Grip Change as an Information Side Channel for Mobile Touch Interaction

ABSTRACT

In order to reach targets on ever increasing smartphone touch displays, users tilt and shift the device in their hands. In this work, we use this grip change as a continuous information stream, useful for detecting where the user will touch while their finger is still en-route. We show that grip change detected using standard mobile motion sensors produces similar in the air touch point predictions to techniques that use auxiliary sensor arrays, even in varying physical scenarios (e.g. interacting in a moving vehicle). Finally, we show that a model combining grip change and the resulting touch point is able to predict where users *intended* to land, lowering error rates by 41%.

Author Keywords

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

With the growing popularity of mobile devices, researchers have searched for ways of improving input accuracy on touch displays. Of note, researchers have identified a number of factors that affect touch patterns in systematic ways: hand posture used to touch a target, for example, using the dominant thumb versus two thumbs to type [1]; the location of the target on the screen [4]; and even the physical activity, such as walking [3] versus sitting at a desk, and whether or not users are interacting while carrying items in the offhand [6]. These are all significant factors that impact touch patterns and therefore touch accuracy on mobile devices. Adaptive systems take advantage of systematic patterns of user touch interaction

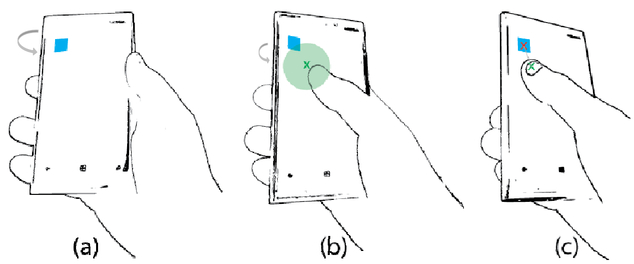

Figure 1. Grip change as the user reaches for the target (blue square): (a) device lays flat in palm, then (b) device is tilted towards the thumb by the other fingers enabling in the air prediction of an area (green circle) around a likely touch point (green X), and finally (c) at thumb touchdown, prediction can adjust actual input to intended target (red X)

under these conditions to adjust where the user touched to where they *intended* to touch. Yin et al. demonstrated improved touch accuracy using an adaptive soft keyboard that takes into consideration systematic offsets caused by hand posture and individual differences [7].

The need for these adaptations has, if anything, increased with the variability of form factors spanning compact 3.5 inch displays (e.g. Nokia M8) and phone/tablet hybrids with near-6 inch touchscreens (e.g. Samsung Galaxy Note 3). As mobile displays get larger, the ability to accurately target all around the screen suffers. Bergstrom-Lehtovirta et al. adaptively modeled user's reach to define a functional area describing the maximum reach of their current hand posture [2].

When users need to reach targets on the edge and outside their functional area, they change their grip of the phone. As Figure 1 shows, users tend to shift the device in their hand, extending the thumb's reach. As Noor et al. showed, this grip change can be used to predict the touch point [5]—or rather a likely area around it, Figure 1b in green – while the finger is still en-route, a state we refer to as *in the air prediction*. Beyond in the air prediction, grip change can also be an indicator of where the user *intended* to land at the moment they touch down—which we refer to as *on touchdown prediction*. As the target in Figure 1c is outside their functional area, even at maximum reach, the user's final touch point might miss a target but be close to one or more viable targets. The grip change motion preceding the touch is indicative of intent to reach to further targets. In other words, a model that uses the grip change and the touchdown point together, can better predict where the user intended to hit than models that only consider one of these factors.

Noor et al. detected this grip change through *adding extra hardware* – a sensor array on the back of the device [5]. The authors used the grip change—as measured while the user reaches for the target—to predict the resulting touch down point while the finger is still en route (in the air). However, we wondered whether an extraneous sensor array is critical to in the air prediction, or whether internal motion sensors (i.e. accelerometers and gyroscopes common to present devices) can achieve similar results. Given the potential predictive power of grip change on touch accuracy, it is additionally important to understand whether detection of grip change is differentially affected by under common physical scenarios such as walking or interacting while in a moving vehicle. This is not addressed in the literature.

Our work proposes using solely the motion sensors internal to present devices to detect grip change and use it to make in the air predictions of touch endpoints in four different physical scenarios (sitting at a desk, standing without support, interacting while walking, and while on a moving bus). Our prediction rates are similar to Noor et al. Additionally, we show that a model that combines both grip change and the actual touch point to adjust the final landing point lowers touch errors by 41% consistently throughout the four physical scenarios tested.

