Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 69 (2011) 401–414
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Air pointing: Design and evaluation of spatial target acquisition with and without visual feedback

A. Cockburn[a,*], P. Quinn[a], C. Gutwin[b,1], G. Ramos[c,2], J. Looser[a]

[a] *Department of Computer Science and Software Engineering, Private Bag 4800, University of Canterbury, Christchurch, New Zealand*
[b] *Department of Computer Science, University of Saskatchewan, 110 Science Place, Saskatoon, Saskatchewan, Canada*
[c] *Microsoft Corporation, One Microsoft Way, Redmond, WA, USA*

Received 16 February 2010; received in revised form 17 February 2011; accepted 21 February 2011
Communicated by D.A. Bowman
Available online 25 February 2011

Abstract

Sensing technologies such as inertia tracking and computer vision enable spatial interactions where users make selections by 'air pointing': moving a limb, finger, or device to a specific spatial region. In addition of expanding the vocabulary of possible interactions available, air pointing brings the potential benefit of enabling 'eyes-free' interactions, where users rely on proprioception and kinaesthesia rather than vision. This paper explores the design space for air pointing interactions, and presents tangible results in the form of a framework that helps designers understand input dimensions and resulting interaction qualities. The framework provides a set of fundamental concepts that aid in thinking about the air pointing domain, in characterizing and comparing existing solutions, and in evaluating novel techniques. We carry out an initial investigation to demonstrate the concepts of the framework by designing and comparing three air pointing techniques: one based on small angular 'raycasting' movements, one on large movements across a 2D plane, and one on movements in a 3D volume. Results show that large movements on the 2D plane are both rapid (selection times under 1 s) and accurate, even without visual feedback. Raycasting is rapid but inaccurate, and the 3D volume is expressive but slow, inaccurate, and effortful. Many other findings emerge, such as selection point 'drift' in the absence of feedback. These results and the organising framework provide a foundation for innovation and understanding of air pointing interaction.

Keywords: Target acquisition; Proprioception; Spatial memory; Eyes-free interaction.

---

1. Introduction

Many interactions in the physical world take place without the need for visual feedback—we quickly and accurately flick a light switch in the dark, or manipulate a familiar gear stick without glancing. Interacting with graphical user interfaces is normally dependent on visual feedback; however, our experiences with interaction in the physical world suggest that computer use could also take advantage of eyes-free interaction. One way this is possible is through interfaces that allow item selection by pointing to a specific region in the space around the user ('air pointing'), ultimately without need for visual feedback. There are many air pointing techniques that can leverage human spatial cognition and proprioceptive memory to help users acquire commands or data held at consistent egocentric locations. For example, Fig. 1 portrays the three designs we study in this paper, which are based on 'raycasting', moving across a 2D plane, and moving through a 3D volume.

Eyes-free interaction offers many benefits to users (Li et al., 2008; Oakley and Park, 2007; Zhao et al., 2007). It can be rapid because it does not require continual conscious monitoring (for example, reaching to keys while touch typing), and convenient because users do not need to shift their visual attention (for example, they can maintain

---

[*] Corresponding author. Tel.: +64 3 364 2987x7768; fax: +64 3 364 2569.
E-mail addresses: andy@cosc.canterbury.ac.nz (A. Cockburn), gutwin@cs.usask.ca (C. Gutwin), gonzalo@microsoft.com (G. Ramos).
[1] Tel.: +1 306 9668646; fax: +1 306 9664884.
[2] Tel.: +1 425 502 8418.

1071-5819/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2011.02.005

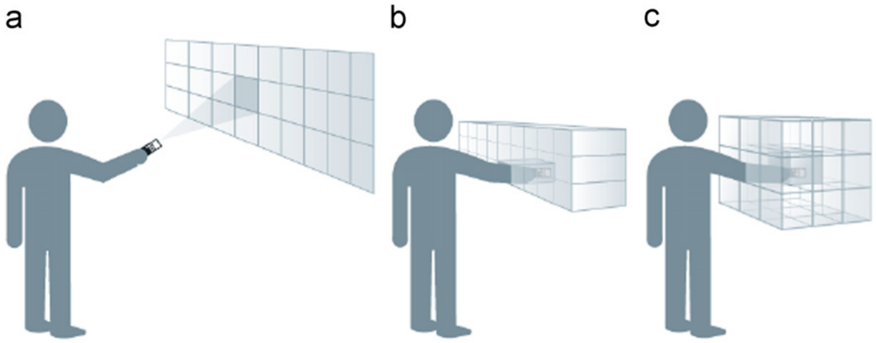

Fig. 1. Three different air pointing interfaces: (a) aiming the pointer at a virtual screen; (b) translating the pointer across a 2D plane; and (c) translating the pointer within a 3D volume. (a) Raycasting, (b) 2D plane, and (c) 3D volume.

eye contact with another person). These potential benefits have prompted several researchers to examine eyes-free interaction (Kurtenbach and Buxton, 1994; Kurtenbach et al., 1993; Zhao et al., 2007), typically through multimodal feedback or gestural input that exploits small-scale human kinematics on the small surface of a mobile device.

Our research, in contrast, explores eyes-free interaction possibilities when the comparatively larger space around the user is available. Large-scale spatial interaction should allow fast and accurate eyes-free acquisition due to human proprioception (based on findings in the psychology literature, reviewed in the following section). Techniques in this area are also becoming practical, due to the range of devices supporting motion sensing, including the XWand (Wilson and Shafer, 2003), Apple iPhone,[3] Nintendo Wii Remote (Wingrave et al., 2010), and Logitech MX Air mouse.[4]

Spatial interactions are a prominent component of research in 3D user interfaces and virtual reality environments, with many studies examining techniques that use pointing, reaching, and gestures to navigate, partition, and select items within a virtual space (Bowman et al., 2005; Mine et al., 1997). These techniques are designed with an assumption that visual feedback will be continually available during the manipulation. Our research explores techniques that share many of the interaction qualities; however, our goal with air pointing is to relax the dependence on visual feedback through the development of proprioceptive memory and expand the range of applications beyond virtual environments.

The following design scenario exemplifies our area of interest and loosely guides our investigation:

> A car manufacturer has developed tracking technology that can accurately measure the spatial location of the driver's fingertips. They intend that drivers will select and manipulate controls by moving their index finger to locations in the air around them, and then tapping the tips of their index finger and thumb to select. Visual feedback of controls and cursor location will initially be displayed on the windshield using heads-up projection, but this will gradually fade to nothing as drivers learn the controls' spatial locations.
>
> The designers believe this interaction style will be safer than current technologies that require users to look away from the road.

Such designs should promote spatial learning and maximum proprioceptive expressivity. The visual feedback is *supplementary* to the interface—it is not designed to be the primary driver of interaction, but a passive support mechanism; ideally, users should develop the ability to perform interactions confidently without feedback to guide them.

In this paper, we explore the design space and associated human factors for air pointing, presenting a framework that identifies important concepts in designing eyes-free air pointing techniques and highlights the interaction qualities that different design decisions can affect. The first part of the framework sets out the qualities of interaction that designers can attempt to achieve with air pointing—including learnability for novices, speed for experts, accuracy, expressiveness, cognitive simplicity, and physical comfort. The second part of the framework identifies the input dimensions of air pointing – reference frame, input scale, degrees of freedom, and feedback – which can be used to try and achieve a desired interaction.

We conducted an experiment that compared three different air pointing techniques that sample different points in the design space (Fig. 1). For each design approach, we tested how well users were able to acquire a set of arbitrary spatial locations as visual feedback was progressively removed. Results show that the 2D plane interface is fast, accurate, and popular; raycasting is fast but inaccurate, and the 3D volume is expressive but slow, inaccurate, and effortful. The study provides initial data for expanding on the relationships outlined by the framework, and identifies an initial set of design factors that are important in developing air pointing techniques.

---

[3] http://apple.com/iphone.

[4] http://logitech.com/mice-pointers/mice/devices/3443.

2. Related work

The following sections briefly review prior work on air pointing input systems, human factors research on proprioceptive target acquisition, and the use of spatial memory as a lever for improved interaction. Hinckley et al. (1994) provide a thorough review of the broader topic of spatial input for interested readers.

2.1. Air pointing input systems

Many researchers have investigated interaction methods that are activated by either holding a device in the air or by tracking the users' limbs or eyes. For example, Bolt (1980) coupled gesture tracking with voice commands to allow interactions such as "put that there", and eye-gaze tracking for window selections. Baudel and Beaudouin-Lafon (1993) described the *Charade* system, which used a DataGlove to support natural, terse, direct interaction for tasks such as giving presentations, and Pierce et al. (1997) describe a variety of image plane interaction techniques that allow users to interact with 2D projections of 3D spaces using head-tracked immersive virtual environments. Other systems have coupled hand and body gestures to enhance expressiveness and playfulness (Brewster et al., 2003; Krueger, 1991).

