ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Int. J. Human-Computer Studies ▮ (▮▮▮▮) ▮▮▮–▮▮▮

International Journal of Human-Computer Studies

www.elsevier.com/locate/ijhcs

Motion marking menus: An eyes-free approach to motion input for handheld devices

Ian Oakley[a,*], Junseok Park[b]

[a]Department of Mathematics and Engineering, University of Madeira, Campus Universitário da Penteada, Funchal, Portugal
[b]Smart Interface Research Team, Electronics and Telecommunications Research Institute, 161 Gajeong-Dong, Yuseong Gu, Daejeon, Republic of Korea

Received 15 August 2007; received in revised form 3 February 2009; accepted 4 February 2009
Communicated by P. Pendharkar

Abstract

The increasing complexity of applications on handheld devices requires the development of rich new interaction methods specifically designed for resource-limited mobile use contexts. One appealingly convenient approach to this problem is to use device motions as input, a paradigm in which the currently dominant interaction metaphors are gesture recognition and visually mediated scrolling. However, neither is ideal. The former suffers from fundamental problems in the learning and communication of gestural patterns, while the latter requires continual visual monitoring of the mobile device, a task that is undesirable in many mobile contexts and also inherently in conflict with the act of moving a device to control it. This paper proposes an alternate approach: a gestural menu technique inspired by marking menus and designed specifically for the characteristics of motion input. It uses rotations between targets occupying large portions of angular space and emphasizes kinesthetic, eyes-free interaction. Three evaluations are presented, two featuring an abstract user interface (UI) and focusing on how user performance changes when the basic system parameters of number, size and depth of targets are manipulated. These studies show that a version of the menu system containing 19 commands yields optimal performance, compares well against data from the previous literature and can be used effectively eyes free (without graphical feedback). The final study uses a full graphical UI and untrained users to demonstrate that the system can be rapidly learnt. Together, these three studies rigorously validate the system design and suggest promising new directions for handheld motion-based UIs.

Keywords: Motion input; Gesture; Eyes free; Mobile interface; Evaluation

1. Introduction

Handheld computers, in their most successful instantiation as mobile phones, are a widely available computational platform providing a broad range of advanced features and services to users all over the world. They are used not only to make calls but also to send and receive email, browse the Internet, perform bank transactions, act as navigation aids and capture, manage, organize and display digital media. As their computational and communicative abilities have grown, there has been a similar expansion in the sophistication of the applications they support.

In contrast to this rapid technological development, there have been few major changes to the user interfaces (UIs) of handheld devices. In many systems, users press directional keys to navigate through menu hierarchies and observe the results of their actions on small graphical screens. This paradigm dates back to the earliest handsets and is now straining to deal with phones that feature hundreds of commands organized in dozens of menus (St Amant and Horton, 2007). In response to this mobile devices have begun to feature richer interfaces, often modeled closely on those of desktop computers. Such Smartphones typically sport touch-sensitive screens that occupy large portions of their front surfaces and a point-and-click interface style directly adopted from the windows systems dominant in desktop computers.

*Corresponding author. Tel.: +351 291 705 117; fax: +351 291 705 199.
E-mail addresses: ian@uma.pt (I. Oakley), parkjs@etri.re.kr (J. Park).

1071-5819/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2009.02.002

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

However, it is far from clear whether an interface style developed for desktop computing is well suited to mobile use scenarios (Pirhonen et al., 2002). It is important to design for particular use contexts, and there are substantial differences between mobile and desktop scenarios (Holtzblatt, 2005). Fundamentally, desktop computing tasks are performed in stationary, stable, spacious and well-lit environments using relatively large-scale input and output devices such as keyboards, mice, large screens and hi-fidelity speakers. On the other hand, mobile tasks can occur in almost any situation: in the office, relaxing at home, traveling or otherwise out and about in venues such as shops, restaurants, bars and meeting rooms (Tamminen et al., 2004). There are also severe practical limits to the size and form of the input and display technologies that can be feasibly deployed. Given this qualitative discrepancy, there is a growing interest in developing rich new interaction techniques explicitly for mobile use (e.g. Schwesig et al., 2004; Zhao et al., 2007). Just as windows systems were developed to fit the scenario of a single user at a desk, mobile interaction techniques should also be designed from the ground up to fit mobile contexts. The movement towards this approach can be observed not only in the research community, but also in the marketplace where it is instantiated most famously in the novel wheel interface of Apple's trendsetting iPod (Apple iPod, 2007).

One technique that has attracted considerable attention within this domain is that of sensing the physical motions applied to a mobile computer (e.g. Rekimoto, 1996; Harrison et al., 1998). In this paper, this modality is termed *motion input* and is defined as the act of moving a handheld device in free space in order to issue commands or specify parameters. Examples of this kind of motion include rotating a device, pointing with it or making complex gestural movements such as tracing out a circle or square. One key advantage of this input modality is that it has the potential to support input based on proprioceptive feedback, the innate awareness of bodily movement, position and posture. This has the potential to enable *eyes-free interaction*, a term this paper uses to signify interaction unsupported by graphical feedback (Brewster et al., 2003) and performed by expert users. The inclusion of this last clause is significant—it encompasses the use of graphical interfaces intended to support novices as they initially experience and learn a system, eventually allowing them to attain a level of expertise that allows eyes-free interaction. This paper argues that the support of users of all skill levels is a practical concern that has been largely overlooked in the literature on motion input and that needs to be tackled before approaches using this interaction modality are sufficiently mature for widespread commercial deployment.

This paper addresses this concern through the design and thorough evaluation of a novel approach to motion input based on rotation of a handheld device. The main goal of this work is to empirically demonstrate a motion input system with the following three properties. Firstly, it must be usable by novices when they encounter it for the first time. Secondly, it must support rapid, eyes-free use by experienced expert users. Thirdly, it must enable novices to seamlessly develop into experts through nothing more than extended use of the system. The contents of this paper can be summarized as follows. The subsequent section details related work and focuses on the practical importance of designing motion input systems to suit users of all expertise levels: ease of use for novices, speed of use for experts and a smooth path for the former to develop into the latter. This is a fully inclusive approach to system design that has made scant appearance in the motion input literature in key domains such as gesture recognition (e.g. Baudel and Beaudouin-Lafon, 1993) and menu navigation (e.g. Poupyrev et al., 2002).

The remainder of this paper is dedicated to the design, evaluation and discussion of a novel menu system explicitly addressing these concerns. The system is inspired by marking menus (Kurtenbach et al., 1993), an established technique that successfully embodies the combination of a full graphical interface intended for novices with input fundamentally designed to be used by experts eyes free. Fig. 1 provides an overview of the operation of the

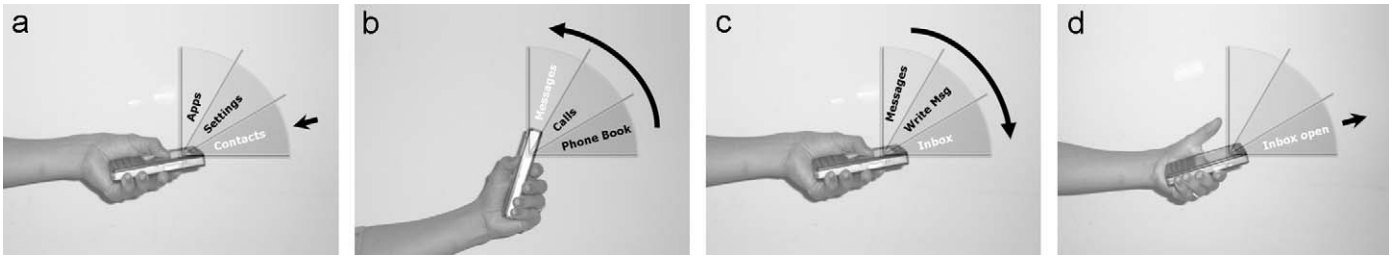

Fig. 1. Practical illustration and overview of the motion-controlled mobile device menu system introduced in this paper. In this example, 90° of rotational space (from holding a device vertical to holding it horizontal) is divided into three targets. These are shown overlaid on a series of photographic images. Commands are issued by making motions that start, optionally turn and then terminate in particular targets. The text depicts the commands within targets; white text indicates the currently selected item. In this scenario, the commands were chosen to represent common mobile device functionality. Black arrows show user actions. In (a), the main menu is displayed and the user presses against the device touch screen to select the Contacts menu (horizontal). In (b), the Contacts menu is shown and the user has rotated the device to select the Messages sub-menu (vertical). In (c), the Messages sub-menu is displayed and the user has returned the device to the horizontal to select the Inbox command. In (d) the touch screen is released to issue the Inbox command.

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

proposed technique in a scenario based on a typical mobile phone menu system. Commands are issued by rotating a handheld device between a series of targets spread between the postures of holding it horizontally and holding it vertically. Three studies of this system are presented. The first two studies use an abstract interface and focus on exploring the design space of the menu system and capturing basic performance data to contrast against that reported using other input techniques and user models. The final evaluation uses a more realistic graphical interface to the menu system and looks at learning rates. In this study, practice time is kept to a minimum and the question of whether novices can use the system from the outset is addressed. These studies form a comprehensive examination of the technique and, as a whole, are supportive of the further development of the ideas it embodies.

2. Related work

Motion input possesses a number of inherent advantages as a modality for controlling a handheld device. First among these is that physical motion is an expressive channel composed of six independent degrees of freedom. Furthermore, it is one that is conveniently available in mobile scenarios—typically users are already holding a device in their hands. Finally, it is technologically feasible as accelerometers, the sensors most commonly used to measure motions, are small, robust, cheap, low power and can be contained entirely within the casing of a device. Reflecting these facts, motion-sensing functionality has appeared in several mobile phones (such as the Samsung SCH-S310 (Cho et al., 2006)) and applied to tasks such as recognizing when a numeral has been drawn in the air in order to control speed dialing functions. However, it is also worth noting that these advanced implementations have not transitioned to widespread availability and indeed tended to remain as single model concept designs. For instance, at the time of writing, Samsung has not included the technology from the SCH-S310 in other phone models. The causes underlying the lack of success of such systems are explored in this review.

In general, the use of motion-sensing technology has been studied in a variety of domains, including context awareness (in which a device uses information about the motions that affect it to alter its behavior automatically; Hinckley et al., 2005), text entry (Partridge et al., 2002; Wigdor and Balakrishnan, 2003), gesture recognition and interaction with menu systems. These last two topics are of particular relevance to mobile device use as command entry is a fundamental operation and deep menu systems are the dominant interface with which it is achieved on current mobile phones (St Amant et al., 2004).