In a nutshell, grip change can be used to continuously provide information about users' intent, in the air and on touchdown, beyond where they actually touch. Predicting where the user will touch prior to touchdown allows for the potential of a virtual hover space for mobile devices, while touchdown models take into consideration user intent in lowering touch errors.

STUDY 1 – IN THE AIR PREDICTION USING MOBILE MOTION SENSORS

In order to verify that hand grip change can be reliably detected using solely mobile motion sensors (i.e. accelerometer and gyroscope) in the air, we ran a study that largely replicates Noor et al.'s touch target prediction experiment. Participants were asked to touch the single on-screen square target to complete a trial. The target size was chosen to be physically identical to Noor et al.'s setup on the mobile device ($1cm^2$). The target was randomly placed on the surface of the display for each trial with a uniform distribution. As in Noor et al., once the user successfully clicked on the target, the next target was displayed after an enforced 500ms delay.

Participants and Apparatus

Eight participants (mean age 25.7, 7 males), all right-handed, were asked to perform 1000 trials each (8000 trials in all), using the phone in their right hand and interacting with the thumb (i.e. a one-handed posture). All participants were regular and frequent mobile phone users

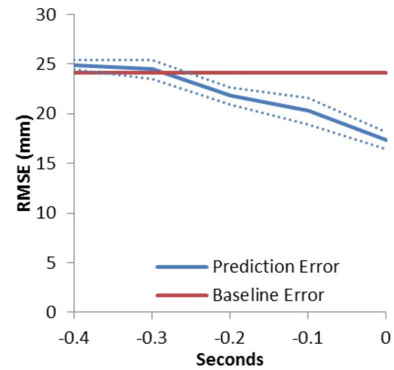

Figure 2. Root mean square error (and ± standard error in dashed lines) of touch point prediction before touch contact using mobile sensor detected hand grip change.

Participants were presented with a custom application on a Nokia Lumia 920 with a 4.5" display at a 1280x720 resolution. As the trial start button always covered the bottom 14% of the device's touchscreen the total application area was instead 1094x720. Of note is that this device is significantly larger than the one used in Noor et al. (a Nokia N9 with a 3.9" display), which makes a direct comparison of prediction accuracy non-trivial. The application recorded 2D touch points on the surface of the display as well as 3-axis accelerometer and gyroscope readings at a sampling rate of 50Hz.

Results

In order to predict where participants touched based on hand grip change, we used the raw, time ordered accelerometer and gyroscope measures as features. As did Noor et al. [5], we used Gaussian Process (GP) Regression. It generated a nonlinear model that maps motion sensor values to 2D screen coordinates. Due to individual differences in how users grip and move the phone in their hand, the model was created on a per-user basis. Participant data was split between training and test trials using 10-fold cross validation.

As Figure 2 shows, the resulting models (in blue) provide a root mean square error (*RMSE*) smaller than a baseline model that simply picks the center of the available area for its prediction (in red). Predicting 0.1s before the touch provides a RMSE of 20.3mm, and 21.8mm at 0.2s. Though our much larger device makes comparison difficult, our grip change model provides competitive results to Noor et al.'s 18mm (at 0.2s before touch) but without the added external sensor array. The ability to predict a touchdown area while the finger is still en route opens up the possibility of developing a hover space for touch interaction and improved feedback.

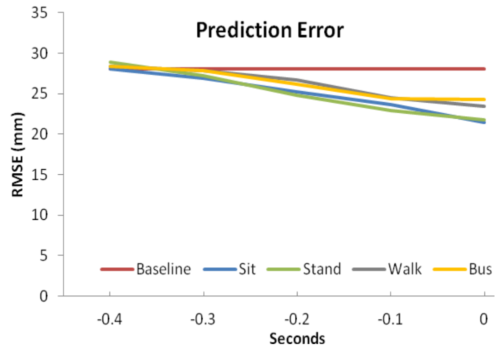

Figure 3. Root mean square error for the predicted touch points in Experiment 2's four different physical conditions.

STUDY 2 – GRIP CHANGE ON THE MOVE

The second experiment had two goals: (1) show how in the air prediction using grip change detected from mobile motion sensors performs in a more representative task and under a variety of physical conditions (e.g. while walking); and (2) show how grip change can provide information on users' intent.

