed how we used deictic pointing to vaguely point out objects in our environment using gestures that differed from those reported in the interaction literature. However, we intentionally avoided giving any indication or examples to the participants on what we understood as casual pointing to not introduce any bias. That said, just mentioning that they should point casually might already bias them as it has the assumption that prior pointing was not casual.

Similarly, our study was still in a lab environment and with the experimenter in the room. We also used markers for maximum accuracy when tracking the participants' hands and fingers because the markerless motion capture system used did not yet support finger tracking. However, it is possible that the use of markers could have limited comfortable hand shapes or had an effect on how participants aligned their fingers to the targets. Therefore, based on the results from this study, a more ecologically valid exploration that also considers additional user contexts, such as in non-static contexts (e.g. walking) or non-standing contexts (e.g. while seated or lying down), targets placed at distances beyond and closer than 2m, targets with size and shape beyond point LEDs, different social contexts, and which removes all instrumentation, would be an interesting avenue for future work towards fully natural pointing-based interaction. Additionally, while we know that interaction intent and target properties can bring-out casual pointing features, it is unclear whether conversation would draw out additional behaviours when pointing casually. It would however be of interest to explore such pointing interactions in a collaborative environment to try an elicit such casual gestures without instruction.

Finally, the pointing task had no feedback that users would have in an interaction context that could have changed their pointing behaviour. The frequency of interactions was much higher than one would potentially expect in a deployed system and the abstract pointing-only task mundane, as also noted in the subjective responses. Future work should consider how the behaviours may vary and affect pointing performance in-situ, such as in a smart-home setting, when collaborating, or while using a mobile system, where targets may not be in fixed positions and the number of interactions would be given by a realistic task. Similarly, many HMDs used for spatial computing constrain the user's field of view which might also affect how our results carry over to these platforms when used in actual scenarios.

In our analysis of pointing ray accuracy, we applied systematic offset correction models Mayer et al. to the rays calculated for the pointing gestures. The coefficients proposed by Mayer et al. [45] are different to the coefficients derived from this dataset, likely because ours contains much larger pointing variability and range of target positions. Further research is also required to explore and validate whether systematic offsets differ between behaviours. For example, due to small sample sizes, it is possible the coefficients derived for the whole dataset under-fit the smaller behaviours. Additionally, our results show behaviours, such as hip fire and unoccluded pointing, which are much less accurate with existing ray casting approaches. Further exploration may reveal other approaches that better support these behaviours for accurate target selection.

7 CONCLUSION

Prior works exploring pointing for communication show that pointing gestures are not performed with a single pose, however this is not something previously studied for pointing in interaction. In this paper we collected and labelled motion capture data of unconstrained, cursorless, freehand pointing gestures to understand how variability affects pointing performance. With this work we empirically derive three behavioural traits with which one can characterise a pointing gesture: 1) Arm Pose, 2) Arm-Torso Alignment, and 3) Finger Incorporation. With these traits, we find that pointing for interaction is typically performed by raising the arm and aligning the index finger with the direction of the target (outstretched ADS pointing), consistent with prior works. We find that pointing behaviour varies depending on whether a user is pointing in an accurate or casual manner, whether pointing is the user's primary focus, and due to individual preferences. However, we show how there are other behaviours for pointing accurately which can be done with a less fatiguing pose by using a bent elbow to raise and align the index finger (proximal pointing). Additionally, people can point quickly and accurately without having to align the finger with the target, instead covering the target with the finger-tip but pointing in a different direction to the target (occluded pointing). Cursorless pointing is a fast and efficient method of interacting, and we demonstrate that users are quicker when pointing either casually or while focusing on another task. Casual pointing can also lead to a much less fatiguing partial pointing pose (hip fire), which is also much less accurate than other poses, with the most suitable ray casting approach switching depending on the height of the target.

Cursorless pointing systems should incorporate user variability, allowing people to point in the way they prefer rather than defining specific ways the user should point. For this to happen, systems should support and recognise the range of different pointing poses highlighted in this paper. Support for casual and distracted pointing requires further investigation to improve their accuracy, with common ray casting approaches, even with application of systematic offset compensation models, being significantly less accurate. In particular, future work is needed to understand the inherent inaccuracy of specific behaviours, such as hip fire and unoccluded pointing, or whether alternate approaches to ray casting are required to better support these types of pointing.

8 ACKNOWLEDGEMENTS

The authors thank Alvaro Favaratto Santos of the Science Technical Services Department at the University of Bath for their expert technical input and guidance in this work. James Whiffing's PhD is funded by a UKRI EPSRC DTP training grant (EP/W524712/1). This work was supported and partly funded by the Centre for the Analysis of Motion, Entertainment Research and Applications (CAMERA 2.0) - EP/T022523/1.