2.1. Gesture recognition

At first glance, gesture recognition seems an ideal mechanism by which free motions in space could be used as the interface to a command set. Indeed, this concept has attracted considerable research attention. An early example of such a system is CHARADE (Baudel and Beaudouin-Lafon, 1993), which used motions and hand postures captured by a data glove to control navigation through a HyperCard stack displaying a presentation. CHARADE was a desktop application and relied on simple, metaphorical gestures such as moving left to right to advance to the next slide and right to left to return to the previous one. The technical, algorithmic and implementation issues underlying the use of general gesture recognition through motion input on standalone handheld devices were discussed by Benbasat and Paradiso (2001). They demonstrated that this technology could be realistically deployed in mobile scenarios. However, despite the fact that this interaction concept is relatively mature, recent research on free motion gesture control continues to tackle essentially algorithmic issues (e.g. Schlömer et al., 2008). The remainder of this section explores the fundamental factors underlying why this is the case.

In a perfect gesture recognition system, users would make specific, natural, fluid motions, which would be captured and accurately translated into commands. However, systems that have attempted to achieve such a vision have suffered from a number of practical constraints. These can be summarized by the fact that gesture recognition is a challenging problem with no simple algorithmic solution, and that the difficulties this presents are compounded by the fact that motion is a relatively noisy input channel. In a typical system, users require some explicit training to learn the gestures (both to remember what they are and to physically practice their execution; Long et al., 1999), and the system requires additional time to adapt each user's individual performance (Mantyjarvi et al., 2004). Furthermore, while expert users can perform well, there is a steep learning curve and little support for novice users as they attempt to acquire the required skills. Training sessions can be conducted and false positives and negatives reported, but such binary feedback is minimally informative and offers little potential for rapid improvement. While development of the algorithms underlying such systems is proceeding steadily (e.g. Kela et al., 2006), the difficulties experienced on initial user exposure represents a fundamental problem to which there is currently no generally applicable solution. How can a completely novice user know what commands are valid, and exactly how (and how not) they should be issued? Researchers are only just beginning to consider the semiotics of depicting gestural commands, or displaying interactive performance in a gestural task to facilitate learning. For example Hinckley et al. (2007) discuss this issue in the context of screen-based stylus gestures (concluding that novices require explicit visual instructions) and Kallio et al. (2006), motivated by the need to provide feedback to users, present initial work on the visual depiction of motion input gestures. The successful resolution of these issues will be key steps in popularizing

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

gestural interfaces as they hold the promise of providing the same "no manual" ease-of-use that has become the standard since the advent of rich graphical UIs.

2.2. Scrolling

A number of designs have also been proposed based on the general concept of scrolling (e.g. Bartlett, 2000), and in particular, navigating a list-like menu structure by adjusting the orientation of a device in order to select and issue commands. Poupyrev et al. (2002) describe and evaluate a system that involves mapping variations in orientation to the rate at which menu items are traversed; the steeper the incline, the faster the traversal. The authors conducted an empirical evaluation of their system, concluding that users were able to select menu items rapidly, but suffered from problems of overshooting targets. Indeed, recent work (Cho et al., 2007) looking at a broadly similar interaction technique reinforces this observation and suggests that using a set of dynamic equations to mediate scrolling speed may alleviate this problem. Oakley and O'Modhrain (2005) proposed an alternative menu system based on mapping changes in orientation directly to menu position. This system presented a fixed number of menu items (between 6 and 15), each of which was allocated a small, identically sized segment of rotational space. Users selected a menu item by rotating the handheld device to the appropriate orientation. They empirically compared their system with Poupyrev et al.'s and concluded that their technique offered significant improvements in both task completion time and error rate, suggesting that users may find this simpler mapping easier to use.

2.3. Eyes-free interaction

Menu systems controlled by motion input are generally reliant on continuous graphical feedback. This is problematic in the case of mobile devices as the act of moving the device fundamentally conflicts with the act of viewing its screen. In the menu systems described in the previous section, users must rotate the device, sometimes by as much as 90°, while monitoring its display in order to make an accurate selection. This is a physically challenging and potentially annoying task. Furthermore, this style of interaction is a poor fit to many mobile scenarios in which users may be busy, distracted or simply unwilling to direct their visual attention towards their handheld device. A number of authors have suggested that eyes-free interfaces that can be operated in the absence of graphical feedback are especially well suited for mobile interaction (Pirhonen et al., 2002; Brewster et al., 2003; Zhao et al., 2007). The physical buttons featured on the exterior of most handheld devices represent an enduring example of this. They can be operated by feel alone and many are tied to commonplace tasks such as answering or refusing calls, engaging or disengaging communication functionality, adjusting volume and halting alarms. Users require that these kinds of commands be initiated easily, rapidly and in potentially distracting or demanding situations, and an important aspect of this is that they do not demand a user's visual attention. Indeed, the usefulness of eyes-free interfaces has been passively acknowledged in other motion-controlled interfaces. Perhaps the most prominent example of this is Wigdor and Balakrishnan (2003), a motion interface supporting mobile text entry, which achieves high levels of performance and focuses entirely on atomic, gestural, motions that can be performed without looking at the handheld device's screen. In this system, text is entered into a phone by pressing one of its numerical keys and rotating the device forward, leaving it stationary or rotating it backwards to choose one of the three letters associated with that key.

2.4. Gestural menu systems

Extending the approach seen in TiltText has considerable promise for motion-sensing interfaces designed to issue specific commands. It may be able to avoid the weaknesses and combine the strengths of both current gestural systems and menu systems. The use of simple gestures based on conditional logic removes reliance on complex recognition technologies and more importantly opens the door to continuously displaying system state in an interactive interface. This should enable novice users to operate the system immediately. Equally, focusing on essentially gestural motions should allow expert users to eventually learn and issue commands autonomously without referring and to referencing the UI.

Gestural menus leveraging these concepts have appeared previously. They are most comprehensively incarnated as marking menus (Kurtenbach et al., 1993), a form of interface typically used in conjunction with styli and rarely seen in other contexts. Marking menus are stroke driven and users interact with them by making a rapid series of one or more lines (or marks) in a zig-zag pattern. Initially they are guided by interactive visual feedback in the form of graphical items displayed radially around the menu origin (in a similar manner to a pie menu). However, as users become more experienced, the movements required to access particular commands are internalized and these intermediate graphical representations of system state are no longer required. For example, in such a system a user might tap to activate the menu, move down to select a sub-menu (which then appears centered at the current cursor position) and then move towards the right to select an item in that sub-menu. At first, screen-based feedback enables selection of the appropriate items, but given sufficient repetition, this process becomes automated and the user can simply activate the menu and draw an L shape, mechanically moving down then right. This innate learnability, or support for the seamless transition from a supervised, novice mode to a gestural expert mode is a key conceptual feature of marking menus. It represents a significant advantage over the majority of other

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

gesture-based systems that rely on the user learning a corpus of commands explicitly and by rote.

A weakness of this approach is the limited number of commands that can be accessed, both in terms of the number of items that can be presented in each menu (typically either 4 or 8; Kurtenbach et al., 1993) and the number of menu hierarchies that can be supported (typically 2 or 4; Kurtenbach and Buxton, 1993). There have been numerous extensions to the making-menu concept in order to address these issues. The depth limitation has been significantly alleviated by multi-stroke marking menus (Zhao and Balakrishnan, 2004) while zone and polygon menus (Zhao et al., 2006) and flower menus (Bailly et al., 2008) have significantly improved the number of items that can be presented in each menu level. The motion-based menu system presented in this paper is based on, and shares many of the same objectives as, the body of work exploring stylus-based marking menus.

3. System design

The system design studied in this paper was based on rotations of a handheld device around a single axis in a 90° range starting with the device vertical and ending with it horizontal. This space was divided into a small number of identically sized targets. Fig. 2 depicts an example of this space featuring 3 targets. In the system, commands were issued by rotating the device into one of the targets, and pressing a button (or against a touch screen). Releasing the button immediately issued what we termed a *no-stroke* command. Rotating the device into another target (and therefore making a stroke or mark) before releasing the button resulted in a *one-stroke* command. Similarly, a *two-stroke* command could be produced by changing the direction of the rotation (at the level of targets) and executing a second mark before releasing the button. Fig. 3 illustrates these three command types in systems divided into 2–4 targets. Additional strokes or targets could be used to extend the system, at the cost of increasing the complexity of the required movements.

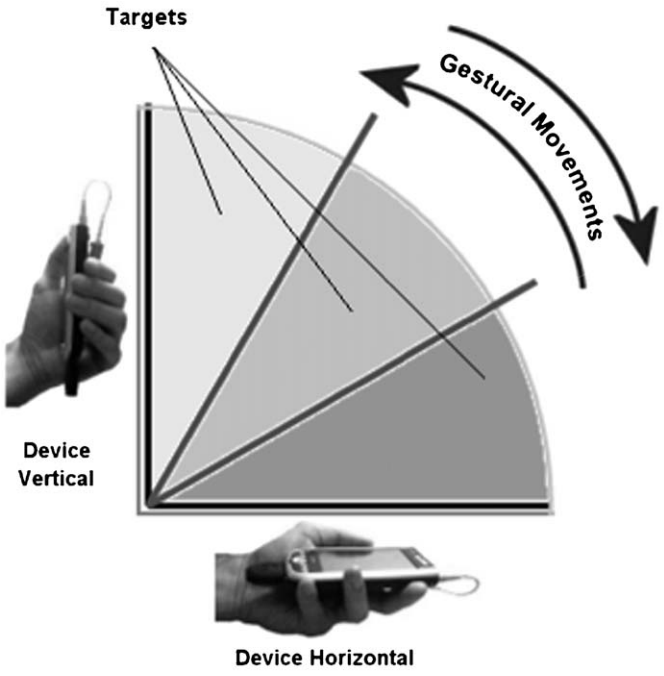

Fig. 2. Range of motion used in menu system, showing the example of dividing this space into three discrete targets (shown in grey shades).

