Whack Gestures: Inexact and Inattentive Interaction with Mobile Devices

Scott E. Hudson[1] **Chris Harrison**[1] **Beverly L. Harrison**[2] **Anthony LaMarca**[2]

[1] Human-Computer Interaction Institute
Carnegie Mellon University, Pittsburgh, PA 15213
{scott.hudson, chris.harrison}@cs.cmu.edu

[2] Intel Labs Seattle
Seattle, WA 98105
{beverly.harrison, anthony.lamarca}@intel.com

ABSTRACT
We introduce *Whack Gestures*, an inexact and inattentive interaction technique. This approach seeks to provide a simple means to interact with devices with minimal attention from the user – in particular, without the use of fine motor skills or detailed visual attention (requirements found in nearly all conventional interaction techniques). For mobile devices, this could enable interaction without "getting it out," grasping, or even glancing at the device. This class of techniques is suitable for a small number of simple but common interactions that could be carried out in an extremely lightweight fashion without disrupting other activities. With *Whack Gestures*, users can interact by striking a device with the open palm or heel of the hand. We briefly discuss the development and use of a preliminary version of this technique and show that implementations with high accuracy and a low false positive rate are feasible.

Author Keywords
Physical Interaction, Sensing, Lightweight Interaction, Adaptive User Interfaces, Gesture-Based Interfaces.

ACM Classification Keywords
H5.2 [Information interfaces and presentation]: User Interfaces.

INTRODUCTION
Advances in small and low-powered electronics have created new opportunities for mobile computing, generating an explosion of useful new devices. These devices have tremendous potential to bring the power of computation and communication to a wider audience and to more aspects of our lives. However, with this potential comes new challenges for interaction design. Individual device use must be balanced with other activities and, in general, be appropriate in a wide variety of contexts, both environmental and social. Without new interaction approaches that are attuned to this emerging modality, we are likely to increasingly experience the kind of aggravation typified by cell phones ringing in meetings and theaters. Much of the potential of mobile devices could be lost to their human costs.

TEI 2010, January 25–27, 2010, Cambridge, Massachusetts, USA.

In this note, we describe a simple example of a larger category of techniques we call inexact and inattentive interactions. In situations where interaction with a device represents a disruption from another activity (e.g., disturbing a social situation, such as a conversation with a coworker, or an attentionally demanding activity like driving), it is important that basic interactions can be carried out with minimal attention demand so that users can remain focused on their primary task or situation. To achieve this, these interactions need to avoid physical actions requiring, for example, fine motor control or continuous visual attention – characteristics typical of most conventional interaction techniques. In short, such interactions can benefit from being inexact in nature – requiring only gross motor control and little or no visual assistance.

As part of our investigations within this category, we developed a simple interaction technique called *Whack Gestures*. This technique allows a small number of very simple but common interactions to be handled with gross movements including, for example, coarsely striking ("*whacking*") a mobile device. This interaction can be achieved without "taking out" the device (e.g., removing it from a belt-clip or pocket). Indeed, users need not even grasp or glance at the device.

We were motivated to explore inexact and inattentive interaction because of problems that arose in previous work using the experience sampling method (ESM) [1,2,6,11,12]. This technique allows in-situ information to be gathered about subject's everyday activities by periodically asking them to fill out a survey – often administered with a mobile device such as a PDA or cell phone. It has been instrumental in collecting training data to drive the creation of learned statistical models of user activities [4,10,13,14]. However, both observation of our subjects and personal experience from serving as pilot subjects indicated that responding to even very short questionnaires (e.g., one question with only a few possible responses) could be remarkably annoying and difficult. This limits responsiveness and/or increases the costs incurred in using this technique, and makes it difficult to use in deployed products.