REFERENCES

[1] Ferran Argelaguet and Carlos Andujar. 2009. Efficient 3D Pointing Selection in Cluttered Virtual Environments. *IEEE Computer Graphics and Applications* 29, 6 (2009), 34–43. https://doi.org/10.1109/MCG.2009.117

[2] Adrian Bangerter and Eric Chevalley. 2007. Pointing and describing in referential communication: When are pointing gestures used to communicate?. In *MOG 2007 workshop on multimodal output generation*. 17–28.

[3] Gemma Barberà and Martine Zwets. 2013. Pointing and Reference in Sign Language and Spoken Language: Anchoring vs. Identifying. *Sign Language Studies* 13, 4 (2013), 491–515. https://doi.org/10.1353/sls.2013.0016

[4] Dhruv Bihani, A. K. M. Amanat Ullah, Charles-Olivier Dufresne-Camaro, William Delamare, Pourang Irani, and Khalad Hasan. 2024. Exploring Pointer Enhancement Techniques for Target Selection on Large Curved Display. *Proc. ACM Hum.-Comput. Interact.* 8, ISS, Article 535 (2024), 22 pages. https://doi.org/10.1145/3698135

[5] Richard A. Bolt. 1980. "Put-that-there": Voice and gesture at the graphics interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques* (Seattle, Washington, USA) *(SIGGRAPH '80)*. Association for Computing Machinery, New York, NY, USA, 262–270. https://doi.org/10.1145/800250.807503

[6] Gunnar Borg. 1998. *Borg's perceived exertion and pain scales.* Human kinetics.

[7] Sebastian Boring, Marko Jurmu, and Andreas Butz. 2009. Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7* (Melbourne, Australia) *(OZCHI '09)*. Association for Computing Machinery, New York, NY, USA, 161––168. https://doi.org/10.1145/1738826.1738853

[8] Choongho Chung and Sung-Hee Lee. 2024. Continuous Prediction of Pointing Targets With Motion and Eye-Tracking in Virtual Reality. *IEEE Access* 12 (2024), 5933–5946. https://doi.org/10.1109/ACCESS.2024.3350788

[9] Rachel B Clancy, Matthew P Herring, and Mark J Campbell. 2017. Motivation Measures in Sport: A Critical Review and Bibliometric Analysis. *Frontiers in Psychology* 8, Article 348 (2017), 12 pages. https://doi.org/10.3389/fpsyg.2017.00348

[10] Hélène Cochet and Jacques Vauclair. 2014. Deictic gestures and symbolic gestures produced by adults in an experimental context: Hand shapes and hand preferences. *Laterality: Asymmetries of Body, Brain and Cognition* 19, 3 (2014), 278–301. https://doi.org/10.1080/1357650X.2013.804079

[11] Kensy Cooperrider. 2014. Body-Directed Gestures: Pointing to the Self and Beyond. *Journal of Pragmatics* 71 (Sept. 2014), 1–16. https://doi.org/10.1016/j.pragma.2014.07.003

[12] Kensy Cooperrider. 2016. The co-organization of demonstratives and pointing gestures. *Discourse Processes* 53, 8 (2016), 632–656. https://doi.org/10.1080/0163853X.2015.1094280

[13] Kensy Cooperrider, Jordan Fenlon, Jonathan Keane, Diane Brentari, and Susan Goldin-Meadow. 2021. How pointing is integrated into language: Evidence from speakers and signers. *Frontiers in Communication* 6 (2021), 567774. https://doi.org/10.3389/fcomm.2021.567774

[14] Kensy Cooperrider, James Slotta, and Rafael Núñez. 2018. The preference for pointing with the hand is not universal. *Cognitive Science* 42, 4 (2018), 1375–1390. https://doi.org/10.1111/cogs.12585

[15] Tor-Salve Dalsgaard, Jarrod Knibbe, and Joanna Bergström. 2021. Modeling Pointing for 3D Target Selection in VR. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (Osaka, Japan) *(VRST '21)*. Association for Computing Machinery, New York, NY, USA, Article 42, 10 pages. https://doi.org/10.1145/3489849.3489853

[16] Brendan David-John, Candace Peacock, Ting Zhang, T. Scott Murdison, Hrvoje Benko, and Tanya R. Jonker. 2021. Towards gaze-based prediction of the intent to interact in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 2, 7 pages. https://doi.org/10.1145/3448018.3458008

[17] M Etoh, A Tomono, and Y Kobayashi. 1989. Direct Finger Pointing as a Man-machine Interface. *IFAC Proceedings Volumes* 22, 12 (1989), 125–130. https://doi.org/10.1016/S1474-6670(17)53066-9

[18] Jordan Fenlon, Kensy Cooperrider, Jon Keane, Diane Brentari, and Susan Goldin-Meadow. 2019. Comparing sign language and gesture: Insights from pointing. *Glossa* 4, 1 (2019), 1–26. https://doi.org/10.5334/gjgl.499

[19] Charles Golden. 1978. Stroop Color and Word Test: A Manual for Clinical and Experimental Uses. (1978).

