# Sensing Foot Gestures from the Pocket

*Jeremy Scott, David Dearman, Koji Yatani, and Khai N. Truong*
University of Toronto
Toronto, ON M5S 3G4, Canada
jer.scott@utoronto.ca, {dearman, koji, khai}@dgp.toronto.edu

**ABSTRACT**
Visually demanding interfaces on a mobile phone can diminish the user experience by monopolizing the user's attention when they are focusing on another task and impede accessibility for visually impaired users. Because mobile devices are often located in pockets when users are mobile, explicit foot movements can be defined as eyes-and-hands-free input gestures for interacting with the device. In this work, we study the human capability associated with performing foot-based interactions which involve lifting and rotation of the foot when pivoting on the toe and heel. Building upon these results, we then developed a system to learn and recognize foot gestures using a single commodity mobile phone placed in the user's pocket or in a holster on their hip. Our system uses acceleration data recorded by a built-in accelerometer on the mobile device and a machine learning approach to recognizing gestures. Through a lab study, we demonstrate that our system can classify ten different foot gestures at approximately 86% accuracy.

**Author Keywords:** Mobile devices, eyes-free interaction, hands-free interaction, foot-based gestures

**ACM Classification Keywords:** H5.2 [Information interfaces and presentation]: User Interfaces - Input devices and strategies.

**General Terms:** Design, Human Factors, Experimentation

**INTRODUCTION**
User input on a mobile device is typically supported through direct touch on a display or through physical buttons. Despite the variety of input methods, output still heavily relies on visual feedback. The visual demand of a mobile device's interface can diminish the user experience by monopolizing the user's attention when they are focusing on another task and impede accessibility for visually impaired users. In such cases, eyes-free interaction can improve the user experience.

Many eyes-free interaction techniques have been developed to improve access to and awareness of the features offered by a mobile device. These techniques use auditory or vibrotactile feedback as well as redesign conventional input techniques. Using different input modalities, such as speech [22, 33], or physical gestures performed by various parts of the body [15, 23, 30], are other approaches to eyes-free interaction. Because mobile devices now often have integrated sensors (*e.g.*, accelerometers), they can be leveraged to sense physical gestures performed by a user [10, 11]. These physical gestures can be used to provide eyes-free access to a device's features [19]; however, these gesture-based interfaces often still require the user to hold the device in her hand. In contrast, Patel *et al.* [26] used the placement of a mobile phone in a pocket or bag to develop a gesture-based authentication technique for securely connecting to other computers. Building on this premise, we envision that a mobile device placed in a pocket of the user's pants can recognize simple foot gestures using a built-in accelerometer. Although the foot does not offer the same precision and dexterity for selection as the wrist and hand, the foot is appropriate when used to perform simple coarse gestures [24, 28]. For example, the foot is a robust input method in a variety of applications (*e.g.*, driving, musical instruments, and audio transcription).

In this work, we first explore the foot-based interaction space through a study similar to the investigations of Crossan *et al.* [7] and Rahman *et al.* [29] for the wrist-based interaction space. We designed the first study to understand the human capabilities of performing foot-based interactions involving lifting and rotation of the leg when pivoting on the toe and heel. The results of this study show that:

- Users are more capable of performing accurate foot gestures with plantar flexion (heel lift with 6.31° error) than gestures with dorsiflexion (toe lift with 11.77° error). Selecting targets above 30° for gestures with lifting the toe resulted in a significant increase in error.
- Selection error is relatively consistent across all the gestures which involve rotating the foot when pivoting on the toe (8.55° error) and the heel (8.52° error). In addition, participants tended to overshoot smaller angle targets close to the starting position of their foot.
- Participants preferred gestures which involve rotating the toe and lifting the toe in terms of comfort.

Building upon the results of our first study, we developed a system to recognize foot gestures using a single built-in accelerometer on a commodity mobile phone placed in the user's pocket or in a holster on their hip. Our system takes a machine learning approach to recognizing gestures, using features extracted from acceleration data. In our second study, we determined that our system could classify ten different foot gestures at approximately 86% accuracy.


*UIST'10*, October 3–6, 2010, New York City, NY, USA.

## RELATED WORK

Mobile devices typically rely on visual feedback to provide an awareness of their available services and how to interact with the services. However, visual feedback on a mobile device can distract the user from their primary task and is a significant challenge for users with a visual impairment. In such cases, eyes-free interactions enabled through the use of audio, tactile feedback and physical gesturing can improve the user experience and allow for greater focus on their primary task. In this section, we review eyes-free interaction techniques for mobile devices and methods for gesture recognition using inertial sensors.

### Eyes-free Interaction for Mobile Devices

In some mobile applications, the complexity and variability of tasks are often low enough to be mapped to a small set of interactions. In these cases, visual feedback can be replaced with audio cues [36], tactile feedback [35], and the improved placement of physical buttons [20] for eyes-free interaction. A simple audio cue (*i.e.*, a click sound) along with an auditory depiction of menu items can assist the user in navigating and understanding the structure of a radial menu [36]. Vibrotactile feedback can inform the user as to the presence of on-screen objects and additional context information concerning interactions with the objects. Using five vibration motors positioned along the extremity and middle of a device, Yatani and Truong demonstrated that users could differentiate between ten vibration patterns [35]. Li *et al.*'s BlindSight placed buttons on the back of a phone to allow access to the phone's applications (*e.g.*, calendar) when the user is engaged in a conversation with the phone placed against her face [20]. The user could access information on the phone in parallel through audio feedback heard only by the user.

Eyes-free interaction can also leverage the most common method used for interpersonal communication: speech. Sawhney *et al.*'s nomadic radio allowed access to and control of a mobile device through spoken commands [33]. Although speech-based interaction is a common feature for most mobile devices [22], its use is not always socially appropriate in some situations, and the technology may fail to recognize commands in noisy environments [33].

The foot is an input mode that is used successfully in many different activities (*e.g.*, driving, playing musical instruments, and controlling the audio for transcription), but has not been explored heavily as an input channel for computers except in specific application domains such as physiotherapy [25], ambient awareness [32] and multi-modal input [8]. A problem associated with foot-based input is that the fine motor skills associated with the arm, hand and fingers cannot be similarly achieved by the foot in homing and selection tasks [24, 27, 28]. However, the foot is much quicker when performing simple coarse gestures [24, 28]. In a study comparing multiple techniques for parallel text entry and selection on mobile devices, Dearman *et al.* found the foot to be a fast medium for selecting on-screen widgets using a simple tap gesture on a foot-pedal [8]. In this work, we explore the foot-based interaction space with a similar systematic investigation to the wrist design space conducted by Crossan *et al.* [7] and Rahman *et al.* [29].