There has been extensive work on 2D cursor movement using devices or hands held in the air. Vogel and Balakrishnan (2005) provide a recent review of pointing techniques, and present an evaluation of several gestural pointing and clicking interfaces on large displays. They found that raycasting methods were faster than other techniques, but that imprecision made it ineffective for small targets. Myers et al. (2002) examined pointing with a handheld laser-pointer, and found it to be slow and inaccurate due to jitter—even when noise reduction techniques were used. Yee (2003) investigated 'Peephole Displays', which use handheld interactive techniques (like air pointing) to control the display's viewport; Cao et al. (2008) subsequently modelled user performance in similar 'peephole pointing' activities.

3D cursor control has also been considered, with Zhai (1998) providing a strong review of the performance of various isotonic 'flying mouse' designs and identifies several problems with them—including their limited movement range, problems with co-ordination (in particular, clutching), fatigue, and the difficulty of maintaining persistent locations.

The broader field of 3D user interfaces and interaction methods is reviewed by Bowman et al. (2005), which includes a classification and review of selection and manipulation techniques that employ spatial gestures and manipulations. Steed (2006) and Poupyrev et al. (1997) also review selection techniques in virtual environments.

Raycasting interfaces (pointing with a virtual ray from the user's hand) have been examined in several studies of 3D interfaces and virtual environments. Poupyrev et al. (1998) compared raycasting and 'virtual hand' techniques (where there is a mapping between the user's hand position and the selector in the virtual environment), finding performance to be generally comparable. However, they argue that performance and suitability of the techniques is task-dependent: for example, raycasting hindered performance when selecting small or distant objects due to the high level of angular accuracy required. Bowman et al. (1999) present a methodology and evaluation of nine selection and manipulation techniques—including raycasting, linear, and non-linear 'virtual-hand' styles. They found raycasting to perform significantly faster than arm-extension techniques for selection tasks and to be less effortful than occlusion techniques; they also found males performed faster than females. Wingrave et al. (2005) compared performance with raycasting and occlusion techniques with self-report data and measures from psychological aptitude tests. They found a number of positive correlations, including between performance with raycasting techniques and aptitude test scores, and between raycasting performance and participant height and arm length.

None of these research projects use eyes-free air pointing (although, Mine et al. (1997) discuss the possibility of using proprioception in virtual environment interaction). However, Li et al. (2009) recently proposed 'virtual shelves' for 2D eyes-free pointing (similar to Fig. 1(b), but across a curved surface, rather than a plane); their results showed that participants could acquire up to 28 items with reasonable accuracy, although mean selection times were slow at ~3 s per item.

2.2. Eyes-free and proprioceptive target acquisition

Eyes-free interaction is achieved in one of the three ways: (1) by enabling proprioceptive target acquisition; (2) using commands that do not need visual feedback for selection (such as gestures or speech); or (3) by providing guidance on non-visual output modalities. Our interests lie primarily in proprioceptive target acquisition, described below. We refer readers interested in multi-modal feedback for eyes-free interaction to the recent review by Oakley and Park (2007).

A number of models have been developed to describe how feedback is used to mediate rapid, aimed pointing movements (first studied by Woodworth (1899), and briefly reviewed by Elliot and Lee (1995)). Prablanc et al. (1979) and Elliot and Madalena (1987) studied the importance of visual target information in pointing tasks; they found that not only did pointing errors increase as visual feedback was removed during pointing, but also when a visual representation of the environment was removed prior to pointing.

Proprioception is our sense of the relative location of the parts of our body and is a key component in muscle memory (Sherrington, 1907). As with air pointing systems, there have been human factors studies of 2D and 3D locations. In 2D (using a $30 \times 30\,\text{cm}^2$ plane), Hocherman (1993) examined the ability to reach to proprioceptively defined targets (using acoustic cueing during training), and found that people can quickly and accurately reproduce reaching movements without visual guidance. Additionally, he found no significant difference between the accuracy of the open-loop motion phase and the final target acquisition, suggesting that people can rapidly draw on proprioception alone for targeting. Soechting and Flanders (1989) conducted a series of related experiments in 3D, finding that people reproduced target directions accurately, but made significant errors in distance. Similarly, Medendorp et al. (1999) found that people systematically undershot distant targets, while sometimes overshooting nearby ones when pointing to memorised targets.

Crossman and Goodeve (1963/1983) discussed the interaction between visual and proprioceptive feedback when carrying out pointing actions in several experiments. They concluded that proprioceptive systems can serve as a good fallback for detecting accuracy errors when visual feedback is unavailable.

2.3. *Spatial learning, reacquisition, and interfaces*

Psychology literature shows that humans learn spatial locations without paying particular attention to them (Andrade and Meudell, 1993; Postma and De Haan, 1996), and learning is improved with focussed attention (Naveh-Benjamin, 1987). Furthermore, it also shows that humans can reacquire spatial locations with whole-body locomotion when the walking time to target is less than 8 s (Thomson, 1983). Elliot and Lee (1995) and Elliot and Madalena (1987) showed similar effects for manual aiming (pointing with a pencil) if the target's icon is available in memory from a visual scan up to 2 s before the action. Elementary psychological issues of how spatial information about body and object location relationships is encoded, retrieved, and manipulated (without vision) has also been examined (Easton and Sholl, 1995).

There is extensive research demonstrating that spatial memory is a valuable resource in efficient interaction. Several studies agree that measures of spatial aptitude correlate well with efficiency when using document editors (Egan and Gomez, 1985), computer games (Gagnon, 1985), and file browsers (Vicente et al., 1987). Several studies also show that interfaces that exploit spatial memory aid performance over those that do not (Cockburn et al., 2006, 2007; Czerwinski et al., 1999).

With the exception of Li et al.'s (2009) 'virtual shelves' and Ängeslevä et al.'s (2003) 'body mnemonics' we are unaware of previous interfaces that explicitly leverage spatial memory to help users select items, eyes-free, from spatial locations in the air around them.

3. A design framework for air pointing

3.1. *Part 1: interaction qualities*

Several interaction qualities can be identified as potentially desirable characteristics for air pointing techniques, based on the domains in which they are used. All of these areas of use involve eyes-free interaction—including activities such as driving, mobile computing with handheld devices, and interfaces for the visually impaired. Therefore, the first part of the framework states six basic qualities that form the endpoints and goals of the design space:

1. *Learnability for novices*: With sufficient practise, humans can eventually learn sophisticated motor activities like juggling and complex procedural tasks like navigating arcane user interfaces. Good interface design, however, promotes a rapid transition from novice to expert performance, and we wish to explore how interfaces can exploit spatial memory and proprioception to achieve this.
2. *Selection speed for experts*: Expert users should be quickly able to acquire targets without disrupting their visual focus on salient items in the environment. The power law of practise (Newell and Rosenbloom, 1981) suggests that users can rapidly capitalise on experience, enabling them to quickly reach a performance asymptote with interfaces that promote learnability. Consequently, users are likely to spend proportionately more time expert than novice, so expertise is critical.
3. *Accuracy*: Selection errors caused by slips or control imprecision are time consuming, frustrating, and can cause dangerous distractions in drawing the user's attention away from critical activities. Traditional mouse-driven item selections are minimally susceptible to errors because users can visually confirm their target before clicking the button, and importantly the mouse is a self stabilising device due to static friction with the surface it rests on. Consequently, error rates remain relatively low unless users are deliberately rushed (Wobbrock et al., 2008).

   Air pointing devices, however, are not self stabilising and must therefore address the challenge of human tremor. Furthermore, our eyes-free objective means that our designs will not be able to rely on visual target confirmation. It is unclear how accurate users can be with these constraints.
4. *Expressivity*: The number of spatial locations that users can rapidly and accurately retrieve pivotally influences the effectiveness of air pointing designs. For example, if an interface has n possible commands, but users can only accurately acquire two spatial locations (say, 'left' and 'right'), then command retrieval will demand a sequence of $log_2 n$ discrete acquisitions. However, if users can instead acquire 27 different locations (such as a $3\times3\times3$ cube of 3D items) then only $\log_{27} n$ actions are needed, allowing access to 729 commands with two discrete actions.
5. *Cognitive effort*: An important motivation for eyes-free interaction stems from the safety and convenience of eyes-free interaction in activities like driving or walking. However cognitive distraction negatively impacts on safety (Harbluk and Noy, 2002; Lee and Strayer, 2004). Again, we hope that proprioception and spatial memory will enable cognitively lightweight interactions.

