

Instrumented Usability Analysis for Mobile Devices

Andrew Crossan,^{1,2}
Roderick Murray-Smith,^{1,2}
Hamilton Institute,¹
NUI Maynooth,
Ireland.
{ac, rod}@dcs.gla.ac.uk

Stephen Brewster,²
Dept. of Computing Science,²
University of Glasgow,
Scotland
stephen@dcs.gla.ac.uk

Bojan Musizza,³
Dept. of Systems & Control,³
Institut Jozef Stefan
Slovenia
bojan.musizza@ijs.si

Abstract

Instrumented usability analysis involves the use of sensors during a usability study which provide observations from which the evaluator can infer details of the context of use, specific activities or disturbances. This is particularly useful for the evaluation of mobile and wearable devices, which are currently difficult to test realistically without constraining users in unnatural ways. To illustrate the benefits of such an approach, we present a study of touch-screen selection of on-screen targets, whilst walking and sitting, using a PocketPC instrumented with an accelerometer. From the accelerometer data the user's gait behaviour is inferred, allowing us to link performance to gait phase angle, showing there were phase regions with significantly lower error and variability. The chapter provides examples of how information acquired via sensors gives us quantitatively measurable information about the detailed interactions taking place when mobile, allowing designers to test and revise design decisions, based on realistic user activity.

INTRODUCTION

Mobile and wearable devices are becoming increasingly important in our daily lives, and there is a correspondingly large activity in the design of interaction for these devices. It is obviously very important to be able to evaluate their usability, but by their very nature, these devices are intended for use in mobile settings, not for use by someone seated in a usability lab.

As described by Kjeldskov and Stage (2004), there is a wealth of guidelines for running laboratory-based usability studies, but these studies will lack realism for mobile devices. To test mobile devices in mobile settings, however, we are required to use field-based evaluations, which are far from straightforward to implement. Kjeldskov and Stage's review of the literature points out three difficulties: 1. It is difficult to define a study that captures the use-scenario, 2. It is hard to use many established evaluation techniques, and 3. Field evaluations complicate data collection, and limit experimental control. Examples of papers where researchers have proposed additional techniques such as distance walked and percentage preferred walking speed to assess usability include Brewster (2002), Petrie *et al.* (1998), and Pirhonen *et al.* (2002), using a mix of qualitative questions and manual recording of walking pace. Mizobuchi *et al.* (2005) examine the effect of key size on handheld devices while walking.

Barnard *et al.* (2005) review the differences between desktop and mobile computing, and they observe “*for researchers aiming to isolate the effects of motion from other contaminants, the idea of such uncontrolled studies can be daunting. Control is critical for empirical data collection methods employing the scientific method.*” Roto *et al.* (2004) discuss the use of Quasi-experimentation based on best possible control over nuisance variables, coupled with recordings of the user, interaction with the device and environment. The innovation in their

recordings was the use of multiple cameras worn around the body of the user, and attached above the screen of the mobile device. This does make the recording process obtrusive and might change both user behaviour and that of people in the environment around them. It is also time-consuming to analyse after the experiment. This recording arrangement has been used successfully in (Oulasvirta *et al.* 2005) to investigate the fragmentation of attention in mobile interaction.

Instrumented Usability Analysis

Here, we define '*Instrumented usability analysis*' as the use of sensors during a usability study which provide observations from which the evaluator can infer details of the context of use, or specific activities or disturbances.

Sensors such as accelerometers, magnetometers and GPS systems have been added to mobile devices, and are now in mass-production in mobile phones. These have been included for informing the user (about location, number of steps taken), or giving the user novel input mechanisms, such as gesture recognition or input for game playing.

There are many examples of both prototype and commercially available sensors and sensor packs for motion or context sensing. Fishkin *et al.* (2004) describe a system for detecting interactions with RFID technology and suggest it can be used to infer user movement by examining signal strengths from a sensor network. Gemmell *et al.* (2004) describe the SenseCam system used to capture life experiences without having to operate complex recording equipment. SenseCam combines a camera with a group of sensors including an accelerometer, infrared, light and temperature sensors and a clock to automatically detect, photograph and map out changes in context or events during a persons day. Kern and Schiele (2003) describe a hardware platform combining multiple wearable accelerometers in order to infer the user's context and actions. They demonstrate how these acceleration signals can be used to classify user activity into actions such as sitting, standing, walking, shaking hands and typing.

More recently, the general purpose Bluetooth SHAKE (Sensing Hardware Accessory for Kinesthetic Expression) inertial sensor pack, described in (Williamson *et al.* 2007), which is available for general use by the research community, from SAMH Engineering Services. It features a tri-axis accelerometer, tri-axis magnetometer, dual channel analogue inputs, dual channel capacitive sensing and an internal vibrating motor. Communications are over a Bluetooth serial port profile. SHAKE includes a powerful DSP engine, allowing real time linear phase sample rate conversion. These capabilities allow rapid prototyping of inertial-sensing-based interfaces with real-world hardware. But the small size, and the onboard processor and memory mean that the device can be used completely separately from the implementation on a mobile device, and can be used to log movement at multiple points around a user in a wide variety of situations. It can be attached to the back of their device, or could be attached to their belt, or elsewhere on the body, to detect activity without wires to restrict a person's movements.

In this chapter we suggest that such sensors and sensors packs can be used in an indirect fashion, to better understand what was happening to the device and user at any point in time during a usability experiment. Figure 1 demonstrates how such a system would work. There could potentially be multiple sensors placed on the user or mobile device. Outputs from the sensor would be run through one or many classifier algorithms that could infer the user's actions or context at any one time.