### Gesture Recognition for Mobile Devices

Sensors integrated within a mobile device can be used to support alternative modes of interaction [10, 11]. An accelerometer can be used to identify the angular orientation (*i.e.*, tilt) of the mobile device, and allows for gestural navigation and selection of menu items [11, 23]. Rekimoto's GestureWrist used a tilt sensor embedded in a wrist worn device to infer the orientation of the arm [30]. Similarly, Brewster *et al.* developed a 3D audio radial pie menu system that uses head gestures to select items positioned in a radial menu around the user's head [4] and tilt for target selection [6].

In general, accelerometers are used primarily to sense and infer a user's activity. Bao and Intille found that five accelerometers attached to different parts of the user's body could differentiate between 20 activities at 70-90% accuracy [1]. Similarly, Iso *et al*. used a commodity mobile device with a built-in accelerometer for detecting the user's gait and different activities, such as walking, running, and taking stairs, at 80% accuracy [14]. Although use of accelerometers offers promising results, the majority of projects exploring this space focus on a specific application domain: tapping and shaking of the leg [13], shaking of a device for authentication purposes [12, 26] and identifying incorrect movements for physiotherapy [25]. We extend their work by formally exploring the human capabilities for performing foot gestures and designing a gesture recognition system that utilizes only a mobile phone placed in a user's pocket.

## STUDY OF FOOT-BASED INTERACTION SPACE

We conducted a laboratory study to explore human capabilities and limitations involved with performing foot-based gestures. In particular, we examined foot gestures using a target selection task that required participants to select discrete targets along three axes of foot rotation.

### Axis of Rotation

The three axes of rotation evaluated include ankle, heel and toe. Based on results from an initial pilot study, we separated the rotation of the ankle into two distinct conditions in which the toe is either moving closer or farther away from the shin. The four experimental conditions include:

- *Dorsiflexion*: rotation of the ankle such that the angle between the shin and foot decreases (Figure 1a).
- *Plantar flexion*: rotation of the ankle such that the angle between the shin and foot increases (Figure 1b).
- *Heel rotation*: internal and external rotation of the foot and leg with respect to the midline of the body while pivoting the rotation on the heel of the foot (Figure 1c).
- *Toe rotation*: internal and external rotation of the foot and leg while pivoting the rotation on the toe of the foot (Figure 1d).

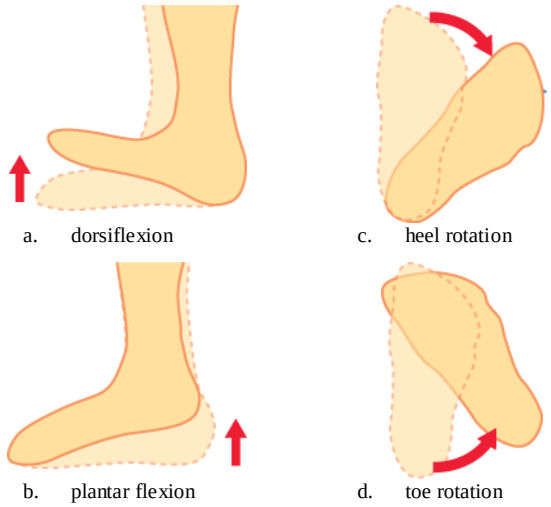



**Figure 1. The four conditions. The perspective for (c) and (d) are from above the right foot.**

We intentionally excluded another possible axis of rotation—the rolling of the foot around the vector from heel to toe—because of possible injuries to the foot and ankle as a result of that particular movement (*e.g.*, a sprain resulting from over-rolling one's ankle).

### Selection Task

We asked participants to perform a target selection task with their dominant foot while standing. All participants were right-footed. Targets were placed at 10° increments (Figure 2) and were selected by rotating the foot from the start position (0°) to the respective angle of the target along the condition's axis of rotation:

- *Dorsiflexion*: four targets placed between 10° and 40° inclusive (Figure 2a).
- *Plantar flexion*: six targets placed between 10° and 60° inclusive (Figure 2b).
- *Heel & toe rotation*: 21 targets (each), with 9 internal rotation targets placed between -10° and -90° inclusive, and 12 external rotation targets placed between 10° and 120° inclusive (Figure 2c and Figure 2d).

The range of selection (ROS) for each condition intentionally extends beyond the ranges of motion (ROM) for the ankle and leg [3, 21]. For example, the ROM for plantar flexion is 0° - 50°, but the ROS we evaluated is 10° - 60°. We intentionally extended the ROS beyond accepted ROM values because it is possible to demonstrate greater flexibility when standing by utilizing the knee, hip and lower back.

### Apparatus

We used six M-Series Vicon Motion Capture cameras to accurately capture and log the movement of a participant's foot. To account for the different sizes and shapes of each participant's foot, we developed a rigid foot model that was fitted on top of the right foot (Figure 3: top) and modeled in the Vicon IQ 2.0 software. The movement of the participant's foot was reported to a laptop running the experiment's

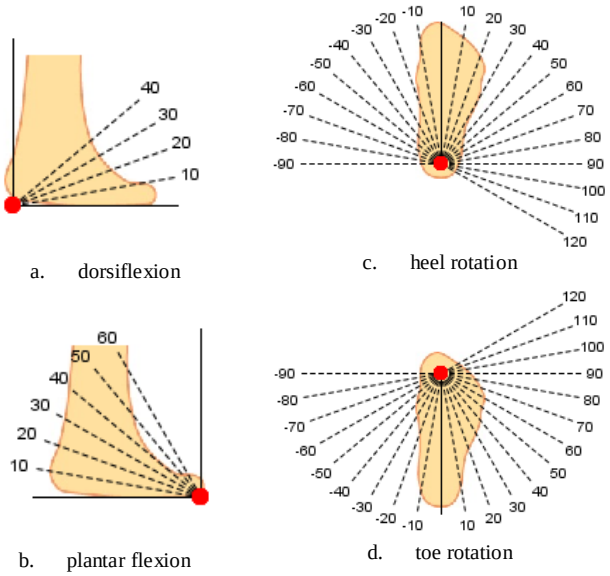



**Figure 2. Target placements along the axis of rotation (pivot-point: red dot) for (a) dorsiflexion, (b) plantar flexion, (c) heel rotation (top view), and (d) toe rotation (top view).**

software. The laptop was also used to direct the experiment, presenting to the participants the targets they were to select. The presentation for dorsiflexion and plantar flexion consisted of the first-quadrant of a circle with angular targets appearing as red lines from 10° to 60° (Figure 3: bottom). The presentation for heel rotation and toe rotation consisted of a semi-circle or inverted semi-circle (similar to Figure 2: right) with the angular targets appearing as red lines from -90° to 120°. Participants were given a wireless mouse that they used to indicate the start and end of a selection and to respond to the experiment software's prompts.

