Putting Your Best Foot Forward: Investigating Real-World Mappings for Foot-based Gestures

Jason Alexander[1], Teng Han[2], William Judd[2], Pourang Irani[3], Sriram Subramanian[2]

[1]School of Computing and Communications
Lancaster University, UK
j.alexander@lancaster.ac.uk

[2]Department of Computer Science
University of Bristol, UK
{han, sriram}@cs.bris.ac.uk

[3]Department of Computer Science
University of Manitoba, Canada
irani@cs.umanitoba.ca

ABSTRACT

Foot-based gestures have recently received attention as an alternative interaction mechanism in situations where the hands are pre-occupied or unavailable. This paper investigates suitable real-world mappings of foot gestures to invoke commands and interact with virtual workspaces. Our first study identified user preferences for mapping common mobile-device commands to gestures. We distinguish these gestures in terms of discrete and continuous command input. While discrete foot-based input has relatively few parameters to control, continuous input requires careful design considerations on how the user's input can be mapped to a control parameter (e.g. the volume knob of the media player). We investigate this issue further through three user-studies. Our results show that rate-based techniques are significantly faster, more accurate and result if far fewer target crossings compared to displacement-based interaction. We discuss these findings and identify design recommendations.

Author Keywords

Foot-based interaction; foot gestures; mobile device interaction

ACM Classification Keywords

H5.2 [Information interfaces and presentation]: User Interfaces – Input Device and Strategies.

General Terms

Human factors.

INTRODUCTION

Foot and lower-leg gestures open a broad design space to the problem of interacting with a mobile device when the users' hands are busy or otherwise unavailable. Such scenarios include instances when the user's hands are dirty (e.g. a farmer operating a device on the field, Figure 1), when the user's hands are occupied with other tasks (such as carrying shopping bags in one hand and holding a child with the other), in environments where upper-body space is

limited or inappropriate for use (such as during a meeting) or when it is too cold to remove gloves (many touch-screens do not work without finger contact). In most of these scenarios the user's hands can hold the device giving them visual access to information on screen but it is difficult for the user to perform precise interactions with the touch-screen or buttons required for device operation.

The use of feet for non-mobile device interaction is already common: the pedals in a modern car are often connected to a computer rather than directly to the engine, arcade games such as *Dance Dance Revolution* (konami.com/ddr) use only the feet for interaction and home entertainment systems such as Microsoft Kinect and Nintendo Wii use foot-based interaction for gaming purposes. Passive interaction systems such as Nike+ (nikeplus.com) connect an accelerometer embedded in the shoe to a mobile device to record running or walking information.

The literature has considered various foot gestures for different contexts and one can find three emerging categories: kicking [6], foot tapping [4] and ankle rotations [22]. Prior work has focused on providing a thorough understanding of human abilities for input using foot interaction. Less attention has been paid as to how these types of interactions map to real-world commands.

Figure 1: (left) An example scenario (from iStockphoto) where it is undesirable for the hands to touch the screen but holding the device is still possible. (right) Foot gestures as an alternative to interaction when hands cannot be used.

This work therefore aims to understand how user would expect foot gestures to be mapped to common mobile device functions and to investigate interaction parameters associated with these mappings. We achieve this through a guessability-style study, alike [24, 19], that probes users on suitable commands. This resulted into two major groupings, discrete and continuous foot mappings. Through three user-

studies we further explored the design space for continuous foot gestures in the context of 2D spatial document navigation, one that requires moving around with fluidity and precision.

Our investigation results in the following contributions: (1) a novel foot-based gesture set for over thirty common mobile device commands; (2) a validation that our gesture set can be reliably detected with commonly used sensing devices; (3) a thorough examination through three user study's investigating the design factors surrounding the mapping of continuous foot interaction (in the context of spatial navigation); (4) design recommendations for implementing foot-based map navigation

RELATED WORK

To find ideal mappings for foot interaction we first examine some of the salient properties of feet motion. We then review prior foot-based interactions and group these into two camps, discrete and continuous mappings. Finally, related to this work are methods for defining a new catalog of gestures, which we briefly review.

Foot Motion

The human foot is a highly dexterous system with advanced movements using multiple joints that increase in movement complexity from the hip to the ankle. Advanced movements such as balancing on one foot to kick can be mastered as early as nine months [5]. Each leg joint provides multiple movements with varying ranges of motion. While the lower limb allows for a variety of maneuvers, studies comparing it to the arms show that it is not as precise as human hand and fingers [13]. This is particularly the case for fine and precise movements, such as moving a trackball with the foot to select on-screen targets [16]. However, studies have also shown that the foot is better suited at more coarse level movements, and can be quick in performing such actions [13]. It is for such types of coarse and less precise motions, that we investigate the use of foot gestures.

Continuous and Discrete Foot Interaction

Early work in foot-based input concentrated on the design of suitable foot controls [16] and sensor embedded shoes for entertainment purposes [14]. Recent work has extended foot input to specific application contexts, such as to retrain the lower limbs in physiotherapy [15], for ambient awareness [18], and for user identification through gait [8] and foot imprints [3]. In relation to our work, we broadly categorize existing literature into discrete and continuous foot-based input.

Discrete foot input has primarily focused on command invocation. The most common type consists of foot tapping [4] which was shown to provide sufficient input breadth to interact with mobile devices, even when the latter are buried in a user's pant pockets [22]. Crossan et al. [4] investigate menu command invocation via a single or double-foot tap. With accelerometers attached to the foot,

they found foot tapping to be as accurate as visually interacting with a menu, but with a speed trade-off. Scott et al. [22] further investigate the use of foot gesturing, such as toe tapping (dorsiflexion), and heel tapping (planar flexion) for interacting with a device residing in the user's pockets. With accelerometers on the mobile device, their trained system could classify ten different foot gestures with approximately 86% accuracy. In addition to tapping, which commonly occurs when the user is stationary, harnessing discrete foot input has also seen some success when jogging. Foot-Step [21] can recognize a command based on the type of side-stepping (right or left) a user makes while jogging. Foot stepping can be captured with as high as 95% accuracy but was useful for a limited input range of up to four commands.