6. *Comfort*: Finally, the interaction must be socially and ergonomically satisfactory. For example, an extended reaching action might be satisfactory for rarely reoccurring actions, but entirely unsuited for frequent ones and undesirable to perform in many social contexts.

3.2. Part 2: interaction dimensions

The second part of our framework identifies the major ways in which air pointing interactions can differ—specifying the possibilities for the design of new techniques. Like the interaction qualities described in Part 1 (Section 3.1), the interaction dimensions are dictated by the fact that air pointing will use human spatial memory and proprioception to help users acquire specific regions in space, ultimately without need for visual feedback. The five dimensions of the framework, summarised in Fig. 2, are the reference frame for the air pointing technique, the scale of input control, the input degrees of freedom, the feedback modality, and the feedback content:

1. *Reference frame for spatial input*: Spatial locations can be absolute, relative to an external object, relative to the body, relative to the device, or some hybrid combination. *Absolute location* (or *relative to world*) targets are acquired by motioning towards the real-world location of the target, relative to the user/pointing device. Specifying locations *relative to an external object* establishes an external object as the origin for which motions are interpreted as being relative to; unlike a *relative to world* reference frame, this origin moves with the external object. *Relative to body* locations (such as ‘three o’clock high’ or ‘just in front of my nose’) utilise the user’s own body as the origin for their motions to be interpreted relative to—essentially a *relative to an external object* reference frame, with the ‘object’ being the user themselves. Finally, *relative to device* locations utilise the local space around a device for actions to be interpreted relative to—a *relative to an external object* reference frame, with the ‘object’ being the initial location of the pointing device itself.

   Absolute locations are largely constrained to one physical location as motions must be towards the actual location of the target as it exists in the world. While this may be useful in common locations that have a relatively static layout (such as a building plan, home, or office), it requires the user to perform spatial transformations to derive the correct motion vector from their current location to the target (Easton and Sholl, 1995). The user can leverage their spatial memory in familiar locations (for example, sending a file a printer by ‘pointing’ at its fixed location from your office), but may require pause to do so in unfamiliar locations (for example, the same task from somebody else’s office), or difficult and error-prone as the distance to the object increases (for example, from another building).

   Relative to external object locations, typified by touch-typing, may be useful in contexts such as driving or piloting, where items remain in fixed locations relative to the vehicle. For example, adjusting the stereo while driving or adjusting a music player while it remains in your pocket leverages spatial knowledge of the local environment and device. The vehicle becomes the origin (regardless of its absolute position) for targeting the items within it. However, as with *absolute locations*, it requires spatial transformations when the user is in an unfamiliar orientation to the external object—for example, when in the passenger seat of a car that you typically drive.

   Relative to body locations are unconstrained by specific locations or objects; they do not require spatial transformations around some external object and can leverage proprioceptive memory. However, it is susceptible to ambiguity. First, it is unclear which part of the body provides the ‘best’ frame of reference (eyes, head, shoulder, torso, etc.). For example, if a user knows their calendar resides at a particular location (say, up and right), should the gesture be produced with respect to the current orientation of the torso, or the head? In different contexts either would make sense, but only one can be interpreted as the origin. Second, a *relative to body* frame of reference complicates determining the user’s actual intentions. For example, if the head provides the frame of reference, should the head’s roll and pitch be considered when determining direction, or only yaw? A mismatch between the user’s expectations and the system’s interpretation could lead to confusing selection errors.

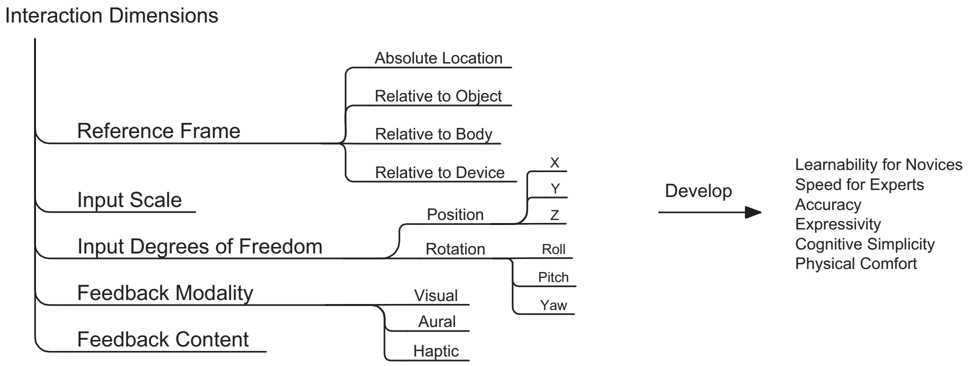

Fig. 2. Summary of the air pointing design framework.

Relative to device locations can allow small-scale, rapid movements and are well-researched with gesture-controlled interfaces (Kurtenbach and Buxton, 1994; Wobbrock et al., 2003; Zhao et al., 2007). Mental transformations are only required when operating the device in an unusual orientation and there are opportunities for leveraging spatial and proprioceptive memory. However, an origin must be signalled manually before each interaction and there may be some ambiguity about how the orientation of the device should be interpreted.

Hybrid frames of reference are also possible, where a context or 'origin' is established in one frame for subsequent actions to be performed in another; or the reference axis is specified in one frame, but manipulated in another. This is exemplified in our experimental interfaces, which use a three step process—first, the user establishes a movement origin relative to the body, followed by a consistent body referenced human movement to the target space, terminated with a selection action that measures the device's displacement from the origin.

2. *Scale of spatial input control*: The physical movements for controlling spatial input can be small (such as twisting the wrist or flicking a finger) or large (such as reaching an arm). Small control movements allow rapid and subtle actions that are socially acceptable in most settings (not substantially different to the frenetic thumb actions used by advanced text messagers). However, subtle movements demand a relatively fine granularity of control, which may influence the number of items that can be proprioceptively acquired. Finally, neurophysiology literature has found large undershoots (falling short of the required movement distance) for more distal targets (Medendorp et al., 1999), which may harm the accuracy/expressivity of large spaces.
3. *Degrees of freedom in spatial input*: The six degrees of freedom of 3D spatial input consist of three translations (movements on the x, y, and z axes) and three rotations (*pitch*, *roll*, and *yaw*). Translations allow users to specify points in 3D space (such as placing a device 'high', 'left', and 'back'). Rotations readily allow remote pointing in 2D space (such as casting a 'ray' onto a wall with a laser-pointer by manipulating pitch and yaw), but the rotational specification of depth is less obvious (for instance, using wrist rotations to control cursor depth; see Grossman and Balakrishnan (2006) for a discussion of remote 3D point specification). If a handheld device is used to display feedback, rotational input methods may also move the display into an orientation that frustrates viewing. Finally, selections demand an additional input control such as a button click, screen tap, or dwell timeout, which must not interfere with the spatial specification.
4. *Feedback modality*: Users need feedback to learn the association between locations and commands, and the feedback modality may influence how well this happens. Visual feedback is the dominant modality, but researchers have explored audio (Brewster et al., 2003, 1993) and haptic feedback (Brewsterand and Brown, 2004) for situations where the eyes are otherwise engaged or opportunities for visual feedback are limited (such as handheld devices with small screens).
5. *Feedback content*: The information content is also important to consider for supporting novices in searching and navigating the information space and for helping users confirm selections. Issues include: number of items to display at once (such as only the current item, or the current item and all its immediate neighbours); the way to communicate the user's location within the dataset (such as an overview widget); and whether or not the connection between device-space and information-space can be relaxed (for example, allowing users to browse item locations by scrolling a display without air pointing).

4. Experiment: spatial acquisition with and without visual feedback

To explore some of the factors in the design framework we carried out an experiment with three different air pointing techniques. The study was designed to investigate four questions about air pointing that arise from the relationship between interaction dimensions and resulting interaction qualities:

1. How quickly and accurately can users acquire targets at known locations with and without visual feedback?
2. How is performance influenced by the input degrees of freedom? For example, what are the differences between acquisition in 2D and 3D spaces?
3. How is performance influenced by the scale of input movement?
4. Are there clear subjective preferences or workload differences between spatial layouts?

To begin exploring these issues, we implemented three air pointing interfaces, as summarised in Fig. 1. All three interfaces use a hybrid "relative to body/relative to device" reference frame, which requires a lightweight orientation action to establish a reference point for each selection. Acquiring an item with any of the interfaces is a three stage process. First, the user establishes a relative movement origin by holding the device in a comfortable neutral position and activating a control. Second, the user positions the device (through translation or rotation, as appropriate for the interface) to a point in space that corresponds with the target. Third, they specify the selection point with an interface action.