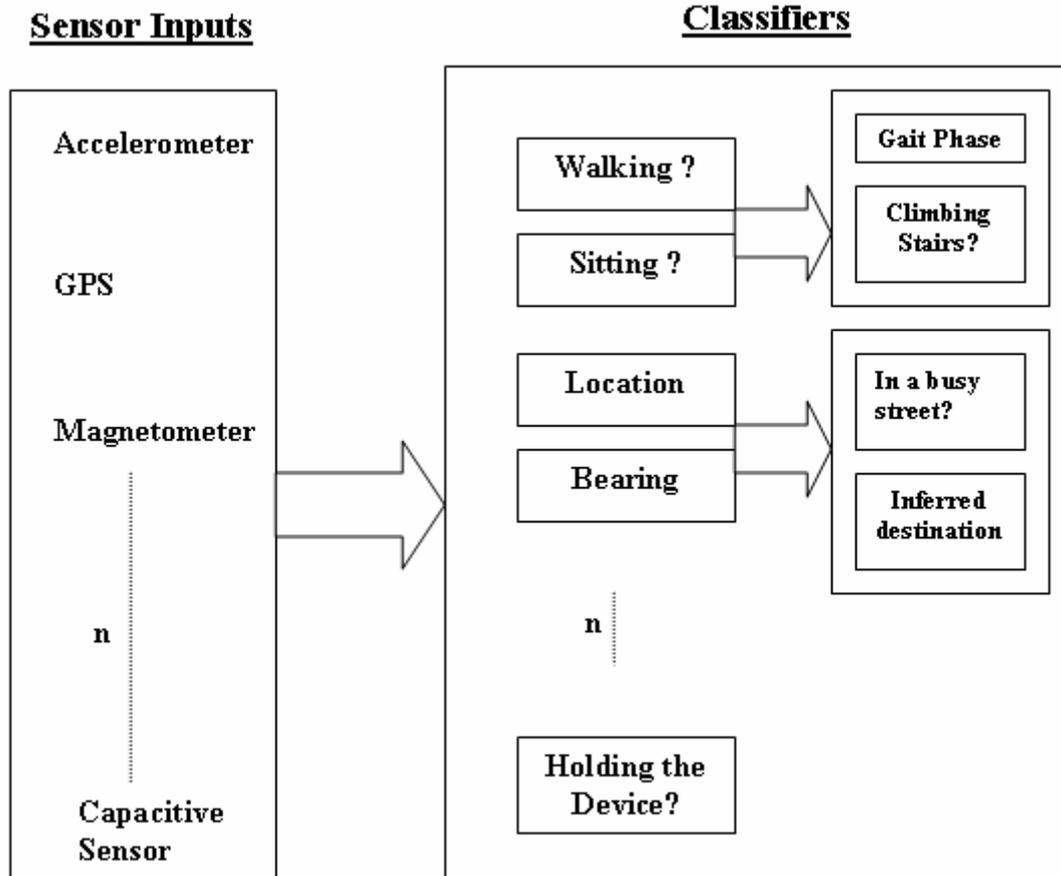


Figure 1. An example of how sensors and classifiers could be combined to infer user actions during a usability study, and store these as an annotated log file, allowing developers to correlate different states with user interaction behaviour. Raw readings are interpreted in a hierarchical fashion by a range of plug-in classification or signal transformations. These can be arranged hierarchically, so e.g. only if the person is classified as “walking” do we infer gait phase angle, and whether they are climbing the stairs or not.

Of course, as the algorithms for automatically inferring context of use from sensors develop to a level of robustness which allows them to be used online, they can be used in everyday mobile situations to subtly adjust the nature of the interaction – the “background interaction ... using naturally occurring user activity as an input that allows the device to infer or anticipate user needs” described in Hinckley *et al.* (2005). For example, if tapping is less accurate when a user is walking, the display could adjust to a mode which had fewer, but larger, buttons. In this chapter, however, we will concentrate on the use of such information to analyse user behaviour in greater temporal detail than is typical in mobile usability trials. Rather than performing a trial, and then asking subjective questions, or analyzing video footage, we will classify activity from sensors on the device or user, and relate these to any log of explicit interaction activity during the evaluation. If a user has an unusually high error rate, can we better determine exactly what was happening at that point in time in each case?

This approach is obviously related to research in context-dependent interaction which used information from sensors to infer context of use (walking, running, in car, inside, outside), to allow more appropriate behaviour from the device. Yi *et al.* (2005) used an accelerometer in evaluation, but used mean activity over the whole period in different conditions, rather than detailed results during the evaluation.

These techniques can be combined with a system such as Replayer described by Morrison *et al.* (2006). This is a system designed to aid usability evaluation and provides tools that allow evaluators from different backgrounds to easily view, annotate, and analyse multiple streams of heterogeneous data logged during a usability study. These data streams could potentially be obtained from multiple sensors attached to a mobile device. A first prototype of this, using the MESH device used in this paper, is described in Morrison *et al.* (2007)

Example: Mobile Text Entry

An illustration of how the method could be used is that of mobile text entry. The questions we might want to answer for a given method could be: do people use it on the go, as well as when stationary? How much slower are they when they use it while walking? Do they enter text continuously, but slowly, or do they stop every few metres to enter more text? How is their error rate related to their walking speed? Do they link the entry of a new character with a new step? What is the effect on walking speed, when entering text? If the user enters text in a car, or bus, how are they affected by movement of the vehicle? Do they wait until the bus stops then enter text?

Figure 2 below shows a time-series of accelerometer readings while a user enters text, while seated as a passenger in a car in urban rush hour traffic with frequent stop-start activity.

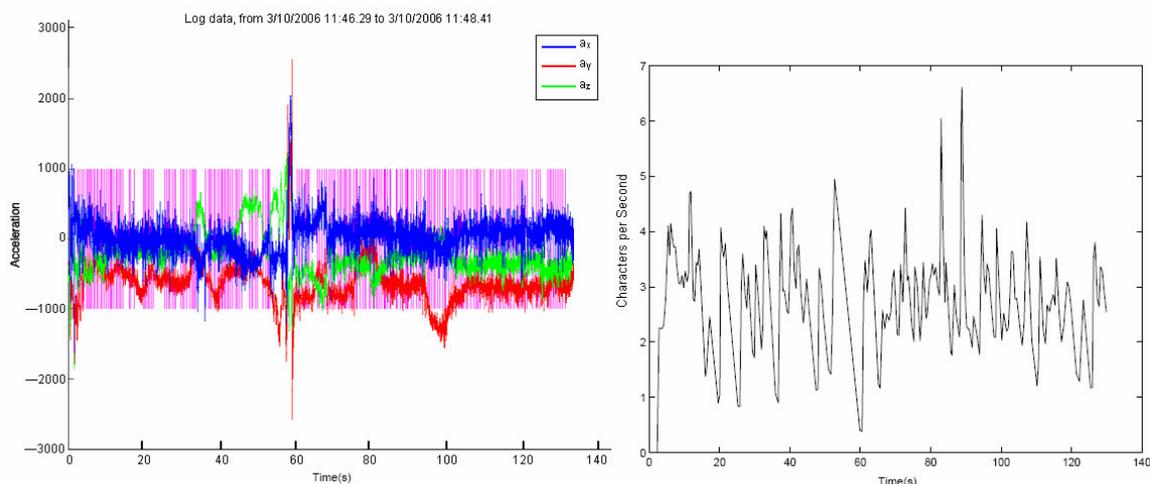


Figure 2. Text entry while seated in a car. The number of characters entered per second is plotted to the right of the acceleration time series (the vertical lines in the left plot are individual key-press events).

Figure 3 shows the example of a user entering text in various contexts. The user started to enter text while sitting, stood up, walked around some narrow corridors avoiding objects, down stairs, along a straight corridor, up stairs, then returned to seated position, and entered more text while resting his hand on a table. The plots show the overall activity of the user, along with the throughput of characters entered at each point. In the walking case, we can see text entry pause, as the user takes a seat just after 140s, and we see faster entry rates while seated in the car, compared to walking. The text entry rates while walking are nevertheless fairly constant. This illustration acts as an example of how the accelerometers can give us extra information from which we can infer more about what was happening at each point in the interaction.

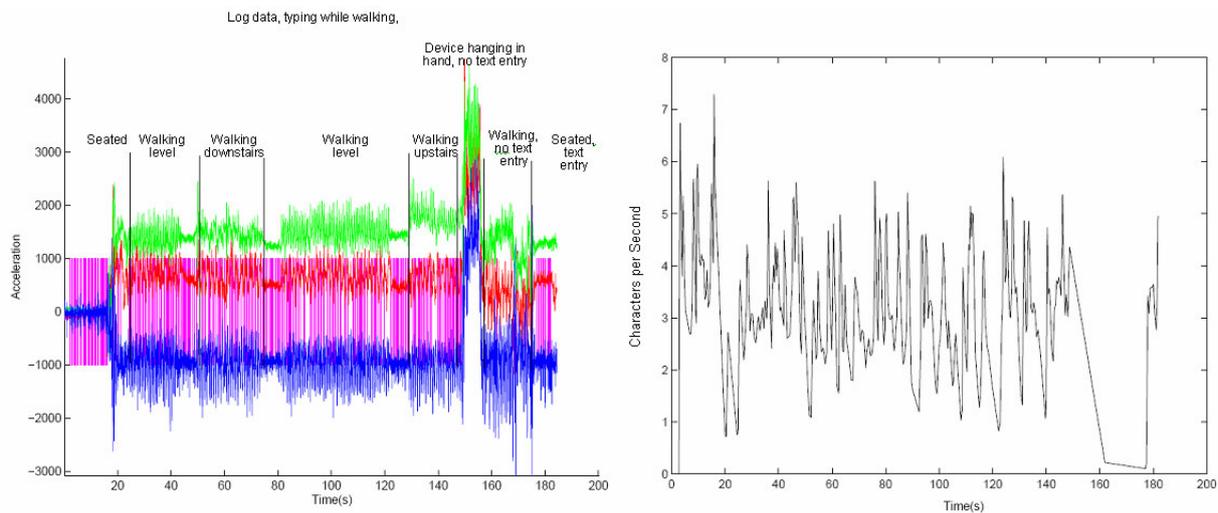


Figure 3. Text entry in a range of conditions. The number of characters entered per second is plotted to the right of the acceleration time series (the purple vertical lines in the left plot are individual key-press events, and the black ones delineate the different walking conditions).