As it relied on humans making rotational movements in free space, the system design encompassed two techniques to provide feedback and increase reliability. The first of these was the display of tactile cues on the transition between 2 targets, an approach that has been previously shown to be valuable (e.g. Poupyrev et al., 2002; Oakley and O'Modhrain, 2005). The cues took the form of a 100 ms sample composed of a 250 Hz sine wave (the frequency to which the skin is most sensitive; Verrillo, 1966) with a curved amplitude envelope. This resembles the feel of a small impact or brief click. The second involved the use of a dynamic resize to minimize unintentional transitions between targets. This was achieved by simply enlarging the currently occupied target by 2° in each direction. This ensures that on entering a target, to immediately exit it again requires a reversal of direction and a movement of 4°, a procedure of sufficient scale that it is unlikely to occur without a user's conscious intent.

The expressivity of this system, in terms of the number of commands it can issue, can be mathematically determined and is dependent on both the number of targets present and number of strokes used to select them. The equations below can be used to calculate the number of unique no-stroke, one-stroke and two-stroke commands that can be issued (where n is the number of targets):

$$f\text{-}no\text{-}stroke = n \tag{1}$$

$$f\text{-}one\text{-}stroke = n(n-1) \tag{2}$$

$$f\text{-}two\text{-}stroke = \sum_{i=2}^{n}(2(n-1)^2) \tag{3}$$

For example, a system with 3 targets has 3 no-stroke commands, 6 one-stroke commands and 10 two-stroke commands for a total of 19 possible commands. A 2-target system is considerably less expressive with a total of just 6 commands, while a 4-target system supports a total of 44 commands.

3.1. Design rationale

A number of design goals shaped this system. First and foremost was the requirement that it (like marking menus) feature a learnable gesture interface. Such a system requires a rich interface component that allows it to be used immediately by a novice. In order to support eyes-free operation by expert users, it must also be composed of fundamentally simple motions, which can be easily and seamlessly learnt. A secondary goal was domain based; the

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

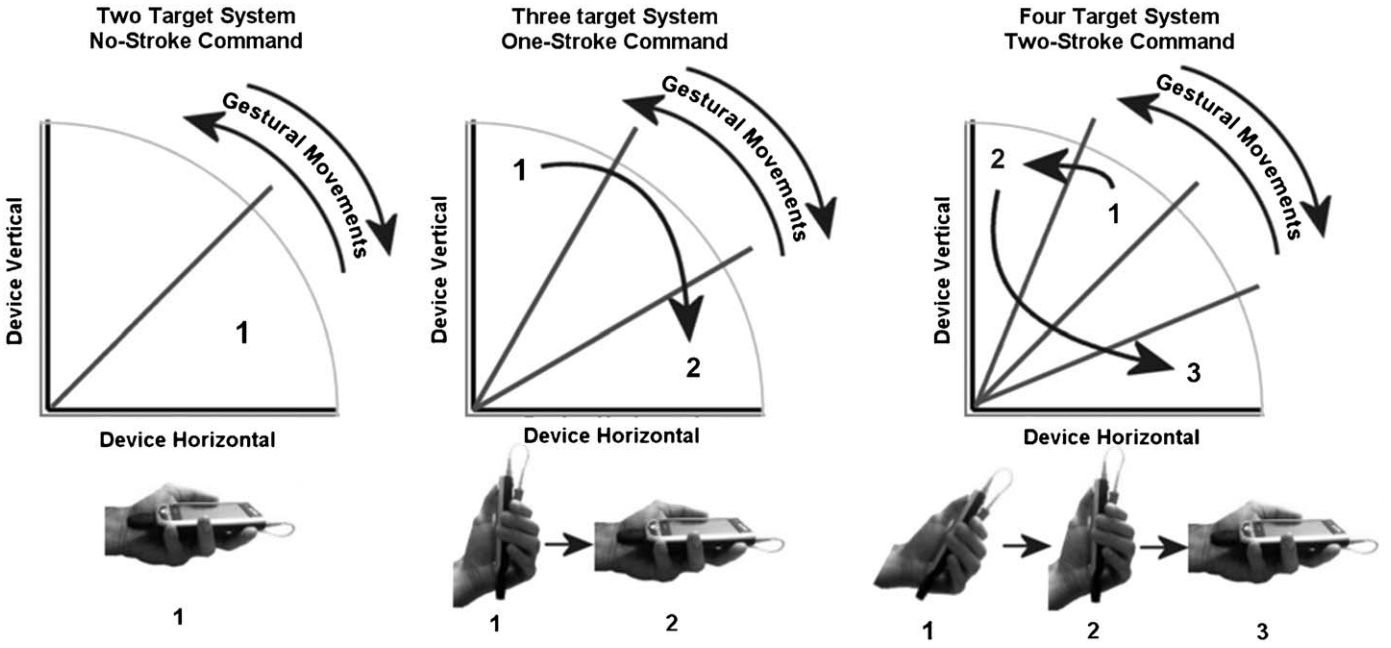

Fig. 3. Usage of the menu system. Left to right shows two, three and four target variations of the system overlaid with numbers and arrows indicating no, one- and two-stroke gestures, respectively. The gestures commence in the target labeled "1" and follow the path described by subsequent arrows and numbers. They finish on the final number. The device motions required to produce these gestures are shown at the base of the figure.

system was intended to address only a small set of commonly accessed functionalities. The objective was to provide access to applications under shortcut keys rather than support the navigation of an arbitrarily sized and customizable address book.

These goals informed the selection of orientation input (motions made by rotating the handheld device) over translation input (movements made by moving the handheld device along one or more spatial axis). Orientation input has been well studied by pervious authors (e.g. Poupyrev et al., 2002; Oakley and O'Modhrain, 2005) and has the advantage of simplicity; the motions it includes are generally of small scale, easily understood and un-taxing. Device orientation is also easy to measure through monitoring the vector of gravity. The decision to restrict input to rotations around a single axis further simplifies the input and reflects informal observations (e.g. Hinckley et al., 2000) that simultaneous control of device orientation in two axes is challenging. This may be due to the fact that accelerometer interfaces based on orientation input operate relative to the vector of gravity rather than a device-centric set of axes. Consequently, when the device is rotated away from gravity on one axis, it is cognitively complex to accurately adjust its orientation (relative to gravity) on a second.

The expressivity (in terms of the number of discrete commands that can be specified) of the intentionally simple motions used in the system was maintained through the division of the rotational space into separate targets in a manner inspired by marking menus. However, it should be noted that this design violates several of the key properties identified in Kurtenbach et al.'s (1993) original vision. For example, in the system described here, strokes that commence at different start points, but are otherwise identical, result in the activation of different commands. This kind of variation has been previously shown to offer benefits when used in stylus-driven marking menus (e.g. Zhao et al., 2006). However, perhaps more seriously, the system described in this paper also relies on the length of the strokes to activate different commands, thereby violating the principle of using scale-independent marks highlighted in Kurtenbach's original work (1993). Although necessary to increase system expressivity, this is a significant departure, which may affect long-term usability of the system, particularly in reference to the ballistic production of strokes by expert users.

3.2. System implementation

The system was implemented on a 624 MHz Dell Axim X51v under MS Windows Mobile Version 5. Motions were captured using the TiltCONTROL, a sub-$100 device built by PocketMotion (PocketMotion, 2007). It features a 2-axis accelerometer packaged with a microprocessor that provides an RS232 interface (some rewiring was required to make it compatible with the Dell X51's serial connector). A simple API allows it to be easily integrated into an application, and functions exist to provide orientation (filtered with a simple 8-sample rolling average algorithm). Vibrotactile cues were integrated into the system using a VBW32 skin transducer (Tactaid VBW32, 2007), a device commonly used by the tactile research community. The transducer was attached to the back of the PDA with a

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

Fig. 4. Final experimental device, showing motion sensor and vibrotactile transducer.

Velcro strip, and driven from the PDA's headphone jack. The final device is illustrated in Fig. 4.

4. Initial evaluation

Two fundamental studies were conducted to assess user performance with this system. The first featured a UI on the screen of the PDA, while the second's interface was displayed on a separate monitor. The two studies are subsequently referred to as the *local-UI* and *distant-UI* studies, respectively, and this experimental design was selected to allow explicit comparison of user performance with motion input supported by on-device and off-device graphical UIs. The local-UI study also varied the number of targets used in the system (featuring 2–4) and considered only no-stroke and one-stroke commands. The distant-UI study featured only a 3-target system, but used two different sizes of target (*original* and *expanded*) and also included two-stroke commands. In this way, each study explored different aspects of the design space. They were intended to provide a broad window onto user performance with this technique, allowing it to be meaningfully compared with that reported in literature. A final important aspect of the distant-UI study was a closing condition with no graphical feedback. This was termed the *blind* condition and tested the system's performance in an eyes-free situation.

4.1. Hypotheses

These two experiments were exploratory, and intended to capture the raw magnitude of user performance for comparison against that reported in literature. However a number of concrete hypotheses were formed. In the local-UI study the following predictions were made:

H1: performance will decrease as the number of targets increases;
H2: performance will decrease from no-stroke trials to one-stroke trials.

In the distant-UI study, it was predicted that:

H3: performance will decrease from no-stroke to one-stroke and two-stroke trials;
H4: performance will increase from the original to expanded targets;
H5: performance will decrease from the two conditions with visual feedback to the blind condition.

These experimental hypotheses use the term performance to encapsulate both task completion times and error rates (see Section 4.6 for further details).

4.2. Participants

The local-UI study featured 12 participants (6 male, 6 female, with a mean age of 32), while the distant-UI study featured 8 participants (4 male, 4 female, mean age 28). The majority were employees at our institute, the remainder were members of the general public sampled via the snowball method from this initial group. Each participant completed only one study and they were not compensated. All in all, 18 were right handed, 2 left handed and none reported any physical impairment in their dominant arms.

4.3. Experimental design

The local-UI study had three conditions containing 2, 3 or 4 targets. In each condition the trials were an exhaustive set of all possible no-stroke and one-stroke commands. The conditions were set to be of approximately the same length; the 2-target and 4-target conditions featured 80 trials (respectively, 20 and 5 times for each possible menu item) while the 3-target condition featured 81 (9 times for each possible menu item). All trials were delivered in a random order. The study was fully balanced, with 6 order conditions, and participants distributed equally among them.

The distant-UI study used 3 targets throughout but varied their size; original and expanded targets were used. The original targets were identical to those in the 3-target condition of the local-UI study (30° wide), while the targets in the expanded condition were larger, 45° wide (the same size as in the 2-target condition of the local-UI study). In both conditions, the menu system remained centered at the same 45° orientation, as illustrated in Fig. 5. Each condition featured 114 trials (each possible command, 6 times) delivered in a random order. Order effects between these two conditions were fully balanced. However, participants had to also complete a brief closing condition without graphical feedback representing system state. This blind condition was 38 trials in length (each possible command, twice) and targets were kept at the same size as those each participant experienced in the latter of the two main conditions. This tested if the system could be used eyes free.