The primary differences between the three techniques are in the *degrees of freedom in spatial input* and the *scale of spatial input control.* Two interfaces use 2DOF input, encouraging users to conceive of items located on a virtual plane in front of them: one using small scale movements to 'raycast' to the item using pitch and yaw rotations similar to using a laser pointer ('raycasting', see Fig. 1(a)); the other using large scale x and y translations ('2D plane', see Fig. 1(b)). The other interface uses 3DOF input, encouraging users to conceive items located in 3D volume, locating items with x, y, and z translations ('3D volume', see Fig. 1(c)). All of these three interfaces allow users to conceive of targets as being placed in a large scale space around them (as shown in Fig. 1), but they offer substantially different interaction mechanics to specify the spatial point associated with each target. Our experimental objective is to characterise how these broadly different interaction mechanics influence the participants' ability to develop and draw on their proprioceptive memory to achieve eyes-free target acquisition.

Our method for examining this development and use of proprioceptive memory involved training and testing participants' target acquisition through a series of conditions that gradually reduced the amount of visual feedback until there was none. We analyse performance during the reduction of feedback as well as during absent feedback (eyes-free) to characterise how the interfaces influenced the participants' ability to refine their proprioceptive memory.

4.1. Participants and apparatus

Fifteen postgraduate students (two female) took part in the experiment, which lasted approximately 1 h. They stood approximately 2 m in front of three $241 \times 183\ \text{cm}^2$ rear-projected displays at a resolution of 1024×768 pixels each (see Fig. 3). The side displays remained off, and there were no obvious visual markers in the participant's field of view (which might have been used as 'landmarks' for target locations). Room lights were dim. The experimenter sat at a desk behind the participant's left shoulder. Participants held a wireless pistol-grip handle (from a joystick) in their dominant hand, using its trigger to specify an origin point before moving the handle to the target location and re-clicking the trigger. The handle's location was tracked with millimetre precision using an ART infrared motion tracker.[5]

4.2. Interfaces

The same interface cued trials with all interfaces (shown in Fig. 3). It displayed the target name on the left, while the main interactive region displayed varying levels of feedback about the location of the cursor and target.

When visual feedback was displayed, the targets were shown as transparent green circles of 60 pixel diameter at 75% opacity, with a solid centre circle of 14 pixels in a marked 650×650 pixel area ($156 \times 156\ \text{cm}^2$, with the origin at the centre). The cursor was identically displayed, except light blue. The 3D volume interface used the same display for items on its front plane, but the targets and cursor diminished with depth, to a minimum of ~ 6 pixels at the maximum target depth (there was no stereoscopy, motion parallax, or other depth cues); the name labels on targets in all conditions were at a constant size of 60 pt.

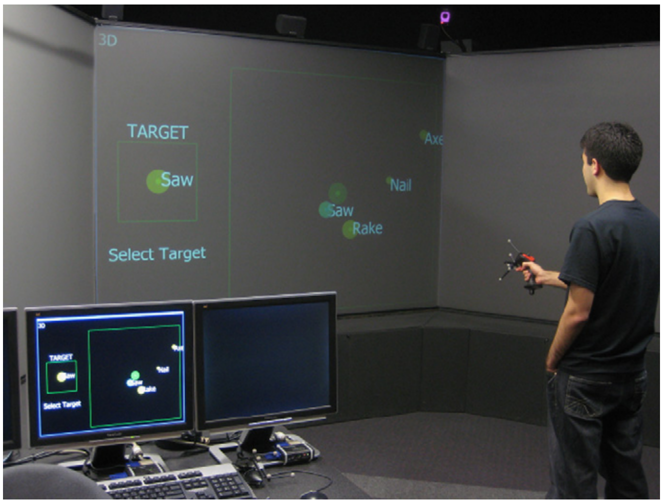

Fig. 3. A participant using the 3D volume interface. The cueing interface is on the left of the main screen, the area on the right provides feedback about target and cursor locations within the pointing space.

4.2.1. Raycasting

The raycasting interface was controlled with small angular movements of the handle, predominantly at the wrist. Participants were encouraged to think of the interface as laser pointing onto a wall (as in Fig. 1(a)). On the x-axis, wrist flexion points left and wrist extension right, and on the y-axis radial deviation points up and ulnar deviation down. NASA anthropomorphic measures (NASA, 1995) show that 5th and 95th percentile maximum movements for males and females are as follows (using tuples of 5th, 95th male, 5th, 95th female)—wrist flexion: 61.5°, 94.8°, 68.3°, 98.1°; wrist extension: 40.1°, 78.0°, 42.3°, 74.7°; radial deviation: 16.9°, 36.7°, 16.1°, 36.1°; ulnar deviation: 18.6°, 47.9°, 21.5°, 43.0°.

We chose to limit targets to those that could be achieved through wrist movements alone to reduce the possibility of interference through pointer translation that could occur if movement of the elbow or arm joints was required. Allowing participants to utilise a full range of arm motion would increase the expressible space of raycasting, but may impact on its performance and accuracy characteristics (further discussed in Section 6). Additionally, there are usage scenarios where small wrist-only movements may be better than those requiring large scale movements—on a bus, in a crowd, or driving your car. The wrist-only technique allows movement with the arm kept close to the body.

To ensure all targets are readily attainable with wrist movements, the maximum target displacement from the origin on either axis is 30°—within the fifth percentile maximum movements on the x-axis, and close on the y-axis (and elbow rotation accommodates rare target/participant couplings that prohibit acquisition with the wrist alone).

---

[5] http://ar-tracking.de.

4.2.2. 2D plane

Fig. 1(b) portrays the 2D plane interface in which items are conceived as being located on a flat surface in front of the user, extending from comfortable arm's reach top-left to bottom-right. Although comfortable arms reach is best described by an arc (NASA, 1995; Sengupta and Das, 2000), we used a flat plane to help leverage user's familiarity with flat working surfaces and to match the visual feedback presented on a flat screen. Information from the *z*-axis is ignored, making items selectable in a 'tunnel'. The plane extends one cubit (a measure based on forearm length, 45.72 cm)[6] horizontally and vertically from the origin.

4.2.3. 3D volume

The 3D volume, depicted in Fig. 1(c), extends the 2D plane into three dimensions. The origin lies at the centre of the front plane, extending one cubit horizontally and vertically from the origin (like the 2D plane), but also one cubit in depth. Target and cursor depth is depicted by adjusting visual size: their size diminishes with depth.

4.3. Method

All participants completed five experimental stages with each of the three interface types. Interface order was balanced using a Latin square. The first stage was for interface familiarisation (logs discarded): participants were given brief instructions on how to select items, and they then had 90 s of free-form interaction using the cuing and selection interface. The remaining four stages involved repeatedly reacquiring four targets with progressively less visual feedback. In all stages, target acquisition involved clicking to establish the origin, then moving to the target, and finally clicking to acquiring it. There was no notion of a successful or unsuccessful acquisition—the second click completed the trial regardless of the final position, and automatically cued the next. To help participants maintain a consistent movement origin, the previous origin was displayed on the screen with pointer location feedback prior to each trial, allowing users to move to that point before their initial click. Participants could take breaks between each experimental stage and between each interface. Interfaces were rated on NASA TLX workload sheets (Hart and Staveland, 1988) and commented on between interfaces participants ranked their interface preference after the final interface. The four experimental stages progressively reduced feedback, as follows:

1. *Training with full visual feedback*. This stage trains participants on target locations and provides baseline accuracy measures. Four random target locations were calculated for each participant with each interface. The locations could be anywhere within the available pointing space ($60^\circ \times 60^\circ$ for raycasting, 91×91 cm^2 for 2D, and $91 \times 91 \times 46$ cm^3 for 3D), but a location was recalculated if it was closer than 5 cm (2D/3D) or 5° (raycasting) to another target on any axis or any edge of the pointing space. The four targets were then randomly assigned single-syllable names from the sets "Ant, Bee, Cat, Dog", "Axe, Nail, Rake, Saw", and "Cup, Fork, Pan, Wok", with the name sets rotated between interfaces.