In this chapter, which expands on our earlier work in (Crossan *et al.* 2005), we work towards a quantitative understanding of the detailed interactions taking place, via additional sensors on the mobile device and user, so that we can better understand how users interact with the devices, and so further improve the designs.

DETAILED CASE STUDY OF WALKING AND TAPPING

Here, we present a detailed example of an instrumented usability study to demonstrate the benefits of this approach. Standard usability time and error metrics are gathered, while the instrumentation allows us to gain a greater insight into the users actions and disturbances during the study. Although this example specifically examines users walking patterns sensed through an accelerometer, the techniques discussed can be applied to a wide range of contexts.

Given the importance of devices being used while the user is walking, and the difficulty researchers have had about getting detailed insight into user behaviour, down to the level of each step taken, we now concentrate on the example of tapping buttons or other widgets on a touch screen. This is a common form of input, and is effective when seated, but difficult when walking. Brewster (2002) showed a more than 30% reduction in performance tapping buttons on the display of a PDA when walking, compared to sitting. If we can gain more detailed insight into how and when users tap during walking, we might be able to adjust the design of the interface to improve robustness.

Here we show how sensors, like accelerometers, can be used in ways other than for explicit interaction. In this case we use the acceleration data to infer the user's gait, and we investigate whether the rhythm of walking affects the tap timing and error rate of a user selecting targets on screen, while walking and sitting.

Experiment

Introduction

This study examines in detail the behaviour of users tapping on the screen of a mobile device. It analyses behaviour in two different situations that a user might perform this task: while sitting and while walking. The sitting condition will be used to provide a performance

baseline for the walking condition. Disturbances to the device and the users stylus due to the user's walking will affect how and how well the user performs the task. By instrumenting the device with a sensor (in this case an accelerometer), we will be able to gain a deeper understanding of how these disturbances affect performance.

Equipment

This system was developed using an HP 5550 PDA with the Xsens P3C 3 degree of freedom linear accelerometer attached to the serial port, as shown in Figure 4. Its effect on the balance of the device is negligible (its weight is 10.35g). The accelerometer was used to detect movement of the device, sampling at a rate of approximately 90Hz.



Figure 4. PDA with the Xsens P3C accelerometer attached to the serial port.

Task

The interface used for the study is displayed in Figure 4. Participants were asked to tap on a series of cross-hair targets (drawn 30 pixels high and wide) that were displayed on the screen. There were 15 possible target positions spaced equally around a 3 wide by 5 high grid of positions on the screen. Every second target presented to the participants was the target in the centre of the screen. This ensured that the user must return the stylus to the centre of the screen such that when a target other than the central target was tapped, the path to that target was always from the centre. The other 14 targets were displayed to the user in a random order four times each. The accuracy and speed of tapping were both emphasised as equally important. The position of the tap was recorded as the initial stylus down position on the screen. Once one target had been selected, the next target was displayed a random time interval from 0.5 to 1.5 seconds after the previous selection. This was to prevent rhythm effects affecting the tapping phase information in the mobile condition. There were no restrictions on the accuracy that was required by the user. A tap anywhere on the screen regardless of the position of the target counted as a selection.

There were two experimental conditions: tapping while sitting and tapping while walking and 20 users performed both conditions in a counterbalanced order, with 18 participants being right handed and 2 participants being left handed. All participants tapped with their dominant hand while holding the device in their non-dominant hand. For the walking condition, the participants navigated a quiet triangle of paths on the university campus (of total length approximately 200 metres).

Calibration of the screen becomes an issue when looking at accuracy of tapping in a pen based interface, as an error in the calibration can lead to a consistent and unwanted bias in the

results. The screen was calibrated once at the start of the experiment, and the same device was used throughout the experiment. Three participants tested the screen calibration. The device was placed on the desk and users performed a similar task to the tapping study for four separate sessions. In this case accuracy was heavily emphasised as the most important aspect of the study. This was borne out by the much closer concentration of points than in the final results, with mean standard deviation of the error for each participant for all targets being less than a pixel. After each session, the device was rotated by 90 degrees (additive for each section) to negate any systematic tapping bias. Mean values were recorded for each screen target position and were subtracted from the final results. This method provides a closer match between the position the user actually tapped in and the recorded tap position.

Metrics

Standard usability metrics were used for assessing user performance in the task. Comparisons were made between time to tap and accuracy of tap for each of the groups. Time to tap was taken from the time that the target was displayed on the screen to the time of the stylus down event. The hypotheses were that users would be more accurate and faster in the seated condition. The effect of screen position of the target on accuracy of the tap was also examined.

The instrumented usability approach also allows us to gain further insights into the users actions during the study. The interactions of participants' tapping and step patterns were examined.

Gait Detection

As a mobile user walks while holding a mobile device, his or her arm will oscillate as a result of the user's gait. If we examine only the vertical axis of this oscillation, there will be one oscillation per step. Figure 5 shows a time series for the vertical acceleration axis. A Fast Fourier Transform is used to determine the frequency at which the peak amplitude occurs, between 1 and 3Hz in the spectrum. For the controlled conditions in this study, this corresponds to the walking step rate. In practice, this is the frequency of maximum power in the spectrum as the users are looking at the screen and therefore trying to hold the device relatively still with respect to their body as they walk. The vertical axis acceleration signal is then zero phase shift filtered using a narrow bandpass Butterworth filter centred around this frequency. Figure 5 demonstrates the filtered signal. As the user walks with the device held steady in one hand, an approximately regular oscillation is formed in the vertical axis. One oscillation corresponds to one step.

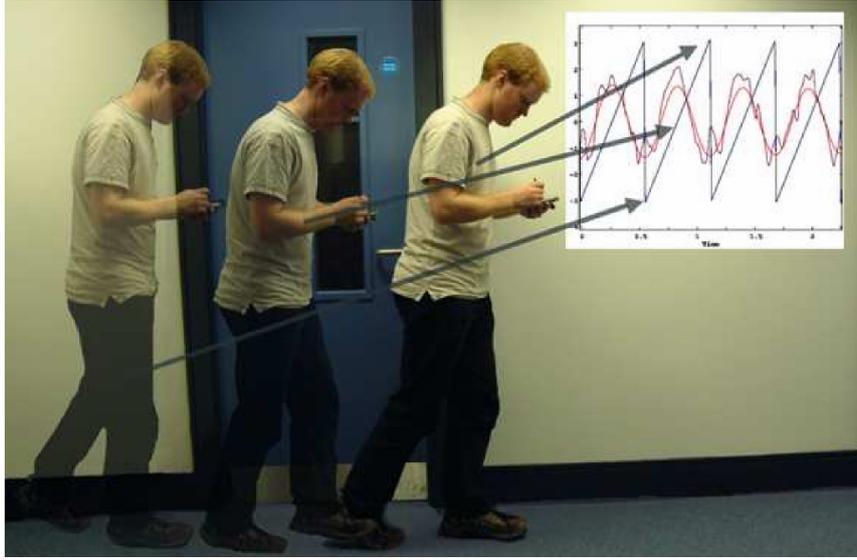


Figure 5. A user walking with the device and corresponding acceleration trace. The unfiltered vertical acceleration signal (rough sinusoid), the filtered signal (smooth sinusoid) and the phase estimate (in radians) for the signal (saw-tooth).