In both studies, practice came in two forms. At the start of each session, participants completed a shortened version of the experiment with one trial for each possible command under the supervision of an experimenter. The goal of this

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

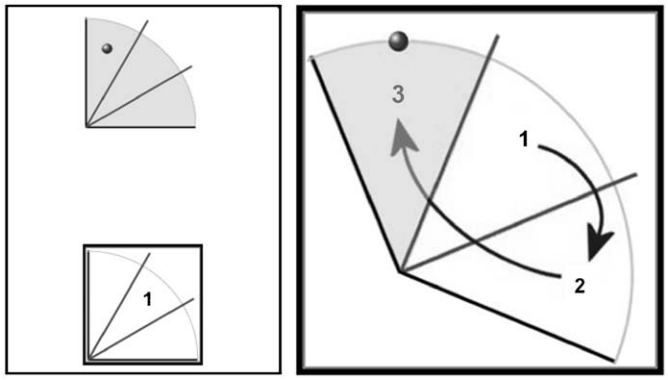

Fig. 5. Experimental screen shots from local-UI (left) and distant-UI (right) studies. The distant-UI shot shows the size of the expanded targets. Both screen shots include highlighting to show the current target (the ball icon and dark target in the distant-UI shot) and a series of numbers and arrows depicting the required gesture. The interfaces differed between these studies due to the specific display requirements of the two conditions, as discussed in Section 4.4 of the main text.

familiarization stage was to ensure that participants were fully comfortable with the experimental procedure. They were free to ask questions, clarify problems and encouraged to experiment with the interface rather than complete the trials. It typically took 5 min. Half-length practice sessions also took place immediately prior to each experimental condition.

4.4. Materials

Both studies featured similar graphical interfaces, customized for their target display. In the local-UI study, there were two small graphical visualizations, one at the center top of the PDA's screen indicating the currently selected target and one at the center bottom showing the sequence of targets required to complete the current trial. This split system was chosen to avoid overloading a single graphical visualization with too much information. Additional reasons were that the center top and bottom zones can be observed equally well by both right- and left-handed users and are unlikely areas to be obscured by thumb presses against the screen (which were used to mediate the gestures and tended to take place in the center of the display, see Section 4.5). In the distant-UI study these two glyphs were merged into one larger visual presentation, which was better suited to display on a relatively large remote screen. Fig. 5 shows screen shots of both of these interfaces. One important implication of using different interfaces in each study is that this may act as a confounding variable, compromising the reliability of inter-study comparisons. To alleviate this concern, this issue is explicitly addressed in the discussion of the results (Section 4.8).

All the visualizations featured a graph depicting the rotational space and targets currently being used. The horizontal and vertical axes corresponded to horizontal and vertical orientations, respectively, of the PDA. The instructions were presented as a numbered sequence of strokes required to complete the gesture. A "1" marked the initial target, an arrow and a "2" marked the path, second target (if required) and a final arrow and a "3" the last target (if required). The highlighting used to display the currently selected target varied between the studies. In the local-UI study, a circular highlight indicated the currently active target and changed color (from grey to red) when a gesture was in progress. In the distant-UI study the presence of this circular object again indicated that a gesture was underway, but the currently active target was also shown by darkening the current graph segment. This interactive feedback was deactivated in the blind condition.

4.5. Procedure

Both experiments took place in an unused office. The only variation between the two studies was the site of the graphical UI; in the local-UI study it was on the PDA itself and in the distant-UI study, it was on a laptop screen positioned in front of the user at approximately head height. All participants were instructed to stand naturally and to hold the PDA in their dominant hand for the duration of the study. Each trial began by informing the user that they could rest briefly, and to tap the PDA screen to proceed. When they did so, a fixation spot was displayed for 500 ms, followed by the experimental interface (as described in the materials section). Participants then had to rotate the PDA to the starting orientation indicated, initiate the gesture by pressing anywhere on the device's screen with their thumb and then (if required) rotate the PDA through the series of movements shown before releasing the screen, completing the gesture and ending the trial. These thumb presses tended to be in the central portion of the PDA screen. Explicit breaks between each condition were enforced and participants were kept informed about the details of the upcoming condition.

4.6. Measures

The primary measures were task completion time and error rate. Task completion time was broken down into two discrete stages, termed *planning time* and *execution time*. Planning time referred to the span between when a trial was first displayed and when a user initiated a gesture. It therefore included not only reaction time, but also the time to interpret the experimental instructions and to move to the initial target. Execution time referred to the period in which participants pressed against the screen, actually making a stroke. The sum of these measures is the total task completion time. Errors were defined as trials featuring an inaccurate initial or final target, or a change in direction at the level of targets. For instance, in a system divided into 3 targets a stroke that began in the horizontal target, and ended in the vertical target had to enter and exit the central target once, and only once, for it to be

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

considered valid. Multiple entries and exits to any target that deviated from the sequence indicated in the instructions led to potentially ambiguous input. The temporal data from trials in which participants made errors were included in the timing results, and these erroneous trials were not re-run.

4.7. Results

The task completion times from the two studies are presented in Figs. 6 and 7. They show the individual measures of planning and execution time and break the data down into that from no-stroke, one-stroke and two-stroke trials. The error data are presented in Figs. 8 and 9. Analyses on these data used two-dimensional ANOVA (experimental condition against number of strokes in each trial) followed by post-hoc *t*-tests incorporating Bonferroni confidence interval adjustments.

In the local-UI study, both planning and execution times increased with the number of targets ($F(2, 11)=10.7$, $p<0.001$ and $F(2, 11)=5.34$, $p<0.01$) and the number of strokes ($F(1, 11)=51.01$, $p<0.001$ and $F(1, 11)=256.1$, $p<0.001$). Execution time showed an interaction between these factors ($F(2, 1)=3.38$, $p<0.05$) while planning time did not ($F(2, 1)=0.85$, $p=0.43$). Error rates increased with the number of targets ($F(2, 11)=6.02$, $p<0.005$), but not strokes ($F(1, 11)=0.09$, $p=0.76$) and did not result in an interaction ($F(2, 1)=0.06$, $p=0.94$). Post-hoc *t*-tests revealed that the 2-target condition yielded improvements over the 3- and 4-target conditions in planning time and error rate and the 4-target condition in execution time (all at $p<0.05$ or lower).

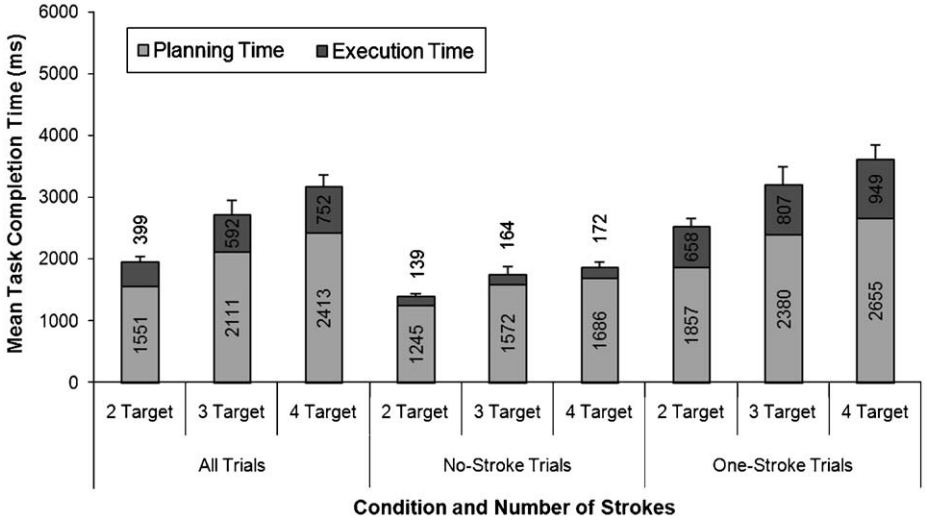

Fig. 6. Mean planning and execution times from the local-UI study, broken down to show data from no- and one-stroke trials.

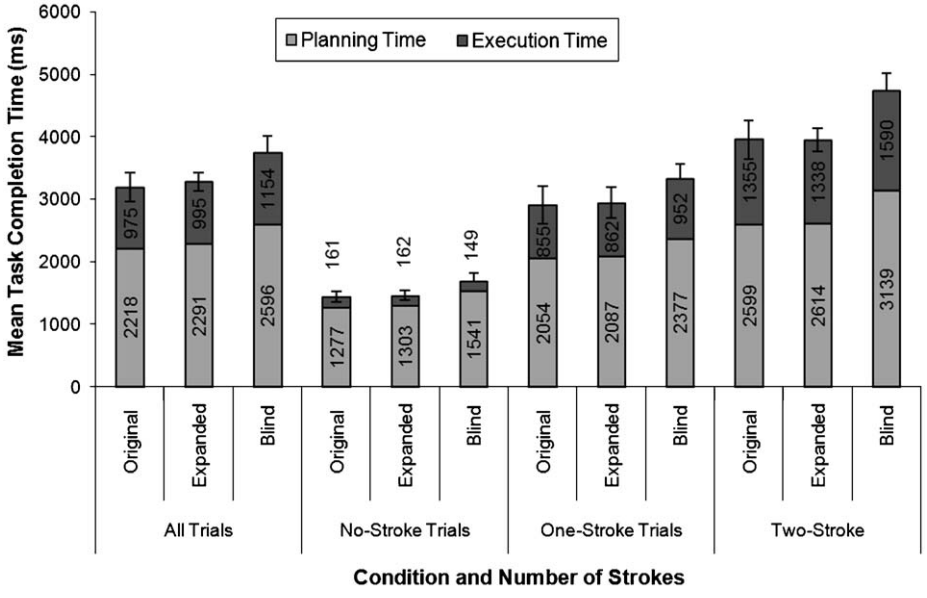

Fig. 7. Mean planning and execution times from the distant-UI study, broken down to show data from no-, one- and two-stroke trials.