   Participants then completed twelve blocks, with each block consisting of four trials (one for each target). There was full visual feedback throughout, with all targets, the origin point, and cursor all continually displayed (see Fig. 3). Once all 48 trials (12 blocks × 4 targets) were complete, the participants moved to the second experimental stage.
2. *Repeated trials without cursor feedback ('no cursor')*. This stage is intended to support development of kinaesthetic and proprioceptive memory for each target. It consisted of 48 trials in four blocks, with each block involving twelve repetitions of the same item. The target was shown in the cueing interface, and its location displayed on the screen. Unlike the previous stage, however, all feedback was removed as soon as the participant clicked to establish the movement origin. They then moved to the target with only kinaesthetic and proprioceptive memory to guide them (as well as vision of their limb position). After clicking to select the target the cursor and target locations were displayed for 2 s, allowing the user to refine their movement on successive trials.
3. *Trials without initial locations ('no location')*. This stage provides final location training. It consisted of 48 trials in twelve blocks, with each block containing one trial for each target. Target, origin, and cursor locations were not shown until after participants made their selection. Participants therefore had to remember target locations and reproduce acquisition actions without any display feedback.
4. *Eyes-free trials without visual feedback ('blank')*. The objective of this stage is to measure and characterise speed and accuracy of 'eyes-free' spatial target acquisition. It consisted of 48 trails in four blocks, with each block consisting of twelve rapid repetitions of the same target. There was no feedback of origin, target, or cursor location.

4.4. Design

Dependent measures are acquisition time (between the origin and selection clicks) and distance between the selection point and the target' actual location. Subjective workload and preferences are also measured.

Dependent measures are analysed using a 3×4 within-subjects analysis of variance for factors *interface* (raycasting, 2D plane, and 3D volume) and *feedback* type (visual

---

[6] The same value, 45.72 cm, was used for all participants to avoid differences in arm length influencing the resolution of the pointing space. Motion was not restricted, so participants could rotate or tilt their torso to reach extreme targets comfortably.

training, no cursor, no location, and blank). Euclidean distance measures need to be normalised to allow small raycasting movements to be equitably compared with large 2D plane and 3D volume movements. We use pixel distances to do so. All interfaces used visual feedback to train participants, displaying targets inside a 650 × 650 pixel boundary. The 650 pixels on either axis correspond to 91 cm movements with 2D/3D and to 60° angular movements with raycasting. The 3D interface used 46 cm movements for maximum depth, corresponding to 325 pixels. Distance, therefore, is calculated as the Euclidean pixel separation between the target's location and the selection point.

One critical aspect of this experimental design is that targets are not discrete—there was no notion of hitting or missing the target. Instead, every selection provides data on selection accuracy (the distance between the ideal target location and the user's estimation). Although unlike traditional target acquisition, this important experimental design decision allows us to characterise the underlying human factors of spatial target acquisition with the three interface styles, whereas traditional discrete targets would not because there would be no basis for selecting any particular number or layout of targets in each spatial arrangement. For example, if we had chosen to place 27 discrete targets in a 3 × 3 × 3 matrix for 3D and in a 9 × 3 arrangement for 2D and raycasting, then the results would not necessarily generalise beyond the selected layout—3 × 3 × 3 might coincidentally be the 'sweet spot' (or its antithesis) for 3D.

5. Results

Results are organised below in terms of the interaction qualities from the design framework that were explored in the study: speed, accuracy, expressivity, and effort.

5.1. Selection times (speed)

Fig. 4(a) shows mean acquisition times with the three interfaces in each feedback type. Selections were generally very rapid, at just over 1 s with feedback and just under without, except for the 3D interface. There is a significant main effect of *interface* ($F_{2,28} = 73.07$, $p < 0.001$), with raycasting fastest (mean 1078 ms, SD 495 ms), closely followed by 2D (1116 ms, 501 ms), and 3D substantially slower (1922 ms, 1017 ms). *Feedback type* also showed a significant effect ($F_{3,42} = 28.8$, $p < 0.001$), with the initial training stage much slower than the others (as expected), with means of 1975, 1097, 1295, and 1121 ms, respectively, for training, no cursor, no location, and blank. A significant *interface × feedback* interaction ($F_{6,84} = 19.5$, $p < 0.001$) is best attributed to the marked slow performance of 3D during training (Fig. 4(a)), in which participants had to match targets on three axes rather than two.

5.2. Accuracy

Mean normalised target 'miss' distances are shown in Fig. 4(b). Selections were most accurate with 2D (mean distance 35.5 pixels), followed by raycasting (43.8 pixels) and 3D (72.1 pixels), giving a significant effect of *interface* ($F_{2,28} = 52.8$, $p < 0.001$). Naturally, participants became less accurate in the absence of feedback (significant effect of *feedback*, $F_{3,42} = 146.3$, $p < 0.001$), increasing from 14.0 pixels (SD 8.2 px) with complete dynamic feedback during training, through 43.5 (22.8), 68.3 (30.9), and 76.0 (31.1) pixels as feedback decreased across respective stages. Fig. 4(b) also suggests the cause of a significant *interface × feedback* interaction ($F_{3.4,47.6} = 5.8$, $p < 0.001$; fractional degrees of freedom stem from Greenhouse–Geisser correction for violated sphericity assumption)—miss distances increase rapidly across stages with 3D and raycasting, but much less rapidly with 2D.

To further characterise accuracy, Fig. 5 shows one participant's distribution of selections around the four targets in each of the interfaces during training and blank feedback types (these results are representative of other participants). Two columns are used in the figure for the 3D interface, displaying selection distributions on the (x, y) and (z, y) planes. The plots in the Training row show that participants were able to make accurate selections on the

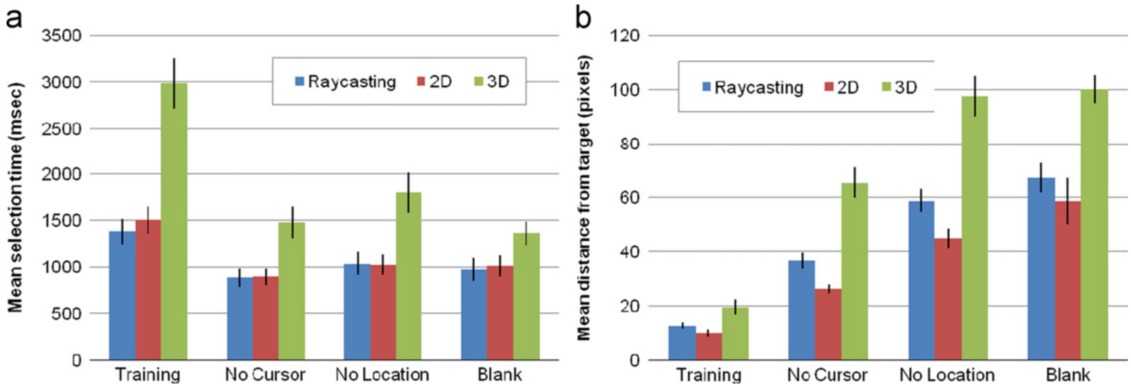

Fig. 4. Speed and accuracy results for the three interfaces across feedback type (error bars ± 1 standard error). (a) Speed: mean selection time. (b) Accuracy: mean distance from target.

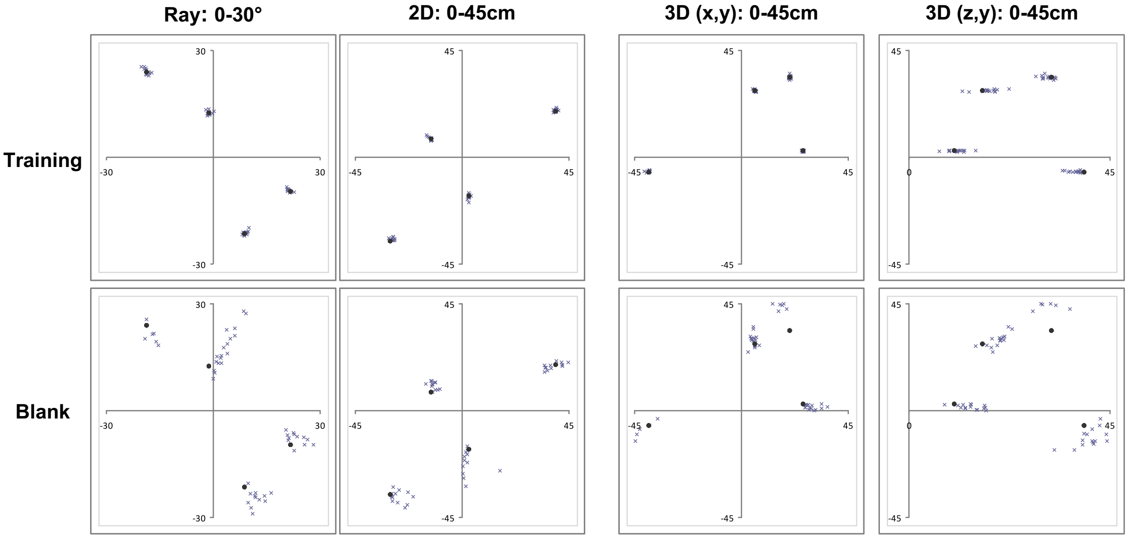

Fig. 5. A participant's dot-cluster of selections with training and blank feedback types. 3D columns show (x, y) and (z, y) planes.