Methodology

The experiment was designed to gather touch and motion data from participants in a common task from every-day device use: selecting a particular target from an array of homescreen icons. In contrast to the first study, the participant selected one target, highlighted red, out of the always-visible array of 5x4 targets that simulate a Google Android and Apple iOS homescreen. Each target was sized 99x99 pixels, which corresponds to the Google Android recommended size for icons for the Lumia 920's resolution and physical size [8].

In order to see how well a grip change prediction model works under common motion scenarios, participants performed the study in four physical conditions:

a) *Sit*: seated at a table, with their phone-handling hand partially resting on the table for stability

b) *Stand*: free standing without any stability support

c) *Walk*: walking around a closed outdoors course at a constant, researcher defined comfortable walking speed

d) *Bus*: seated on a moving bus

Each trial highlighted one out of the 20 targets available at random with a uniform distribution. In order to simulate realistic, continuous interaction, the application moved to the next trial as soon as the participant touched the screen, without the enforced delay of the first study. The study was designed to be approximately 1 hour in duration, which allowed 200 trials per physical condition. During the Bus scenario, data was only collected while the bus was moving – participants were asked to pause while the vehicle was stationary.

Participants and Apparatus

In all, 12 participants were recruited from a local university (mean age 26.1, 7 males, all right handed). All completed the study for a total of 9600 touch trials (200 trials for each of the four conditions per participant) using a one-handed posture. All participants were regular and frequent mobile users.

The custom experiment application ran on the same Nokia Lumia 920 device as in the first study. The application collected 2D touch coordinates for each trial, and 3-axis accelerometer and gyroscope readings sampled at 50Hz.

Results – In the Air Prediction on the Move

In order to verify whether target prediction is possible in different physical conditions, we once again trained user-specific GP regression models mapping mobile motion sensor sequences to touch coordinates. Figure 3 shows touch prediction results using 10-fold cross validation for each condition, with the baseline model – always guessing in the center of the application area – shown in red. Note that the baseline error is higher (28mm) than the first experiment due to the static set of available targets to select from. First, we note that, as in Study 1, the prediction models created in the stationary conditions (Sit and Stand) have a lower RMSE than the baseline model. At 0.1s before touch, the GP regression model created in the Sit condition had a RMSE of 23.7mm (with Stand at 22.9mm), still lower than the baseline model (28.3mm).

Specifically comparing the results between the two studies shows that Study 1's was better; it produced an RMSE of 20.3mm at 0.1s before touch, whereas in Study 2 for Sit it was 23.7mm. This can be explained by our more conservative approach in Study 2. We used a more realistic task and did not enforce a delay between trials; at times, more than one target was clicked within the 500ms prior to touch down. Due to this lack of clear segmentation and a more limited data set (200 trials per user, per condition) prediction RMSE in the Sit condition is higher than the first study. This more realistic scenario shows how in the air prediction using grip change degrades in a freeform homescreen task.

Lastly, unpredictable motion such as interacting in a bus, leads to higher prediction RMSE (24.4mm at 0.1s before touch) than sitting at a table (23.7mm at 0.1s), though critically, still lower than baseline models (28.3mm). This shows that, even with a relatively small amount of prior data, touch prediction using hand grip change is viable in high motion, unpredictable physical conditions.

Results – On Touchdown Prediction of Touch Intent

The secondary goal of the study was to verify whether grip change can improve touch performance by providing insight into targeting intent (i.e. on touchdown).

In order to separate user intent from their inherent touch patterns, we first built a baseline regression model

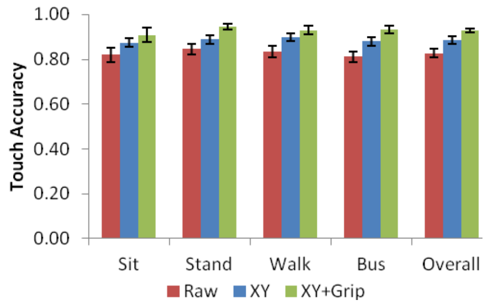

Figure 4. Touch accuracy comparing the three models per physical condition; Overall aggregates across all the data.

mapping the 2D touch coordinates to the center of the highlighted target (referred to as the *XY Model*). We use the center of the highlighted target as an approximation of where the user intended to touch. This model takes into consideration individual touch behavior but uses no other information other than touch down to infer user intent.