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

The distant-UI study showed significant differences between the three conditions in planning time ($F(2, 7)=6.11$, $p<0.005$) but not execution time ($F(2, 7)=1.63$, $p=0.2$) or error rate ($F(2, 7)=1.07$, $p=0.35$). All three of these measures increased with the number of strokes in a command (respectively, $F(2, 7)=69.2$, $p<0.001$, $F(2, 7)=170.3$, $p<0.001$ and $F(2, 7)=8.73$, $p<0.001$), but no interactions were found ($F(2, 2)=0.34$, $p=0.84$, $F(2, 2)=0.78$, $p=0.54$ and $F(2, 2)=1.86$, $p=0.12$); t-tests on this data showed than planning times in the blind condition were significantly reduced compared to the other two conditions (both at $p<0.05$) and all pairwise comparisons relating to the number of strokes in a command attained significance (at $p<0.01$ or lower) except for that between the error rate in one- and two-stroke trials.

4.8. Discussion

4.8.1. General discussion

The local-UI study showed a clear decrease in performance with increased number of targets and both studies showed that additional strokes also have a strong impact on performance. This confirms the first three experimental hypotheses (H1, H2 and H3) and is unsurprising; as targets became more densely packed or motions became more complex, performance dropped. In the distant-UI study, the expanded targets did not offer a performance improvement over the original targets, causing this hypothesis (H4) to be rejected. Finally, the blind condition led to a modest reduction in performance (confirming H5). This discussion presents a more detailed analysis of these results before moving on to a thorough performance comparison with the data reported in literature on motion input and mobile interaction.

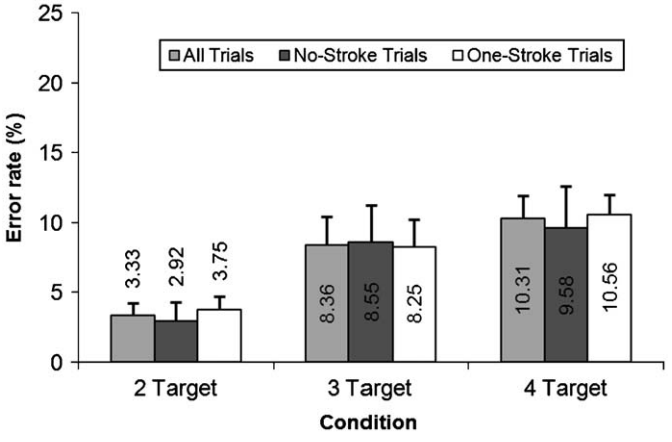

Fig. 8. Mean error rate from the local-UI study, broken down to show data from no- and one-stroke trials.

The experimental interface was composed of abstract graph-like visualizations and the results show that the mental process of interpreting these images was time consuming. The measure of planning time encompasses this activity (as well as the relatively static components of reaction time and a single targeting operation) and accounted for up to 75% of total trial time. Furthermore performance was measurably slower with more complex instructions; planning time increases with both the number of strokes and the number of available targets. The high time cost of interpreting the instructions is therefore an artifact of the stimuli used in this study and unlikely to transfer to an ecologically valid situation in which expert users issue commands according to their needs. The development of simpler instructions may also alleviate this time-consuming aspect of the study and more accurately reflect the time it takes users to interact with the system rather than absorb the instructions.

In contrast to the increase in planning time, the linear nature of the increments in execution time between no-, one- and two-stroke trials in the distant-UI study suggests that users found issuing a short sequence of strokes little more challenging than issuing a single one. This observation is reinforced by the error data. In the local-UI study, a greater number of targets results in a greater number of errors, but there is no difference between no-stroke and

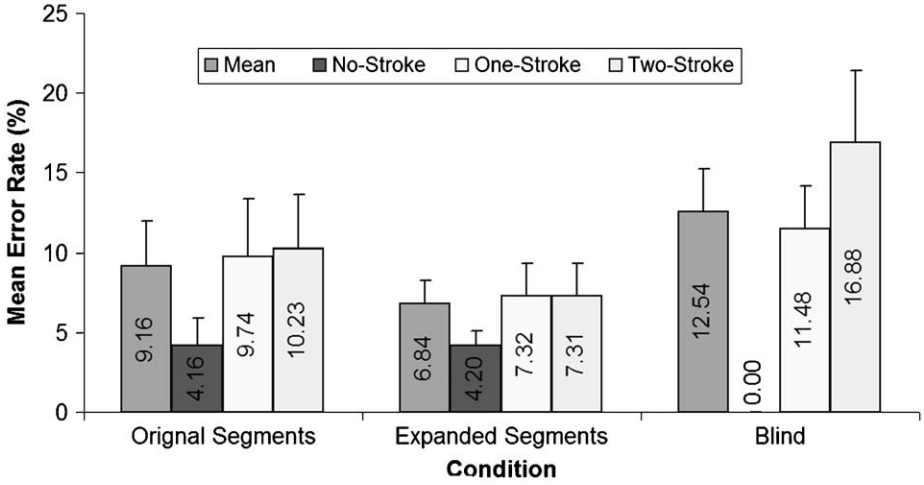

Fig. 9. Mean error rate from the distant-UI study, broken down to show data from no one- and two-stroke trials.

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

one-stroke trials. Although this result is not replicated in the distant-UI study, an examination of the raw data suggests that this is mainly due to increased error rates in the one- and two-stroke trials in the blind condition (although no interaction was detected). Indeed, an ANOVA which excludes data from the blind condition indicates that additional strokes do not result in a statistical increase in errors ($F(1, 7) = 2.29$, $p = 0.11$). Together, this suggests that, when novice users are supported by a graphical interface, they can relatively reliably make multiple strokes using this system.

Varying the number of targets in the local-UI study and the size of the targets in the distant-UI study allowed the interplay of these factors to be observed. The presence of differences in the former but not the latter clearly indicates that the number of targets is the key factor determining performance in this system. Fitts law (1954), a mathematical model for targeting performance that has been widely studied in HCI (Soukoreff and Mackenzie, 2004) offers some insights into this result. Two of its predictions are that larger targets are more rapidly acquired (something not seen here) and also that targets that are at a boundary are effectively infinitely deep, and therefore can be reached extremely quickly. The most common example of this is the edges or corners of the screen in a graphical UIs; they can be reached simply by a ballistic and unmonitored flick of the mouse. The horizontal and vertical targets in the system described here are at a boundary and are therefore infinitely deep; the system does not record when a user moves beyond their outer edge, and so it can be approached with confidence at great speed. They therefore afford rapid targeting. Additional central targets do not possess this quality and cannot be targeted as rapidly, irrespective of their size.

While reasonable, this explanation is complicated by a consideration of the kinesthetic elements of the input. The basic postures of device horizontal and device vertical are easily identifiable kinesthetic landmarks, while targets in between these two may be somewhat more ambiguous. As evidenced by the blind condition in the distant-UI study, kinesthesia or muscle memory (Clark and Horch, 1986) can be used to accurately issue commands in this system. However, it is unclear how this sensory input might distort a typical Fitt's law account of performance. Beyond these issues, the implications of this analysis for the design of a menu system are straightforward; within the limits imposed by expressivity, use fewer targets and more strokes. For example, a reasonably sized 19-item lexicon can be achieved using a 3-target, two-stroke system. It is this system configuration that the remainder of this paper focuses on.

Contrasting the two studies also allowed comparison of performance of a handheld UI, which is potentially difficult to observe, with one that is always clearly displayed on an external screen. This comparison may have been influenced by the use of different graphical interfaces in the studies. It was achieved with *t*-tests on the planning and execution times and error rates gathered from the 3-target condition in the local-UI study and those drawn from all no-stroke and one-stroke trials in the distant-UI study. In this way, these tests compared all no- and one-stroke trials conducted on a 3-target system. The raw data for these tests are shown in Table 1. They resulted in no significant differences: planning time ($t(18) = 1.71$, $p = 0.051$), execution time ($t(18) = 0.93$, $p = 0.18$) and error rate ($t(18) = 0.72$, $p = 0.24$). However, the strong trend and raw magnitude of the difference observed in the planning time are indicative of a Type II error and suggest that a larger experiment would unearth a significant result. In the setup of these experiments it is impossible to ascertain whether this trend was due to the location of the UI, the differences between the UIs or a combination of these factors. However, the main conclusion of this test is that that although participants appeared to find it more time consuming to read the experimental instructions with the local-UI interface, they were able to successfully achieve this, and these difficulties did not affect subsequent aspects of task performance (execution time and error rate). This result suggests that an on-screen graphical interface is feasible for this system; users would be able to successfully use it to learn the system before transitioning to expert phase, where they require little or no visual cueing.

Table 1
Mean results from all no- and one-stroke gestures based on three targets in the local-UI and distant-UI studies.

	Local-UI study	Distant-UI study
Planning time (ms)	2110	1680
Execution time (ms)	591	509
Error rate (%)	9	7

Indeed, the most important conclusion from the distant-UI study relates to this aspect of performance. There is a modest but clear dip in performance in the blind condition; task times increase, but remain respectable, while error rates do not change significantly (although there is a noticeable numerical increase as the gestures become more complex). Indeed, the temporal increases and the trend in the error rates may simply be attributable to the participant's relatively short experience with the system—a weakness common in the lab-based evaluation of gestural interfaces. Although the blind condition exhibited reduced performance compared to the two visual conditions, the relatively small magnitude of this dip is strongly suggestive that the system can be effectively used eyes free. This sets it apart from other motion-based menu systems, which are typically tightly coupled to their visual interfaces and do not consider such an interaction style.

4.8.2. Comparison with state of the art

It is also informative to contrast the results of these studies with the literature on motion-controlled menus. Poupyrev et al. (2002) reports a task time of 3.1–3.7 s for

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

menu commands 6 and 12 items distant from the user start point; 3.4 s is a likely extrapolation to a 19-item system used in the distant-UI study. No error rates are given, but Oakley and O'Modhrain (2005) re-ran this study and reported a rate of 19% for a 15-item menu. Hence, Poupyrev's temporal results are broadly similar to those attained in this paper, but the error rate is considerably worse. Oakley and O'Modhrain's (2005) own technique leads to a 2.75 s task time and an error rate of 9.8% with a 15-item menu; this technique is somewhat faster than the one proposed in this paper with a comparable error rate. However, these differences may simply be due to the complexity of the instructions used in the studies here (as evidenced by the lengthy planning times recorded) compared to the textual (and numerically ordered) instructions Oakley and O'Modhrain used. Furthermore, one of the main goals of this work is to create a system that can be used eyes free, whereas Oakley and O'Modhrain's system is tightly coupled to its graphical UI. Indeed, they report that several subjects complained about the conflict between looking at the device and moving it in order to interact. Adopting an eyes-free design, and therefore freeing visual attention (Pirhonen et al., 2002), affords advantages that cannot be adequately expressed by a comparison of temporal statistics.