### Procedure and Design

Participants were asked to select targets presented on a laptop placed on a table in front of them. Before starting each

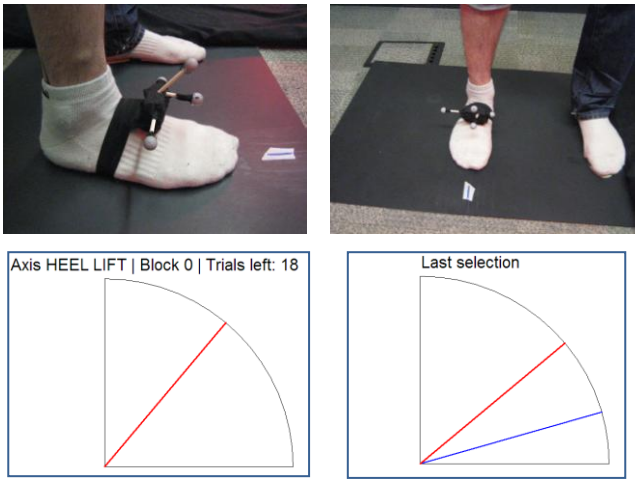



**Figure 3. The rigid foot model used to track the foot (top) and the visual feedback used to train participants (bottom) for the dorsiflexion and plantar flexion conditions.**

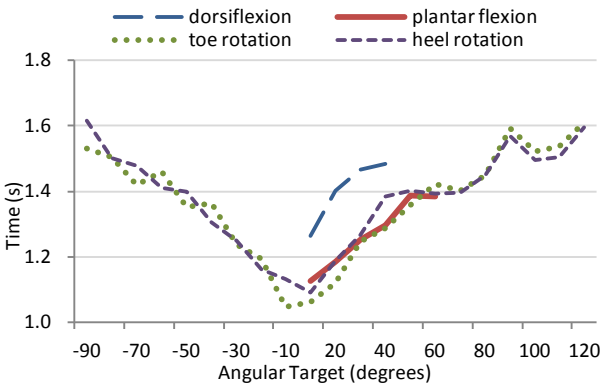



**Figure 4. Plot of the mean selection time for each of the four experimental conditions' target angles.**

trial, the participants were instructed to move their foot back to the origin, defined as 0° lift and 0° rotation with a ±2.5° threshold on each axis. Participants were then prompted to hold down the left-click button on a wireless mouse to begin the selection trial. After the button press, a red line appeared at an angular target, and the user moved her foot to the target angle along the axis instructed by the system. When making a selection, no visual feedback was given in order to simulate an eyes-free interaction. The participant released the mouse button to complete their selection when they believed that their foot was at the target angle.

The order of the four conditions was fully counterbalanced. Each condition included practice and testing phases with the practice phase conducted before the testing phase. Before starting the practice phase for each axis, we explained the condition and the range of targets to be selected. During the practice phase, participants were provided with visual feedback to confirm that they selected the target correctly. Each participant completed three practice selections for each angular target within a condition. After participants completed the practice phase, they started the testing phase. In the testing phase, participants performed three blocks of selection tasks. Each block consisted of three selections of each angular target without any visual feedback: participants completed 36 selections for dorsiflexion, 54 selections for plantar flexion, and 189 selections for both heel rotation and toe rotation. In total, each participant completed 156 selections in the training phase and 468 selections in the testing phase.

After all testing phases were completed, we conducted semi-structured interviews asking the participants to rank the gestures in order of preference and indicate their overall level of comfort on a 7-point Likert scale for each condition. The participants were thanked, but not compensated for their time.

### Participants

Sixteen right-footed participants (8 female and 8 male) recruited from our university and its surrounding community took part in the study. The age of participants varied between 20 and 29 years of age, with a median age of 24. Because range of motion changes with age, we studied only individuals belonging to the same age group. All the participants were screened to ensure that they had no current or prior injuries that would limit the range of motion for their right foot, ankle, leg and lower back. We further confirmed the footedness and normal range of motion (ROM) for all participants along each axis using a medical standard goniometer.

## RESULTS

Statistical analyses of the four experimental conditions were conducted independently of one another. Failed trials due to clicking mistakes and trials that exceeded three standard deviations from the mean selection error were removed as outliers—2.3% of trials were removed from the analysis. The analysis of the selection error was conducted on the absolute median error using repeated measures ANOVA; we applied the Greenhouse-Geisser correction when sphericity was violated. Event-count measures, such as the number of overshot and undershot selections, were analyzed using nonparametric Friedman tests. Post-hoc pair-wise comparison of event-count measures were conducted using the Wilcoxon signed-rank test with the Holm-Bonferroni correction.

### Target Selection Time

Targets closer to the origin were selected more quickly than targets at the extremity of the range of selection (see Figure 4). Our analysis revealed a significant main effect of target angles on the selection time for all four conditions: dorsiflexion ($F_{1.94,27.13}$=10.56, $p$<0.001), plantar flexion ($F_{2.58,36.05}$=16.07, $p$<0.001), toe rotation ($F_{20,280}$=27.48, $p$<0.001), and heel rotation ($F_{20,280}$=24.78, $p$<0.001). There is no observed difference in selection time by block.

The selection time for the 10° target for dorsiflexion was significantly faster than the other targets (all at $p$<0.05). The selection time for the 10° and 20° targets for plantar flexion was faster than the 40° (p<0.05), 50°, and 60° targets (both at $p$<0.005). We do not report the detailed results of our post-hoc pairwise comparisons for the toe rotation and heel rotation targets due to the limited space. However, as presented in Figure 4, we found that the selection time for these two rotations proportionally increased as the target angle became larger. For example, the -90° target for toe rotation was significantly slower than all the targets between -50° and 40° (inclusive; all at $p$<0.05), and the -90° for heel rotation was significantly slower than all the targets between -60° and 40° (inclusive; all at $p$<0.05).

### Target Selection Error

The median selection error across all targets was 11.77° for dorsiflexion, 6.31° for plantar flexion, 8.55° for toe rotation, and 8.52° for heel rotation. The mean selection error and standard deviation are illustrated in Figure 5.

