

Foot Tapping for Mobile Interaction

Andrew Crossan, Stephen Brewster, Alexander Ng

Department of Computing Science

University of Glasgow

Glasgow, G12 8QQ

+44 141 330 8430

{ac, Stephen, ngaw}@dcs.gla.ac.uk

ABSTRACT

In this paper we present an initial investigation of foot tapping as a mechanism for interacting with a mobile device without removing it from a pocket. We compare a foot tapping technique for menu interaction with two more traditional situations: one where the user has the phone in hand, and one where the user must remove it from an inside pocket before interacting. Results show that over the course of the full study, all conditions allowed a high level of accuracy in selections. The visual and in pocket conditions were overall faster and more accurate. However, for short selections requiring four or less foot taps or button presses the foot tap condition was faster than the in pocket condition.

Categories and Subject Descriptors

H.5.2 [User Interfaces]: *Input devices and strategies*

General Terms

Human Factors.

Keywords

Mobile, accelerometer, foot tap, hands-free interaction.

1. Introduction

The methods in which people are interacting with mobile devices are constantly growing and changing. Driven both by new input and output technologies coming to market and greater processing, battery and storage space on the devices, there is a lot of interest in investigating novel techniques that allow users to interact with their device in engaging and low effort ways. After many years of promise, we are finally starting to see gesture interaction become a standard part of mobile interfaces. Touch screens allow a user to quickly and easily perform a limited set of operations using quick flicks of the finger to scroll, or accept or cancel an action for example. Touch screens are a very visually intensive modality. The lack of tactile feedback inherently present in physical button-based devices restricts the functionality when the visual channel is reduced or unavailable. This forces the user to look at the device when interacting which means that when input is required in response to an event, the user must locate the device (which may be in a pocket or a bag) and look at the screen while

interacting. This work examines foot tapping as an alternative input mechanism that would potentially allow users to interact with their mobile device without removing it from their pocket.

As well as touch screens – that are now appearing in almost all high end mobile devices – other sensors are now small and cheap enough to be a common feature in mobile phones. Many newer smart phones contain multiple sensors such as accelerometers, magnetometers, and GPS that allow the programs running on the phone to extract context information that can be used in applications. For example, accelerometers have been used to estimate phone orientation and automatically orient photographs, where as GPS and magnetometers have been used in many location sensitive and mapping applications. For the majority of applications, the sensors are embedded in the phone. The idea of a multi-part phone however opens up different mechanisms for interacting with a mobile device. There are an increasing number of devices that allow input and output to occur on a separate device to the phone. For example, Bluetooth headsets allow phone calls to be conducted with the phone in the pocket and the Nike+ running device attaches to users feet to allow them to receive information about step rate through a mobile device.

2. Related work

There is a large literature on mobile applications of gesturing using the hands and fingers. Much of the work talks about uses for sensors such as accelerometers and capacitive sensors to provide additional information about the context of use. Hinckley *et al.* [4] use multiple sensors to allow automatic adaptation of a user interface where the user activates functionality by tilting the device or holding it in certain positions. Early work from Rekimoto [8] expands on the variety of tasks that could be performed such as scrolling and menu navigation. There are now many examples of device-based accelerometers allowing interaction with applications. For example, Bodyspace described by Strachan and Murray Smith use an accelerometer embedded in a mobile device to access information by moving the device to different areas of the body [10]. A common technique is to use device tilt for gesturing or moving a cursor. For example Rahman *et al.* [7] examine wrist tilt as an input technique for mobile situations and show how it can be used to make menu selections. There is less research examining the possibilities that sensors separate from the device but attached to the body can bring, particularly for mobile settings. Constanza [1] proposes electromyogram (EMG) interaction for mobile situations. A sensor mounted on the upper arm detects muscle flex through changes in the EMG signal. The advantage of this system is that it allows discreet, motionless gesturing. Rekimoto [9] describes GestureWrist, a system that allows users to interact through

changes in hand and wrist posture. This system allows users to gesture without encumbering their hands with devices or bulky sensors. Crossan *et al.* describe two studies involving sensors attached to the user's wrist [2] and the user's head [3]. In both papers, they demonstrate how the user can target in static and mobile situations by tilting his/her head or wrist. An accelerometer is used to estimate orientation and control the movement of a cursor on the screen. They propose these techniques as a method of interacting with a mobile device without removing it from the pocket.