The algorithms used in this chapter were developed in research on synchronisation effects in nature. The oscillations involved in many natural systems are often irregular, ruling out simple strategies. In some cases, such as respiratory examples, or electrocardiogram data, there are clear marked events with pronounced peaks in the time-series which can be manually annotated, or automatically detected. One practical advantage of the use of synchronization theory is that often we have a quite complex nonlinear oscillation, which might be sensed via a large number of sensors. The phase angle of that oscillation is however a simple scalar value, so if we are investigating the synchronization effects in two complex systems, the analysis can sometimes be a single value, the relative phase angle $\phi_1 - \phi_2$.

The Hilbert Transform

How do we find the phase angle from the data? A common approach is to use the Hilbert transform introduced by Gabor in 1946, which gives the instantaneous phase and amplitude of a signal $s(t)$ (Pikovsky et al. 2001). The Hilbert transform signal $s_H(t)$ allows you to construct the complex signal

$$\zeta(t) = s(t) + is_H(t) = A(t)e^{i\phi(t)} \quad (1)$$

Where $\phi(t)$ is the phase at time t , and $A(t)$ is the amplitude of the signal at time t . The Hilbert transform signal of $s(t)$ is

$$s_H(t) = \frac{1}{\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{s(\tau)}{t - \tau} d\tau \quad (2)$$

Although $A(t)$ and $\phi(t)$ can be computed for an arbitrary $s(t)$ they are only physically meaningful if $s(t)$ is a narrow-band signal. For the gait analysis, we therefore filter the data to create a signal with a single main peak in the frequency spectrum around the typical walking pace (between 1 and 3Hz).

This phase plot signal is again shown as the saw-tooth waveform in Figure 5 and Figure 6 and can be seen to reset at the lowest point in the signal. This corresponds to the lowest point of the hand in the oscillation.

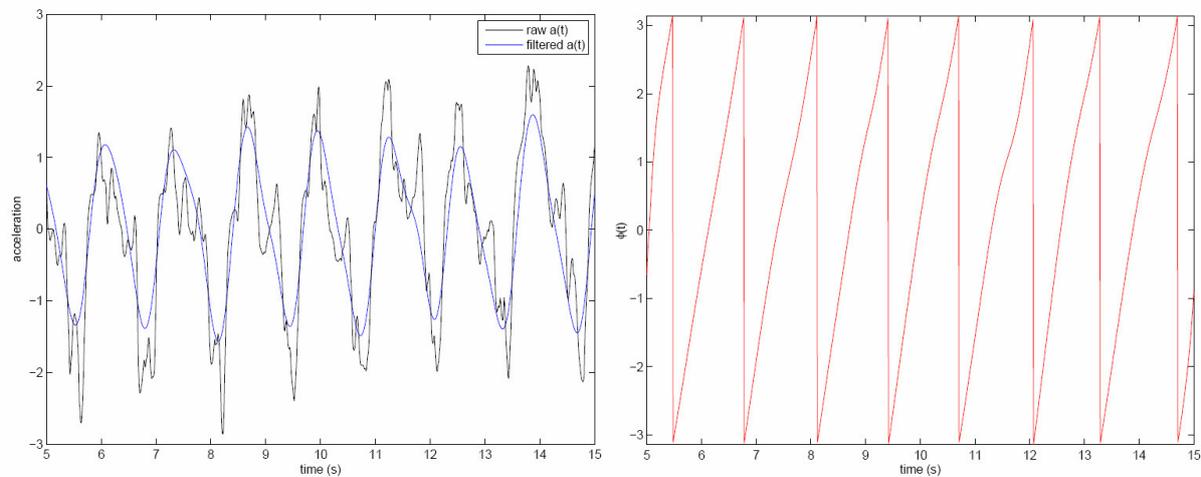


Figure 6. Generating the phase angle $\phi(t)$ from observed acceleration data $a(t)$ from a user walking.

Details of the Hilbert transform and filtering are included here for completeness, however, this functionality is easily accessible in many standard data analysis programs such as Matlab through simple function calls and understanding these equations is not essential for understanding the remainder of this chapter.

Standard Usability Results

Time to Tap

The mean time to tap was lower in the sitting case than the walking case as would be expected. The mean time to tap a target in the walking condition was 0.79s (std dev = 0.18) compared to 0.70s (std dev = 0.22) in the seated case. This can be further broken down into tapping the centre target and outer targets. The mean time to tap the centre target was 0.75s (std dev = 0.23) when walking and 0.65s (std dev = 0.19) while sitting. This compared to 0.82s (std dev = 0.22) while walking and 0.75s (std dev = 0.20) while sitting to tap the outer targets. This difference between centre and outer targets is indicative of users predicting the appearance of the centre target since it consistently appeared every second target.

Tap Accuracy

A graph of tapping accuracy is shown in Figure 7. The graph demonstrates that as expected, users were more accurate tapping in the seated condition with 78% of taps being within 5 pixels in the seated case compared to 56.5% in the walking case. Participants remained more accurate in the seated case and reached 98% of taps within 15 pixels in the seated condition compared to 25 pixels in the walking condition. Separating these into x and y pixel error showed little difference between accuracy in vertical or horizontal error.

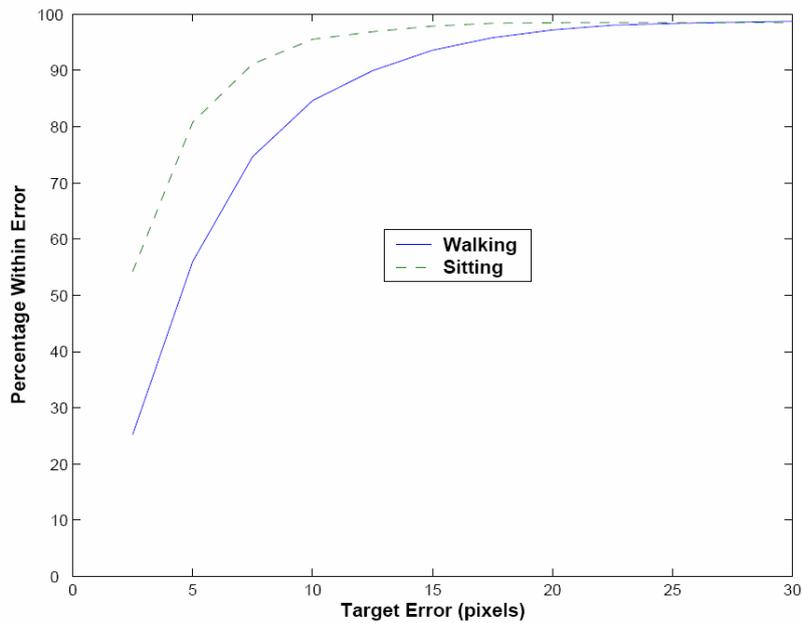


Figure 7. Percentage of taps with the given pixel radius for sitting and walking users.