Our informal examination of this issue found that even the act of getting out the device to turn it off was disruptive in settings like meetings, and subjects often felt compelled to remark on it or apologize. Similar behavior is readily ob-

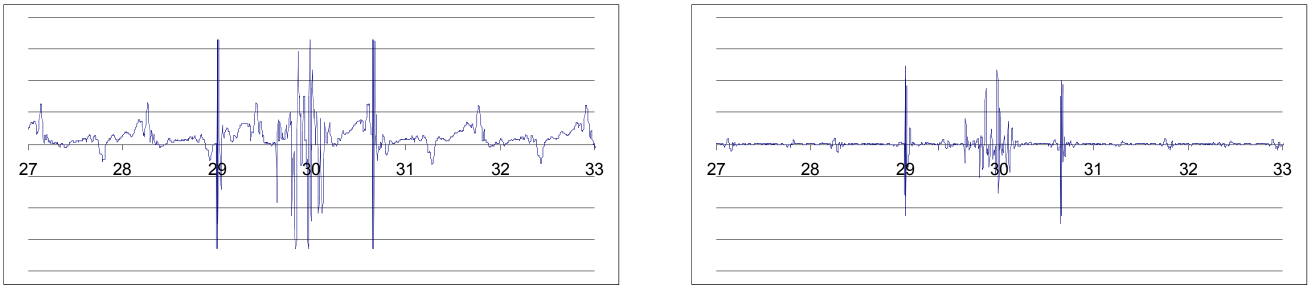

Figure 1. Whack gesture accelerometer data captured while walking: Z-axis raw data (left) and "change from background" feature (right). Horizontal scale in seconds.

served for people who receive cell phones calls in similar settings. This indicated a difficult design challenge since our target level of attention demand for the interaction had already been exceeded before the interaction had really begun. This led us to consider radically different approaches, including schemes that involved quickly manipulating a device without retrieving it or orienting it sufficiently to find and press a button, and eventually to the specific approach described here.

While we were initially motivated by ESM applications, we believe other interactions could also benefit from this approach. Perhaps most compelling of these is quickly responding to (or silencing) a ringing cell phone. Another scenario is ambient information displays, where users might optionally respond to the device by simple actions such as dismissal, deferral, or delegation – saying in essence "yes I've seen that, you can move on", "show that to me again at a better time", or "send that to my assistant." Such simple responses could make ambient information displays much more useful, but only if they were minimally disruptive. Finally, in continuous recording situations, these simple interactions might be useful for a "remember this" (or privacy preserving "don't keep the recording of that") signal.

For these exemplar applications, a small number of stylized responses are sufficient. For example, consider a cell phone in conjunction with custom ring tone - responses might be "shut off ringing" and answer, then play one of two "please hold" messages. In our ESM work, we have often used questions with only two or three possible responses. While these interactions are simple and limited, they still cover many common cases for applications. It is important to note that the type of interaction described here is intrinsically limited in its total expressive power because of the nature of the physical actions used, and more generally, the limits on attention demand motivating them. Also, the more complex operations become, the more likely they are to be efficiently accomplished with button presses.

WHACK GESTURES

We designed a small vocabulary of gestures intended to interact with a small mobile device worn at the waist, e.g., in a holder attached to a belt. While this clearly does not cover all of the ways such devices are carried, it does provide an initial feasibility test for the approach. Other locations, such as in the pocket or in the bag or purse, are also possible, and would likely only marginally affect the quality of the accelerometer-driven data we use for gesture recognition. Gestures are performed by firmly striking the device, e.g. with an open palm or heel of the hand moved towards the waist – an action we refer to simply as a *whack*. We feel this interaction is both simple and compelling for users, and is confirmed in a video scenario and survey study [16], where this form of interaction was rated most positively among six alternative interaction styles.

Whacks were also selected in part because they can be recognized in a straightforward and high-accuracy manner. Figure 1 shows accelerometer data (captured while walking) that illustrates this clarity – the distinctive leading and trailing sharp peaks are whacks. Furthermore, accelerometers are becoming increasingly prevalent in mobile devices, and have been shown to be well suited for this style of interaction (e.g., [5,7,18]).