[20] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2017. Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 23, 6 (2017), 1706–1724. https://doi.org/10.1109/TVCG.2016.2543720

[21] Darren Guinness, Alvin Jude, G. Michael Poor, and Ashley Dover. 2015. Models for Rested Touchless Gestural Interaction. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction* (Los Angeles, California, USA) *(SUI '15)*. Association for Computing Machinery, New York, NY, USA, 34––43. https://doi.org/10.1145/2788940.2788948

[22] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*. Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[23] Oliver Herbort and Wilfried Kunde. 2018. How to point and to interpret pointing gestures? Instructions can reduce pointer–observer misunderstandings. *Psychological research* 82, 2 (2018), 395–406. https://doi.org/10.1007/s00426-016-0824-8

[24] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1063–1072. https://doi.org/10.1145/2556288.2557130

[25] Yichao Huang, Xiaorui Liu, Xin Zhang, and Lianwen Jin. 2016. A Pointing Gesture Based Egocentric Interaction System: Dataset, Approach and Application. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 370–377. https://doi.org/10.1109/CVPRW.2016.53

[26] Humane Inc. 2025. *Humane AI Pin*. https://humane.com/

[27] Robert JK Jacob. 1995. Eye tracking in advanced interface design. *Virtual environments and advanced interface design* 258, 288 (1995), 258–288. https://doi.org/10.1093/oso/9780195075557.003.0015

[28] Ricardo Jota, Miguel A. Nacenta, Joaquim A. Jorge, Sheelagh Carpendale, and Saul Greenberg. 2010. A comparison of ray pointing techniques for very large displays. In *Proceedings of Graphics Interface 2010* (Ottawa, Ontario, Canada) *(GI '10)*. Canadian Information Processing Society, CAN, 269—-276. https://doi.org/10.5555/1839214.1839261

[29] Adam Kendon and Laura Versante. 2003. Pointing by hand in "Neapolitan". In *Pointing: Where Language, Culture, and Cognition Meet*. Lawrence Erlbaum Associates Publishers, 117–146.

[30] Daehwa Kim, Vimal Mollyn, and Chris Harrison. 2023. WorldPoint: Finger Pointing as a Rapid and Natural Trigger for In-the-Wild Mobile Interactions. *Proc. ACM Hum.-Comput. Interact.* 7, ISS, Article 442 (2023), 19 pages. https://doi.org/10.1145/3626478

[31] Sotaro Kita. 2003. Interplay of gaze, hand, torso orientation, and language in pointing. In *Pointing*. Lawrence Erlbaum Associates Publishers, 315–336.

[32] Sotaro Kita (Ed.). 2003. *Pointing: Where Language, Culture, and Cognition Meet.* Lawrence Erlbaum Associates Publishers.

[33] Alfred Kranstedt, Andy Lücking, Thies Pfeiffer, Hannes Rieser, and Marc Staudacher. 2006. Measuring and reconstructing pointing in visual contexts. *Proceedings of the Brandial* (2006), 82–89.

[34] Lisa-Marie Krause and Oliver Herbort. 2021. The observer's perspective determines which cues are used when interpreting pointing gestures. *Journal of Experimental Psychology: Human Perception and Performance* 47, 9 (2021), 1209.

[35] Jaewook Lee, Jun Wang, Elizabeth Brown, Liam Chu, Sebastian S. Rodriguez, and Jon E. Froehlich. 2024. GazePointAR: A Context-Aware Multimodal Voice Assistant for Pronoun Disambiguation in Wearable Augmented Reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 408, 20 pages. https://doi.org/10.1145/3613904.3642230

[36] Martin Lemay, Christopher P Bertram, and George E Stelmach. 2004. Pointing to an allocentric and egocentric remembered target. *Motor control* 8, 1 (2004), 16–32.