Above the range of 30 pixels, structure can be seen in the errors where tap position corresponds to the position of the previous target (shown in Figure 8). This indicates a tap when the user did not mean to tap. This is most likely the result of a user accidentally double tapping in position of the previous target. These taps were viewed as outliers and discounted from the final analysis.

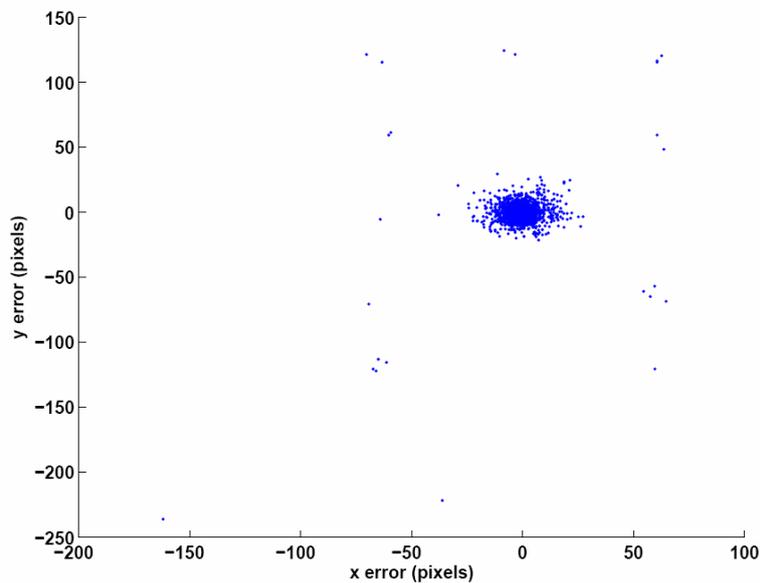


Figure 8. The x- y pixel errors for all users for all targets. The structure of the 3 by 5 grid of targets can be seen indicating users mistakenly double tapping.

Observation in the walking condition showed that when tapping, all participants immediately adopted the strategy of grounding the side of their hand holding the stylus on the hand holding the device to reduce independent movement of the hands and thereby improve accuracy. Targeting therefore involve pivoting the hand about the grounded position.

Figure 9 shows the mean variability and covariance of the x and y target errors for all users for each of the 15 targets. In almost all cases, the variability in tapping is smaller in the seated condition than in the walking condition. Due to the controlled conditions of this study, the movements to the outer targets were always from centre target. The variability in tap position for the centre targets is less than that of the outer targets. This is due to the fact that the stylus over the centre target position was the default position for most users. Covariance of the x and y tap positions can be seen to be along the direction of movement for most of the targets. This is particularly true for the corner targets.

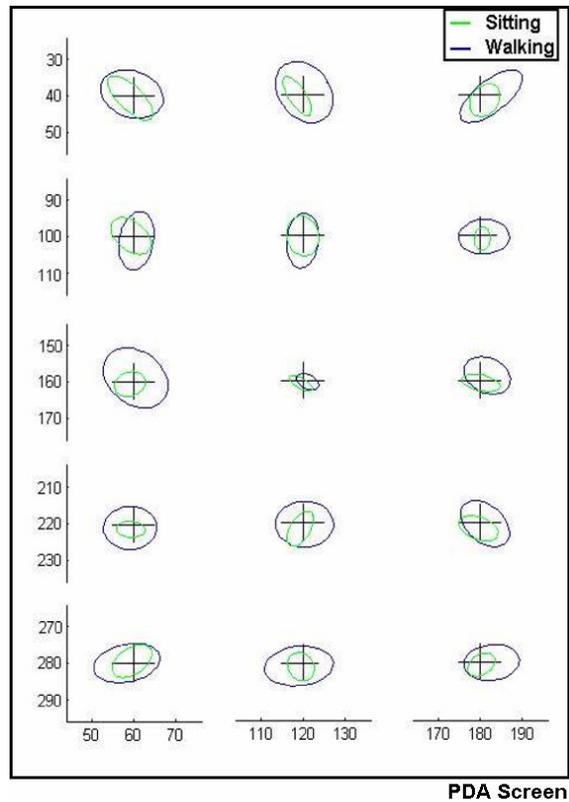


Figure 9. Ellipses show 2 standard deviations of a Gaussian fit to the spread of mean tap positions (from 4 points per participant) from all 20 participants, for each target. In each case the smaller ellipse shows the results for the seated condition and the larger ellipse shows the results for the walking condition. The crosses represent the target positions.

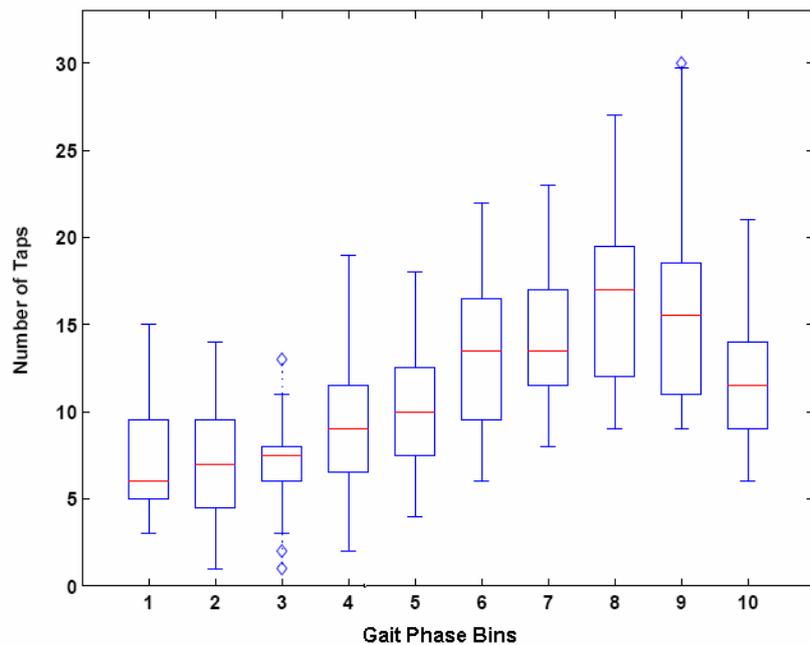


Figure 10. Box plot visualising the distribution of tapping times. Median phase in which the user taps (split into 10 sections) with the reset position for the phase corresponds to the lowest point of the arm which occurs just after a step.

Instrumented Usability Analysis Results

Tap Phase

The method for obtaining the phase of step that the tap occurred at is described above. Figure 10 splits one step into 10 equal sections and plots the median of the number of taps in each section for each participant. The reset phase position corresponds to the lowest point of the vertical accelerometer trace. Bins 1 to 5 correspond to the arm as it moves upwards to its peak, and bins 6 to 10 correspond to the arm moving downwards.

A bias is clearly shown towards tapping in the second half of the oscillation. This bias is not present when analysing the phase at which the targets are displayed and must therefore have been introduced by the user. The phases when most taps occur correspond to when the device is moving downwards with the arm. As soon as the device begins to move upwards in the hand again towards the stylus, the number of taps on the screen decreases. When questioned after the experiment, none of the participants was aware that a bias existed.

Figure 11 shows the median of the mean magnitude tap error for each participant, for each of the step phase bins above. This figure shows that users were more accurate when tapping in the second half of the phase - the time when most taps occurred. The mean error is 7.1 pixels in the first section (just when the arm starts to rise again), compared to a mean of 5.6 pixels in the fourth section when the hand is moving downwards.