(x, y) plane when visual feedback was available. However, 3D depth selections were relatively inaccurate (rightmost column of Fig. 5). Clearly, matching object sizes (used to communicate depth) was more difficult than matching target and cursor centres. This is an interesting challenge for 3D air pointing, which we return to in the discussion.

The plots in the lower row of Fig. 5 show selection distributions in the *blank* condition, during which no display feedback was provided. The selections are clearly much more widely distributed around targets (particularly in the (z, y) 3D plane), but generally selection clusters are apparent in target vicinities. Target 'drift' is also apparent in several participants' data, as shown in the figure by the Ray and 2D targets that lie close to the y-axis—it shows selections gradually moving away from the target, up and right with Ray, and vertically downwards with 2D. In the absence of anything confirming the success or failure of their selections, it seems participants sometimes continually 'reprogrammed' their memory of target locations.

5.3. Theoretical expressivity

To estimate the theoretical eyes-free expressivity of the interfaces we further examined the accuracy data from the 'blank' condition. Due to the gradual 'drift' from target centres reported above, the distribution of selections around the targets is not normally distributed on any axis (D'Agostino–Pearson normality test, $K^2 > 100$, $p < 0.01$ in all cases). Given the absence of normality, we calculated 99th, 95th, 90th, and 85th percentile accuracy values for each axis with each interface, and used these to calculate the theoretical number of discrete items selectable without visual feedback within the 650×650 ($\times 325$, for 3D) pixel space used in the experiment.

Fig. 6 shows the expressivity values, which suggest that 12 items can be selected with 99% accuracy with the 2D and 3D interfaces, but only 6 with raycasting. At 95% accuracy, the 2D, 3D, and raycasting values increase to 24, 50, and 25, respectively. The large volume supported by 3D provides good theoretical expressivity even though it had the least accurate selections in terms of miss distance. The predicted values for 2D are supported by Li et al. (2009), who found that 'virtual shelves' users could acquire 28 items with good accuracy.

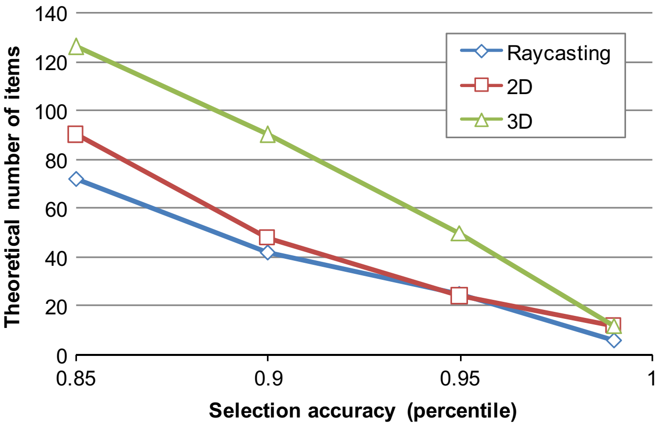

Fig. 6. Prediction of the total number of items that users can acquire eyes-free with the three interfaces based on percentile distance errors along each axis.

This analysis *only* considers mechanical issues of acquisition. Further work should consider issues of whether humans can effectively learn this number of items, the impact of the constrained movement range used in the study, and tolerance to long term drift effects.

5.4. Cognitive effort, physical comfort, and preferences

Participants' subjective workload assessment and rankings (presented in Table 1) amplified the quantitative results, clearly showing that participants found the 3D interface slow, inaccurate, effortful, and mentally

Table 1
Mean (and standard deviation) subjective workload assessments and ranking for each interface. Significance via Friedman χ^2test for ranks, df = 2.

NASA-TLX (1–5)	Raycasting	2D	3D	*p*
Mental load Low–high	2.33 (1.1)	2.20 (0.8)	3.86 (0.9)	0.0002
Physical load Low–high	2.47 (1.0)	2.73 (1.1)	3.73 (1.0)	0.005
Temporal load Low–high	2.0 (0.8)	1.8 (0.9)	2.27 (0.9)	0.26
Performance Poor–excellent	3.4 (0.7)	3.73 (0.7)	2.1 (0.8)	0.0001
Effort Low–high	2.87 (0.8)	2.93 (1.1)	4.0 (1.0)	0.0039
Frustration Low–high	2.47 (1.1)	2.1 (1.0)	3.27 (1.0)	0.003
Ranking (1–3) Best–worst	2.0 (0.8)	1.27 (0.5)	2.73 (0.5)	0.0003

demanding. Mean NASA-TLX worksheet responses for mental load and physical load with the 3D interface were 3.9 and 3.7, much higher than raycasting and 2D (all under 3.8). Eleven participants ranked 3D as their least preferred interface, with a mean ranking of 2.73. Comments included "3D was hard to use" and "hard to remember 3D locations".

Assessments of the raycasting and 2D interfaces were positive, with a general preference for 2D. The 2D interface was ranked top by eleven participants (mean ranking 1.27), while raycasting was favoured by the other four (mean 2.0). However, there were preference tradeoffs within 2D and raycasting. For example, participants stated that they liked the low effort of small raycasting movements, but found it hard to control ("too amplified an effect") and particularly difficult without feedback in the *blank* condition—"without feedback I felt most unsure of myself with this [raycasting] interface" and "I found it harder to remember locations in terms of pointing direction rather than actual positions". 2D also had tradeoffs with high workload ("very physically demanding for large distances"), but eased control ("I felt much more in control because you get to move your hand/arm a bit more"), and improved motor memory ("easier to associate a movement to a target and consequently easier to remember").

6. Discussion

The results provide insights into the comparative performance of three air pointing techniques that represent different design approaches to spatial targeting interfaces. Prior research in HCI, psychology, and neurophysiology (reviewed earlier) has already demonstrated that people can make spatial selections in these environments, but it has not given insights into their comparative merits—especially when there is a lack of visual feedback.

The 2D plane interface generally outperformed the others. Participants were able to make rapid (less than 1 s) and relatively accurate selections, even in the absence of any guiding feedback. Raycasting allowed similar acquisition speeds, but with much less accuracy; however, it was rated as the least physically demanding and would be a good pointing solution for rapid coarse selections where there are few candidate targets. All of the interfaces suffered from drift, with selections progressively moving further from targets when no feedback was provided. The following sections will discuss how these results fit into our design framework and describe some experimental limitations and opportunities for developing these interfaces.

6.1. Fitting the results into the framework

The study provides initial results that allow us to begin filling in the details of the design framework. The results indicate several relationships in terms of the concepts identified earlier (we note that these are speculations rather than confirmed relationships, and require further study).

Reference frame: Our interfaces all used a hybrid of device and body referenced locations, which required users to specify a movement origin. Any error in reacquiring the origin caused a corresponding movement in the target's spatial location. The effect of this error depends on how participants remembered target locations: either as relative proprioceptive actions, or as absolute spatial positions.

Participants were assisted in finding their previous origin by displaying it on the screen, and Fig. 5 makes us believe that this error was a minor issue with 2D and raycasting (because participants accurately acquired visual targets during training). However, origin drift may have influenced 3D more because participants were poor at controlling depth, even with continuous visual feedback (Fig. 5).

Input degrees of freedom: The difference in quantitative and qualitative assessment of the 2D and 3D interfaces suggests that 3D is more difficult to learn, slower, and less accurate. However, this finding may have been influenced by the difficulty our participants had in depth targeting, which may be eased by improved feedback (below).

Input scale: The raycasting and 2D interfaces primarily differed in input scale. Both allowed rapid selections, but the 2D plane was much more accurate. Participants also stated that large scale 2D movements helped them remember locations, but that small movements were (naturally) less physically demanding. Small scale movements are probably preferable when the number of targets is low, allowing coarse grained selections.

Feedback: Participants had particular difficulty acquiring targets on the 3D depth plane. As there was no stereoscopy, participants could only differentiate depths through the size of the target/cursor; although pilot testing indicated that this feedback was better than a separate display for depth, the visual feedback of matching cursor and target sizes may have been insufficient. In particular, the resolution of the feedback was fairly coarse (a 1 pixel change in cursor diameter corresponded to ~0.7 cm of physical movement along the *z*-axis, compared to ~0.14 cm along *x* or *y*).