[37] Heng Li and Yu Cao. 2019. Hands occupied: Chinese farmers use more non-manual pointing than herders. *Lingua* 222 (2019), 1–9. https://doi.org/10.1016/j.lingua.2019.02.006

[38] Yi Li, Benjamin Tag, Shaozhang Dai, Robert Crowther, Tim Dwyer, Pourang Irani, and Barrett Ens. 2024. NICER: A New and Improved Consumed Endurance and Recovery Metric to Quantify Muscle Fatigue of Mid-Air Interactions. *ACM Trans. Graph.* 43, 4, Article 102 (2024), 14 pages. https://doi.org/10.1145/3658230

[39] Chiuhsiang Joe Lin, Sui-Hua Ho, and Yan-Jyun Chen. 2015. An investigation of pointing postures in a 3D stereoscopic environment. *Applied ergonomics* 48 (2015), 154–163. https://doi.org/10.1016/j.apergo.2014.12.001

[40] Mathias N. Lystbæk, Peter Rosenberg, Ken Pfeuffer, Jens Emil Grønbæk, and Hans Gellersen. 2022. Gaze-Hand Alignment: Combining Eye Gaze and Mid-Air Pointing for Interacting with Menus in Augmented Reality. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 145 (2022), 18 pages. https://doi.org/10.1145/3530886

[41] R Mangera. 2013. Static gesture recognition using features extracted from skeletal data. (2013).

[42] Tomohiro Mashita, Koichi Shintani, and Kiyoshi Kiyokawa. 2020. Improving pointing direction estimation by considering hand-and ocular-dominance. *IEICE TRANSACTIONS on Information and Systems* 103, 10 (2020), 2168–2177. https://doi.org/10.1587/transinf.2019EDP7320

[43] Sven Mayer, Jens Reinhardt, Robin Schweigert, Brighten Jelke, Valentin Schwind, Katrin Wolf, and Niels Henze. 2020. Improving Humans' Ability to Interpret Deictic Gestures in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376340

[44] Sven Mayer, Valentin Schwind, Robin Schweigert, and Niels Henze. 2018. The Effect of Offset Correction and Cursor on Mid-Air Pointing in Real and Virtual Environments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1—-13. https://doi.org/10.1145/3173574.3174227

[45] Sven Mayer, Katrin Wolf, Stefan Schneegass, and Niels Henze. 2015. Modeling Distant Pointing for Compensating Systematic Displacements. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 4165––4168. https://doi.org/10.1145/2702123.2702332

[46] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport* 60, 1 (1989), 48–58. https://doi.org/10.1080/02701367.1989.10607413

[47] David McNeill (Ed.). 1992. *Hand and mind: What gestures reveal about thought.* University of Chicago Press.

[48] Jörg Müller, Antti Oulasvirta, and Roderick Murray-Smith. 2017. Control Theoretic Models of Pointing. *ACM Trans. Comput.-Hum. Interact.* 24, 4, Article 27 (2017), 36 pages. https://doi.org/10.1145/3121431

[49] Christian Müller-Tomfelde. 2007. Dwell-based pointing in applications of human computer interaction. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-Computer Interaction* (Rio de Janeiro, Brazil) *(INTERACT'07)*. Springer-Verlag, Berlin, Heidelberg, 560—-573.

[50] Shu Nakamura, Yasutomo Kawanishi, Shohei Nobuhara, and Ko Nishino. 2023. DeePoint: Visual Pointing Recognition and Direction Estimation. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 20520–20530. https://doi.org/10.1109/ICCV51070.2023.01881

[51] Kai Nickel and Rainer Stiefelhagen. 2003. Pointing gesture recognition based on 3D-tracking of face, hands and head orientation. In *Proceedings of the 5th International Conference on Multimodal Interfaces* (Vancouver, British Columbia, Canada) *(ICMI '03)*. Association for Computing Machinery, New York, NY, USA, 140—-146. https://doi.org/10.1145/958432.958460

[52] Jeni Paay, Jesper Kjeldskov, Dimitrios Raptis, Mikael B. Skov, Ivan S. Penchev, and Elias Ringhauge. 2017. Cross-device interaction with large displays in public: insights from both users' and observers' perspectives. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (Brisbane, Queensland, Australia) *(OzCHI '17)*. Association for Computing Machinery, New York, NY, USA, 87—-97. https://doi.org/10.1145/3152771.3152781

[53] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for

Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[54] Katrin Plaumann, Matthias Weing, Christian Winkler, Michael Müller, and Enrico Rukzio. 2018. Towards accurate cursorless pointing: the effects of ocular dominance and handedness. *Personal Ubiquitous Comput.* 22, 4 (Aug. 2018), 633–646. https://doi.org/10.1007/s00779-017-1100-7

[55] Richard M Ryan. 1982. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology* 43, 3 (1982), 450–461. https://doi.org/10.1037/0022-3514.43.3.450

[56] Eri Sato, Toru Yamaguchi, and Fumio Harashima. 2007. Natural interface using pointing behavior for human–robot gestural interaction. *IEEE transactions on Industrial Electronics* 54, 2 (2007), 1105–1112. https://doi.org/10.1109/TIE.2007.892728