We can expect any mobile device to be bumped or jostled fairly often. As a result, a central concern – in addition to factors such as intuitiveness and ease of physical action – will be false positive invocations where ordinary jostling is incorrectly recognized as a gesture. For example, a preliminary implementation of a related technique described in [16], reports "on average, 1.5 false taps per minute" while walking. Such false positives are a major concern with mobile devices, even for conventional button-based interactions. For instance, many mobile devices default to automatically lock the keys after a preset period of inactivity. Even with this precautionary measure, there are many anecdotal reports of what people's cell phones "have done all by themselves" while stored in a user's pocket or purse.

Whacks alone, however, do not provide a very expressive gestural design space (one tends to devolve into Morse-code-like interactions). Further, even with high accuracy recognition of individual whacks, we believe false positives will be a concern because "bumps" with similar acceleration profiles can be expected occasionally. To counter this, our design uses a pair of whacks, which serve as a framing feature. The user performs the sequence <whack> <signal gesture> <whack> for one of several possible signal gestures. For the gesture to be accepted, the framing whacks must being of comparable magnitude and occur within a limited time window of each other. After consideration of a range of alternatives, we settled on an alphabet of three specific gestures to evaluate in this initial work:

whack-whack — Using an empty signaling gesture.

whack-whack-whack — Using an extra whack as the signaling gesture.

whack-wiggle-whack — Using a quick shaking motion as the signaling gesture (this produced the data for Figure 1).

RECOGNITION AND CLASSIFICATION

To develop a proof of concept implementation of this technique, we used accelerometer data collected by a Mobile Sensor Platform (MSP) [13] (see Figure 2). The MSP is a pager-sized, battery powered computer with sensors chosen to facilitate a wide range of mobile sensing applications. The MSP has ten sensors including: 3-D accelerometer, barometer, humidity, visible light, infrared light, temperature, 44kHz microphone, and compass. It includes a low power XScale microprocessor, 32MB RAM, 2GB of removable flash memory for storing programs and logging data, and a Bluetooth radio. We selected this platform due to its easy programmability, large non-volatile memory capacity, and appropriate form factor.

For our experiments, we sampled the three axes of the accelerometer 256 times per second and recorded the raw sensor values on the flash card for later analysis. When attached to a belt as shown in Figure 2, the MSP's accelerometer has its z-axis pointing into the user’s waist, y-axis pointing down, and x-axis pointing forward. For this implementation we use this case aligned coordinate system. For other applications, one could use a gravity-aligned coordinate system (using a “gravity tracking” feature already implemented in the platform software), or employ a coordinate system independent set of features.

To recognize Whack Gestures, a simple ad-hoc recognizer was developed. To emphasize large z-axis magnitude changes, a “change from background” feature is computed by subtracting an exponential decaying average (with a decay rate of ½ per time step) from the raw z-axis data. Figure 1 (right) shows magnitude of this feature computed from the raw data shown (left). Whacks are characterized by short, high-intensity pulses within this feature. The threshold for detection of a whack peak is set to two standard deviations below the user’s mean whack magnitude, provided during a training phase (six or fewer samples are needed). Once a pulse of appropriate magnitude is detected, the next 300ms are monitored, and the highest encountered magnitude is recorded as the pulse’s true peak. A framing whack must occur within three-seconds and can vary no more than ±33% in intensity from the opening whack. If more than one whack is detected within the three-second period, the final one is considered the closing gesture. As a final filter, we compute a simple energy measure between framing whacks and discard frames which are above a threshold set to detect periods of high activity or noise (outside the range of the signaling gestures in use).

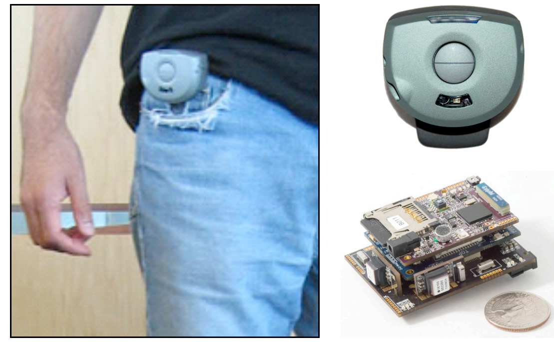

Figure 2. The Mobile Sensor Platform

Once a candidate gesture frame has been identified, the data is passed to a secondary recognizer. For the *whack-wiggle-whack* gesture, we use a simple energy measure looking at the average absolute acceleration within the signal frame. The *whack-whack* and *whack-whack-whack* gestures are recognized by counting peaks (two and three respectively), which must be separated by no less than 200ms to be considered discrete. If the signal does not conform to one of the three supported gestures, the frame is ignored.