An analysis of accuracy during the training condition across the 3D axes shows that the mean miss distance on the *z*-axis (17.9 pixels) was much higher than the *x* and *y* axes (3.7 and 3.2, respectively), suggesting that participants had problems in visual targeting in depth. However, prior experiments using volumetric displays (which do not suffer the parallax, stereoscopy, and motion parallax limitations of our study) have also shown that movement on the *z*-axis is slower and more error prone than *x* and *y* axes (Grossman and Balakrishnan, 2004). In the eyes-free 'drift' condition the difference across axes was small (*z*, *x*, and *y* means of 51.5, 48.9, and 52.9 pixels, respectively).

6.2. Exploring the framework for future research

Our experiment presents a preliminary investigation into these types of interaction techniques; our design framework allows us to understand the results we have obtained, and also the dimensions against which our techniques can be further developed.

Reference frame: The nature of the memory people form about proprioceptively defined targets and the decay of this memory (resulting in drift) require further investigation. A better understanding of proprioceptive memory development can inform the development of better training methods and how reference frames influence recall ability.

One method for alleviating the effect of drift would be to require a static reference frame, in which movements are relative to the world rather than body or device; however, while static reference frames are potentially useful in situations where users are either physically restrained or where the environment is stable (and external landmarks can be used to guide recall), they cannot be used while mobile.

Input degrees of freedom: While the 3DOF input utilised in the 3D interface presents a higher theoretical expressivity, the difficulties participants had in manipulating the depth component suggests that there are differences between the types of movement afforded by different degrees of freedom. Even within 2DOF movement, the translation motions of the 2D interface enabled higher accuracy than the rotational motions of the raycasting technique.

The interaction between degrees of freedom of both input device and human biomechanics requires further investigation. For example, our raycasting technique limited targets to those achievable with wrist motion; however, allowing fully developed pointing actions utilising the elbow and shoulder may allow users to leverage a greater degree of expressivity or introduce a cumulative error in precision. Similarly, while 3DOF interaction was reported to be more effortful than 2DOF, interaction styles that encourage the internalisation of 3D points as motion vectors or pointer steering gestures (rather than three separate translations) may help reduce the cognitive effort of performing 3DOF interactions.

Input scale: Our experiment mapped items onto Cartesian planes (raycasting and 2D) and rectangular cuboid volumes (3D). While this matched the visual feedback given to participants, it does not match their natural range of motion. This may have produced interference between the direction or position of targets and participants' natural intuition of locations (Poupyrev et al., 1997). We intend to investigate other co-ordinate systems (such as spherical or cylindrical polar) in future research.

Individual differences in pointing capabilities may also impact performance. Our experimental interfaces used fixed physical dimensions based on average anthropomorphic measures (to avoid differences in space resolution that would occur if we calibrated the interfaces per-participant), but it would be desirable for air pointing applications to allow calibration for differently sized users. The relationship between individual differences (e.g., in reach and range of movement) and accuracy and performance is an important topic for future research.

Feedback: Although the goal of air pointing interfaces is to eliminate the reliance on external feedback, it is still a critical component during the learning stages of expertise development, and may be periodically required to maintain expert performance. Our experiment only examined feedback during training stages; however, feedback during selection (such as a 'click' sound on target edges) may help combat drift issues and provide reinforcement through non-attentive feedback. Haptic and audio feedback modalities (for training and continued maintenance) are possibilities for further research.

With visual feedback, there are interesting challenges in addressing the inconsistency between proprioceptive locations and their visual display. Participants were instructed to try to learn the physical location of the wand they were holding, but one participant reported that during the *blank* condition he conceptualised targets on the screen rather than in motor space around him. Although the same disparity existed in all conditions, there is a risk that it may have differently affected the interfaces. This raises interesting design challenges in promoting learning of physical rather than visual spatial locations. We intend to investigate spatial audio as a feedback method for reinforcing spatial learning.

Other issues: The majority of participants in our experiment were young male postgraduate students. Using a homogeneous participant pool gains experimental control, but weakens results generality. A broader participant pool should be examined in further work. Finally, our studies concern target acquisition where selections are made with a single movement to a point in space, and further research is necessary for continuous eyes-free interactions such as steering and gesturing.

7. Conclusions

New forms of sensing in handheld devices are making it possible to develop eyes-free 'air pointing' techniques, in which users move their hand or a handheld device to a specific spatial region. In this paper, we explored the design space for air pointing, and introduced a framework that sets out input dimensions and resulting qualities for this new type of interaction. The framework provides a set of underlying concepts that aid in thinking about the air pointing domain, characterizing and comparing existing solutions, and evaluating novel techniques. We investigated some of the initial questions raised by the framework in an empirical comparison of three air pointing interfaces. Results showed that the 2D-plane technique performed best; raycasting is rapid but inaccurate, and 3D volume is slow, inaccurate, and requires more effort. The study provides initial insights into the relationships that exist between the two sides of the framework.

Our future work will continue in three directions. First, we will refine the framework, considering new input dimensions and interaction qualities: for example, the study suggests *feedback-free stability* as a possible quality (where a stable technique would not drift over repeated trials). Second, we will carry out further studies to test other factors and relationships in the framework. In particular, we will test the effects of using different reference frames and different feedback techniques on accuracy and learnability. Third, we will use the results of our studies to improve our existing air pointing techniques, and develop new ones; these will be tested in laboratory and realistic settings to further explore and understand the design space.

References

Andrade, J., Meudell, P., 1993. Short report: is spatial information encoded automatically? Quarterly Journal of Experimental Psychology 46A, 365–375.

Ängeslevä, J., Oakley, I., Hughes, S., O'Modhrain, S., 2003. Body mnemonics: portable device interaction design concept. In: Adjunct Proceedings of the UIST 2003: ACM Symposium on User Interface Software and Technology, Vancouver, CanadaACM Press.

Baudel, T., Beaudouin-Lafon, M., 1993. Charade: remote control of objects using free-hand gestures. Communications of the ACM 36 (7), 28–35.

Bolt, R.A., 1980. Put-that-there: voice and gesture at the graphics interface. In: Proceedings of the Seventh Annual Conference on Computer Graphics and Interactive Techniques, Seattle, Washington, United States, pp. 262–270.

Bowman, D.A., Johnson, D.B., Hodges, L.F., 1999. Testbed evaluation of virtual environment interaction techniques. In: Virtual Reality Software and Technology, ACM Press, London, pp. 26–33.

Bowman, D.A., Kruijff, E., LaViola Jr., J.J., Poupyrev, I., 2005. 3D User Interfaces: Theory and Practice. Addison-Wesley, Boston, MA.

Brewster, S., Brown, L.M., Tactons: structured tactile messages for non-visual information display. In: Proceedings of the Fifth Conference on Australasian user Interface, vol. 28, Dunedin, New Zealand, 2004. Australian Computer Society, Inc., Conference Paper, 976313. pp. 15–23.

Brewster, S., Lumsden, J., Bell, M., Hall, M., Tasker, S., 2003. Multimodal 'eyes-free' interaction techniques for wearable devices. In: Proceedings of the CHI'03: Conference on Human Factors in Computing Systems, Ft. Lauderdale, Florida, ACM, pp. 473–480.

Brewster, S., Wright, P., Edwards, A., 1993. An evaluation of earcons for use in auditory human–computer interfaces. In: Proceedings of the INTERCHI'93 Conference on Human Factors in Computing Systems, Amsterdam, pp. 222–227.

Cao, X., Li, J., Balakrishnan, R., 2008. Peephole pointing: modeling acquisition of dynamically revealed targets. In: Proceedings of the CHI'08: ACM Conference on Human Factors in Computing Systems, Florence, Italy, ACM Press, pp. 1699–1708.

Cockburn, A., Gutwin, C., Alexander, J., 2006. Faster document navigation with space-filling thumbnails. In: Proceedings of the CHI'06 ACM Conference on Human Factors in Computing Systems, Montreal, Canada, ACM Press, pp. 1–10.

Cockburn, A., Gutwin, C., Greenberg, S., 2007. A predictive model of menu performance. In: Proceedings of the CHI'07: ACM Conference on Human Factors in Computing Systems, San Jose, CA, ACM Press, pp. 627–636.

Crossman, E., Goodeve, P., 1963/1983. Feedback control of hand-movement and Fitts' Law. Paper presented at the meeting of the Experimental Psychological Society, Oxford, July, 1963. Reprinted in Quarterly Journal of Experimental Psychology 35A, 251–278.

Czerwinski, M., van Dantzich, M., Robertson, G., Hoffman, H., 1999. The contribution of thumbnail image, mouse-over text and spatial location memory to web page retrieval in 3D. In: Proceedings of the Interact '99, Edinburgh, Scotland, IOS press, pp. 163–170.

Easton, R.D., Sholl, M.J., 1995. Object-array structure, frames of reference, and retrieval of spatial knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition 21 (2), 483–500.