In contrast to discrete command invocation, continuous foot input has been explored for a number of activities such as artistic expressions [14], game playing [12] and document navigation [20]. Paelke et al [12] successfully report on user satisfaction in playing games when using a mobile phone's camera (even with limited frame rates) to detect foot movements and control a soccer ball. They apply their results to kick-up menus for selecting items on a mobile device with the foot. Similarly, Han et al. [6] propose kicking as a method for controlling objects on a mobile device, by examining the direction and speed of the foot movement. Schoning et al. [20] apply foot pressure input to navigate a spatial document as a means to complement hand interaction, such as panning faster when leaning on one side. Drawing inspiration from work on wrist control [17], Scott at al. [22] exploit continuous ankle movements such as toe and heel rotations. Similar to their results on feet tapping [22] they show that heel and toe rotations provide a reasonable range-of-motion and can be successfully detected by a device placed in a user's pocket.

The above work provides a frame of reference and motivates our systematic study on foot gestures. What is less known is what types of tasks map well to either discrete, continuous or both input types. Given the dexterity of the lower limb, we explore the use of specific real-world mappings for foot based gestures, by first developing a catalog of novel foot gestures.

Defining a Mapping of Gestures

Numerous interplaying factors influence the design of any new system of gestures. Designers have to provide optimally usable features within a fixed number of constraints, such as the type of sensing device used, limitations of human motor function, and finally the robustness and distinctiveness of the gesture set for easy and sound classification. Researchers have proposed several methods to provide rigor to the design of gestures.

The design of most gestural syntax begins with a basic knowledge of human motion capabilities, to determine what is mechanically possible. For such systems the onus is on

the user to learn and familiarize themselves with the gestures. Alternatively, the user can be included in the process [23]. Participatory design incorporates users in the design process and carefully iterates on numerous proposals before settling on a given syntax. When possible, researchers can observe user interaction with physical objects, as gestures are largely based on our interaction with items around us. For example, observing user behavior in passing sheets of paper around a table [10] or rotating and dragging objects [7] has given impetus to the design of novel gestural techniques for tabletops. In the absence of an implemented system, designers can get quick feedback on their newly defined catalog using Wizard of Oz approaches.

Recently, researchers have elicited user input and defined methods for converging on a given gestural mapping through metrics such as agreement scores [24]. Such an approach, referred to as guessability, has worked successfully in defining a gestural mapping for one or two handed input on a tabletop [24] and for gestures with mobile devices [19]. The ultimate objective of democratizing the design process is to funnel the user's input into a sound set of implementable gestures. This provides some level of assurance that a good number of users may intuitively agree on the gesture set and above all, that these are physically possible. We adopt this approach in identifying a suitable mapping of real-world gestures for foot-based input, one that has not been explored in this context. Additionally, since users may provide broad mappings from task to gestures, to complete the design loop we also further refine user suggested mappings and consider how best to design for cases where the mappings may not be so straightforward.

MAPPING LOW LEVEL GESTURES TO COMMANDS

The literature provides a solid understanding of the various methods of conducting foot-based interaction. However, there is little knowledge of how these gestures may be mapped to device commands. To investigate this we did a study to understand users' perceptions on how foot gestures should be coupled to common mobile device functions.

To capture these perceptions we wished to create a user-defined gesture set for a range of mobile device interactions. To do this we conducted a guessability study, similar in style to Wobbrock et al. [24] and Ruiz et al. [19] where we asked participants to provide gesture suggestions for a variety of mobile device commands. This type of study is particularly suitable for foot-based interaction, as users currently have few, if any, preconceptions of this mapping—the gestures they provide are likely to be natural or logical mappings. The remainder of this section describes our experimental design and results.

Participants

We recruited 19 participants (seven females), within the ages range 18–25 years of age. Eight participants indicated that they use hand-based gesture devices (such as the Nintendo Wii) on a weekly basis and none regularly used any foot-based gesture systems.

Experimental Design and Methodology

The study primarily consisted of a researcher presenting participants with a range of mobile interaction scenarios (first column of Table 1) for which each participant was requested to perform a foot-based gesture that they believed was appropriate for the required situation.

During the study, participants stood upright on a marked area on the floor. They wore a Texas Instruments EZ430-Chronos Sports watch on their dominant lower leg, just above their ankle. The watch contains an accelerometer that was set to sample at 400Hz. This was used to record their movement while performing gestures, with each session also video-recorded for later verification. Although participants only wore the watch on their dominant leg, we placed no restrictions on which leg (or if they wished to use both legs) they could use for performing the gestures.

We choose a range of mobile-device interactions that we believed were representative of the commands that users issue to their mobile devices, as shown in Table 1. Many of these were also used in the study by Ruiz et al. [19]. For each of these scenarios the experimenter placed the command into context, for example, "your phone is ringing and you want to answer it. What foot gesture would you perform to answer the phone?" and the participant then performed a gesture they believed appropriate for this task. Once all gestures were performed, participants completed a questionnaire regarding their interactions.

Participants were told to consider each gesture independently and not to worry about the technical issues of gesture detection. The categories of gestures were always presented in the same order (phone control, media control, map navigation, browser navigation); the gestures within those categories were presented randomly.