EVALUATION

To provide a preliminary test of the effectiveness of our proof-of-concept implementation, we collected data from a set of 11 volunteer participants (6 female, mean age 29.4). Each participant was given the MSP, instructed on how to attach the device, and told to simply go about their normal routine while wearing it for a period of two hours. This provided 22 hours of baseline data. Following this collection period, each participant was given very brief demonstration on how to perform the three test gestures. Participants then performed these gestures three times each, which was recorded for later analysis.

We measured two accuracy outcomes from this data. First, we considered the number of false positive recognitions within our baseline data (where our participants had presumably not intended to perform any gestures). Then we consider the (true positive) recognition rate for the gestures known to be demonstrated by the subject. To provide a fair test, we set aside the data from six randomly selected subjects (our holdout test set) and did not consider it while developing our recognizers, e.g., to pick specific thresholds. However, it is important to note that our recognizers are not tightly bound to the specifics of any user or data set.

A vital question to be resolved in the evaluation of this technique was its ability to avoid false positive activations. At first, we looked at just the accuracy of detecting the framing gestures, independent of the signaling gesture that occurred in between. This allows us to gauge how robust the technique might be when used with other signaling gesture alphabets. When measuring false positives occurring in our holdout test set (6 participants), we found one false positive (yielding a rate of one false positive per 12 hours of in-the-wild use). No false positives were found within the remaining subjects, giving an overall rate of one false positive in 22 hours of use. (We note that the single false positive occurred during the first two minutes of recording and was likely associated with the subject attaching the device. However, we left these periods in our data since real devices will likely need to operate in similar conditions.) Our recognizers correctly classified 100% of the known framing gestures (performed at the end of the collection period) from both the testing and holdout data sets.

When considering accuracy for the full recognizer, including classification of the signaling gesture, we found that the gesture recognizer was unable to reject the single false positive framing from our baseline test set, instead recognizing it as a *whack-whack* gesture. Our recognizer correctly classified all but one *whack-wiggle-whack* signal gesture in our positive test set (seeing it as a *whack-whack instead*) resulting in an overall true positive rate of 97%.

RELATED WORK

Whack Gestures builds on some of the ideas put forth in Ronkainen et al. [16]. Much of the latter work is dedicated to the idea of providing "tap" input while the device is located in the user's hand (using fingers on the other hand for tapping). Orientation is important, as taps on different sides of the device control different functions (e.g., increase/decrease volume). There is also discussion of free space gestures, such as writing an 'X' to activate silent mode. Far briefer is the discussion of simple, omnidirectional physical actions for triggering special functions, such as the ability to silence an incoming call. It is this initial idea that we build upon and more rigorously evaluate.

One of the key findings in Ronkainen et al. was that the proposed *tap* and *double-tap* driven interactions were extremely sensitive to false positives (90 per hour while walking). Our intensity-matched framing gestures are a direct solution to this central issue, and reduce the false positive rate by three orders of magnitude. Furthermore, the authors noted single tap false positives were so high, they discarded the gesture from their second study entirely, leaving *double-tap* as the sole gesture for all functionality, heavily limiting the technique's utility. Our solution, however, enables considerable gestural expressivity beyond *taps* and *double-taps* through the use of inter-whack signal gestures, such as wiggles, twists, taps, and flicks.

Numerous other systems have also made use of motion-based gestures using accelerometers or other sensors (see e.g., [5,7,18]). Course shaking gestures or contact gestures (sensed in unison from two sensors) have been used to establish proximity and/or spatial relationships [8,9,15].