[57] Immo Schuetz and Katja Fiehler. 2024. Object center of mass predicts pointing endpoints in virtual reality. *Frontiers in Virtual Reality* 5 (2024), 1402084. https://doi.org/10.3389/frvir.2024.1402084

[58] Robin Schweigert, Valentin Schwind, and Sven Mayer. 2019. EyePointing: A Gaze-Based Selection Technique. *Proceedings of Mensch und Computer 2019* (2019), 719–723. https://doi.org/10.1145/3340764.3344897

[59] Hyunggoog Seo, Jaedong Kim, Kwanggyoon Seo, Bumki Kim, and Junyong Noh. 2021. Overthere: A Simple and Intuitive Object Registration Method for an Absolute Mid-air Pointing Interface. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 127 (2021), 24 pages. https://doi.org/10.1145/3478128

[60] Dadhichi Shukla, Ozgur Erkent, and Justus Piater. 2015. Probabilistic detection of pointing directions for human robot interaction. In *International Conference on Digital Image Computing: Techniques and Applications*. IEEE. https://doi.org/10.1109/DICTA.2015.7371296

[61] Jaisie Sin and Cosmin Munteanu. 2020. Let's Go There: Combining Voice and Pointing in VR. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) *(CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 31, 3 pages. https://doi.org/10.1145/3405755.3406161

[62] Maurício Sousa, Rafael Kufner dos Anjos, Daniel Mendes, Mark Billinghurst, and Joaquim Jorge. 2019. Warping Deixis: Distorting Gestures to Enhance Collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300838

[63] Michael Tomasello, Malinda Carpenter, and Ulf Liszkowski. 2007. A new look at infant pointing. *Child development* 78, 3 (2007), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

[64] Rafael Veras, Gaganpreet Singh, Farzin Farhadi-Niaki, Ritesh Udhani, Parth Pradeep Patekar, Wei Zhou, Pourang Irani, and Wei Li. 2021. Elbow-Anchored Interaction: Designing Restful Mid-Air Input. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 737, 15 pages. https://doi.org/10.1145/3411764.3445546

[65] Daniel Vogel and Ravin Balakrishnan. 2005. Distant freehand pointing and clicking on very large, high resolution displays. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology* (Seattle, WA, USA) *(UIST '05)*. Association for Computing Machinery, New York, NY, USA, 33—-42. https://doi.org/10.1145/1095034.1095041

[66] Uta Wagner, Mathias N. Lystbæk, Pavel Manakhov, Jens Emil Sloth Grønbæk, Ken Pfeuffer, and Hans Gellersen. 2023. A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3544548.3581423

[67] Mengxin Wang, Paul McGraw, and Timothy Ledgeway. 2019. Individual variation in inter-ocular suppression and sensory eye dominance. *Vision research* 163 (2019), 33–41. https://doi.org/10.1016/j.visres.2019.07.004

[68] James Whiffing, Tobias Langlotz, Christof Lutteroth, Adwait Sharma, and Christopher Clarke. 2025. *Dataset for "Understanding Freehand Cursorless Pointing Variability and Its Impact on Selection Performance"*. Bath: University of Bath Research Data Archive. https://doi.org/10.15125/BATH-01594

[69] David Wilkins. 2003. Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In *Pointing: Where Language, Culture, and Cognition Meet*. Lawrence Erlbaum Associates Publishers, 179–224.

[70] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[71] Ting Zhang, Zhenhong Hu, Aakar Gupta, Chi-Hao Wu, Hrvoje Benko, and Tanya R. Jonker. 2022. RIDS: Implicit Detection of a Selection Gesture Using Hand Motion Dynamics During Freehand Pointing in Virtual Reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 4, 12 pages. https://doi.org/10.1145/3526113.3545701

A TRACKING RATES FOR THE DATA COLLECTED FROM THE MOTION CAPTURE SYSTEM

Table 11 shows the tracking rates of the system for the dataset we collected. This is referenced in subsection 3.3.

Table 11. Tracking rate for the points on the participant's Cyclops eye and the pointing finger, calculated during the labelled pointing gestures.

Point	Capture Rate
Cyclops Eye*	99.533% (3.626)
Thumb Tip	97.813% (12.250)
Index Finger Tip	99.886% (1.218)
Middle Finger Tip	83.644% (32.694)
Ring Finger Tip	81.917% (29.515)
Pinky Finger Tip	80.394% (29.992)
Index Finger Base Knuckle	99.969% (1.232)
Pinky Finger Base Knuckle	98.819% (5.571)
Inner Wrist	99.612% (3.664)
Outer Wrist	98.399% (7.226)

* Derived from the tracking of four markers affixed to glasses worn by the participant.