Our analysis did not reveal a significant main effect of target angles on the selection error, except for dorsiflexion ($F_{1.3,17.2}$=237.96, $p$<0.001). A closer examination of each target selection error for dorsiflexion revealed that participants were significantly more accurate when selecting the 10° (2.91° ± 2.16°) and 20° (3.73° ± 3.90°) targets, $p$<0.001, and significantly less accurate when selecting the

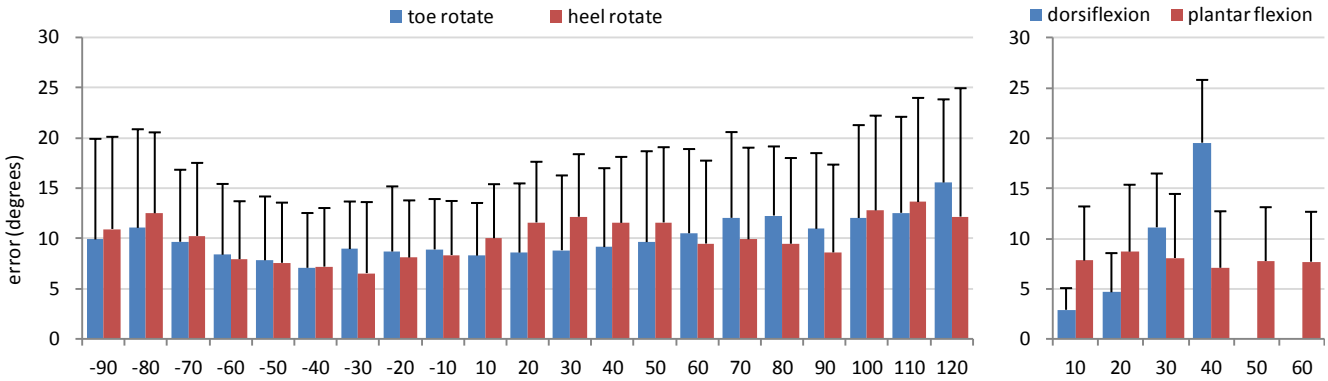



**Figure 5. The mean and standard deviation selection error (degrees) for each of the four experimental conditions' target angles (x-axis). Presented left is toe rotation and heel rotation, and right is dorsiflexion and plantar flexion.**

30° (11.34° ± 5.35°) and 40° (29.82° ± 6.58°) targets, $p<0.005$. Note that the 30° and 40° targets were located at the upper bound of the typical range of motion (0° – 30°) for dorsiflexion. However, a similar effect for targets outside the typical range of motion (ROM) impacting selection accuracy was not observed for the other three conditions. For example, there was no observed difference in the selection error for any target in the plantar flexion condition ($p=0.84$), despite the range of selection (ROS) between 10° – 60° extending beyond the typical ROM (0° – 50°). We postulate that because we did not enforce strict rotation around the ankle only, plantar flexion and the two leg rotation conditions were able to extend beyond their typical ROM because of additional rotation in the knee, hip and lower back. Dorsiflexion did not benefit from additional freedom in these joints, because more rotation would require the knee to inflect in a physically impossible way.

Our analysis also revealed a decrease in accuracy across the blocks. We observed a significant main effect of trial block on the selection error for the dorsiflexion ($F_{1.5,19.3}=13.25$, $p<0.001$), and toe rotation ($F_{2,26}=10.05$, $p<0.001$) conditions. With respect to the dorsiflexion, participants were significantly more accurate in the first block of trials than the second ($p<0.05$), and third block ($p<0.005$). A similar trend was observed for heel rotation, but only for the first and second block ($p<0.005$). We postulate that the effect of the block on the selection error was the result of fatigue. Many of the participants expressed being tired after completing all trials. In addition, it is possible that participants may have become increasingly less accurate with each subsequent trial because they did not receive enough training to develop the kinesthetic memory needed for consistently selecting targets accurately.

### Numbers of Overshot and Undershot Selections

We counted the number of overshot and undershot selections by using the absolute values of the target angle and actual selection. For example, for a 20° target, a selection of 30° is a 10° overshoot. Similarly, for a -20° target, a selection of -30° is a 10° overshoot. There was a significant main effect of target angle on the number of overshot and undershot selections with respect to dorsiflexion ($\chi^2_{(1,N=14)}=14.0$, $p<0.001$), toe rotation ($\chi^2_{(1,N=14)}=4.57$, $p<0.05$), and heel rotation ($\chi^2_{(1,N=14)}=7.14$, $p<0.01$). No significant difference was observed for plantar flexion.

In all conditions except dorsiflexion, we observed that participants tended to overshoot the small angular targets more frequently than the large angular targets. For plantar flexion, participants significantly overshot the 10° and 20° targets (both at $p<0.001$), and the 30° target ($p<0.05$). For toe rotation, participants significantly overshot the targets between -20° and 20° (all at $p<0.001$), and the -30° and 30° targets (both at $p<0.005$). For heel rotation, participants overshot fewer targets for internal rotation (-10°) and more targets for external rotation (10°, 20°, 30°, and 40°) in comparison to toe rotation (all at $p<0.001$).

In the dorsiflexion condition, participants significantly undershot all of the targets ($p<0.001$) except the 10° target, for which no significant difference was observed. Although different than the other three conditions, this result is in line with our previous finding concerning the ROM for dorsiflexion. The limited ROM for dorsiflexion resulted in participants underestimating the 20° and 30° targets (which lie within the ROM) and not being able to reach the targets beyond the ROM, which resulted in the larger error values.

### User Preference

In the post-experimental interview, participants identified heel rotation as the most comfortable gesture (5.8 on 7 point Likert scale with 7 for strongly comfortable gestures and 1 for strongly uncomfortable gestures; SD = 0.4), followed by plantar flexion (4.9, SD = 0.4), toe rotation (4.5, SD = 0.5) and dorsiflexion (2.58, SD = 0.6).

For heel rotation, external rotation was preferred over internal rotation. For toe rotation, internal rotation was preferred over external rotation. Although both movements focus around a different pivot point, they are effectively the same: the toe always points external to the body. Participants expressed less comfort for large angles, but clarified that rotating the foot to the targets smaller than 100° was not difficult.

**Figure 6. Three placements of the mobile phone: front right pocket (left), right side hip-mounted (middle) and back right pocket (right), pinned to maintain consistent orientation.**

For plantar flexion, there was no difference across all the targets; however, there was a slight decrease for the 50° and 60° targets. For dorsiflexion, participants expressed strong discomfort for targets above 30°. These targets were beyond their range of motion and therefore difficult to select.