Next, we built a GP regression model that uses both where the user touched, and the hand grip change as measured by the motion sensor readings of the 500ms prior to touch down (*XY+Grip Model*). Once again, the model attempts to map these features to the center of the highlighted target. We test the two models against the raw touch points (*Raw*). The models' goal is to improve on the Raw accuracy rates by inferring where the user intended to touch, rather than where they actually touched. To build and test the models, we use 10-fold cross validation.

An analysis of variance with *Model* (XY, XY+Grip, Raw) and *Physical Condition* (Sit, Stand, Walk, and Bus) as within-subject factors found a significant main effect of *Model* on touch accuracy ($F_{2,22} = 25.162$, $p < 0.001$). Bonferroni correction showed significant differences between all pairs of *Models*. As Figure 4 shows, participants had a Raw accuracy of 83% over all physical conditions. The XY Model which took into consideration user-specific touch behaviours improved accuracy to 88%. Our XY+Grip Model further boosted accuracy to 93% consistently over all physical conditions – a 41% reduction in error rates over raw touch performance.

In summary, hand grip as measured by the phone's motion sensors right before a touch provides valuable insight into where the user intended to touch. Leveraging this information in addition to touchdown XY coordinates significantly improves touch accuracy when compared to a model that adjusts based solely on the touch down point.

CONCLUSIONS AND FUTURE WORK

In contrast to more static models that take into consideration individual touch behaviour (e.g. hand posture), we consider grip change as a continuous information stream, valuable at multiple points in the interaction sequence. In the air, grip change detected using standard mobile motion sensors can help estimate a landing point which may be used as a virtual hover space and provide interactive support—e.g. constant feedback of a user's landing point. On touchdown, grip change that preceded a touch point can be a useful predictor of where users intend to hit, rather than where they landed. Our results show that, although performance degrades in high motion scenarios, hand grip change can make reasonable landing predictions and reliably improves touch accuracy, reducing error rates by 41%.

Importantly, our study trained a model to infer intent by having full knowledge of the intended target. A question remains of how well such models can be trained and perform *online,* with some user clicks being unintended (i.e. an erroneous click relative to their intended target).

Furthermore, as our in the air and touchdown models are adaptive, they are not immune from a potential user feedback loop. Our models were evaluated offline, and it remains to be seen how users will adapt their touch when encountering a system that is itself adapting based on inferred intent (i.e., online). Continuous visual feedback of the model's predicted touch point may be critical to help users correct their targeting motion.

Finally, grip change was detected through data gathered on a relatively large smartphone. While the current trend is one of ever increasing displays, more research is needed to verify how grip change varies on different phone-to-hand ratios, and whether smaller devices have detectable and useful grip variations.

In conclusion, our studies show that grip change is a promising information side channel that is detectable with internal motion sensors, reliable to physical movement, and makes valuable predictions both in the air and on touchdown to improve touch interaction.

REFERENCES

1. Azenkot, S. and Zhai, S. Touch behavior with different postures on soft smartphone keyboards. *Proc. of MobileHCI'12*, ACM (2012), 251–260.
2. Bergstrom-Lehtovirta, J. and Oulasvirta, A. Modeling the Functional Area of the Thumb on Mobile Touchscreen Surfaces. *Proc. of CHI'14*, ACM (2014), 1991–2000.
3. Goel, M., Findlater, L., and Wobbrock, J. WalkType: using accelerometer data to accomodate situational impairments in mobile touch screen text entry. *Proc. of CHI'12*, ACM (2012), 2687–2696.
4. Henze, N., Rukzio, E., and Boll, S. 100,000,000 taps: analysis and improvement of touch performance in the large. *Proc. of MobileHCI'11*, ACM (2011), 133–142.
5. Mohd Noor, M.F., Ramsay, A., Hughes, S., Rogers, S., Williamson, J., and Murray-Smith, R. 28 Frames Later: Predicting Screen Touches from Back-of-device Grip Changes. *Proc. of CHI'14.* ACM (2014), 2005–2008.
6. Ng, A., Brewster, S.A., and Williamson, J.H. Investigating the Effects of Encumbrance on One- and Two- Handed Interactions with Mobile Devices. *Proc. of CHI'14*, ACM (2014), 1981–1990.
7. Yin, Y., Ouyang, T.Y., Partridge, K., and Zhai, S. Making touchscreen keyboards adaptive to keys, hand postures, and

individuals: a hierarchical spatial backoff model approach. *Proc. of CHI'13*, 2013, 2775–2784.
8. Iconography
http://developer.android.com/design/style/iconography.html.