B FULL BREAKDOWN OF EXCLUDED TRIALS

We had to remove pointing trials due to data collection failures, data processing failures, or invalid gestures. Table 12 provides a full breakdown of the trials that were excluded. This is referenced in subsubsection 3.4.1. With the remaining dataset, trials are distributed between conditions as such: 2733 Accurate-Distracted, 2911 Accurate-Focused, 2702 Casual-Distracted, and 3021 Casual-Focused, resulting in a discrepancy of each condition's contribution to the dataset being at the minimum 23.8% and at the maximum 26.6%.

Table 12. Breakdown of trials removed from dataset.

Removal Reason	Count	Notes
Study Overrunning*	180 (4 recordings)	Some studies overran, either due to resolving technical issues or setup and briefing taking longer than normal, with participants being unable to extend the session to complete all trials.
Crash During Recording*	315 (7 recordings)	The motion capture system recording crashed, either during the session or when saving the recording, resulting in the loss of all trials for that recording.
Trial Not Captured*	2	Two trials were lost due to de-sync between the motion capture system and LED target system, resulting in either the first or last trial in the recording to be lost.
Markerless Data Processing Failure†	225 (5 recordings)	For unknown reasons, the videos could not be processed by Theia to compute the markerless data.
Marker Tracking Errors†	64	Markers could not be accurately tracked due to confusion with nearby markers or from markers coming dislodged from participant.
Wrong Target‡	40	Pointing to the wrong target was typically observed when the participant was distracted.
Multiple Pointing Gestures‡	21	Some trials contained multiple pointing gestures, typically observed in the distracted condition.
No Pointing Gesture‡	47	No pointing gesture was found, typically due to focus on the Stroop test in the distracted conditions.
Incomplete Pointing Gesture‡	35	Pointing gestures that continued into the subsequent trial, exceeding 2 seconds of the target LED switching-off.
Corrected Pointing Gesture‡	124	Pointing to the wrong target but correcting with a second ballistic movement. This was most common in the Accurate-Distracted (60 Trials) condition.
Total	1053	With the removal of these trials, we are left with a dataset containing gaze data with 11,367 trials out of a possible 12,420 (91.5%).

* Data Collection Failure, † Data Processing Failure, ‡ Gesture Outlier

C SYSTEMATIC OFFSET COMPENSATION MODEL COEFFICIENTS

Table 13 shows the coefficients for the ordinary least squares models, trained using the `lm` function from the `stats` package in `R version 4.3.1`. These are referenced in subsection 4.1.

Table 13. Coefficients for systematic offset compensation models [45] fitted to the rays in our dataset.

	EFRC (Cyclops)		HFRC		IFRC		FRC	
Coef.	Yaw	Pitch	Yaw	Pitch	Yaw	Pitch	Yaw	Pitch
a	2.259e-07	-4.916e-08	4.425e-07	1.386e-07	2.007e-07	-5.169e-08	-6.785e-08	-7.623e-08
b	7.302e-08	-1.753e-06	-3.556e-06	-9.030e-07	-7.201e-07	-8.401e-07	7.658e-07	-2.195e-06
c	7.095e-07	-9.223e-08	1.156e-07	-1.589e-07	-7.526e-07	-3.666e-08	-6.349e-07	-7.852e-07
d	-1.462e-06	-5.047e-07	2.488e-06	-1.212e-06	7.091e-07	9.984e-07	4.348e-06	2.417e-08
e	-2.205e-06	8.375e-06	-1.029e-05	-4.300e-06	9.146e-06	-2.638e-06	2.861e-05	-7.428e-06
f	2.705e-06	1.608e-04	-1.812e-04	1.941e-04	-1.466e-05	1.651e-04	-1.999e-04	3.369e-04
g	1.347e-06	2.386e-07	1.499e-06	3.027e-07	2.811e-06	-3.772e-09	3.756e-06	-1.625e-06
h	-2.650e-05	-9.955e-06	-5.865e-07	-8.965e-06	4.692e-05	-2.716e-05	4.357e-05	4.119e-05
i	-1.501e-04	-8.569e-06	-1.595e-05	-2.108e-05	-1.338e-04	-3.571e-05	-2.548e-04	2.534e-05
j	-1.077e-03	-4.936e-04	-1.146e-03	-5.737e-04	-7.425e-04	6.577e-04	1.750e-04	1.288e-03
k	7.910e-04	5.556e-03	-3.981e-05	3.212e-03	-1.310e-03	3.223e-04	3.434e-03	-6.841e-03
l	-2.377e-04	3.969e-04	-2.285e-03	1.173e-03	-2.629e-03	-3.777e-05	-6.319e-03	-1.140e-03
m	3.663e-02	2.889e-03	-4.683e-02	1.669e-02	8.859e-02	-2.478e-02	1.894e-02	-6.341e-02
n	2.905e-02	-1.826e-02	1.040e-01	-9.887e-02	-7.237e-02	4.161e-02	4.610e-02	-8.401e-03
o	1.649e+00	-1.805e+00	2.637e+00	3.014e-01	-7.210e+00	9.276e+00	-1.109e+01	1.589e+01

D FULL RM-ANOVA RESULTS FOR SUBJECTIVE MEASURES

Table 14 shows the full statistical results from the subjective measures. These are referenced in subsection 4.2.

