PVT-Touch: Adapting a Reaction Time Test for Touchscreen Devices

Matthew Kay[*], Kyle Rector[*], Sunny Consolvo[†], Ben Greenstein[*], Jacob O. Wobbrock[†], Nathaniel F. Watson[§], Julie A. Kientz[¶]

[*]Computer Science & Engineering, [†]Information School, [§]Neurology, [¶]Human-Centered Design & Engineering

DUB Group, University of Washington

Seattle, WA

{mjskay, rectorky}@uw.edu, {sunnyinseattle, bengreenstein}@gmail.com, {wobbrock, nwatson, jkientz}@uw.edu

Abstract—**The Psychomotor Vigilance Task (PVT) is a validated reaction time (RT) test used to assess aspects of sleep loss including alertness and sleepiness. PVT typically requires a physical button to assess RT, which minimizes the effect of execution time (the time taken to perform a gesture) on RT. When translating this application to mobile devices, a touchscreen version is useful for widespread in situ sleepiness assessments to produce more ecologically valid data. We describe the Android-based implementation of a touchscreen version of PVT, called PVT-Touch. In an evaluation (N=20), we compared four different touchscreen input techniques to a physical button: *touch down*, *finger lift*, *finger tilt*, and *goal crossing*. We found that *touch down* was comparable to the physical button approach used in traditional PVT in execution time and in several measures associated with sleepiness, and was preferred by most participants. We also found that *finger lift* may be a more precise but less intuitive measure, which may warrant further study.**

Keywords—reaction time, psychomotor vigilance task, PVT, sleep, health assessment, health, touchscreen, personal informatics.

I. INTRODUCTION

The Psychomotor Vigilance Task (PVT) is a reaction time (RT) test commonly used by sleep clinicians and researchers to assess a person's alertness. During the test, a visual stimulus is shown at random intervals, and the user presses a single physical button in response to each stimulus over a 5–10 minute period. Reaction times are reported in milliseconds, and various statistical summaries of these times have been shown to correlate well with wakefulness [1], [2]. Thus the test is often used to assess sleep deprivation and fatigue for different populations and as part of an overall picture of health and wellbeing.

The original implementation of PVT was on purpose-built hardware designed for lab use only [1]. However, researchers are moving toward mobile implementations in order to assess sleepiness in everyday life and to gather more ecologically valid data. An implementation for Palm OS [3] that has users press a physical button on a PDA is currently the most widely adopted version. Unfortunately, Palm OS devices are no longer being manufactured as of 2009[1], so use of this version is unsustainable for researchers and requires users to carry an additional device. Implementations for more current mobile touchscreen platforms, such as iOS and Android, would allow these assessments to be used in research studies and for con-

---

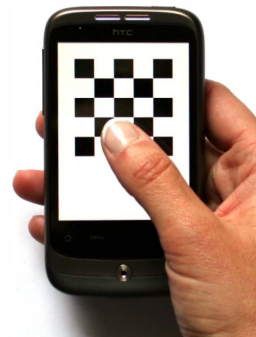[1] http://www.brighthand.com/default.asp?newsID=14896

Fig. 1. PVT-Touch running on an Android phone, with stimulus visible. The checkerboard pattern provides a high-contrast, obvious signal to participants.

sumer-level personal informatics applications; in addition, integration of a well-validated sleepiness assessment into existing mobile health frameworks (such as [4]) would also offer exciting possibilities for connecting sleep to overall health.

While PVT is a fairly simple test, the translation to smartphones is not as straightforward as it might seem, as precise timing (down to ~10 milliseconds) is important to sleep experts. Traditional implementations have used devices with a physical button: users rest their thumb on the button and depress it when they see the stimulus. This minimizes variability introduced by *execution time*: the time taken to target and press the button. However, modern smartphones often do not have a centrally-placed button; physical buttons on recent Android phones are usually limited to volume or camera buttons placed along the edge of the phone. We used an older phone model with a central button for our tests, but these models are limited. Side buttons may be uncomfortable for either left or right handed users to press with the thumb of their dominant hand, as PVT is typically administered [3]. By contrast, touch-based interaction is increasingly becoming the primary, most familiar, and most comfortable mode of input on modern smartphones; thus, we consider it worthy of investigation for use with PVT.