### Summary of Results

Users are more capable of performing consistent, accurate foot gestures with plantar flexion than dorsiflexion. The median selection error for plantar flexion (6.31°) was lower than the error for dorsiflexion (11.77°) and participants expressed greater comfort for a larger range of motion. In addition, the error for plantar flexion was relatively consistent across the target angles we tested, whereas the median error for dorsiflexion grows significantly from 3.73° for the 20° target, to 11.13° for the 30° target and 19.56° for 40° target. The growth in error limits the gestures that can be defined for dorsiflexion. However, the gesture space for plantar flexion is more expressive because it offers a greater range of movement that is easily and comfortably accessible. Informed by the observed error, we believe that plantar flexion can be comfortably segmented into three distinct gestures: low (0° - 20°), middle (20° - 50°), and high (> 50°). We chose the above range between gestures because it segments the selectable space into regions that accommodate two times the median selection error, and includes additional flexibility for instances when the participants overshoot.

In addition, we found that the selection error for toe rotation (8.55°) and heel rotation (8.52°) were comparable, but participants expressed greater comfort when performing the heel rotations. Across both conditions, participants expressed difficulty when selecting targets beyond -90° and 90°, and were significantly more likely to overshoot smaller targets between -40° and 40°. Informed by these results, it is evident that both heel and toe rotation offer similar performance, but heel rotation is a more preferred gesture. Similar to plantar flexion, we believe that both toe and heel rotation can be comfortably segmented into three distinct gestures for both internal and external rotation: low (internal: -25° to -0°; external: 0° to 25°), middle (internal: -60° to -25°; external 25° to 60°) and high (internal: less than -60°; external: more than 60°). The low and high ranges capture two times the median selection error for these conditions, and the middle range is broad enough to encapsulate the transition between regions where there is equal probability for over- and undershooting a selection.

## SENSING FOOT GESTURES WITH A MOBILE DEVICE

The Vicon Motion Capture technology we used to accurately measure the participants' foot movements is not practical in a mobile setting. Accelerometers, however, are a common sensor available in most commodity mobile devices, and have been used successfully to support gesture [13, 26] and activity recognition [1, 14]. In this section, we describe an implementation and evaluation of a technique to recognize foot gestures using only the internal accelerometer of a mobile phone placed in a pant pocket or a hip-mounted holster. The technique allows for hands-free and eyes-free interaction without requiring additional sensors. We envision that the user would perform a foot gesture as follows:

1. The user carries their 3-axis accelerometer-equipped mobile phone in a pocket of their pants.
2. The user initiates a foot gesture by positioning their foot at the origin (flat and facing forward) and performing a quick double-tap with their foot.
3. The user performs a foot gesture along a single axis stopping at the desired angular target. The phone recognizes the gesture, and executes a linked command.

### Preliminary Study of Features for Machine Learning

We used machine learning to recognize the foot gestures. For this purpose, we first determined features which might be useful for gesture classification. During the first experiment, we placed three iPhone devices in the participants' pockets: one in the front right, one in the back right and one in a pouch safety-pinned to the side of the user's leg to emulate a hip-mounted phone holster (Figure 6). All phones were initially oriented upwards such that the iPhone screen faced the user's body. Each device logged acceleration from its built-in 3-axis accelerometer while the participants were performing the first study's target selection tasks. The built-in accelerometer could record acceleration within ±2g at a 50 Hz sampling rate.

### Classification Features

Examination of the acceleration data revealed 34 features that we could utilize in building the gesture classifier.

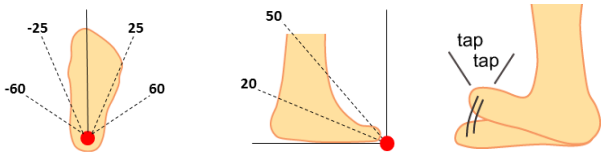
#### *Time-domain Features*

In the time domain, we computed the *mean*, *standard deviation*, *minimum*, *maximum*, and the interval time between the minimum and maximum points of accelerations (*max-min interval*) for each accelerometer axis.

#### *Frequency-domain Features*

In the frequency domain, we first computed a 64-point FFT over the acceleration data from each trial. If the length of the windowed sample was less than 64 points, the samples were zero-padded before computing the FFT. We then calculated the following features:

- *FFT Coefficient Values:* every 4$^{th}$ FFT coefficient.




a. Heel rotation b. Plantar flexion c. Double-tap

**Figure 7. The gesture set for the second experiment.**

- *Max DFT Value and Index:* the maximum FFT coefficient and its index in the 64-point FFT, excluding the DC component.
- *Spectral Energy*: the sum of FFT coefficients.

Upon examining the impact of these features across all axes of rotation, we used only the Y-axis because no noticeable difference was observed in these features for the other axes.

### Gesture Space

Based on the design implications from the first study, we focused specifically on plantar flexion and heel rotation gestures. We chose not to include dorsiflexion because of the limited gesture space resulting from the high degree of error. In addition, we introduced a double tap gesture that can be used to initiate a gesture. Informed by the selection error we outlined in the summary of results for the previous study, we chose the following gesture set (Figure 7):

- *Heel rotation*: internal high (-Hi: less than -60°), middle (-Md: -60° to -25°), and low (-Lo: -25° to 0°) and external low (+Lo: 0° to 25°), middle (+Md: 25° to 60°), and high (+Hi: more than 60°).
- *Plantar flexion:* low (0° to 20°), middle (20° to 50°), and high (more than 50°).

### Classification Algorithm

We used Naïve Bayes to perform our classification of the user's foot movements. Naïve Bayes is based on the Bayesian theorem [2], and assumes that features are conditionally independent. This is often a strong assumption for problems like ours, but it is also commonly acknowledged that Naïve Bayes still performs well even for such problems. Naïve Bayes is not necessarily a strong classifier compared to other supervised learning methods, like *k*-Nearest Neighbor and Support Vector Machine [5]. However, the time complexity of a Naïve Bayes classification is small for both training and testing, and it is generally faster than other machine learning methods [34]. This is an advantage when the system has limited computational capacities and requires real-time recognition. For example, Korpipää *et al.* have previously discussed the feasibility of using Naïve Bayes in a mobile system to recognize user contexts [16].

We used a single Gaussian distribution for each feature to accommodate for the continuous values calculated from the acceleration data. We implemented our classifier using the WEKA Data Mining Software package [9].

## CLASSIFICATION TEST

To evaluate how accurately we could recognize foot gestures with accelerometers embedded in a mobile device, we conducted another laboratory study. This study aimed to gather acceleration data for each foot gesture described in the previous section, and to classify gestures using a Naïve Bayes classifier.

### Data Gathering Procedure and Apparatus

Each participant was asked to complete a series of foot-gesture tasks as quickly as possible without any kind of feedback. We used the same procedure and apparatus as in the first experiment with the targets described above. We also placed three iPhones (Figure 6) in their front right pocket (*front*), back right pocket (*back*), and in a pouch securely fastened to the side of their upper right leg (*side*). We used Vicon Motion Capture cameras to measure the location of the participant's foot to ensure it was positioned at the origin.