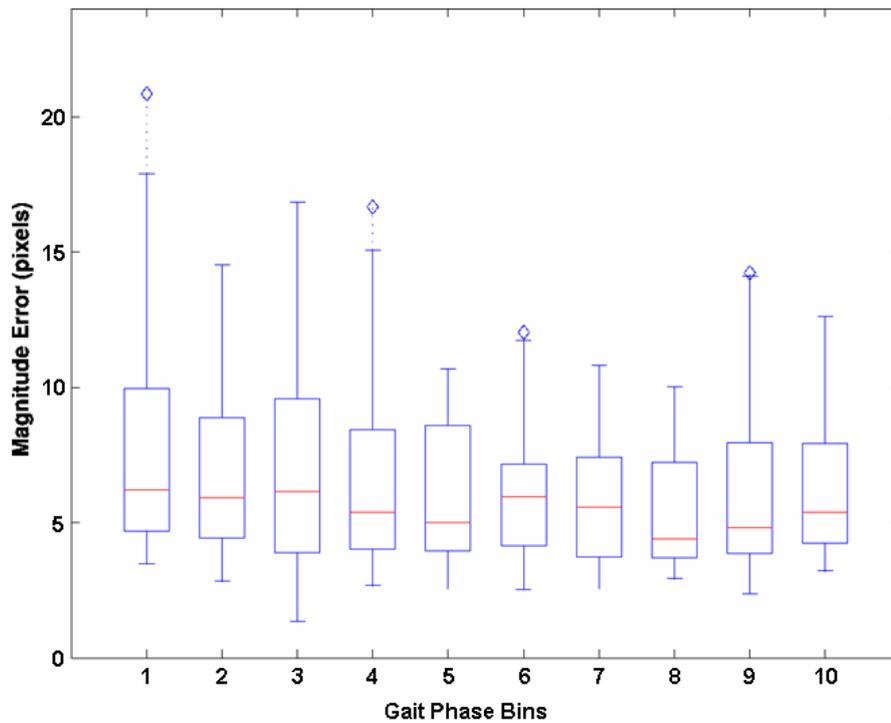


Figure 11. Median target tap error in pixels for each phase of the motion (split into 10 segments) with the reset position for the phase corresponding to the lowest point of the arm which occurs just after a step.

Further to this, if we consider just the three most probable tap phase bins (P_{HP}) and the three least tap probable tap phase bins (P_{LP}) a clearer indication of this is given. Figure 12 shows a box plot of the tap error in P_{HP} and P_{LP} . P_{HP} has median tap error of 4.6 pixels compared to 5.7 pixels for P_{LP} . A Mann Whitney test showed that this difference was highly significant ($p < 0.002$). If we consider the timing data for the same phase regions, it can be seen that users take significantly longer to tap in the high probability regions. Users took a median of 0.69 seconds to tap in P_{LP} compared to 0.73 seconds for P_{HP} . This difference was again tested using a Mann Whitney test and was shown to be significant ($p = 0.05$). Figure 13 shows the corresponding skew plot for high and low tap probability regions.

When combined with the results shown in Figure 10 above, these data suggest that users were able to subconsciously alter their behaviour in the task in order to improve their accuracy by tapping at a time in their step when it was easier to tap more accurately. The longer time to tap in the high probability region indicates that users tended to subconsciously wait for that particular phase region to tap in.

Left-right step analysis

For the previous set of results, each step has been treated as one cycle. However, we could also choose to separate out the left foot steps from the right foot steps. As the user walks, the vertical acceleration sensed through the device will complete one phase cycle at every step the user takes. The lateral acceleration can also be seen to be oscillatory. However, one oscillation in the lateral direction will now correspond to a combination of one left foot step and one right foot step. The dominant frequency of the lateral oscillation is therefore half that of the vertical oscillation. The device will therefore undergo consistently different disturbances depending on whether the user is stepping with the left or right foot.

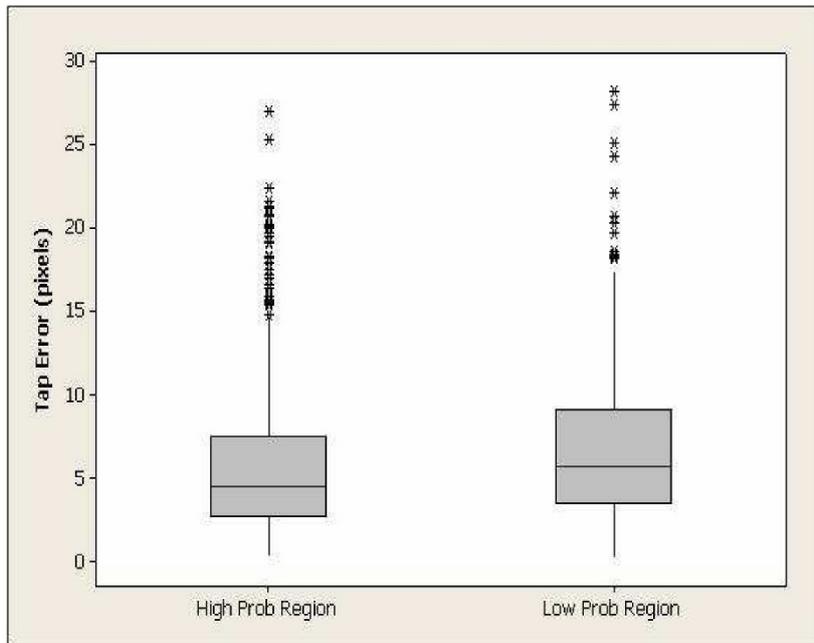


Figure 12. Target tap error in pixels for the high probability tap phase region and the low probability tap phase region

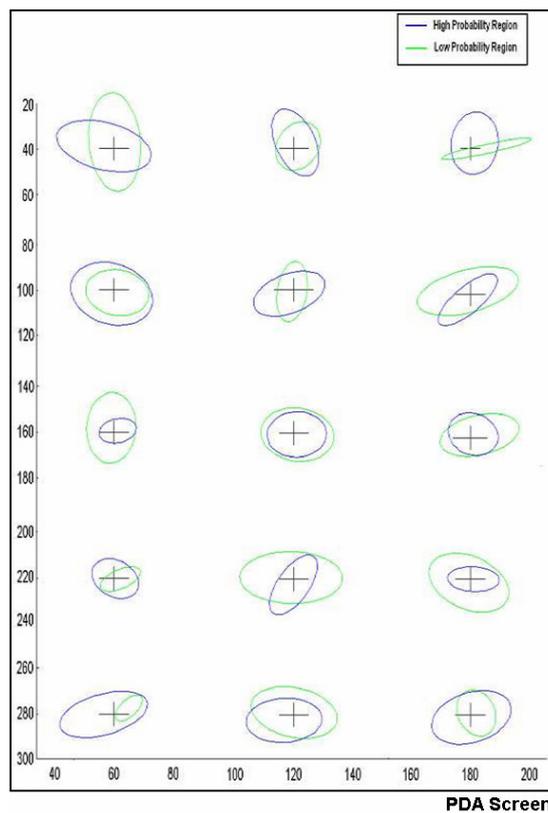


Figure 13. Target tap error in pixels for the high probability tap phase region and the low probability tap phase region.