Given the affordance and timing differences between a touchscreen and a physical button, we investigated several touchscreen-based input techniques to determine if reaction times are comparable. In particular, we wished to determine which techniques minimize execution time in order to best approximate physical button timing. To do so, we developed an implementation of PVT for Android-based smartphones called

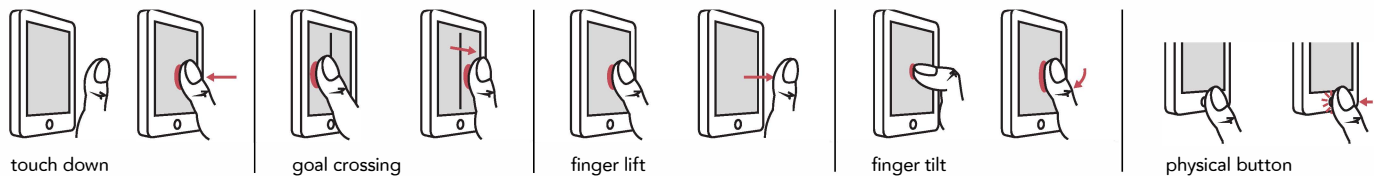| touch down | goal crossing | finger lift | finger tilt | physical button |

Fig. 2. Diagram of the 4 touchscreen input techniques and the physical button.

PVT-Touch, and conducted an in-lab test to compare four touchscreen input techniques against a physical button: *touch down*, *finger lift* [5], *finger tilt*, and *goal crossing* [6].

The contributions of this paper are: (1) a new Android implementation of PVT that can assess sleepiness on consumer-level devices, (2) the results of an evaluation comparing PVT-Touch to a traditional physical button test, and (3) a comparison of execution times across four touchscreen input techniques. This work has implications for those interested in sleep, in-situ health assessments, and reaction time for touchscreen interfaces. A validated reaction time test may also have implications for mobile games or assessing users' phone use habits.

II. DESIGN & IMPLEMENTATION OF PVT-TOUCH

In this section, we summarize the implementation of PVT-Touch and describe the four input techniques we evaluated.

A. Implementation

As with traditional PVT, a single PVT-Touch *test* consists of multiple *trials* run back-to-back. In each trial, the screen starts out blank (white). After a random delay (the *foreperiod*), a high contrast checkerboard pattern (Fig. 1) appears, at which point the participant provides a response (e.g., pressing a button on the phone or touching the screen). Our implementation uses the 5-minute version of PVT with random foreperiods from 2 to 10 seconds [7]. The time between the stimulus appearing and the response is the participant's reaction time (RT) for that trial. A number of metrics (e.g., median RT) are used to assess the overall results of a test, detailed later.

Because human reaction time can be fairly quick (~250–500 ms), the ability to measure timing accurately is critical to PVT. Hardware used with previous versions, such as PalmPVT, had a timing resolution of ~10 ms [3], less than 5% of typical reaction time. We ran timing experiments with our implementation on three Android phones running versions 2.1, 2.2, and 2.3. We found these phones had a timing resolution of ~2 ms, suggesting Android is a suitable platform for PVT.

We also took great care to reduce the effects of other sources of latency on our measurements. Profiling of our implementation suggested the primary sources of latency were graphics drawing latency and garbage collection (GC). Graphics latency was kept low (<2ms) by obtaining a dedicated drawing surface on the screen that we update outside of the graphical toolkit's main event loop. In versions prior to 2.3, the Android GC caused pauses of up to 130ms in our timing experiments (2.3 introduced a concurrent garbage collector, fixing this problem). For versions prior to 2.3, we heavily optimized memory use to ensure that the GC runs at most once during a 10-minute test, in which case the single affected trial can be dropped without skewing results.

B. Alternative User Input Techniques

We considered several touchscreen-based techniques as alternatives to a physical button (see Fig. 2):

Touch down. When the stimulus is shown, the user touches anywhere on the screen. This is the most commonly used technique in touch-based user interfaces.

Finger lift. The user holds their thumb down on the screen. When the stimulus is shown, they lift it off [5]. While uncommon, we included this technique because we hypothesized it may have the lowest execution time.

Goal crossing. [6] The user rests their thumb on either side of a dividing line. When the stimulus is shown, they swipe their thumb to the other side. We divide the entire screen vertically to minimize targeting.

We also considered a pressure-based input technique, but none of the phones we tested had pressure sensitivity, making it inappropriate for current widespread deployment. By contrast, capacitive touch screens can report the area of the contact surface, suggesting the following technique in place of pressure:

Finger tilt. The user places the tip of their thumb on the screen. When the stimulus is shown, they tilt the thumb so that the pad of the thumb contacts the screen. When the size of the contact surface crosses a threshold, a response event is registered. We used a fixed threshold that was selected through piloting the application with several users. While a dynamic threshold may perform differently, we allowed a training period for users to become accustomed to each technique, so we do not believe this would substantially affect our results.

These touch-based techniques were compared against the *physical button* technique due to its use in traditional versions of PVT [1], [3]: the user is instructed to rest their thumb on the button and then press down when the stimulus is shown.