Results

We observed a total of 537 commands: each of the 19 participants provided a foot gesture for the 30 commands, with 33 instances where no gesture was performed. Based on the participants' actions, we performed a series of analyses to understand participant preferences for mapping commands to gestures.

Analysis One: Participant Agreement

To begin, we analysed the agreement between participants for the gesture sets they selected using Wobbrock et al's methodology [24]. High agreement values indicate many participants selected the same mapping, low value indicates a large diversity in the selected gestures.

Overall, we found a large diversity in the gestures participants selected for each command. Across all commands there was a mean agreement value of 0.13 (s.d. 0.08), with 90% of the commands having agreement values below 0.2. The gestures for *shuffle* (shake foot, 0.46), *rotate*

clockwise (trace clockwise circle, 0.29) and *rotate counter-clockwise* (rotate foot counter-clockwise, 0.29) were the only three commands where a clear preference between participants was observed.

To further understand *why* we had observed such disagreement between participants, we performed a second, deeper analysis of the suggested gestures.

Command	Generalised gesture	% part.	Assigned gesture
Phone control			
1. Answer incoming call	Tap foot	58%	Double tap
2. Ignore incoming call	Shake foot	42%	Shake foot
3. End current call	Tap foot	68%	Double tap
4. Activate voice search	Trace circle	26%	Trace circle with toe
5. Lock phone	Tap	26%	None (clash)
6. Unlock phone	Tap	26%	None (clash)
7. Home screen	Tap	42%	None (clash)
8. Next item	Rightwards movement	42%	Kick right
9. Previous item	Leftwards movement	37%	Kick left
10. Activate silent mode	Tap	32%	None (clash)
11. Activate loud mode	Tap	37%	None (clash)
12. Activate gesture recognition	Trace symbol	37%	Trace clockwise circle with foot
13. Deactivate gesture recognition	Trace symbol	37%	Trace counter-clockwise circle with foot
Media control			
14. Play/pause	Tap	68%	Double tap foot
15. Next track	Rightwards movement	74%	Kick right
16. Previous track	Leftwards movement	68%	Kick left
17. Shuffle	Shake foot	68%	Shake foot
18. Volume up	Forwards movement	47%	Kick forward
19. Volume down	Backwards movement	42%	Kick backward
Map navigation			
20. Pan left	Leftwards movement	68%	Kick left
21. Pan right	Rightwards movement	74%	Kick right
22. Pan up	Forwards movement	74%	Kick forward
23. Pan down	Backwards movement	68%	Kick backwards
24. Zoom in	Tap	32%	None
25. Zoom out	Tap	26%	None
26. Rotate left	Counter-clockwise movement	100%	Trace arc counter-clockwise
27. Rotate right	Clockwise movement	100%	Trace clockwise arc
Browser navigation			
28. Select address bar	Tap	74%	Double tap forward
29. Back	Tap to the left	47%	Double tap forward & left
30. Forward	Tap to the right	47%	Double tap forward & right

Table 1: Gesture set for common mobile device commands. "None" refers to cases without convergence and/or clashes with previously assigned gestures.

Analysis Two: Gesture Set Generalisation
In the first analysis we strictly classified gestures into sets that were *identical*—for example, gestures that used two foot taps and gestures that used three foot taps were classified as different. To better understand the disagreement between participants, we relaxed the gesture classification criteria from being strictly identical to those that are 'similar'. Similar gestures were those classed as having a common property. Examples of these properties include: the direction of leg movement and the type of movement, e.g. tapping on the ground, tapping heals together or tracing an arc. For those gestures with multiple common properties, we grouped gestures in such a way that we formed the largest groups possible. The coding of identical (analysis one) and 'similar' gestures was performed independently by two coders with discrepancies reviewed and corrected.

A much clearer pattern of gesture to command mappings emerged from the more generalised groupings. We

observed several occasions where participants wanting tapping actions (e.g. answering an incoming call) or direction movements (e.g. next item) that were not immediately visible in the initial groupings. In the generalised gesture set classification, we now saw a mean agreement across all commands of 0.42 (s.d.0.2), with all commands now having agreement values above 0.2 (greater than the mean of the previous classification).

This approach improved our ability to create a coherent gesture set based on user inputs. A summary of the process for gesture selection is shown in Table 1: for each command we indicate the most common generalised gesture feature, the percentage of participants who performed a gesture with this feature and then finally the selected gesture for each command. Note that for some gestures there was still significant disagreement between participants and/or the same gesture selected for multiple commands meaning that some commands did not have a gesture assigned.

We were able to define gestures for 23 of the 30 commands in the command set. Several specific gestures appear multiple times in the defined gesture set (defined as a clash). However, contextually, the gestures are unique: gestures from one set will not be used at the same time in another, for example, media control gestures will not be required while navigating maps.

Discussion

Gesture Classification
Our findings suggest that many of the gestures are logical mappings from commands participants are already familiar with. For example, tapping to select or a rightward movement to move to the 'next' item, seem to be direct translations of what occurs when using touch on mobile devices. Interestingly, we find a consistent left-to-right pattern for distinguishing commands along the time dimension. Finally, for browser control, such as moving forward or backward, users apply spatial positioning to their taps. These results overall reveal intuitive mappings that seem sound and logical.

Gesture Set Customisability
In generating the mappings for common phone commands to foot-based gestures, our initial analysis resulted in a large diversity in participants' gesture selections. By performing a generalised set analysis we could better represent the gestures of a greater number of participants. This required generalisation and is perhaps representative of the need for gestures to be flexible and/or customisable. If a user finds a particular gesture uncomfortable or the mapping unnatural the option should exist for reconfiguration.