Egan, D., Gomez, M., 1985. Assaying, isolating, and accommodating individual differences in learning a complex skill. In: Dillon, R. (Ed.), Individual Differences in Cognition, pp. 173–217.

Elliot, D., Lee, T., 1995. The role of target information on manual-aiming bias. Psychological Research 58 (1), 2–9.

Elliot, D., Madalena, J., 1987. The influence of premovement visual information on manual aiming. Quarterly Journal of Experimental Psychology 39A (3), 541–559.

Gagnon, D., 1985. Videogames and spatial skills: an exploratory study. Educational Communication and Technology 33 (4), 263–275.

Grossman, T., Balakrishnan, R., 2004. Pointing at trivariate targets in 3D environments. In: Proceedings of the CHI'04: ACM Conference on Human Factors in Computing Systems, pp. 447–454.

Grossman, T., Balakrishnan, R., 2006. The design and evaluation of selection techniques for 3D volumetric displays. In: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, Montreux, Switzerland, ACM Press, pp. 3–12.

Harbluk, J., Noy, Y., 2002. The Impact of Cognitive Distraction on Driver Visual Behaviour and Vehicle Control. Transport Canada. Ergonomics Division of the Road Safety Directorate and Motor Vehicle Regulation Directorate.

Hart, S., Staveland, L., 1988. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P., Meshkati, N. (Eds.), Human Mental Workload, pp. 139–183.

Hinckley, K., Pausch, R., Goble, J., Kassell, N., 1994. A survey of design issues in spatial input. In: Proceedings of the UIST'94 Symposium on User Interface Software and Technology, California, pp. 213–222.

Hocherman, S., 1993. Proprioceptive guidance and motor planning of reaching movements to unseen targets. Experimental Brain Research 95, 349–358.

Krueger, M., 1991. VideoPlace and the interface of the future. In: B., Laurel (Ed.), The Art of Human Computer Interface. Addison Wesley, pp. 417–422.

Kurtenbach, G., Buxton, W., 1994. User learning and performance with marking menus. In: Proceedings of the CHI'94 Conference on Human Factors in Computing Systems, Boston, April 24–28, pp. 258–264.

Kurtenbach, G., Sellen, A., Buxton, W., 1993. An empirical evaluation of some articulatory and cognitive aspects of marking menus. Human–Computer Interaction 8 (1), 1–23.

Lee, J., Strayer, D., 2004. Preface to the special section on driver distraction. Human Factors 46 (4), 583–586.

Li, F., Dearman, D., Troung, K., 2009. Virtual shelves: interactions with orientation aware devices. In: Proceedings of the UIST 2009: The 22nd

Annual ACM Symposium on User Interface Software and Technology, Victoria, Canada, ACM Press.

Li, K., Baudisch, P., Hinckley, K., 2008. BlindSight: eyes-free access to mobile phones. In: Proceedings of the CHI'08: ACM Conference on Human Factors in Computing Systems, Florence, Italy, ACM Press, pp.1389–1398.

Medendorp, W., Van Asselt, S., Gielen, C., 1999. Pointing to remembered visual targets after active one-step self-displacements within reaching space. Experimental Brain Research 125, 50–61.

Mine, M.R., Brooks Jr., F.P., Sequin, C.H., 1997. Moving objects in space: exploiting proprioception in virtual-environment interaction. In: Proceedings of the International Conference on Computer Graphics and Interactive Techniques, pp. 19–26.

Myers, B., Bhatnagar, R., Nichols, J., Peck, C., Kong, D., Miller, R., Long, C., 2002. Interacting at a distance: measuring the performance of laser pointers and other devices. In: Proceedings of the CHI'02: ACM Conference on Human Factors in Computing Systems, Minneapolis, Minnesota, ACM Press, pp. 33–40.

NASA, 1995. Man-Systems Integration Standards. Revision B. NASA.

Naveh-Benjamin, M., 1987. Coding of spatial location information: an automatic process? Journal of Experimental Psychology: Learning, Memory, and Cognition 13 (4), 595–605.

Newell, A., Rosenbloom, P.S., 1981. Mechanisms of skill acquisition and the law of practice. In: Anderson, J. (Ed.), Cognitive Skills and their Acquisition. Erlbaum, Hillsdale, NJ, pp. 1–55.

Oakley, I., Park, J., 2007. Designing eyes-free interaction. In: Proceedings of the Second International Workshop on Haptics and Audio Interaction Design, HAID 2007, Seoul, South Korea, Springer Berlin, Lecture Notes in Computer Science, pp. 121–132.

Pierce, J.S., Forsberg, A.S., Conway, M.J., Hong, S., Zeleznik, R.C., Mine, M.R., 1997. Image plane interaction techniques in 3D immersive environments. In: Proceedings of the 1997 Symposium on Interactive 3D graphics, Providence, Rhode Island, United States, ACM, pp. 39–43.

Postma, A., De Haan, E., 1996. What was where? Memory for object locations. Quarterly Journal of Experimental Psychology 49A (1), 178–199.

Poupyrev, I., Weghorst, S., Billinghurst, M., Ichikawa, T., 1997. A framework and testbed for studying manipulation techniques for immersive VR. Virtual Reality Software and Technology, 21–28.

Poupyrev, I., Weghorst, S., Billinghurst, M., Ichikawa, T., 1998. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. Computer Graphics Forum 17 (3), 41–52.

Prablanc, C., Echallier, J.F., Komilis, E., Jeannerod, M., 1979. Optimal response of eye and hand motor systems in pointing at a visual target. Biological Cybernetics 35, 113–124.

Sengupta, A., Das, B., 2000. Maximum reach envelope for the seated and standing male and female for industrial workstation design. Ergonomics 43 (9), 1390–1404.

Sherrington, C., 1907. On the proprio-ceptive system, especially in its reflex aspect. Brain 29 (4), 483–486.

Soechting, J., Flanders, M., 1989. Sensorimotor representations for pointing to targets in three-dimensional space. Journal of Neurophysiology 52, 582–594.

Steed, A., 2006. Towards a general model for selection in virtual environments. In: Proceedings of the IEEE Symposium on 3D User Interfaces, Alexandria, Virginia, USA, pp. 103–110.

Thomson, J., 1983. Is continuous visual monitoring necessary in visually guided locomotion? Journal of Experimental Psychology: Human Perception and Performance 9, 427–443.

Vicente, K., Hayes, B., Williges, R., 1987. Assaying and isolating individual differences in searching a hierarchical file system. Human Factors 29 (3), 349–359.

Vogel, D., Balakrishnan, R., 2005. Distant freehand pointing and clicking on very large, high resolution displays. In: Proceedings of the UIST'05: ACM Symposium on User Interface Software and Technology, Seattle, WA, ACM Press, pp. 33–42.

Wilson, A., Shafer, S., 2003. XWand: UI for intelligent spaces. In: Proceedings of the CHI'2003 Conference on Human Factors in Computing Systems, Fort Lauderdale, Florida, pp. 545–552.

Wingrave, C., Tintner, R., Walker, B.N., Bowman, D.A., Hodges, L.F., 2005. Exploring Individual Differences in Raybased Selection: Strategies and Traits. IEEE Virtual Reality, Bonn, Germany (pp. 163–170).

Wingrave, C., Williamson, B., Varcholik, P.D., Rose, J., Miller, A., Charbonneau, E., Bott, J., LaViola Jr., J.J., 2010. The wiimote and beyond: spatially convenient devices for 3D user interfaces. IEEE Computer Graphics and Applications 30 (2), 71–85.

Wobbrock, J., Cutrell, E., Harada, S., MacKenzie, I., 2008. An error model for pointing based on Fitts' law. In: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, ACM, pp.1613–1622.

Wobbrock, J., Myers, B., Kembel, J., 2003. EdgeWrite: a stylus-based text entry method designed for high accuracy and stability of motion. In: Proceedings of the 16th Annual ACM symposium on User Interface Software and Technology, Vancouver, Canada, ACM Press, pp. 61–70.

Woodworth, R., 1899. The accuracy of voluntary movement. Psychological Review (Monograph Supplement) 3, 1–119.

Yee, K., 2003. Peephole displays: pen interaction on spatially aware handheld computers. In: Proceedings of the CHI'03: ACM Conference on Human Factors in Computing Systems, Fort Lauderdale, Florida, ACM Press, pp. 1–8.

Zhai, S., 1998. User performance in relation to 3D input device design. Computer Graphics 32 (4), 50–54.

Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., Baudisch, P., 2007. earPod: eyes-free menu selection using touch input and reactive audio feedback. In: Proceedings of the CHI 07: ACM Conference on Human Factors in Computing Systems, San Jose, California, USA, ACM, pp. 1395–1404.