### Design

The presentation order of the three types of gestures (*Heel rotation*, *Plantar flexion*, and *Double tap)* was fully counterbalanced in our evaluation. As in the previous study, each condition included a practice phase before the testing phase. Before beginning the practice phase, we explained the gesture type and the range of targets to be tested. During the practice phase for *Heel rotation* gestures, visual cues for the *low*, *middle*, and *high* regions were provided on the floor. During the practice phase for *plantar flexion* gestures, a diagram of the regions was propped vertically against the right side of the foot. These aids were removed during the testing phase. For each gesture type, each target was randomly presented 10 times for training and 50 times for testing. In total, each participant completed 100 gestures in the practice phase and 500 gestures in the testing phase.

### Participants

Six right-footed participants (two female and four male) were recruited from our university. The age of participants varied between 20 and 32 years of age, with a median age of 23. Again, participants were screened to ensure they had no current or prior injuries that would limit the range of motion for their right foot. Participants were also asked to wear a pair of jeans with pockets to hold the iPhones.

| | *All Foot Gestures* | | | | *Between Gesture Types* | | | | *Heel Rotation Gestures* | | | | *Plantar Flexion Gestures* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LOPO | | *WP* | | *LOPO* | | *WP* | | *LOPO* | | *WP* | | *LOPO* | | *WP* | |
| | *Avg.* | *SD* | *Avg.* | *SD* | *Avg.* | *SD* | *Avg.* | *SD* | *Avg.* | *SD* | *Avg.* | *SD* | *Avg.* | *SD* | *Avg.* | *SD* |
| *Front pocket* | 17.5% | 4.5% | 60.2% | 26.9% | 50.4% | 9.0% | 90.3% | 8.6% | 24.8% | 6.0% | 50.8% | 27.5% | 36.3% | 3.0% | 79.1% | 14.5% |
| *Back pocket* | 12.7% | 1.3% | 32.0% | 8.0% | 48.3% | 11.5% | 75.5% | 8.0% | 18.8% | 0.6% | 34.9% | 7.8% | 44.9% | 6.8% | 43.8% | 3.9% |
| *Side pocket* | 30.2% | 7.9% | 85.7% | 4.5% | 62.3% | 12.1% | 92.2% | 3.2% | 39.8% | 10.4% | 86.4% | 5.6% | 53.6% | 16.1% | 82.3% | 5.0% |

**Table 1. Average and standard deviation in classification accuracies using Naïve Bayes based on the three different iPhone locations, for *Leave-one-participant-out (LOPO) Cross-Validation* and *Within Participants (WP) Stratified Cross-Validation.***

| | | Target Range | Avg. | SD |
|---|---|---|---|---|
| heel rotate | –Hi | < -60° | -71.5 | 14.5 |
| | –Md | -60° – -25° | -41.0 | 11.8 |
| | –Lo | -25° – 0° | -18.3 | 8.5 |
| | +Lo | 0° – 25° | 18.2 | 6.4 |
| | +Md | 25° – 60° | 51.0 | 10.1 |
| | +Hi | > 60° | 94.3 | 16.8 |
| plantar flexion | Lo | 0° – 20° | 17.2 | 6.2 |
| | Md | 20° – 50° | 40.3 | 9.3 |
| | Hi | > 50° | 62.4 | 12.0 |

**Table 2. The target range angle for each gesture and the participants' average angle of selection.**

### Classification Procedure

After gathering the acceleration data, we conducted four classification tests: across all gestures, across all gesture types, across all gestures in *Heel rotation*, and across all gestures in *Plantar flexion*. We used the following cross validation protocols for training and testing**:**

- *Leave-one-participant-out (LOPO) cross-validation*: We used the data gathered from 5 of the 6 participants for training, and used the data from the other participant for testing. This was repeated such that each participant's data was used once for validation. This method of validation assumes weak user-dependency.
- *Within-participant (WP) stratified cross-validation*: We used data from only one participant at a time. The data for each participant was split into 10 stratified folds, meaning the ratio of data from each class was equal to the ratio in the total dataset. Using one fold for testing and the other 9 folds for training, tests were repeated such that each fold was used once for testing. The results were then averaged across tests for each participant and summed across participants. This protocol assumes a stronger user-dependency than *LOPO*.

## CLASSIFICATION RESULTS

In most of the classification tests, the *WP* protocol yielded greater accuracy than the *LOPO* protocol (see Table 1). Because of the disparity between the two methods, we only report the results of the *WP* protocol in this section.

### Effect of the Device Placement

The *side* placement of the mobile device resulted in a greater overall accuracy than the other *front* and *back* pocket placements. When simultaneously classifying trials across all types and angles of foot gestures, the Naïve Bayes classifier yielded the highest average accuracy (85.7%, SD 4.5%) across participants when using features extracted from the hip-mounted *side* iPhone. The *front* pocket iPhone gave the next best results (60.2%, SD 26.9%) followed by the *back* pocket iPhone (32.0%, SD 8.0%). Both types of within-gesture-type testing showed a similar trend across sensor locations: *side:* 86.4%, *front*: 50.8%, and *back:* 34.9% for heel rotation gestures; and *side:* 82.3%, *front*: 79.1%, and *back*: 43.8% for plantar flexion gestures.

### Foot Gesture Recognition Accuracy

Table 3 shows the confusion matrix of classification across all gestures using the acceleration data gathered from the *side* mobile device. It reveals that gestures were generally confused with other gestures in the same gesture type, typically between adjacent angular regions.

In contrast, classification of gesture type without considering the target angles (Table 1) resulted in considerably higher classification accuracy across the front and back-pocketed iPhones (*side*: 92.2%, *front*: 90.3%, and *back*: 75.5%).

## DISCUSSION

Naïve Bayes resulted in 82-92% classification accuracy for the gesture space we suggested in the capabilities evaluation with the mobile device attached to the side of the user's leg. This result shows that hands-free and eyes-free interaction for mobile devices is possible using foot gestures recognized by the device's integrated accelerometer. In addition, the level of recognition conforms to the human performance limitations we identified. Although the acceleration profiles used when classifying a gesture do not fully conform to the gesture ranges we define previously, we show (Table 2) a consistent mapping between the two, validating the ability of the accelerometer to sense the gesture space.

Classification with the WP protocol yielded higher accuracies than tests on the LOPO protocol. This implies that foot gestures detected by a built-in accelerometer on a mobile device may have user-dependency. To address this issue, we can train the classifier by asking the mobile device user to perform sample foot gestures when calibrating their device on first use. Completing this calibration stage, similar to how mobile devices used to require a user to calibrate the touch screen, will ensure more accurate recognition. Mobile devices are usually owned and used by a single person. Thus, a user can train the classifier on her personal device to ensure more accurate recognition of these foot gestures.