Classification Lessons
While eliciting user input has given us a useful gesture set, participants may not always be expected to create the full catalogue of gestures on their own. A number of commands clashed in terms of duplicate syntax, leading to ambiguity

in the final gesture set. Additionally, relaxing constraints for classifying the gestures led us to converge more quickly on certain types of gestures. Such generalised classification provides a good starting point for gesture implementation. However, all gestures still require refinement and the ability for users to customise exact input parameters.

VALIDATION OF GESTURE RECOGNITION

The previous section detailed a user-defined gesture set for foot-based control of common mobile device commands. Since users were free to derive any set of gestures, one important design aspect is to ensure that the gestures can be easily recognized and distinguished by an electronic recogniser. To validate this intention, we used an implementation of the Protractor3D gesture recogniser [9] to classify foot gestures detected using an accelerometer.

Equipment Setup

The goal of this validatory study is to determine whether our set of gestures can be accurately detected and classified. For this reason, we asked participants to wear an accelerometer strapped to their leg, combined with the Protractor3D recognizer. While this setup would be impractical for real-world use, we were interested in determining *if* the gestures can be reliably classified, not whether our method is the technologically best implementation. We chose the Protractor3D recognizer as previous studies validating its use showed high recognition rates [9]. Participants wore the same Chronos Sports watch as described earlier. The data from the accelerometer was fed in real time, into the Protractor3D recognizer.

Procedure

Five participants (1 female) with an average age of 22.2 years participated in this validation study. The gestures were presented to users in the sub-categories introduced earlier (phone control, media control, map navigation and browser navigation), using the gestures selected in Table 1. However, each unique gesture was only presented once— identical gestures were not presented multiple times. For each sub-category, participants were first asked to perform each of the gestures five times to train the recogniser. Once the training was complete, gestures from the sub-category were performed, five times, until all gestures were covered.

The gestures were presented to the users in a pseudo-wizard-of-oz style. For example, when the user was required to answer the phone, they were verbally told the context, reminded of the gesture, and were then asked to complete the gesture when they heard a phone ringing—the experimenter played a ringing sound from a PC and manually stopped the sound once the gesture was complete.

Results

The five participants completed all of the tasks for all gestures giving a total of $5 \times 13 = 65$ training gestures and 65 test gestures. The recognition accuracy for each of the gestures is shown in Figure 2. Gestures had a mean overall recognition accuracy of 87%. Recognition of all gestures, except 'shake foot' was successfully with rates above 80%. The 'shake foot' gesture fared poorly during recognition. This is likely due to the ambiguity of the gesture: participants did not perform the same number of left and right shakes each time meaning the classifier misinterpreted the user's intentions.
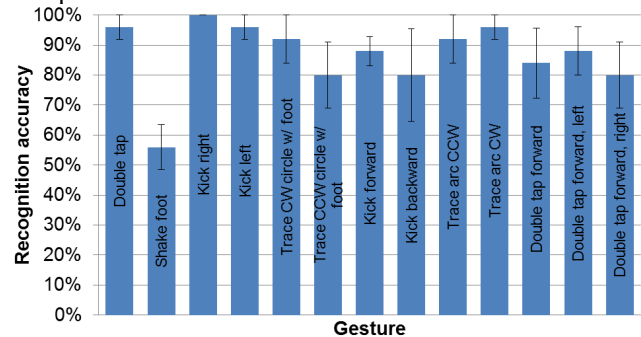

Figure 2: Gesture recognition accuracy

Discussion: Continuous vs. Discrete Actions

This study has shown that the selected gesture set is detectable and individual gestures are distinguishable. These gestures must now be mapped to mobile device functions.

The mapping of *discrete* gestures—single actions often triggering a mode switch, such as answering the phone— are straight-forward to implement. However, those gestures that map to *continuous* actions such as panning a map or adjusting the volume of the audio require further design thought due to the increased number of parameters involved. For example, how should a forward kick map to the size of an upward pan? Should the kick velocity map to distance travelled? Should the distance the foot has moved map to the distance the map moves? Or should the kick initiate a rate-based scrolling action?

To answer these questions we conducted three additional studies that examined the mapping of foot-based gestures to continuous actions.

MAP NAVIGATION: CONTINUOUS FOOT INTERACTION

Many mobile device applications require continuous interaction, such as scanning music playlists, scrolling an address book or navigating a map. Having derived a mapping of foot-based gestures to commands and validated that these gestures can be detected, we wished to examine the issue of mapping foot gestures to continuous interaction. The over-arching goal of these studies is to answer the question: how should a kick action control continuous movement?

To answer this question we implemented four kick-to-movement mappings: one distance-based, two velocity-based and one combined distance and velocity mapping (flick). We then tested the viability of these techniques in three controlled evaluations. The first measures the suitability of these techniques for acquiring targets in all

directions around a compass. The second looks at their suitability for acquiring targets at different distances. While the third combines these two parameters to understand the complete context of continuous navigation.

In these studies we apply our techniques to spatial document navigation; however, we envisage our results will also generalize to other continuous interaction techniques.

Gesture Mapping Techniques

We implemented four mappings of kick gestures to panning actions. These mappings are motivated by techniques found in the literature and commonly employed in desktop or tabletop environments. These are described here.

- **Displacement based (D):** The distance the user kicks from the vertical position directly maps to the distance the map pans. Returning the leg to the vertical position does *not* result in the map panning in the opposite direction, instead the user must move their leg in the opposite direction past the vertical position to achieve reverse panning. A horizontal position delineates movements in left and right directions.

- **Rate-based hold (RH):** The distance the user moves their leg away from the vertical position maps to panning velocity. The user must hold their leg in that position to maintain the panning velocity—moving their leg back towards a vertical position reduces the velocity accordingly.