Researchers have also studied task completion times in cell-phone menu use based on navigation through repeated button selections, currently the dominant interaction metaphor. For instance, St Amant et al. (2004) and St Amant and Horton (2007) evaluate several user-interface models against data gathered from an empirical study and conclude that GOMS is a good predictor of task performance. They decompose menu selection tasks into two discreet actions: scrolling to an adjacent item and selecting an item. They assign, respectively, a time cost of 505 and 616 ms to each of these. Applying these figures to a 19-item non-hierarchical menu that has its central item selected by default (or which wraps around, connecting its head to its tail as most mobile phone menus do) gives a mean selection time of 3504 ms based on 2 item selections and an average of 4.5 scrolling operations. This figure is approximately 10% greater than the mean total task time observed in the two visual conditions in the distant-UI study, and 10% less than that reported in the blind condition. This suggests that users can rapidly learn to use the motion interface to attain levels of performance similar to that possible with more familiar button-based systems.

Finally, there is a wealth of often quite divergent literature that reports on user performance with menus in desktop graphical UIs. For example, in an early comparison between pie and linear, list-style menus, Callahan et al. (1988) reported that the mean time to select a command (from 8 available) was 2.26 s with pie menus and 2.64 s with linear menus. In contrast, Walker and Smelcer (1990) concluded that the mean command selection time for a 9-item pull-down menu was 0.73 and 0.947 s for a similarly sized context menu. The differences between these experiments are likely due to the definition of the task completion time measure—the former study included planning time, while the latter did not. Viewed in this light, figures from both these papers correspond well to the data reported here. For an 8- or 9-item menu, a planning time of approximately 2 s precedes an execution time of less than a second. This analysis suggests that the approach described in this paper is (in terms of task completion time) broadly comparable to performance using a graphical menu on a desktop computer. A caveat to this conclusion is that recent extensions to graphical menus (e.g. Ahlström et al. (2006) for linear menus or Zhao et al. (2006) for marking menus) offer many improvements to the simple systems evaluated in this early work. Making a direct comparison between the technique described in this paper and advanced GUI menu systems displayed on mobile devices is a clear objective for future work.

In summary, the results of these two studies reflect well on the design objectives of this work. They suggest that the technique is both learnable from graphical feedback displayed on the mobile device and also usable eyes free after a relatively short period of exposure. A 3-target, two-stroke version supports a menu system containing 19 commands, sufficient to address a large proportion of commonly used functionality. Its performance compares well with that of previous motion-controlled menu systems and also the wider literature on key-press models for mobiles and on interaction with desktop computers. However, this pair of studies focused solely on an abstract interface used by novices. It is also important to assess performance with a more realistic UI and to explore learning rates. It is to these tasks that this paper now turns.

5. Learning evaluation

The abstract UI used thus far in this paper is unlikely to represent real world performance; it was designed to display the motions required to use the system in the absence of complex contextual background information (such as the names of commands) that might reduce performance. However, it is unclear whether it accurately represents performance, and in particular novice performance, with a realistic graphical UI. The study described in this section was designed to address this issue by evaluating a graphical user-interface front end to the system with a group of untrained users.

The design was based on a menu system functionally identical to the original condition in the distant-UI study. It used 3 targets spread over 90° and gestures of up to two-stroke length. This configuration was chosen as a compromise between expressiveness (it can address 19 items) and the degree to which a user's kinesthetic sense can aid them in identifying targets: one is device vertical, one is device horizontal and the final one simply in between these two poles.

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

5.1. Hypotheses

This study had a single hypothesis:

H6: participant performance will improve steadily as they proceed through the experiment.

This improvement was expected to manifest itself as both a reduction in task completion times and error rates.

5.2. Participants

The experiment featured 10 subjects, 8 male and 2 female, with a mean age of 28. All were right handed. None had participated in either of the previous two studies nor reported any physical impairment in their dominant hands. They were not compensated.

5.3. Experimental design

The goal of this study was to examine initial performance levels and learning rates for novice users. Correspondingly, all subjects experienced 10 identical blocks of trials. Each block featured 19 trials; the complete set addressable using the 3-target, two-stroke version of the menu system. During the first block an experimenter was present and this stage was devoted to demonstration of the system, to ensure that the basic concepts had been grasped. The experimenter made explanations, answered questions and clarified any issues. Participants were encouraged to explore the interface rather than complete the trials. This stage typically lasted less than 5 min and involved successfully completing 19 trials. Experimental data were captured from the remaining nine blocks.

5.4. Materials

Two considerations informed the graphical design of the experimental interface. These were that the orientations used in the system most naturally map to long (Y-) axis of a handheld device, and that the origins of each stroke are important. Correspondingly, the UI was based on a 3-item menu bar continually displayed on one side of the PDA screen (left or right depending on user handedness). Highlighting indicated which menu item (or target) was currently active. When a user engaged the menu (by holding a thumb against the screen) a larger, similarly highlighted 3-item sub-menu appeared. Disengaging the menu at this point resulted in a no-stroke operation. Alternatively, the user could rotate the device to either of the other 2 items on this sub-menu and release the screen (or button) there to perform a one-stroke operation. Furthermore, as they moved between menu items, the content of those they crossed was replaced by new commands. By changing direction, they were able to move back to access these items and perform a two-stroke operation. It was possible to cancel a gesture at any time by rotating around the X-axis by 20° or more, a motion practically instantiated as a sideways flick of the wrist (X-axis cancels). Making additional turns around the Y-axis (and therefore performing unclassified three or more stroke gestures) also resulted in cancelling the gesture (Y-axis cancels). The appearance and behavior of this design are illustrated in Fig. 10 through a series of screen shots depicting a two-stroke gesture.

The menu commands were selected to mimic those on a mobile device. They were organized to take advantage of the hierarchical nature of the menu commands. For example, under the initial menu item "Contacts" were the commands "Call", "Address book" and "Messages". Instructions regarding which menu item to select in each trial were shown on the PDA for the entire duration of each trial. They took the form of a vertical list of command names (either two or three), which indicated not only the final destination to the command, but also the full path that should be taken. An example is also visible in Fig. 10.

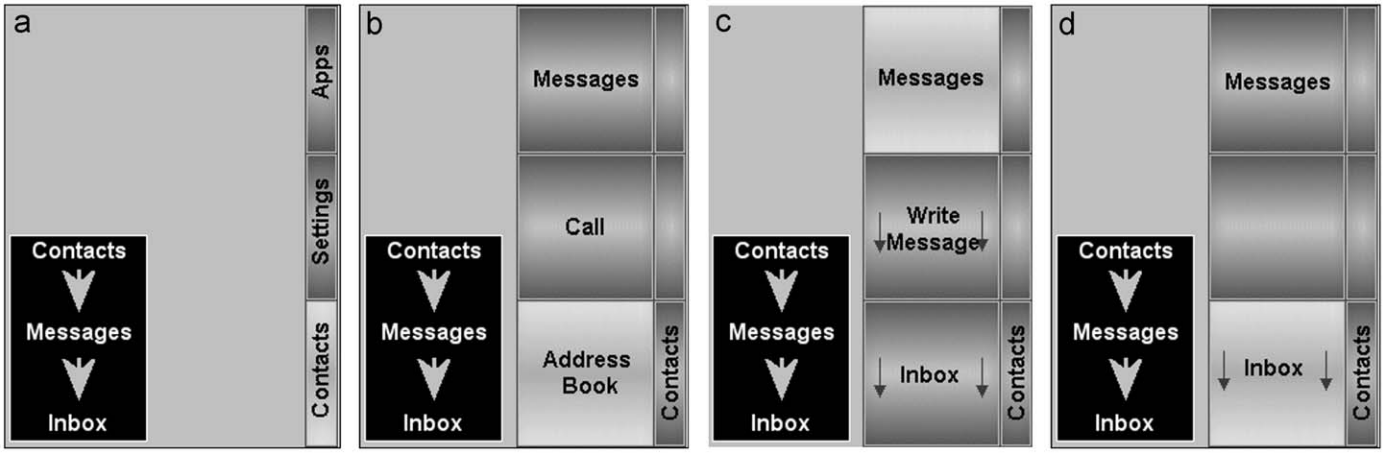

Fig. 10. Four screen shots of interface to learning study showing two-stroke gesture. In (a) the device is vertical, in (b) the screen has been pressed, in (c) the device has been rotated to horizontal and in (d) returned to vertical. The experimental instructions are shown in the black box in the bottom left of each screen shot.

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

The experimental instructions were minimal, consisting of a graphical depiction of three hand positions required to access the three targets, a walk through example of the screen display and hand positions required to make a two-stroke gesture, and an explanation of the experimental procedure, as described below. This approach allowed examination of novice performance.

5.5. Procedures

The study took place in an identical environment to those described previously; participants were in a quiet empty office, standing and holding the PDA in their dominant hand. Trials also had a similar structure; each trial commenced with a message indicating that now was an appropriate time to rest and to tap the screen when ready to continue. This was followed by the display of a fixation spot for 500 ms and then the experimental interface. Selection of a command or cancellation of the gesture completed the current trial and began the following one. Trials in which participants failed to select the correct menu item were repeated in order to explicitly separate timing and error measures.

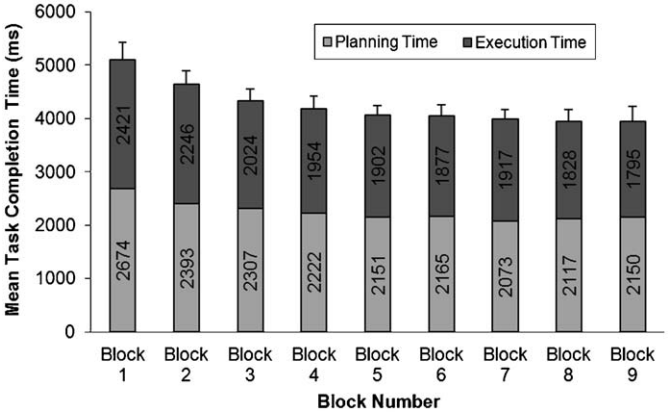

Fig. 11. Mean planning and execution times for each block in the learning experiment.

5.6. Measures

The measures were broadly similar to the previous studies. Task completion time was again broken down into planning and execution time and error rate was defined as performance of a valid, but inappropriate gesture. Command cancellations were also recorded, and separated into *X*-axis and *Y*-axis cancellations.