III. EVALUATION

We wish to choose a technique for PVT with low execution time (the time taken to perform the gesture)—similar to the physical button. PVT-Touch measures reaction time (RT) as the time from when the graphics buffer is flipped to the time an input event is received by the Android stack, using the internal timestamp attached to that event. This definition of RT thus includes additional hardware/software related delay, consistent with previous PVT implementations [3]. Because any delay introduced by the graphics hardware and the touchscreen should be the same across techniques, we expect the primary cause for differences in measured RT between these techniques to be due to execution time. We therefore devised a laboratory experiment to compare the RT of four different touchscreen-based techniques with the traditional physical button approach.
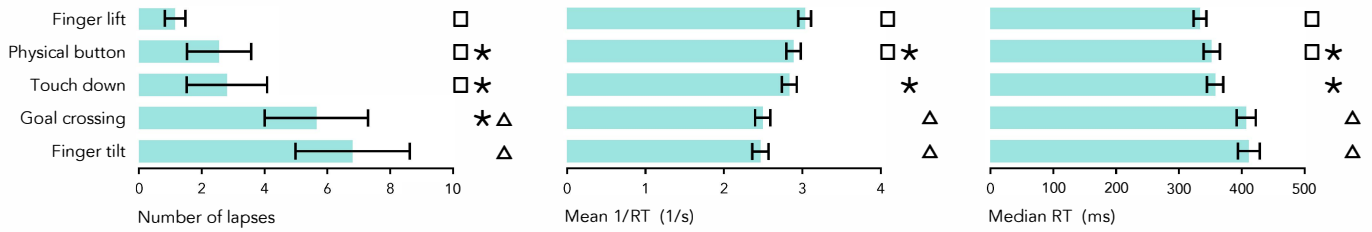
Fig. 3. The means of three PVT metrics for each technique. Results that do not share a symbol were significantly different ($\alpha = .005$).

A. Study Design & Participants

We used a within-subject design where each participant performed a PVT test with each of the 5 input types on an HTC Wildfire mobile phone (Fig. 1). The phone has a physical button centered on the bottom, making it equivalent for left and right-handed participants. The order of the input techniques was counterbalanced using a Latin Square. Participants used their dominant hand for all tasks. Before each test, we demonstrated the technique and had them try a few examples until they were comfortable. We kept the screen brightness constant for all participants. If participants touched the screen too early, the message "too early" was shown. Participants were not shown their reaction times to prevent this from influencing their performance or their responses on the final questionnaire.

After the study, each participant was given a questionnaire on demographics, technology expertise, and which technique they preferred and why. It also included the Epworth Sleepiness Scale [8], a standardized survey for assessing sleepiness. Participants were recruited through flyers and campus mailing lists and offered a $10 gift certificate for participating. We had twenty participants (9 females, 11 males), with a mean age of 28.8 (min=20, max=53, SD=7.7). Most considered themselves technically savvy, with a mean expertise of 4.3 (SD=0.6, Max 5.0); all participants had more than 10 years of computer experience and 90% owned smartphones. The mean Epworth score was within normal limits at 8 (min=0, max=13, SD=3.3), indicating we did not have an overly sleepy study population [9].

IV. RESULTS

Basner and Dinges [2] note a need for improved standardization amongst PVT implementations to compare results across studies; they suggest two measures most sensitive to sleep deprivation: the number of *lapses* (trials with RT > 500 ms [2], [7]) and the *mean 1/RT* (mean of 1000 ms / RT for all reaction times in the test [2]); we therefore report these measures due to their significance in measuring sleep. We also examined *median RT* to estimate differences in execution time. Fig. 3 shows the mean of each metric for each input technique. We saw similar differences across all three metrics, with *finger lift* having the fastest RT and *touch down* performing similarly to *physical button*. All pairwise comparisons are Bonferroni-corrected, giving a significance threshold of $\alpha = .005$.

Lapses. A Friedman test did not find a significant effect of input technique order on lapses ($\chi^2(4, N = 20) = 0.31, p = .99$), indicating adequate counterbalancing. A Friedman test found a statistically significant effect of input technique on number of lapses ($\chi^2(4, N = 20) = 31.24, p < .001$). Post-hoc Wilcoxon Signed-Rank Tests are summarized in Table I.

TABLE I. SUMMARY OF SIGNIFICANT DIFFERENCES

Contrast	Lapses $Z=$	Mean 1/RT $F_{1,76}=$	Median RT $Z=$
finger lift - finger tilt	-3.32**	68.79***	-3.81**
finger lift - goal crossing	-3.36**	62.59***	-3.88**
finger lift - touch down	n.s.	8.66*	-2.80*
physical button - finger tilt	-3.26**	37.67***	-3.77**
physical button - goal crossing	-2.74†	33.12***	-3.58**
touch down - finger tilt	-3.45**	28.63***	-3.40**
touch down - goal crossing	n.s.	24.69***	-2.99*

$\alpha = .005$. Three pairwise comparisons for which all three metrics did not show significant differences have been omitted from the table, but all such comparisons were conducted.