The results also indicate that the form factor of the pant pocket can influence classification accuracy. If the phone is allowed to shift in the user's pocket then its acceleration profile may vary over time, significantly impacting performance. In contrast, the hip mounted *side* position is not subject to displacement within the secure holster and is also tied to upper leg movement. Thus it performed the best among our three device placements to detect foot gestures.

| | Heel Rotation Gestures | | | | | | Plantar Flexion Gestures | | | Dbl Tap |
|---|---|---|---|---|---|---|---|---|---|---|
| | –Hi | –Md | –Lo | +Lo | +Md | +Hi | Lo | Md | Hi | |
| HR –Hi | 286 | 11 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| HR –Md | 11 | 242 | 45 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| HR –Lo | 2 | 39 | 245 | 1 | 0 | 0 | 8 | 0 | 0 | 5 |
| HR +Lo | 0 | 1 | 3 | 266 | 28 | 0 | 0 | 1 | 0 | 1 |
| HR +Md | 0 | 0 | 1 | 33 | 242 | 13 | 2 | 7 | 2 | 0 |
| HR +Hi | 0 | 0 | 0 | 1 | 20 | 273 | 0 | 1 | 5 | 0 |
| PF Lo | 0 | 0 | 12 | 7 | 8 | 2 | 242 | 26 | 0 | 3 |
| PF Md | 0 | 0 | 3 | 1 | 1 | 1 | 22 | 238 | 34 | 0 |
| PF Hi | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 24 | 273 | 0 |
| Dbl Tap | 1 | 2 | 5 | 0 | 1 | 0 | 4 | 0 | 0 | 287 |

**Table 3. The confusion matrix for Naïve Bayes classification across all possible foot gestures based on the *side* iPhone, using the *Within Participant* CV protocol.**

## LIMITATIONS OF THE CURRENT SYSTEM

There are several issues that need to be considered in a practical implementation of a foot gesture recognition system. During the second study, the placement and posture of the mobile devices were fixed in the participants' pocket, which would not always be the case in a real-world setting. This issue can be mitigated in two ways. First, it is possible to identify the position of an accelerometer on the body [17] and compensate for when the sensor drifts [18]. Second, the system can automatically calibrate the acceleration data by sampling the phone's orientation with respect to gravity before the user makes a double tap gesture and calculating a rotation matrix for normalization.

Differentiating foot gestures from other activities like walking and running could be another issue. We believe that gesture initialization with a distinct movement would mitigate this. In our second study, we observed that the acceleration profile for a double-tap had two small spikes separated by ~322 msec, while spikes from walking are separated by~800 msec. In order to validate that these distinct acceleration profiles could be differentiated by our classifier, we ran a short study in which three participants took a casual stroll, which included walking on flat ground and on stairs. Each carried an iPhone in their front right pocket. The acceleration data was processed offline using 40-sample windows with 10 samples of overlap, with one minute of data appended to the foot gesture profiles from the second study in order to create a training set with walking. Testing on the remaining walking data (~1744 windows of samples per participant), the classifier reported 5.3 false positives on average (0.3%). More research is needed to produce a quick way of undoing false positives when they do arise.

Our classifier currently does not run on a mobile device and therefore does not recognize the user's foot gesture in real time. Improving the performance time of gesture recognition may be necessary, particularly for interactive systems. One approach to improving the performance time of recognition is to reduce the number of features. By using Principal Component Analysis, we can compress the current feature set without significantly reducing performance.

| | *Between Gesture Types* | | | | *Plantar Flexion Gestures* | | |
|---|---|---|---|---|---|---|---|
| | *HR* | *PF* | *Tap* | | *Lo* | *Md* | *Hi* |
| *HR* | 1683 | 92 | 25 | Lo | 271 | 28 | 1 |
| *PF* | 91 | 796 | 13 | Md | 24 | 243 | 33 |
| *Tap* | 12 | 2 | 286 | Hi | 1 | 20 | 279 |

| | *Heel Rotation Gestures* | | | | | |
|---|---|---|---|---|---|---|
| | *–Hi* | *–Md* | *–Lo* | *+Lo* | *+Md* | *+Hi* |
| *–Hi* | 285 | 12 | 3 | 0 | 0 | 0 |
| *–Md* | 15 | 238 | 45 | 2 | 0 | 0 |
| *–Lo* | 2 | 40 | 252 | 4 | 2 | 0 |
| *+Lo* | 0 | 1 | 5 | 263 | 31 | 0 |
| *+Md* | 0 | 0 | 2 | 36 | 242 | 20 |
| *+Hi* | 1 | 0 | 0 | 1 | 22 | 276 |

**Table 4. Confusion matrices for Naïve Bayes classification across gesture types and within the two axes of rotation, based on the *side* iPhone, using the *WPCV* protocol.**

## CONCLUSION AND FUTURE WORK

Current user interfaces on mobile devices demand a large amount of visual and cognitive attention from the user. Considering the tendency for primary, real-world tasks to be interrupted by simple yet obtrusive interactions on mobile phones, the development of eyes-and-hands-free interaction techniques can greatly enhance user experience and lower accessibility barriers for the visually impaired.

In this paper, we described the use of foot gestures as eyes-and-hands-free input for mobile devices. We conducted a controlled study to examine human capabilities for performing foot-based interactions. Participants were able to perform four kinds of rotational foot gestures with approximately 10° error, but preferred two in particular: heel rotation and plantar flexion. Building on this result, we developed a system to observe and infer a user's foot gesture using only an accelerometer-enabled commodity mobile phone placed in a pant pocket or hip-mounted holster. With the Naïve Bayes method, the system could achieve a classification of ten foot gestures at approximately 86% accuracy. With this system, it is possible to support scenarios such as allowing a user who is standing at the bus stop with both hands engaged (*e.g.*, carrying bags) to easily switch to the next song on her iPhone without needing to free one hand, pull out the device and explicitly interact with it.

In future work, we will implement a classifier on a mobile device and build a real-time foot gesture recognition system. We are also interested in examining the performance of foot gesture recognition and the acceptability of these foot gestures in more naturalistic settings. We plan to conduct a deployment study for this purpose. As done in Bao and Intille's study [1], we will gather the acceleration data from the day-to-day lives of users and analyze how accurately the system can recognize foot gestures. Rico and Brewster previously examined the social acceptability of various gestures, including foot tapping [31]. They found that participants considered foot tapping socially acceptable because it only involves subtle movements. The foot gestures we explored are not so different from foot tapping, and we believe that they could also be considered acceptable and will evaluate this in the future.