5.7. Results

The mean timing data for each experimental block are shown in Fig. 11. Fig. 12 is similarly organized and shows the mean error and command cancellation data. The Pearson's product–moment correlation coefficient was calculated between each of these five mean measures and the experimental block number (1 ascending to 9). Significant negative correlations (indicating an improvement in performance) were found for planning time ($r(7) = -0.842$, $p < 0.01$), execution time ($r(7) = -0.892$, $p < 0.01$) and error rate ($r(7) = 0.7$, $p < 0.05$) but not *X*-axis cancels ($r(7) = -0.6$, $p = 0.09$) or *Y*-axis cancels ($r(7) = 0.041$, $p = 1$). A regression analysis was also used to test whether the total time (planning plus execution) data follow a power law. This relationship is shown in Fig. 13 and indicates a good fit ($r^2 = 0.97$), a conclusion typical during short-term assessment of tasks that exhibit learning curves.

Table 2 shows the mean timing, error and cancellation data from all 9 experimental blocks in this study, broken down into that captured in no-stroke, one-stroke and two-stroke trials. Table 2 also includes the results of one-way ANOVAs conducted on this data, comparing performance with trials of different stroke complexities. The only measure to attain significance was execution time and post-hoc *t*-tests bore out all differences between these three means (at $p < 0.005$ or better).

5.8. Discussion

The main conclusion from this study is that this kind of gestural design is suitable for novice users. Participants

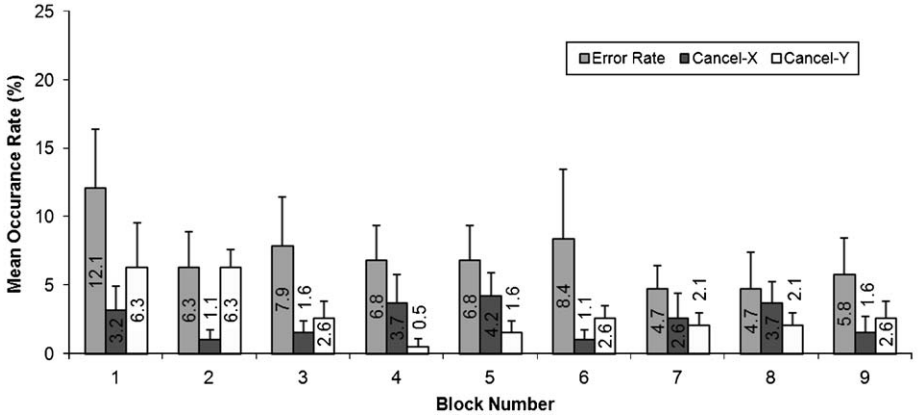

Fig. 12. Mean error rates for each block in the learning experiment.

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

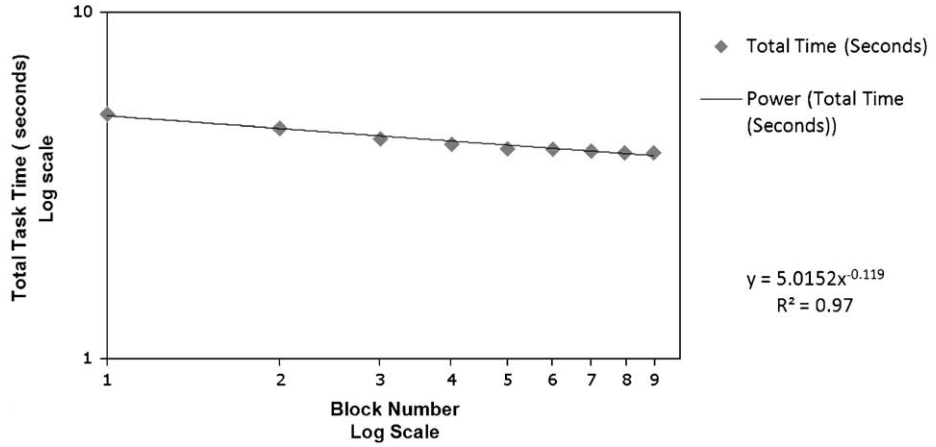

Fig. 13. Total task completion time for each block in the learning experiment shown using log scales and displaying parameters and fit to the power law.

Table 2
Mean results from the learning study, broken down to shown data from no-, one- and two-stroke trials, including results of ANOVAs analyzing these data.

	Planning time	Execution time	Error rate	Cancel-*X*	Cancel-*Y*
No stroke	2142 (134) ms	828 (84) ms	7.4% (4.5%)	5.9% (4.4%)	5.6% (2.8%)
One stroke	2095 (85) ms	1428 (82) ms	5.7% (1.4%)	1.3% (0.4%)	1.7% (0.4%)
Two strokes	2374 (86) ms	2686 (82) ms	7.7% (1.2%)	2.2% (0.6%)	3% (0.7%)
ANOVA result	$F(2, 16) = 2.66$, $p = 0.1$	$F(2, 16) = 101$, $p < 0.001$	$F(2, 16) = 0.4$, $p = 0.67$	$F(2, 16) = 2.31$, $p = 0.13$	$F(2, 16) = 2.46$, $p = 0.12$

were able to use the system effectively after only minimal instruction; error rates in the first experimental block were 12.1%, and task completion time was 5.1 s. This then decreased to 5.8% and 3.95 s after completing several hundred trials, a drop that is statistically correlated with experience with the system. This confirms the experimental hypothesis (H6) and shows that novices can use the system when initially exposed and also improve rapidly with experience. This trend fits a typical power law learning curve.

Contrasting the data from this experiment with that from the two basic studies reported earlier in this paper reveals several key differences. First and foremost, total task completions are considerably slower, although total error rates (including cancellations) remain broadly similar. The fact that each trial involved two or three lexical searches for command items is the likely cause of the temporal increases. Indeed, planning time remains similar with increasingly complex trials, indicating that this interface lacks a complex initial interpretative phase. However execution time increases substantially with more complex trials, suggesting that it now incorporates this cognitive work in the form of additional lexical command searches. This indicates that participants did not internalize the menu structure and accompanying gestures; they searched anew for the commands in each trial. This can be attributed to a combination of the limitations of human short-term memory (generally accepted to be between 5 and 9 items (Miller, 1956) and the brevity of a single experimental session, rather than any more fundamental cause. This study did not examine the performance of experts, rather novices, in the early stages of their experience.

The study also suggests several potential improvements to the graphical UI. In total, the no-stroke trials accrued a relatively large number of errors and cancellations; participants found it hard to complete these trials despite the fact they are physically the simplest (requiring only a press and release of the screen and no device motion). This may have been caused by a preemptive motion after the initial screen press, and before the interface is examined again. This problem could be avoided by incorporating a continuous display of which command would result from a no-stroke gesture. Several participants also suggested that the visual design was overly cluttered and that a single menu bar (rather than the nested one used here) would have been preferable.

In summary, this study indicates that novice users can effectively operate the system with little explanation and no prior training. Furthermore, they rapidly show improvement. Although the initial version of the UI could be refined, it provided sufficient feedback to express both the basic concepts underlying the interaction and the lexical content within the menu system. This result strongly supports the adoption of the interaction approach described in this paper for the future development and deployment of motion-based interfaces; it is imperative to support users from their first contact with a new interaction

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

technique or modality and this system points to how this can be effectively achieved.

6. Conclusions

This paper has presented a novel input system based on motion sensing and marking menus. It addresses fundamental problems with previous motion input techniques that either rely on gestural input, which must be explicitly learnt (e.g. Mantyjarvi et al., 2004; Kallio et al., 2006; Kela et al., 2006) or are tightly coupled to their visual displays (e.g. Poupyrev et al., 2002; Oakley and O'Modhrain, 2005), a paradigm that relies on the contradictory behaviors of simultaneously visually observing and physically moving a device. Two studies exploring its basic parameters are described, and from these positive results, expert user performance is briefly examined, a realistic UI design is proposed and its learnability evaluated. The results from this work compare well those in the wider literature, and suggest that the approach adopted here offers considerable potential for future development. With a lexicon of 19 commands, the final system proposed in this paper is as expressive as a typical gesture recognition system (e.g. Kela et al., 2006) and confers the additional benefits that it is usable immediately by novices, seamlessly learnable and supports rapid, eyes-free use by experts. This system, and the general approach it embodies, tackles world problems that traditional technical approaches such as improving the quality of recognition algorithms are poorly positioned to resolve.

6.1. Limitations

There are a number of important limitations to these conclusions. First and foremost, as the technique is intended for mobile use, evaluations that take place in actual mobile scenarios are an imperative. User performance in real world situations can differ dramatically from that observed in laboratories (Barnard et al., 2005) and if an interface is intended to be used out and about in trains and on streets, it need be tested in such situations. This is especially true of systems that rely on physical input as it has been shown to be susceptible to disruption in mobile scenarios involving active tasks such as walking (Zucco et al., 2006).

A second important limitation is the lack of a longitudinal study. The argument in this paper hinges on assumption that experts will be able to reliably operate the system eyes free. Although this was not formally tested, an exploratory study with a single user (one of the system developers) yielded promising initial results. Each day for two work weeks (10 sessions in total), this user completed a block of 38 trials with the system configured identically to the blind condition described in the distant-UI study. The mean results were a task completion time of 2350 ms (planning time of 1626 ms and execution time of 724 ms) and an error rate of 3.9%, considerably better than that achieved in the formal studies. Although these results have little validity, they do suggest that high levels of performance can be attained after long periods of exposure and use. A formal study of this issue is required to confirm this suggestion.

A final key limitation relates to the fundamental expressivity of this kind of motion input. The use of a small number of large angular targets as building blocks of the technique is what enables the system to be used eyes free. Indeed, broadly similar approaches appear in other eyes-free systems (e.g. Brewster et al., 2003). However, this constraint imposes severe limitations on the number of discrete commands that can be issued. Although this paper argues that the paradigm is expressive enough to support a reasonable "shortcut" style command set of 19 or more items, there are many scenarios in which this would be insufficient. These include common handheld device tasks such as text entry (Partridge et al., 2002), navigation in long lists or address books (Cho et al., 2007) and full access to a complete set of device menu options (St Amant and Horton, 2007). It is therefore worth considering how the expressivity of the technique could be increased. Possible approaches include the use of a second (and perhaps a third) orthogonal axis of motion input, the use of a greater range of angular space and the development of two-handed interaction, possibly in conjunction with a wearable sensor system (e.g. Oakley et al., 2008). Although promising directions for further development, each of these approaches entails an increase in the complexity of the system, which may result in it becoming confusing or unusable in an eyes-free scenario. Further work will be required to explore the viability of these approaches.