The Shoogle system [17] is an interesting example of an inexact and inattentive interaction technique for output. The system simulates a set of virtual balls appearing to be contained within the device. Upon shaking the device, users can infer simulated properties, such as number, mass, and material of the balls. This can be used to express, e.g., the number of outstanding SMS messages. This information exchange can occur with simple, gross motor actions and little visual attention.

CONCLUSION

We introduced *Whack Gestures*, an example of a category of techniques we call inexact and inattentive interaction. Our approach uses coarse physical actions as a mechanism for signaling mobile devices in an extremely lightweight fashion - one where users do not have to "take out the device," use fine motor control or apply visual attention.

ACKNOWLEDGMENTS

This work was supported in part by grants from the Intel Research Council and the National Science Foundation under Grant IIS-0713509.

REFERENCES

1. Barrett, L.F., and Barrett, D.J. (2001). An Introduction to Computerized Experience Sampling in Psychology. *Social Science Computer Review*, *19(2)*, 175-185.
2. Consolvo, S. and Walker, M., (2003). Using the Experience Sampling Method to Evaluate Ubicomp Applications. *IEEE Perv. Comp. Magazine: The Human Experience, 2(2)*, 24-31.
3. Consolvo, S., Harrison, B. L. Smith I, Chen, M. Everitt, K., Froehlich, J., and Landay, J. (2008). Conducting In Situ Evaluations for and with Ubiquitous Computing Technologies. *IJHCI*, 22 (1 & 2), 103-118.
4. Fogarty, J., Hudson, S., Atkeson, C., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J., and Yang, J. (2005) Predicting Human Interruptibility with Sensors. *ACM TOCHI* 12(1), 119-146.
5. Harrison, B. L., Fishkin, K. P., Gujar, A., Mochon, C., and Want, R. (1998). Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. In *Proc. CHI '98*, 17-24.
6. Hektner, J. M., and Csikszentmihalyi, M. (2002). The experience sampling method: Measuring the context and content of lives. In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology*, 233-243. New York: John Wiley & Sons, Inc.
7. Hinckley, K., Pierce, J., Sinclair, M., and Horvitz, E. Sensing techniques for mobile interaction. In *Proc. UIST '00*, 91-100.
8. Hinckley, K. (2003). Synchronous gestures for multiple persons and computers. In *Proc. of UIST '03*, 149-158.
9. Holmquist, L. E., Mattern, F., Schiele, B., Alahuhta, P., Beigl, M., and Gellersen, H. Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artefacts. In *Proc. Ubicomp '01*, 116-122.
10. Horvitz, E., Koch, P., and Apacible, J. BusyBody: creating and fielding personalized models of the cost of interruption. In *Proc. CSCW '04*, 507-510.
11. Intille, S., Rondoni, J., Kukla, C., Ancona, I., and Bao, L. Context-aware experience sampling tool. In *Proc. CHI '03,* 972-973.
12. Larson, R. and Csikszentmihalyi, M. (1983). The Experience Sampling Method. *Naturalistic Approaches to Studying Social Interaction: New Directions for Methodology of Social and Behavioral Science*. San Francisco, CA: Jossey-Bass.
13. Lester, J., Tanzeem Choudhury, T. and Gaetano Borriello, G. A Practical Approach to Recognizing Physical Activities. In *Proc. Pervasive '06*, 1-16.
14. Munguia-Tapia, E., Intille, S. S. and Larson, K. Activity Recognition in the Home Using Simple and Ubiquitous Sensors. In *Proc. Pervasive '04*. 158-175.
15. Patel, S. N., Pierce, J. S., and Abowd, G. D. A gesture-based authentication scheme for untrusted public terminals. *Proc. UIST '04*, 157-160.
16. Ronkainen, S., Häkkilä, J., Kaleva, S., Colley, A., and Linjama, J. Tap input as an embedded interaction method for mobile devices. In *Proc. TEI '07*, 263-270.
17. Williamson, J., Murray-Smith, R., and Hughes, S. Shoogle: excitatory multimodal interaction on mobile devices. In *Proc. CHI '07*, 121-124.
18. Wilson, A. and Shafer, S. XWand: UI for intelligent spaces. In *Proc. CHI '03*, 545-552.