The data gathered from the accelerometer was analysed to separate left and right foot steps. The vertical acceleration was used to delineate the steps with the lateral acceleration used to determine right and left foot steps. Using this method ensures that valid comparisons can be made between the one step-one cycle data and the one step-two cycle data. Figure 14 shows

the distribution of tapping through the phase of the left and right foot steps. The first eight bins correspond to a left foot step, and the second eight bins to the right foot step. It can be seen from this figure that the tapping pattern for the two step per phase cycle data follows the one step per phase cycle data. There are distinct interactions visible between the tapping and the stepping in each step. The pattern for each step is consistent, with the peak tapping phase values occurring at around the foot down phase of the step for both left and right foot. There were no significant differences detected between the tap errors for left foot steps and right foot steps. The median magnitude of error for taps occurring during the left foot phase was 2.9 pixels compared to 2.8 pixels for taps during the right foot phase section. There were no significant differences with either the separated x error or the y error for taps during the left or right steps.

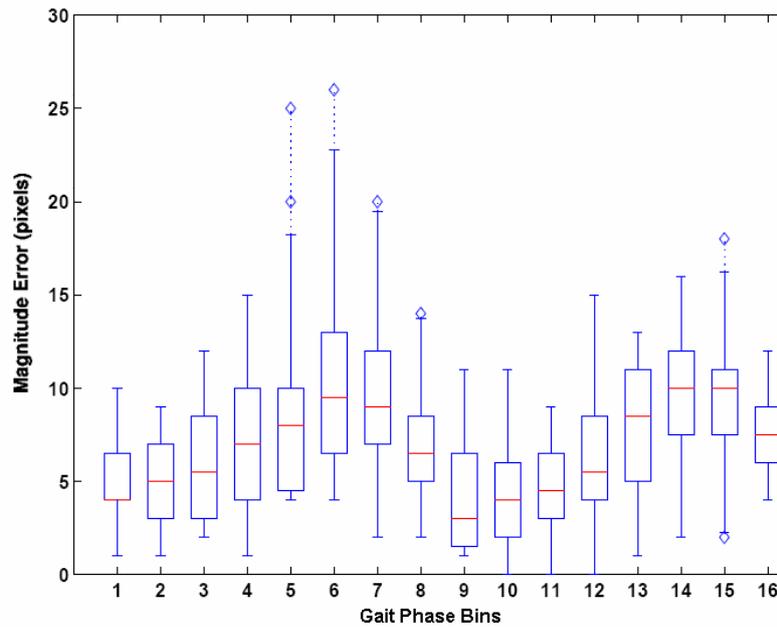


Figure 14. Box plot visualising the distribution of tapping times. Median phase in which the user taps (split into 16 sections). Unlike Figure 10, one phase cycle includes both a left and a right step.

Walking speed analysis

Analysis of the participants' walking speed throughout the experiment showed that the step rate during the study was relatively consistent for all users. Figure 15 shows the estimated step rate for five typical participants over the duration of the study. The task chosen for the study was consistent throughout the task (tapping on a screen). The path that the participants traversed during the experiment was relatively quiet. There was therefore little reason for the participants to speed up or slow down their step rate. In other experiments where the path is more complex or the user must perform different tasks, we would expect the step rate to be more variable, but in this study, mean step rate might actually have been sufficient when analysing the effects of walking rate.

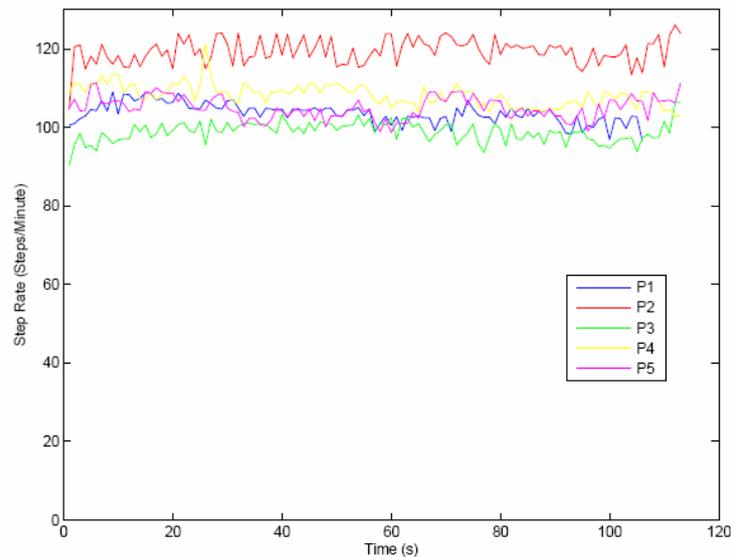


Figure 15. The step rate of five typical participants for duration of the walking condition.

Other Analysis

The results presented so far have involved analysing the acceleration trace to extract information about the users' steps. Now we examine all disturbances affecting the device. If the device is moving around more, we would expect the user to be tapping less accurately. By looking at the magnitude of the of the acceleration trace in x , y , and z we gain an insight into the mean magnitude of disturbance that the device was going through. Figure 16 shows a scatter plot of the magnitude of the tap error plotted against the mean magnitude of the disturbance of the device for the one second previous to the tap. As the figure shows, in this instance there is no simple correlation between tap error and device acceleration.

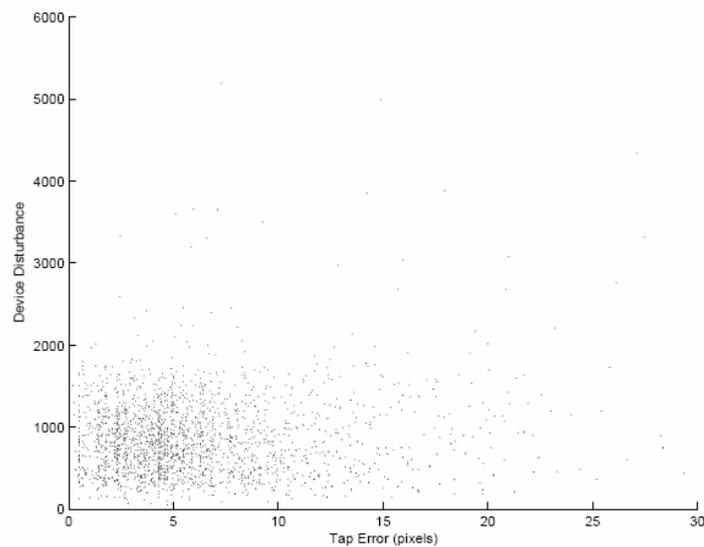


Figure 16. A scatter plot of the magnitude of the device disturbance plotted against the tap error in pixels.

Discussion

Using standard usability metrics, we were able to show that tapping accuracy was, unsurprisingly, typically greater when sitting still, rather than walking. However, the above

results demonstrate the extra insights into user behaviour that were made possible by taking an instrumented usability approach.

Specific experimental observations of this instrumented usability approach are:

- Users' tapping time is significantly correlated with gait phase angle. Users were approximately 3 times more likely to tap at the most favoured tap phase than the least favoured tap phase.

Users' tapping position accuracy is significantly higher (lower mean error and lower variability) at these preferred phase angles. Analysis of the timing data for the different phase regions showed that users subconsciously delayed their target selection in order to tap in one particular phase area rather than any other. There is further structure in the left step-right step tap density, error biases and variability, but even when averaged over all steps, the results are significant.

- The distribution of tapping errors varies both with phase of step, and between walking and sitting and across different screen positions.