Foot-based interaction is used in a number of different activities. It is commonly used when driving a car, or controlling machines (e.g. Adjusting the speed of a sewing machine), and is an integral part of creating different sounds for a musician. A pianist can use pedals to add mute or sustain to the sound of a piano note allowing different moods of music to be played, while guitarists playing live may use foot pedals to switch between different voices and effects for their instrument. There has however been limited investigation of foot tapping as an interaction technique. This is particularly true in the mobile domain. One notable exception is Paelke *et al.* [5] who explored the possibilities of foot-based interaction with mobile devices. Their approach was to allow the users to interact with a game application on a PDA, where the users performed foot-based gestures which mimic real life actions. A soccer game was created such that when the users wanted to kick the ball, they performed a kicking gesture. A camera mounted onto the PDA was used to capture the positioning and movements of the user's foot. The primary goal and motivation was to provide a new and exciting gaming interaction with mobile devices. Pakkanen and Raisamo [6] explored foot-based interaction with a desktop computer. They note that the user's foot is rarely considered for computer interaction but could be used as a supportive input mechanism to, for example, control spatial tasks that do not require a high level of accuracy or have to be completed quickly such as scrolling a web page. By dividing up the task, Pakkanen and Raisamo suggest that foot-based interaction when used in a supporting role could reduce disruptions and increase work flow. Here we investigate foot-based input as a technique for mobile interaction. More specifically, we use a menu navigation task as an initial step in evaluating the potential of foot tap for both short and long interactions.

3. Foot Tap as an Input Technique

The initial stages of this work examined how to distinguish tap events from everyday foot movements.

3.1 Detecting Foot Taps

The challenge in detecting foot taps is to separate out the tapping from the everyday movements that the foot makes. To detect foot motion, we strap a JAKE sensor pack (more information found at <http://code.google.com/p/jake-drivers>) to the top of both of the user's feet. The JAKE is a small Bluetooth sensor pack containing a three axis linear accelerometer and magnetometer. Here, we use the accelerometer to detect foot motion. First, we approximate the rate of change of acceleration by taking the difference between successive sensor values. We then calculate the magnitude of this signal and high pass filter it using a Butterworth filter to remove the low frequencies in the signal that correspond mostly to non-tapping foot movements. The

remaining high frequencies correspond to fast movements, collisions or rapid changes in direction that would be characteristics of tapping or shaking. Taps are then detected using a thresholding algorithm. Evaluation of this system led to a minimum length of tap set to 0.4 seconds before a second tap is detected. This was to avoid the potential problem of a single tap event being recognized twice. Using the above algorithm, more robustness can be built into the recognition by increasing the sample rate of the signal. In this instance due to the limitations of the phone processor, we were restricted to a sample rate of 30Hz for each JAKE, which while being enough for our lab study, would need to be increased for everyday use.

3.2 Foot Tapping Interaction

We propose a scenario where the user can interact with a mobile interface, while the phone remains in the pocket. Short sequences of taps could correspond to commonly access functionality discreetly, like declining to answer a call in a meeting. Alternatively, the evaluation menu task, users would perform all navigation and selections events by tapping their feet. Here, we consider two different styles of interaction using either single foot or double foot selection.

3.2.1 Single foot selection

For single foot selection, one foot is used to navigate, and one to select. In our test system, a right foot tap moves the currently highlighted menu item down to the next item. The menus are cyclical such that moving off the bottom of the menu returns the currently selected item to the top. A left foot tap is used to select a menu item. In the root menu, this will take the user to a submenu, and in a submenu, a it will perform a selection event.

3.2.2 Double foot selection

For double foot selection, the user can navigate down a menu using a right foot tap as before. A left foot tap will however move up to previous menu item. To make a selection, the user taps both feet at the same time.