Table 14. Table showing the main effects from the Pointing Style and Focus on the self reported measures. Significant main effects are highlighted using p < .001***, p < .01**, and p < .05*. Non-significant interactions are not reported. Partial eta-squared (η_p^2) reported with 5% and 95% confidence intervals. The interpretation values are $\eta_p^2 = 0.01$ (small), $\eta_p^2 = 0.06$ (medium), and $\eta_p^2 > 0.14$ (large).

Questionnaires		Effect	DF	F	p	η^2
	BRPE	Pointing Style	1, 66	18.05	<.001***	0.215 [0.085, 1.000]
		Focus	1, 66	45.41	<.001***	0.408 [0.260, 1.000]
NASA-TLX	Effort	Pointing Style	1, 66	6.85	.011*	0.094 [0.013, 1.000]
		Focus	1, 66	93.12	<.001***	0.585 [0.459, 1.000]
	Frustration	Pointing Style	1, 66	0.1395	.709	0.002 [0.000, 1.000]
		Focus	1, 66	45.49	<.001***	0.408 [0.260, 1.000]
	Mental Demand	Pointing Style	1, 66	7.04	.010*	0.096 [0.014, 1.000]
		Focus	1, 66	138.05	<.001***	0.677 [0.571, 1.000]
	Physical Demand	Pointing Style	1, 66	14.92	<.001***	0.184 [0.063, 1.000]
		Focus	1, 66	20.43	<.001***	0.236 [0.102, 1.000]
	Temporal Demand	Pointing Style	1, 66	2.80	0.099.	0.040 [0.000, 1.000]
		Focus	1, 66	82.14	<.001***	0.554 [0.422, 1.000]
	Performance	Pointing Style	1, 66	0.934	0.337	0.010 [0.000, 1.000]
		Focus	1, 66	0.428	0.515	0.006 [0.000, 1.000]
IMI	Effort Importance	Pointing Style	1, 66	8.80	.004**	0.118 [0.023, 1.000]
		Focus	1, 66	42.42	<.001***	0.391 [0.243, 1.000]
	Interest Enjoyment	Pointing Style	1, 22	0.867	0.362	0.038 [0.000, 0.259]
		Focus	1, 22	11.769	.002**	0.349 [0.056, 0.567]
	Perceived Competence	Pointing Style	1, 66	0.362	0.549	0.005 [0.000, 1.000]
		Focus	1, 66	5.20	.026*	0.073 [0.005, 1.000]
	Pressure Tension	Pointing Style	1, 66	2.251	0.138	0.030 [0.000, 1.000]
		Focus	1, 66	46.66	<.001***	0.414 [0.266, 1.000]

E FULL RM-ANOVA RESULTS FOR OBJECTIVE MEASURES

Table 15 shows the full statistical results from the objective measures. These are referenced in the temporal analysis (subsection 4.3), biomechanical analysis (subsection 4.4), and accuracy analysis (subsection 4.5).

Table 15. Table showing the main effects from the Pointing Style and Focus on our dependent variables including significant interactions. Significance is highlighted using p < .001***, p < .01**, and p < .05*. Partial eta-squared (η_p^2) reported with 5% and 95% confidence intervals. The interpretation values are $\eta_p^2 = 0.01$ (small), $\eta_p^2 = 0.06$ (medium), and $\eta_p^2 > 0.14$ (large).