- **Rate-based continuous (RC):** Users are likely to easily tire from holding their leg in the air. In this technique, the user kicks forward to create a velocity mapping—the furthest distance the kick reaches is used as input to the panning velocity. The user can then return their leg to the vertical position. Kicking again changes the pan velocity, tapping their foot on the ground stops the panning motion.

- **Flick (F):** Numerous variations of flick gestures exist, with Aliakseyeu et al. [6] finding Compound-Multi-Flick (CMF) generally out-performed and was subjectively preferred over other implementations. We implemented a version of CMF for kick gestures. For a forward kick, this worked as follows: when the user's leg is moving forward, CMF is in displacement mode. When the user's leg halts or begins to return to the upright position, CMF enters flick mode, with the initial flick velocity derived from the velocity of the forward displacement. A friction element slows the flick, which can also be terminated by the user tapping their foot on the ground.

We fine-tuned the parameters for each of the above interaction techniques through informal pilot studies involving four participants. Participants conducted a series of kicking tasks and provided feedback on the mapping of kicks to parameters for each of the interactions. Using a series of iterations we selected parameters that provided the most 'expected' behavior for each technique.

Pilot Study: Moving the Content or Moving the View?

With all continuous interaction systems, there is the question of whether a navigation action should move the content or move the view of the content (which equate to movements in opposite directions).

To provide an informal answer to this question we implemented two versions of each of the four mapping techniques: one which had the kick actions tied to the content and one which had the kick actions tied to the view.

We gave 10 participants both versions of each mapping technique with a single target location north of the starting position. Participants tried each mapping. Once they had used both mappings of each technique, they were asked to subjectively indicate which they preferred.

Results
Participants overwhelmingly preferred kicks to move the content. This meant that a kick forward should move the map content forward, closer to a southwards target. All participants preferred this mapping for the RH technique, and all but one preferred this mapping for the displacement and flick techniques. Six of the ten participants preferred a kick to move the content for the rate-based continuous technique. In all of our subsequent implementations the kick action is tied to the map—i.e. kicking forward will move the *map* forward, not the view window.

Experiment 1: Kicking Around the Compass

The aim of this experiment is to examine each of the kick interaction techniques when participants have to reach targets located in different compass directions.

Experimental Design and Procedure
We created a 4×8 factor within subjects experimental design with factors *Gesture Mapping* and *Direction*. We used the four gesture mappings (D, RH, RC, and F) described in the previous section. We tested movement in the eight compass directions: north, north-east, east, south-east, south, south-west, west and north-west.

The study had four sections, one for each of the Gesture Mapping techniques. These techniques were provided in a random order to each participant, with all tasks using that technique completed before moving to the next.

The experimental software automatically recorded all aspects of the interaction required to reach a target goal on a map: the task completion time, the number of kicks and the total distance and direction of all movements. We used a spatial workspace, a map, to provide users with a familiar context in which we could test continuous foot gestures.

Participants got two minutes of freeform practice navigation with the techniques followed by four practice trials. Each trial began by the participant pressing a button on the side of the screen and was completed when the target was placed inside the green rectangle (task details are explained further below). If selecting a particular target was difficult the experimenter terminated that task and the study

continued to the next trial. Participants were free to rest at any point when not performing a task. For each factor the participant did three trials resulting in a total of 96 trials per participant (3 repetitions ×4 mappings×8 techniques).

Participants and Physical Setup
We recruited 10 volunteer participants (two female), with an age range of 18–27 years old. All of these participants were right handed; six of whom had experience with on-screen touch gestures.

During the study, participants stood upright on a marked position and held a 7" screen (resolution 640×480px) at waist level, as shown in Figure 1, right. This displayed the experimental interface, as described below.

The participants' foot-based gestures were captured using an XBox Kinect camera mounted 0.75m from the ground and 3m from the participant. This camera detected participants' kick gestures accurately in three dimensions, a 'kick' began when the foot was lifted from the ground. We used the skeleton tracking algorithm provided by the OpenNI (openni.org) libraries and we implemented functions to translate these gestures into the interaction techniques described earlier.
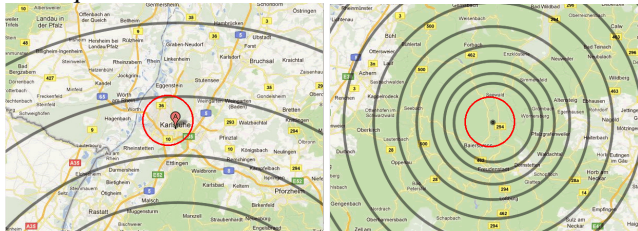

Figure 3: Experimental interface: (left) At the beginning of a task (right) At the end of a task.

Experimental Interface and Tasks
This experiment tests the capabilities of each continuous foot mapping and *not* the users' spatial abilities. For this reason, all tasks began at a random (on land) location. Tasks did not necessarily end at a specific city or location, but were determined by the required direction of travel (in this experiment, all targets were equal distances from the starting location). Each task began with the starting position centered on-screen (Figure 3, left). The target was marked with a black cirlce and had concentric circles extending throughout the whole map (total size 2560 × 1920 px, Figure 3, right) to indicate the required navigation direction. Participants were required to place the target in a rectangle with a green outline (size 128 × 90px) occupying the middle fifth of the screen in order to complete the task. The system automatically detected when the map was stationary and the target was inside the required circle for one second.