3.2.3 Single versus double foot selection

The major issue with single foot selection in our chosen menu task is that it limits the user to moving uni-directionally on the menu. Two foot selection frees up the other foot to allow bi-directional menu movement. However, the two foot selection technique has issues with the selection mechanism. The tap detection algorithm described above continuously looks at both streams of data. It is likely that when the user performs a two foot selection, the taps will not be detected at exactly the same time. It is more likely that a single foot tap occurs first. The algorithm can then group a left and right tap event that occur within a set time interval into a single selection event. We must therefore build in a delay into the system where the feedback on any single tap cannot be provided until we are sure that the event will not be converted into a selection. Without this, we could have situation where audio feedback from the menu item above or below the selected item would be played. By waiting to discover if a selection event occurs, we introduce lag in the audio feedback that occurs for every single tap event. There is trade-off here between robustness of the recognition system and the lag introduced into the audio feedback. To choose between the two different mechanisms, informal pilot testing with 5 volunteers was used. The results from this suggested that the single foot tap selection was both preferred and more robust than the two foot tap selection.

4. Experiment

As a first examination of this technique, we compare it to a standard button-based phone input technique in a menu navigation task. The equipment used was an N95 with headphones for audio and two JAKE sensor packs attached to each of the users shoes (above the toes).

4.1 The Task

The task chosen for the comparison was menu navigation. It allows interactions of different lengths allowing us to examine the number of successive taps appropriate for interaction, and is still a commonly required task on phones. The menu used is shown in Figure 1. It is a two level hierarchical menu where the root nodes represent common tasks performed on a mobile device. As well as a visual representation, the menu item name is read out as the user moves over an item. We use high tempo speech (spearcons) to present the audio. The menus are cyclical such that the currently selected item loops at the bottom and the top of the menus. The final item for each of the sub-menus is 'back', which returns the user to the top of the root menu. As well as these sounds, a short beep was played when a selection event occurred. Users were seated and were prompted through the headphones to select a specific menu item. When the spearcon for the menu item was played, the user navigated to and selected the menu item.

4.2 Conditions

As this technique is aimed at allowing users to interact with the device without removing it from their pockets, we include a button-based condition where the user first must remove the device from a pocket to interact for comparison. We therefore have three different conditions: 'foot tap', 'visual' (a control condition) and 'in pocket'. Visual feedback was present in both the visual and in pocket conditions, but not in the foot-tap condition. Audio feedback was present in all conditions.

4.2.1 Foot-Tap

In the foot tap condition, participants were seated and navigated the menu by tapping their feet. A right foot tap would move down the menu to the next menu item and read this item. As the technique is designed to allow the user to interact without removing the phone from a pocket, no visual feedback was given. To make a selection, the participants tapped their left foot. In the root menu, this would take them to the relevant submenu, while in the submenu (excepting the 'Back' menu items) would make a selection and return the user to the top of the root menu ready for the next trial.

4.2.2 Visual

In the visual condition the participants were seated and held the phone in their dominant hand. To navigate they used the up/down keys on the N95 keyboard, while a selection was made using the central select key. Between tasks, participants rested their hands on their lap holding the phone unlocked. When prompted they looked at the screen and navigated to the appropriate menu item.

4.2.3 In Pocket

For the in pocket condition, the participants all wore the same jacket with an inside breast pocket. Before each selection, the phone is placed in the pocket with the keypad locked. When prompted, the participants removed the phone from the pocket, unlocked the keypad by sliding up the screen of the N95, and navigated to the appropriate menu item as with the visual

condition. Once the item was selected, they closed and replaced the phone in the pocket ready for the next trial.

4.3 Methodology

12 participants took part in all conditions in a counterbalanced order. Before starting the experiment, users were required to familiarize themselves with the layout of the menu and the speech for each menu item by navigating through every item in each menu. Before each condition, five training examples were presented for each mechanism to allow the user to familiarize themselves with the task. In each condition, 40 menu selections were presented. Accuracy, time from the prompt to the selection event, and the number of button/tap events were recorded.