Dependent Variable	Effect	DF	F	p	η^2
Overall Gesture	Pointing Style	1, 66	92.59	< .001***	0.584 [0.457, 1.000]
	Focus	1, 66	64.59	< .001***	0.495 [0.353, 1.000]
	Pointing Style ×Focus	1, 66	12.57	< .001***	0.160 [0.047, 1.000]
Start Phase	Pointing Style	1, 66	43.62	< .001***	0.398 [0.250, 1.000]
	Focus	1, 66	17.04	< .001***	0.205 [0.078, 1.000]
Hold Phase	Pointing Style	1, 66	102.54	< .001***	0.608 [0.486, 1.000]
	Focus	1, 66	43.31	< .001***	0.396 [0.248, 1.000]
	Pointing Style ×Focus	1, 66	21.01	< .001***	0.242 [0.106, 1.000]
Selection Time	Pointing Style	1, 66	102.72	< .001***	0.609 [0.487, 1.000]
	Focus	1, 66	54.97	< .001***	0.454 [0.309, 1.000]
	Pointing Style ×Focus	1, 66	18.68	< .001***	0.221 [0.090, 1.000]
Shoulder Torque (Nm)	Pointing Style	1, 66	4.71	.034*	0.067 [0.003, 1.000]
	Focus	1, 66	0.45	.503	0.007 [0.000, 1.000]
NICER	Pointing Style	1, 66	56.24	< .001***	0.460 [0.315, 1.000]
	Focus	1, 66	28.87	< .001***	0.304 [0.160, 1.000]
	Pointing Style ×Focus	1, 66	5.27	.025*	0.074 [0.005, 1.000]
Norm. Hand Movement	Pointing Style	1, 66	22.93	< .001***	0.260 [0.120, 1.000]
	Focus	1, 66	4.00	.049*	0.060 [0.000, 1.000]
Right Hand Usage	Pointing Style	1, 66	12.85	< .001***	0.163 [0.049, 1.000]
	Focus	1, 66	0.909	.344	0.010 [0.000, 1.000]
EFRC (Cyclops)	Pointing Style	1, 66	15.71	< .001***	0.192 [0.069, 1.000]
	Focus	1, 66	9.28	.003**	0.123 [0.026, 1.000]
	Pointing Style ×Focus	1, 66	6.64	.012*	0.091 [0.012, 1.000]
HFRC	Pointing Style	1, 66	21.85	< .001***	0.249 [0.112, 1.000]
	Focus	1, 66	16.24	< .001***	0.197 [0.073, 1.000]
	Pointing Style ×Focus	1, 66	4.77	.033*	0.067 [0.003, 1.000]
IFRC	Pointing Style	1, 66	15.72	< .001***	0.192 [0.069, 1.000]
	Focus	1, 66	9.55	.003**	0.126 [0.028, 1.000]
FRC	Pointing Style	1, 66	7.22	.009**	0.099 [0.015, 1.000]
	Focus	1, 66	2.756	.102	0.040 [0.000, 1.000]

F BEHAVIOUR CLASS DEFINITIONS

To describe the behaviours beyond rendered examples and written descriptions of the body pose in subsection 5.2, we used Decision Trees to extract the ranges within relevant features which describe a specific pose. Decision Trees are easily explainable as they provide a set of rules that one can apply to data to determine it's class, based on whether a feature's value is greater than or less than some value. Based on preliminary testing, we set the maximum depth to 6, and limit the trees to a maximum of 5 features. We perform 5-fold cross-validation, from which we take the model with the greatest accuracy. The accuracies obtained sit between 82-92%, as a result of misclassification of classes from the fitted features. This may be due to noise in our labelling process and/or the difficulty for a Decision Tree to fit to complicated relationships between multiple features.

F.1 Arm Pose

Classes sizes (proportion) and class weights - Hip fire: 414 (3.6%) 1.0, Outstretched 9857 (86.7%) 0.039, Proximal 1096 (9.6%) 0.338. Feature Importance - Elbow Extension: 0.789, Shoulder Abduction: 0.150, Hand Elevation: 0.061. Overall accuracy = 92.1%.

		True Value:		
		Hip fire	Outstreched	Proximal
Predicted Value:	Hip fire	327	9	50
	Outstreched	17	9782	39
	Proximal	70	66	1007

F.2 Finger Incorporation

Classes Sizes (proportion) and class weights - ADS: 8832 (77.7%) 0.144, Occluded: 2139 (18.8%) 0.536, Unoccluded: 396 (3.5%) 1.0. Feature Importance - EFRC and IFRC alignment: 0.828, EFRC and FRC Alignment (YAW): 0.077, EFRC and FRC Alignment (Overall): 0.065, Finger Flexion: 0.031. Overall accuracy = 82.5%.

		True Value:		
		ADS	Unoccluded	Occluded
Predicted Value:	ADS	7734	4	315
	Unoccluded	396	349	411
	Occluded	702	43	1413

F.3 Hand to Torso Alignment

Classes Sizes (proportion) and class weights - Medial: 1411 (12.4%) 1.0, Aligned: 1574 (13.8%) 0.633, Lateral: 8382 (73.7%) 0.140. Feature Importance: Hand and Near Shoulder Yaw: 1.000. Overall accuracy = 91.7%.

		True Value:		
		Aligned	Lateral	Medial
Predicted Value:	Aligned	1480	95	270
	Lateral	67	8287	0
	Medial	27	0	1141

Received 17 March 2025; revised 12 August 2025; accepted 20 August 2025