Results
With 10 participants and 96 trials per participant we collected 960 data points. All tasks were successfully completed. There was a significant difference in task completion time between techniques: *RH* was the fastest (mean time 3.6sec, s.d. 0.6sec) and *RC* the slowest (mean

time 5.0sec, s.d. 1.1sec), $F_{3,27} = 81.2$, $p < 0.01$. A post-hoc Tukey test ($\alpha = 0.05$) gives an HSD of 0.72 sec indicating the pairs (D, RH), (RH, RC) and (RC, F) are different. These overall results are shown in Figure 4. Visual inspection of the number of target crossings (Figure 4, left) indicates that slower techniques suffered from target crossing (overshooting the target). Interestingly, RC suffered badly from this issue, while RH did not. With RH, users are already poised to drop their foot, while RC users must recognize the end-point, raise and lower their foot to stop, or perform another kick to reduce scrolling speed.
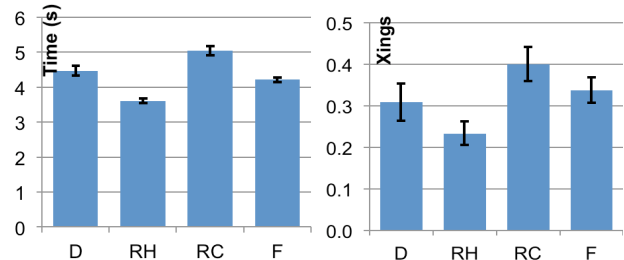

Figure 4: (left) Target acquisition times (sec) for different directions (right) Mean num. of crossings for target selection.
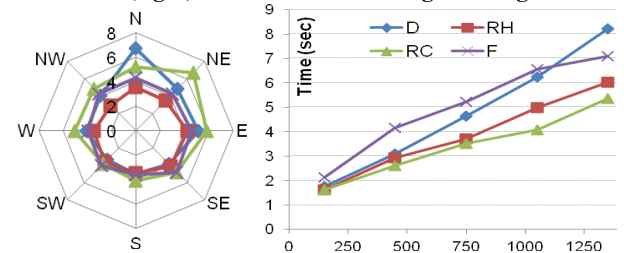

Figure 5: (left) Target acquisition times (sec) for different directions (right) Target acquisition times for different target distances (horizontal axis, px). Figures use the same legend.

However, participants took different times to complete target acquisition in the eight different directions. This is summarized in Figure 5. Participants took the longest to reach targets in the East (5.0 sec) and North-East (5.0 sec) directions, and were the quickest in the South (3.7 sec) and South-West (3.7 sec) directions. There was a significant difference in the target acquisition times of different directions ($F_{7,63} = 33.6$, $p < 0.001$). A post-hoc Tukey test ($\alpha = 0.05$) gives an HSD of 0.82 sec indicating the pairs (N, S), (N, SW), (N, NW), (NE, S), (NE, SW), (E, S), (E, SW), (E, NW) are different.

Overall, participants were more efficient at reaching targets that require kicks in the forward direction (recall in these studies a kick moves the content, equating to targets in the southern half of the compass). This is likely due to participants finding forward kicks more natural—e.g., kicking a ball is often performed with a forward kick.

Experiment 2: Kicking into the Distance
Experiment 1 examined how each of the mapping techniques fared when kicking in various directions. In this experiment we were interested in evaluating the effect of target distance on each of the four mapping techniques. This experiment was conducted using the same

experimental setup as outline in Experiment 1, with the differences noted below.

This study used a 4×5 factor within subjects design, with factors *Gesture Mapping* (D, RH, RC and F) and *Distance*. We selected five distances from the starting position: 150px, 450px, 750px, 1050px and 1350px. These positions represented a range of close and distant targets for a typical map navigation task.

All kicks were made in the same direction—towards a southern target (a forward kick) as this performed the best in experiment one. We ignored all non-north/south direction in participants' kicks. Each participant repeated the kick gestures for each technique/distance combination five times.

Participants
Ten volunteers (two female) aged between 23 and 26 participated in this study. All had previous experience with hand-based gestural interaction (such as using the Nintendo Wii) and two had previous foot-based gestural experience.

Results
We collected data on a total of 10 participants × 4 techniques × 5 distances × 5 repetitions = 1000 target acquisitions. All tasks were successfully completed.

Overall, RC was the fastest technique (mean time 3.4 sec, s.d. 1.3sec) and Flick the slowest (mean 5.0sec, s.d. 1.8sec). There was a significant difference between overall mean acquisition times ($F_{3,27} = 153.2$, $p < 0.01$); a post-hoc Tukey test gave an HSD of 0.53 ($\alpha = 0.05$) indicated the pairs (D, RH), (D, RC), (RH, F), (RC, F) were significantly different.
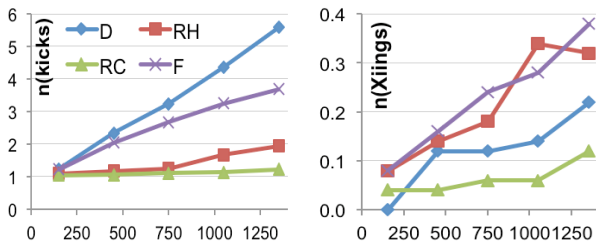

Figure 6: Mean number of kicks and target crossings for various target distances (on the horizontal axis in px). Both figures use the same legend.

Figure 5 shows the expected result of target acquisition time increasing with target distance. There was also a significant difference in target acquisition times for factor *Distance*, with closer targets always quicker than those further away ($F_{4,36} = 602.7$, $p < 0.01$). A post-hoc Tukey test gives an HSD of 0.63 ($\alpha = 0.05$) indicating that all distance pairs were significantly different.

To better understand the acquisition time differences between the four techniques we analysed the number of kicks users performed to reach the required target (see Figure 6, left). The RC technique rarely required more than a single kick to achieve the target, while the Displacment technique required (an almost) linearly increasing number

of kicks as the distance increases. Further analysis also reveals that RC had very few overruns of the targets, meaning time was not wasted 'hunting' for the target (Figure 6, right).