6.2. Future work

Beyond the issues highlighted in the limitations section, future research on this class of interaction technique should include empirical investigations on the fundamental kinesthetic factors in order to reveal general guidelines on how to structure or divide physical space to ensure that targets are identifiable by proprioception alone. Research into training effects in motion input (including retention) would also be valuable. A formal study to compare user performance against that with alternative motion input systems and more general input technologies such as keys, joysticks and touch screens would be an obvious next step. Comparing how recognition of the restricted, simplified motions used in this system performs against that in systems that rely on richer, multi-dimensional motions such those used as general gesture recognition will also be informative. Potential developments to the underlying interaction model include developing versions of the menu system based on ballistic motions.

In conclusion, we believe that motion-based interfaces have a significant role to play in the next generation of mobile devices. However, current systems suffer from the fundamental disadvantages that they either require

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

substantial training to learn a set of gestures or that they depend on the contradictory tasks of simultaneously observing and moving a device. Designing interfaces to be easily learnable by novices and to support eyes-free use by experts is one way to address these problems. Only by considering such real world constraints will motion-based interfaces successfully make the transition out of laboratories into the wider world.

Acknowledgements

This work was supported by the IT R&D program of the Korean Ministry of Information and Communication and the Institute for Information Technology Advancement, Project number 2005-S-065-03, Development of a Wearable Personal Station. Thanks must go to Kiuk Kyung, Jieun Park, Yeongmi Kim and Youngkyu Jung for their help with electronics, translation and mathematics. Vassilis Kostakos was kind enough to provide the text with a last minute polish at short notice.

References

Ahlström, D., Alexandrowicz, R., Hitz, M., 2006. Improving menu interaction: a comparison of standard, force enhanced and jumping menus. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI '06. ACM Press, New York.

Apple iPod, 2007. 〈http://www.apple.com/ipod/ipod.html〉 (accessed July 2007).

Barnard, L., Yi, J.S., Jacko, J.A., Sears, A., 2005. An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. International Journal of Human–Computer Studies 64 (4), 487–520.

Bartlett, J.F., 2000. Rock'n'Scroll is here to stay. IEEE Computer Graphics and Applications 20 (3), 40–45.

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

Benbasat, A.Y., Paradiso, J.A., 2001. Compact, configurable inertial gesture recognition. In: Extended Abstracts of the ACM Conference on Human Factors in Computing Systems, CHI'01. ACM press, New York.

Brewster, S., Lumsden, J., Bell, M., Hall, M., Tasker, S., 2003. Multimodal 'Eyes-free' interaction techniques for wearable devices. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'03. ACM Press, New York.

Bailly, G., Lecolinet, E., Nigay, L., 2008. Flower menus: a new type of marking menu with large menu breadth, within groups and efficient expert mode memorization. In: Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '08. ACM Press, New York.

Callahan, J., Hopkins, D., Weiser, M., Shneiderman, B., 1988. An empirical comparison of pie vs. linear menus. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI '88. ACM Press, New York.

Clark, F., Horch, K., 1986. Kinesthesia. In: Boff, K., Kaufman, L., Thomas, J. (Eds.), Handbook of Perception and Human Performance, vol. 1. Wiley, New York, pp. 13-1–13-62.

Cho, S.J., Choi, E., Bang, W.C., Yang, J., Sohn, J., Kim, D.Y., Lee, Y.B., Kim, S., 2006. Two-stage recognition of raw acceleration signals for 3-D gesture-understanding cell phones. In: Proceedings of the Tenth International Workshop on Frontiers in Handwriting Recognition, IWFHR 10.

Cho, S.J., Murray-Smith, R., Choi, C., Sung, Y., Lee, K., Kim, Y-B., 2007. Dynamics of tilt browsing on mobile devices. In: Extended Abstracts of the ACM Conference on Human Factors in Computing Systems, CHI'07. ACM Press, New York.

Fitts, P.M., 1954. The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology 47, 381–391.

Harrison, B.L., Fishkin, K.P., Gujar, A., Mochon, C., Want, R., 1998. Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'98. ACM Press, New York, pp. 17–24.

Hinckley, K., Pierce, J., Horvitz, E., Sinclair, M., 2005. Foreground and background interaction with sensor-enhanced mobile devices. ACM Transactions on Computer–Human Interactions Special Issue on Sensor-Based Interaction 12 (1), 31–52 (special issue on sensor-based interaction).

Hinckley, K., Pierce, J., Sinclair, M., Horvitz, E., 2000. Sensing techniques for mobile interaction. In: Proceedings of the ACM Conference on User Interface Software Technology, UIST'00. ACM Press, New York, pp. 91–100.

Hinckley, K., Zhao, S., Sarin, R., Baudisch, P., Cutrell, E., Shilman, M., Tan, D., 2007. InkSeine: in situ search for active note taking. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'07. ACM Press, New York.

Holtzblatt, K. (Ed.), 2005. Designing for the mobile device: experiences, challenges and methods. Communications of the ACM, 48(7).

Kallio, S., Kela, J., Mantyjarvi, J., Plomp, J., 2006. Visualization of hand gestures for pervasive computing environments. In: Proceedings of the Working Conference on Advanced Visualization Interfaces. ACM Press, New York, pp. 480–483.

Kela, J., Koripiaa, P., Mantyjarvi, J., Kallio, S., Savino, G., Jozzo, L., Di Marca, S., 2006. Accelerometer-based gesture control for a design environment. Personal and Ubiquitous Computing 10 (5), 285–299.

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

Kurtenbach, G., Buxton, W., 1993. The limits of expert performance using hierarchic marking menus. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93). ACM Press, New York.

Long Jr., A.C., Landay, J.A., Rowe, L.A., 1999. Implications for a gesture design tool. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'99. ACM Press, New York, pp. 40–47.

Mantyjarvi, J., Kela, J., Korpipaa, P., Kallio, S., 2004. Enabling fast and effortless customization in accelerometer based gesture interaction. In: Proceedings of the third International Conference on Mobile and Ubiquitous Multimedia. ACM Press, New York, pp. 25–31.

Miller, G.A., 1956. The magic number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review 63, 81–97.

Oakley, I., O'Modhrain, S., 2005. Tilt to scroll: evaluating a motion based vibrotactile mobile interface. In: Proceedings of World Haptics'05, IEEE Computer Society, pp. 40–49.

Oakley, I., Sunwoo, J., Cho, I.Y., 2008. Pointing with fingers, hands and arms for wearable computing. In: Proceedings of the ACM CHI 2008 Conference on Human Factors in Computing Systems, CHI'08. ACM Press, New York.

Partridge, K., Chatterjee, S., Sazawal, V., Borriello, G., Want, R., 2002. Tilt-type: accelerometer-supported text entry for very small devices. In: Proceedings of the ACM Conference of User Interface Software Technology, UIST'02. ACM Press, New York.

Pirhonen, A., Brewster, S.A., Holguin, C., 2002. Gestural and audio metaphors as a means of control for mobile devices. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'02. ACM Press, New York.

PocketMotion, 2007. 〈www.pocketmotion.com〉 (accessed July 2007).

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002

Poupyrev, I., Maruyama, S., Rekimoto, J., 2002. Ambient touch: designing tactile interfaces for handheld devices. In: Proceedings of the ACM Conference on User Interface Software Technologies, UIST'02. ACM Press, New York.

Rekimoto, J., 1996. Tilting operations for small screen interfaces. In: Proceedings of the ACM Conference on User Interface Software Technologies, UIST'96. ACM Press, New York.

Schwesig, C., Poupyrev, I., Mori, E., 2004. Gummi: a bendable computer. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'04. ACM Press, New York, pp. 263–270.

Schlömer, T., Poppinga, B., Henze, N., Boll, S., 2008. Gesture recognition with a Wii controller. In: Proceedings of the Second international Conference on Tangible and Embedded interaction, TEI'08. ACM Press, New York.

St Amant, R., Horton, T.E., Ritter, F.E., 2004. Model-based evaluation of cell phone menu interaction. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'04. ACM Press, New York.

St Amant, R., Horton, T.E., 2007. Model-based evaluation of expert cell phone menu interaction. ACM Transactions on Computer–Human Interaction 14 (1), 1–24.

Soukoreff, R.W., Mackenzie, I.S., 2004. Towards a standard for pointing device evaluation: perspectives on 27 years of Fitts' law research in HCI. International Journal of Human–Computer Studies 61, 751–789.

Tactaid VBW32, 2007. 〈www.tactaid.com/skinstimulator.html〉 (accessed July 2007).

Tamminen, S., Oulasvirta, A., Toiskallio, K., Kankainen, A., 2004. Understanding mobile contexts. Personal and Ubiquitous Computing 8 (2), 135–143.

Verrillo, R.T., 1966. Vibrotactile thresholds for hairy skin. Journal of Experimental Psychology 72 (1), 47–50.

Walker, N., Smelcer, J.B., 1990. A comparison of selection times from walking and pull-down menus. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'90. ACM Press, New York.

Wigdor, D., Balakrishnan, R., 2003. TiltText: using tilt for text input to mobile phones. In: Proceedings of the ACM Conference on User Interface Software Technologies, UIST'03. ACM Press, New York.

Zhao, S., Agrawala, M., Hinckley, K., Baudisch, P., 2006. Zone and polygon menus: using relative position to increase breadth of multi-stroke marking menus. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'06. ACM Press, New York.

Zhao, S., Balakrishnan, R., 2004. Simple vs. compound mark hierarchical marking menus. In: Proceedings of the ACM Conference on User Interface Software Technologies, UIST'04. ACM Press, New York.

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 ACM Conference on Human Factors in Computing Systems, CHI'07. ACM Press, New York, pp. 1395–1404.

Zucco, J.E., Thomas, B.H., Grimmer, K., 2006. Evaluation of four wearable computer pointing devices for drag and drop tasks when stationary and walking. In: Proceedings of the Tenth International Symposium on Wearable Computers. IEEE Computer Society, pp. 29–36.

Please cite this article as: Oakley, I., Park, J., Motion marking menus: An eyes-free approach to motion input for handheld devices. International Journal of Human Computer Studies (2009), doi:10.1016/j.ijhcs.2009.02.002