It is interesting to note that although there is no simple correlation between tap error and device acceleration, the inferred phase angle, which is based solely on the acceleration observations, does show a strong link between acceleration and tapping accuracy, emphasizing the need for appropriate models in data analysis. One potential reason for this is that the walking route chosen for the study did not require the user to make irregular adjustments to their movements. The path was quiet so that the user only infrequently had to avoid objects. This limited the disturbance of the device to lower than might be expected in a more crowded environment, or in e.g. a moving vehicle. The participants grounded their tapping hand on the device while tapping, which minimised the effect of the external disturbance in this instance, so the main disturbance came from the gait cycle of walking itself.

CONCLUSIONS AND FUTURE WORK

This work has demonstrated that by making fine-grained observations from sensors during a usability study, that we can learn increased detail about the timing and error rates for users. Until now, linking the analysis of, for example walking behaviour, in a realistic setting would typically have required the use of hand scoring videotapes of users' actions – a time-consuming, and potentially subjective and error-prone approach which is also not open to online experimental control. Recent rapid developments in mobile device capacity, and compact sensors, coupled with the use of the analytic tools from synchronization theory, have opened up a new way of investigating gait effects in interaction. The inertial sensors monitor walking patterns throughout the experiment, and can potentially be used together with machine learning classification algorithms, during the experiment to control for experimental stimuli, and adapt the experimental situation online, providing a more stringent method of exploring mobile interaction.

The work opens new directions in both design and usability areas for future work. The specific results gained through the use of the accelerometer data for gait analysis allow us to explore new areas to inform mobile design. For example, one question raised from this study is – does designing an interface such that users tend to tap in preferred phase ranges lead to quantitatively better performance and qualitatively more pleasant user experience? Might it be better to delay user prompts until a particular phase region, in order to sustain rhythmic interaction? (See Lantz and Murray-Smith (2004) for a discussion of rhythmic interaction). This suggests experiments deliberately timing the presentation of prompts, or by using

rhythmic vibrotactile or audio feedback in such a way that the user is pushed towards tapping in the specific phase regions. This sensor-conditional feedback can be generalised, such that specific interventions can be generated in usability experiments, with a frequency which is proportional to the probability of different contexts, allowing users to ‘interact in the wild’ while retaining an increased level of experimental control.

The effects of bias and correlation in tapping errors can be systematically compensated for in real time, improving the tapping accuracy. This information can also be used to automatically adapt screen layout to walking speed, simplifying and spreading out the targets as the speed increases.

Further to that, we have the opportunity to couple the more objective methods of measuring walking speed used in this chapter with the existing literature relating usability to the subjective use of Percentage Preferred Walking Speeds in, e.g. (Pirhonen et al. 2002). For experimental environments that are more difficult for a user to navigate (such as crowded streets), these techniques could potentially provide more information about user disturbances and behaviour. The online recognition of context or situations could be used to have more targeted experiments in realistic environments, where a particular stimulus could be presented when the sensors recognise data compatible with a pre-specified situation.

The experiment described here specifically examines user performance when walking. However, the general approach is applicable to mobile usability studies in general as a method of gaining more information about the moment to moment actions of the user. Specifically, it allows us to gain greater insight into user actions in an uncontrolled environment allowing mobile usability tests to more easily take place in more realistic, less laboratory based circumstances. This work has relevance for tasks such as text entry or menu navigation in mobile settings. While this work was tap-based, similar features might be found in button-pressing, graffiti gestures or tilt-based interaction.

ACKNOWLEDGEMENTS

AC & RM-S are grateful for the support of SFI grant 00/PI.1/C067, and the HEA funded *Body Space* project. RM-S, SB & BM are grateful for the support of EPSRC grant GR/R98105/01, and IRSCET BRG project SC/2003/271, *Continuous Gestural Interaction with Mobile Devices*.

REFERENCES

- Barnard, L., J. S. Yi, J.A. Jacko and A. Sears (2005). ‘An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices’. *Int. J. of Human-Computer Studies* 62, 487–520.
- Brewster, S. A. (2002). ‘Overcoming the lack of screen space on mobile computers’. *Personal and Ubiquitous Computing* 6(3), 188–205.
- Crossan, A., R. Murray-Smith, S. Brewster, J. Kelly and B. Musizza (2005). Gait phase effects in mobile interaction. In ‘ACM SIG CHI 2005, Portland’. pp. 1312–1315.
- Fishkin, K. P., B. Jiang, M. Philipose and S. Roy (2004). I sense a disturbance in the force: Unobtrusive detection of interactions with rfid-tagged objects. In ‘UBICOMP 2004’. pp. 268–282.
- Gemmell, J., L. Williams, K. Wood, R. Lueder and G. Bell (2004). Passive capture and ensuing issues for a personal lifetime store. In ‘ACM CARPE 2004’. pp. 48–55.

- Hinckley, K., J. Pierce, E. Horvitz and M. Sinclair (2005). 'Foreground and background interaction with sensor-enhanced mobile devices'. *ACM Trans. Comput.-Hum. Interact.* 12(1), 31–52.
- Kern, N. and B. Schiele (2003). Multi-sensor activity context detection for wearable computing. In 'European Symposium on Ambient Intelligence'. Eindhoven, The Netherlands.
- Kjeldskov, J. and J. Stage (2004). 'New techniques for usability evaluation of mobile systems'. *International Journal of Human-Computer Studies* 60, 599–620.
- Lantz, V. and R. Murray-Smith (2004). Rhythmic interaction with a mobile device. In 'NordiCHI '04, Tampere, Finland'. ACM. pp. 97–100.
- Mizobuchi, S., M. Chignell and D. Newton (2005). Mobile text entry: Relationship between walking speed and text input task difficulty. In 'MobileHCI'05'. ACM. pp. 122–128.
- Morrison, A., P. Tennent and M. Chalmers (2006). Coordinated visualisation of video and system log data. In '4th International Conference on Coordinated and Multiple Views in Exploratory Visualization, London (UK)'.
A. Morrison, P. Tennent, J. Williamson and M. Chalmers (2007), Using Location, Bearing and Motion Data to Filter Video and System Logs, Proc. 5th International Conference on Pervasive Computing , Toronto, pp109-126.
- Oulasvirta, A., S. Tamminen, V. Roto and J. Kuorelahti (2005). Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In 'CHI 2005'. ACM. pp. 919–928.
- Petrie, H., S. Furner and T. Strothotte (1998). Design lifecycles and wearable computers for users with disabilities. In 'First workshop on human-computer interaction with mobile devices, (Glasgow, UK)'. Glasgow University.
- Pikovsky, A., M. Rosenblum and J. Kurths (2001). Synchronization: A universal concept in nonlinear sciences. Cambridge University Press.
- Pirhonen, A., S. A. Brewster and C. Holguin (2002). Gestural and audio metaphors as a means of control for mobile devices. In 'Proceedings of ACM CHI 2002 (Minneapolis, MN)'. ACM Press Addison Wesley. pp. 291–298.
- Roto, V., A. Oulasvirta, T. Haikarainen, J. Kuorelahti, H. Lehmuskallio and T. Nyysönen (2004). Examining mobile phone use in the wild with Quasi-experimentation. Technical Report 2004-1. Helsinki Institute for Information Technology. P.O. Box 9800, 02015 HUT, Finland.
- Williamson, J., R. Murray-Smith and S. Hughes (2007). Shoogles: Multimodal Excitatory Interaction on Mobile Devices, Proceedings of ACM SIG CHI Conference, San Jose, 2007.
- Yi, J. S., Y. S. Choi, J.A. Jacko and A. Sears (2005). Context awareness via a single device-attached accelerometer during mobile computing. In 'MobileHCI'05'. pp. 303-306.