Summary
The primary finding from this study is that the rate-based continuous technique is most efficient at acquiring targets at a range of distances from the starting position. The rate-based hold technique also performed well for close targets.

Experiment 3: Putting it all Together
The pilot study and the two quantitative experiments have provided us with insights into various aspects of how kick gestures should be mapped to continuous interaction. In this final study, we wished to compare all aspects of the two best performing systems so far: rate-based continuous and rate-based hold. We also wished to gather subjective feedback on these two systems.

This experiment was conducted in the same manner as experiment 1 and 2, with the differences in experimental method listed here. This study used a 2×3×8 factor within subjects design, with factors *Interaction technique* (rate-based hold, rate-based continuous), *Distance* (150px, 750px and 1350px) and *Direction* (eight standard compass points). Each participant repeated the kick gestures for each {Technique, Distance, Direction} combination, three times.

Participants performed all gestures with one technique before moving to the next (order counter-balanced). The presentation of direction/distance pairs was randomised. After performing all kick gestures with one technique, participants completed a subjective evaluation. Overall preferences were gathered at the end of the evaluation.

Results
We collected a total of 1440 data points (10 participants × 2 techniques × 3 distances × 8 directions × 3 repetitions). All tasks were successfully completed. Overall, we found no significant difference in the task completion times between the two techniques. RH acquisitions took on average 3.8 sec (s.d. 1.7sec) and RC 3.7sec (s.d. 1.8sec). As observed in the previous two studies, both target distance ($p < 0.01$, $F_{2,18} = 4079.3$) and direction ($p < 0.01$, $F_{7,63} = 18.8$) yielded significantly different task completion times.
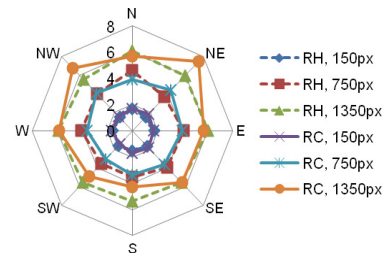

Figure 7: Mean target acquisition time (sec) for the two rate-based techniques

While both techniques had similar average task completion times, they showed weaknesses in different areas (Figure 7).

RH (dotted lines in Figure 7) is more consistent around all directions of the compass. Conversely, RC is faster in achieving southern targets but is generally slower for northern targets (especially those in the NW and NE directions). This skewedness correlates with observations during this and previous studies—participants find it harder to kick backwards than forward, with RC especially suffering. An analysis of the number of kicks and number of overshoots for the NW/N/NE and SW/S/SE sectors showed that both techniques had higher errors and overshoots when heading towards the northern regions.

Overall, seven of the ten participants preferred RC, two preferred RH and one was undecided. After using each system participants completed a NASA-TLX workload assessment form indicating their subjective impressions on 5-point Likert scales. Only the 'physical' category differed significantly between the two techniques: RH had a median ranking of 4.5 (indicating a high physical workload) and RC 3.5 (χ^2=6.4, p < 0.05). This feedback correlates with the requirement for the user to hold their leg in their air when navigation is taking place in the RH condition. User feedback indicated this 'mid-air' holding was tiring.

Discussion
The last experiment assessed the effectiveness of the two rate-based techniques that performed best in comparison to the displacement and flick techniques for distance and direction control. The results overall indicate preference and efficient interaction for rate-based continuous (RC), even though this required participants to move their leg and keep it in the air until they reached their target destination. This result was not echoed for all directions, particularly those requiring the users to kick backwards. As expected time to target grew with distance, but was still within reasonable limits (i.e. the linear growth not exponential).

DISCUSSION AND FUTURE WORK

Continuous vs. Discrete
The guessability study revealed a clear demarcation of gestures that were either discrete or continuous. Discrete gestures were mapped to specific tasks, such as locking or unlocking a mobile phone. Continuous gestures were mapped to tasks with a spatial component, such as moving in one direction in an information space. Such a distinction allows designers to separate their mappings into very clear camps. However, this also requires further design consideration for the continuous mapping, as we performed with the three studies outlined above.

We found good recognition rates of the user suggested gestures using a simple yet efficient recognizer. This suggests that while users were given full freedom for generating best mappings, there should not be issues implementing detection algorithms.

Our work did not fully investigate compound gestures, where a discrete gesture precedes or follows a continuous gesture or where a command is made up from multiple discrete or multiple continuous gestures. Investigating these combinations would prove interesting future work.

Rate-based Continuous Interaction
Our results suggest that for mapping continuous foot gestures to tasks, a rate-based approach works best. This technique results in far fewer kicks than the equivalent displacement based techniques. Users felt most in control, despite having to keep their feet in the air. This finding matches that for other finger-based interactions [1] where rate-based navigation is more efficient in a number of tasks. Finding that this also applies to continuous foot movement reaffirms the strength of rate-based techniques and their transferability to other forms of interaction.

Visual Feedback
In our experiments we assumed the users had visual information available to them at all times through the display of the phone. This we believe is a reasonable assumption. For example, when the user's hands are dirty or wet they can still hold the device on its rim without damaging or dirtying it. In these cases the user would rather use foot gestures to interact with the device than risk touching the screen with their wet or dirty hands.

There are scenarios where the phone is in the user's pocket and interaction is still required (e.g. when in a crowded metro with your hands holding the railing). Combing continuous foot interactions with other output modalities would provide another avenue for future work.

Design Implications
The results of the studies conducted in this paper allow us to draw a number of implications for the design of foot-based gestural interaction.

Context of use can ensure small gesture sets: small gesture sets are easily remembered by users and can be encouraged in foot-based interaction by using the device's context to eliminate ambiguity. The types of gestures should not be restricted as 'off the shelf' recognizers perform well.