## ACKNOWLEDGMENTS

We thank Julie Kientz, Frank Li, and Alyssa Rosenzweig for their feedback on this paper, and Robert Tillman and Mustafa Bilgic for their advice on machine learning. We also thank Frank for his expertise with the iPhone.


## REFERENCES


1. Bao, L. and Intille, S.S. Activity recognition from user-annotated acceleration data. In *Proc.* Pervasive 2004, Springer (2004), 1-17.
2. Bishop, C. *Pattern recognition and machine learning*. Springer Science+Business Media, LLC, New York, NY, USA, 2006.
3. Boone, D.C. and Azen, S.P. Normal range of motion of joints in male subjects. *The Applied Journal of Bone and Joint Surgery, 61*(A), 756-759.

4. Brewster, S., Lumsden, J., Bell, M., Hall, M. and Tasker, S. Multimodal 'eyes-free' interaction techniques for wearable devices. In *Proc.* CHI 2003, ACM (2003), 473- 480.
5. Caruana R., and Niculescu-Mizil, A. An empirical comparison of supervised learning algorithm. In *Proc. ICML 2006*, ACM (2006), 161-168.
6. Crossan, A., McGill, M., Brewster, S., and Murray-Smith, R. 2009. Head tilting for interaction in mobile contexts. In *Proc.* MobileHCI 2009. ACM (2009), 1-10.
7. Crossan, A, Williamson, J., Brewster, S., Murray-Smith, R. Wrist rotation for interaction in mobile contexts. In *Proc.* MobileHCI 2008, ACM (2008), 435-438.
8. Dearman, D., Karlson, A., Meyers, B. and Bederson, B. Multi-modal text entry and selection on a mobile device. In *Proc.* GI 2010, ACM (2010), 19-26.
9. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H. The WEKA Data Mining Software: An Update; SIGKDD Explorations 11 (1), 2009.
10. Harrison, B.L., Fishkin, K.P., Gujar, A, Mochon, C. and Want, R. Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. In *Proc*. CHI 1998, ACM (1998), 17-24.
11. Hinkley, K., Pierce, J. and Horvitz, E. Sensing techniques for mobile interaction. In *Proc.* UIST 2000, ACM (2000), 91-100.
12. Holmquist, L.E., Mattern, F., Schiele, B., Alahuhta, P., Beigl, M. and Gellersen, H.W. SmartIts friends: a technique for users to easily establish connections between smart artifacts. In *Proc.* Ubicomp 2001, Springer (2001), 116-122.
13. Hudson S. E., Harrison, C., Harrison, B. L., and LaMarca, A. Whack gestures: inexact and inattentive interaction with mobile devices. In *Proc TEI 2010*, ACM (2010), 109-112.
14. Iso, T., and Yamazaki, K. Gait analyzer based on a cell-phone with a single three-axis accelerometer. In Proc. *MobileHCI 2006*, ACM (2006), 141-144.
15. Kajastila, R. and Lokki, T. A gesture-based and eyes-free control method for mobile devices. *Ext. Abs.* CHI 2009, ACM (2009), 3559-3564.
16. Korpipää, P., Koskinen, M., Peltola, J., Mäkelä, S., and Seppänen, T. 2003. Bayesian approach to sensor-based context awareness. *Personal Ubiquitous Comput.* 7, 2 (Jul. 2003), 113-124.
17. Kunze, K., Lukowicz, P., Junker, H. and Troster, G. Where am I: recognizing on-body positions of wearable sensors. In *Proc*. LOCA 2005, Springer (2005), 264-275.
18. Kunze, K. and Lukowicz, P. Dealing with sensor displacement in motion-based on body activity recognition systems. In *Proc*. Ubicomp 2008, ACM (2008), 20-29.
19. Li, F.C.Y., Dearman, D. and Truong, K.N. Virtual shelves: interactions with orientation aware devices. In Proc. UIST 2009, ACM (2009), 125-128.
20. Li, K.A., Baudisch, P. and Hinckley, K. BlindSight: eyes-free access to mobile phones. In *Proc.* CHI 2008, ACM (2008), 1389-1398.
21. Luttgens, K., and Hamilton, N. Kinesiology: scientific basis of human motion, 9th Ed., (1997), Madison, WI: Brown & Benchmark.
22. Microsoft Voice Command. http://www.microsoft.com/windowsmobile/en-us/downloads/microsoft/about-voice-command.mspx
23. Oakley, I. and Park, J. Motion marking menus: an eyes-free approach to motion input for handheld devices. *International Journal of Human-Computer Studies 67*, 6 (2009), 515-535.
24. Pakkanen, T. and Raisamo, R. Appropriateness of foot interaction for non-accurate spatial tasks. *Ext. Abs.* CHI 2004, ACM (2004), 1123-1126.
25. Paradiso J.A., Morris S.J., Benbasat A.Y. and Asmussen, E. Interactive therapy with instrumented footwear: *Ext. Abs.* CHI 2004, ACM (2004) 1341–1343.
26. Patel, S.N., Pierce, J.S. and Abowd, G.D. A gesture-based authentication scheme for untrusted public terminals. In *Proc*. UIST 2004, ACM (2004), 157-160.
27. Pearson, G. and Weiser, M. Of moles and men: the design of foot controls for workstations. In *Proc.* CHI 1986, ACM (1986), 333-339.
28. Pearson, G. and Weiser, M. Exploratory evaluation of a planar foot-operated cursor-positioning device. In *Proc.* CHI 1988, ACM (1988), 13-18.
29. Rahman, M., Gustafson, S., Irani, P. and Subramanian, S. Tilt techniques: investigating the dexterity of wrist-based input. In *Proc.* CHI 2009, ACM (2009), 1943-1952.
30. Rekimoto, J. GestureWrist and GesturePad: unobtrusive wearable interaction devices. In *Proc.* ISWC 2001, IEEE Computer Society (2001).
31. Rico, J. and Brewster, S. 2010. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proc.* CHI '10. ACM, New York, NY, 887-896.
32. 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.
33. Sawhney, N. and Schmandt, C. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Human-Computer Interaction 7*, 3 (2000), 252-282.
34. Williams, N., Zander, S., and Armitage, G. A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification. *SIGCOMM Comput. Commun. Rev. 36*, 5 (2006), 5-16.
35. Yatani, K. and Truong, K.N. SemFeel: a user interface with semantic tactile feedback for mobile touch-screen devices. In *Proc.* UIST 2009, ACM (2009), 111-120.
36. Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R. and Baudisch, P. EarPod: eyes-free menu selection using touch input and reactive audio feedback. In *Proc.* CHI 2007, ACM (2007), 1395-1404.