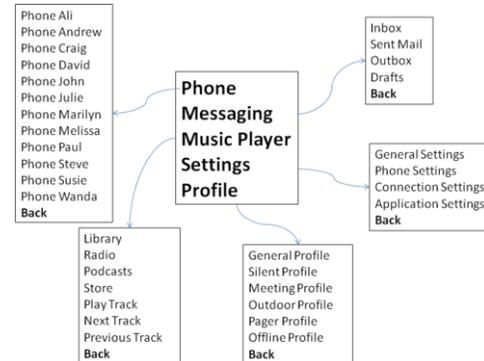


Figure 1. The menu structure used for the experiment. The root menu is shown in the top-centre with submenus around it.

4.4 Results

4.4.1 Over all selections

Figure 2 shows the mean accuracy for selections for all conditions and for all users. These results were analysed using a General Linear Model (GLM) ANOVA with condition and participant as factors showing that the condition was a significant factor in accuracy ($F_{2,11}=19.64$, $p<0.01$). *Post hoc* Tukey tests showed both visual and in pocket conditions were significantly more accurate than the foot tap condition ($p<0.01$ in both cases).

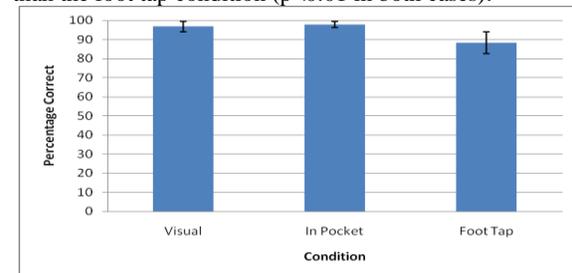


Figure 2. Percentage of correct selections in each condition. Error bars show one standard deviation.

Figure 3 shows the mean time taken to make selection in all conditions (for correct selections only). This again was analysed using a GLM ANOVA with a significant difference being detected ($F_{2,11}=83.0$, $p<0.01$). *Post hoc* Tukey tests showed that the visual condition was significantly faster than the in pocket and foot tap conditions, and the in pocket condition was significantly faster than the foot tap condition ($p<0.01$ in all cases).

4.4.2 Timing for short and long selections

For correct selections in each condition, we can plot the time to target against the number of events (tap or button press) that were

required to make the selection. Figure 4 shows the time required to make all correct selections plotted against the number interaction events required to make the selection. The figure shows that the more events that are required, the more time is needed to select, with the trend being approximately linear. Using linear regression, we extract the gradient (M) and zero crossing (C) of the trend line with values shown in the legend of Figure 4. Using these values it is estimated that foot tapping would be faster than the phone in pocket condition for occasions where four events or less were required to select. For five events and above, removing the phone from the pocket is shown to be faster.

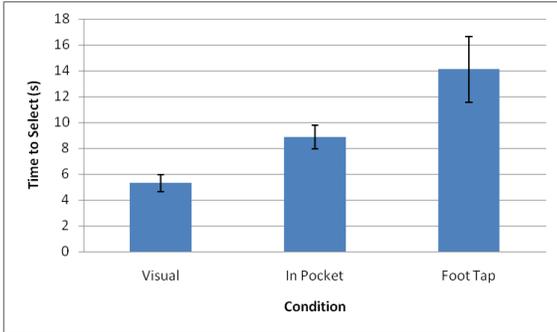


Figure 3. Mean time to select in each condition. Error bars show one standard deviation.

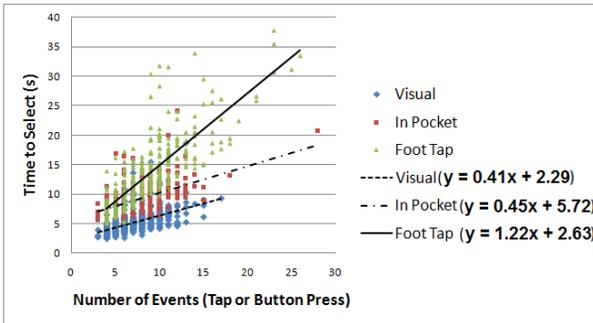


Figure 4. A plot showing how time to select varies with the number of events required to select. The legend gives the gradient and zero crossing values generated through linear regression for each trend line.