Use Rate-Based Continuous for continuous interaction: this technique was subjectively preferred due to the less physical effort even though it suffered from greater acquisition times in the southern direction. Physical effort has the potential to alienate large groups of users, so should be minimized wherever possible.

Avoid backwards selection: experiments 1 and 3 showed that users struggle more with backwards kicks. Where possible, these should be avoided.

Use direction over distance: many designs require users to select from multiple choices. When using foot-based interaction, it is desirable to have these choices made around a compass point rather than by asking users to select items at a distance.

Further Considerations

This work focused on the selection of command-to-gesture mappings and defining parameters for continuous interaction. Real-world deployment of a foot-based gesture system would first require consideration of *context of use* and *social acceptability.* Context of use refers to the user's posture and movement during interaction. For experimental accuracy and reproducibility we only considered stationary users. A real world deployment would need to employ activity recognition algorithms to differentiate walking from phone-interaction gestures. Consideration must also be given to the social acceptability of kick-based gestures in various contexts. It is likely these would be unacceptable in some environments and so alternatives must be provided.

CONCLUSION

This paper examined the mapping of foot-based gestures to mobile device commands and deeply probed continuous interaction gestures. A guessability study derived a mapping for common commands that could be divided into continuous and discrete interaction. We validated that these gestures could be recognized. Through a series of three user studies we examined the most appropriate mechanism to support continuous interaction. Our results showed that users are faster, more accurate and prefer rate-based techniques over displacement based techniques. We believe these results provide a solid foundation for future investigations into the design of foot-gesture interaction techniques for novel mobile applications.

ACKNOWLEDGEMENTS

REFERENCES

1. Aliakseyeu, D., Irani, P., Lucero, A., Subramanian, S. Multi-flick: An Evaluation of Flick-based Scrolling Techniques for Pen Interfaces. *CHI '08*, 1689-1698.

2. Angel O. K., Chan, A., Ng., A., Luk, B. L. A Preliminary Analysis of Movement Times and Subjective Evaluations for a Visually-Controlled Foot-Tapping Task on Touch Pad Device, *IMECS '10*.

3. Augsten, T., Kaefer, K., Meusel, R., Fetzer, C., Kanitz, D., Stoff., T., Becker, T., Holz, C. and Baudisch, P. Multitoe: High-Precision Interaction with Back-Projected Floors Based on High-Resolution Multi-Touch Input. *UIST '10*, 209-218.

4. Crossan, A., S. Brewster, and A. Ng. Foot Tapping for Mobile Interaction. *BCS HCI '10.*

5. Grandjean, E. (1969) Fitting the task to the man: an ergonomic approach. Taylor and Francis, pp. 372.

6. Han, T., Alexander, J., Karnik, A., Irani, P., Subramanian, S. Kick: Investigating The Use Of Kick Gestures For Mobile Interactions. *MobileHCI '11.*

7. Hancock, M.S., Vernier, F., Wigdor, D., Carpendale, S., Chen, S. Rotation and Translation Mechanisms for Tabletop Interaction. *Tabletop '06*, 79-88.

8. Iso, T., and Yamazaki, K. Gait analyzer based on a cellphone with a single three-axis accelerometer. *MobileHCI '06*, 141-144.

9. Kratz, S. and M. Rohs. Protractor3D: A Closed-form Solution to Rotation-invariant 3D Gestures. *IUI '11.*

10. Liu, J., Pinelle, D., Gutwin, C., Subramanian, S. Improving digital handoff in shared tabletop workspaces. *Tabletop '08*, 9-16.

11. Paelke, V., Reimann, C., Stichling, D. Foot-based mobile interaction with games. *ACE '04*, 321-324.

12. Paelke, V., Reimann, C., Stichling, D. Kick-up menus. *CHI EA '04*, 1552.

13. Pakkanen, T., Raisamo, R. Appropriateness of foot interaction for non-accurate spatial tasks. *CHI EA '04*, 1123–1126.

14. Paradiso, J. A., Hsiao, K. Y., Benbasat, A. Interfacing to the foot: apparatus and applications. *CHI EA '00*, 175-176.

15. Paradiso J.A., et al. Interactive therapy with instrumented footwear. *CHI EA '04*, 1341–1343.

16. Pearson, G., Weiser, M. Exploratory Evaluations of Two Versions of a Foot-operated Cursor-positioning Device in a Target Selection Task, *SIGCHI Bulletin 1988*, Vol. 19, Issue 3, 70-75.

17. Rahman, M., Gustafson, S., Irani, P. and Subramanian, S. Tilt techniques: investigating the dexterity of wrist-based input. *CHI '09*, 1943-1952.

18. Rovers, A.F. and Van Essen, H.A. Guidelines for haptic interpersonal communication applications: an exploration of foot interaction styles. *Virtual Reality* 9, 2-3 (2006), 177-191.

19. Ruiz, J., Li, Y., Lank, E. User-defined motion gestures for mobile interaction. *CHI '11*, 197-206.

20. Schöning, J., Daiber, F., Kruger, A., Rohs, M. Using hands and feet to navigate and manipulate spatial data. *CHI EA '04*, 4663-4668.

21. Tetsuya, Y., Tsukamoto, M., Yoshihisa, T. Foot-Step Input Method for Operation Information Devices While Jogging. *SAINT '08*, 173-176.

22. Scott, J., Dearman, D., Truong, K. Sensing Foot Gestures from the Pocket. *UIST '10*, 199-208.

23. Winograd, T, Bringing Design to Software. Addison-Wesley, pp. 234.

24. Wobbrock, J.O., Morris, M.R., Wilson, A.D. User-Defined Gestures for Surface Computing. *CHI '09*, 1083-1092.