4.5 Discussion

When examining the overall results, it is clear that as a general mobile input technique, foot tapping by itself cannot be considered superior to the traditional mechanisms. Both the visual condition and the in pocket condition demonstrated higher accuracy and faster selection times than the foot tapping condition. This is hardly surprising given the high number of taps or button presses to reach some of the menu items and the access to the visual display of the menu, which allows the user to scan all the menu items quickly and move to the correct location without having to wait to hear each in turn. There is also the benefit that the user can move up or down with the button interface but was restricted to one direction only with the foot tap. This was particularly frustrating for users in the longer ‘Phone’ menu. If we examine the trends in the data however, it is possible to see where foot tap could provide some benefit. From the regression values in Figure 4, it can be shown that for less than 5 taps, foot tap can be faster than removing the device from the pocket.

Gradients for the ‘Visual’ and ‘In Pocket’ conditions are similar, with different zero crossings. The difference in zero crossing value here would correspond to the time it took the user to remove the device from the inside pocket. In this case this added approximately 3.4 seconds to the selection time.

4.6 Conclusions

We described an initial study to investigate foot tap as a mechanism for interacting with a mobile phone. Results show that although a high level of accuracy for obtained (~88% correct) it could not compete with traditional methods for longer selections. The results also show however that there is benefit for short selections of less than 5 taps. Our work will now examine whether we can assign short tap gestures to common functionality to allow a user to perform basic operations through simple tap gestures. This might involve specific tap combinations to answer or cancel a phone call. We hope to use these results to as guidance in future studies to develop a simple low effort method of interacting with a mobile phone without removing from a pocket.

5. ACKNOWLEDGMENTS

This work was funded by EPSRC grant EP/F023405 – Gestural and Audio Interactions for Mobile Environments (GAIME), and Nokia.

6. REFERENCES

- [1] Costanza, E., Inverso, S.A., and Allen, R., Toward subtle intimate interfaces for mobile devices using an EMG controller, in Proceedings of the ACM CHI. 2005 Portland, Oregon, USA. p. 481 - 489.
- [2] Crossan, A., Williamson, J., Brewster, S., and Murray-Smith, R., Wrist Rotation for Interaction in Mobile Contexts, in the Proceedings of Mobile HCI. 2008, ACM Press.
- [3] Crossan, A., McGill, M., Brewster, S.A. and Murray-Smith R. Head Tilting for Interaction in Mobile Contexts. in Proceedings of MobileHCI 2009, ACM Press.
- [4] Hinckley, K., Pierce, J., and Horvitz, E. Sensing Techniques for Mobile Interaction. In Proceedings of ACM UIST. 2000, 91-100.
- [5] Paelke, V., Reimann, C., and Stichling, D. 2004. Foot-based mobile interaction with games. In Proceedings of ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, Singapore, 321-324.
- [6] Pakkanen, T., and Raisamo, R. 2004. Appropriateness of foot interaction for non-accurate spatial tasks. In Extended Abstracts of ACM CHI, Vienna, Austria, 1123-1126.
- [7] Rahman, M., Gustafson, S., Irani, P., and Subramanian, S. 2009. Tilt techniques: investigating the dexterity of wrist-based input. In Proceedings ACM CHI, 1943-1952.
- [8] Rekimoto, J. Tilting Operations for Small Screen Interfaces. In Proceedings of UIST. 1996, 167-168.
- [9] Rekimoto, J. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In Proceedings of Fifth International Symposium on Wearable Computers. 2001.
- [10] Strachan, S., Murray-Smith, R., and O’Modhrain, S. BodySpace: inferring body pose for natural control of a music player. In Extended abstracts of ACM SIG CHI 2007