†p<.01 (trend toward significance) *p<.005 **p<.001 ***p<.0001

Mean 1/RT. This metric decreases the impact of outliers so the data more closely approximates a normal distribution [2]. A Shapiro-Wilk test did not find that mean 1/RT deviated significantly from a normal distribution ($W = 0.98, p = .104$), justifying the use of an analysis of variance. A mixed-effects model analysis of variance with test order as a fixed effect and participant as a random effect did not find a significant effect of test order on mean 1/RT ($F_{4,76} = 0.06, p = .993$), indicating adequate counterbalancing. A mixed-effects model analysis of variance with technique as a fixed effect and participant as a random effect found a significant effect of input technique on mean 1/RT ($F_{4,76} = 26.96, p < .001$). Post-hoc contrasts are summarized in Table I.

Median RT. A Friedman test did not find a significant effect of test order on median RT ($\chi^2(4, N = 20) = 0.320, p = .988$), indicating adequate counterbalancing. A Friedman test found a significant effect of input technique on median RT ($\chi^2(4, N = 20) = 40.04, p < .001$). Post-hoc Wilcoxon Signed-Rank Tests are summarized in Table I.

User preference. A Chi-Square test indicates a significant effect for user's self-reported preferred input technique ($\chi^2(4, N = 18) = 11.44, p < .05$). *Touch down* (44%, N=8) and *goal crossing* (33%, N=6) were the most preferred, with *finger lift* (10%, N=2), *physical button* (5%, N=1), and *finger tilt* (5%, N=1) being the least preferred.

V. DISCUSSION

As advocated by Basner and Dinges [2], we report lapses and mean 1/RT and note the results of these measures were similar to those of median RT. As a result, while the first two measures are more significant for measuring sleep, we limit most of our discussion to median RT, as this measure is easiest

to interpret with respect to explaining differences between the techniques. We first discuss a possible explanation for RT differences—execution time—then make recommendations for techniques to use with touch-based forms of PVT.

A. Execution time

We can estimate differences in execution time from median RT and group the touch-based input techniques into three broad categories, each significantly different from the others:

- **Minimal execution time (lifts)**: *finger lift*. This gesture takes little to no execution time.

- **Moderate execution time (presses)**: *touch down*. This gesture takes 24.6ms longer than *finger tilt* on average. Our results suggest that executing a finger press on a touch screen does not take substantially more time than pressing a *physical button* (which was 18.6ms more than *finger tilt*).

- **High execution time (threshold crossing)**: *goal crossing* and *finger tilt*. These gestures respectively take 73.8ms and 77.8ms longer than *finger lift* to execute. Both gestures involve moving from rest to cross an implicit (area of screen contact) or an explicit threshold (the goal line).

B. Recommended input techniques for touch-based PVT

The lower execution time of *finger lift* suggests it might be the best measure to use for PVT. However, there remains a caveat to recommending its use: *touch down*, which has similar performance to the *physical button*, was preferred by more participants. While some appreciated the responsiveness of *finger lift*, others found it awkward, perhaps because smartphones typically require the opposite action: as one participant stated, "The finger lift seemed unintuitive for signaling that I had seen the checkerbox."

In designing PVT-Touch, we strived to measure timing as accurately as possible—with minimum latency introduced by the software—allowing scores to be compared to established metrics (e.g., a lapse time of 500ms, which is only meaningful when reaction times are similar). Ideally, then, we would choose a touch-based technique that yields results comparable to a physical button and is comfortable for users. This suggests that *touch down* may be the most appropriate technique for assessing reaction time, which participants described as "easy", "intuitive", and "comfortable." We note that these results may be affected by variation in hardware not tested here; for example, resistive touch screens may perform differently from capacitive screens, or elicit different preferences from users.

This comfort may be a result of our participants' familiarity with the technology: as noted previously, participants self-reported high levels of computer expertise, were likely more familiar with smartphones than the general population, and were generally younger (max age was 53). Our sample population was also not highly sleep deprived. The *touch down* technique may present fatigue issues among other populations, for whom *finger lift* (which does not require hovering above the screen for long periods) may be more appropriate. That said, we do not believe these limitations substantially restrict our recommendations as they apply to in situ sleep assessment: much of the Pervasive Health and HCI communities' work on supporting sleep in situ has targeted smartphone users, who are most familiar with the *touch down* technique. The potential for fatigue in some populations does suggest future work exploring the use of *finger lift*.

VI. CONCLUSION & FUTURE WORK

This work contributes a new smartphone-based implementation of PVT and an evaluation of four input techniques. We found the familiar *touch down* response was preferred by users and has similar execution time to a physical button, making it a possible replacement for a physical button when measuring reaction time on smartphones, with the caveat that further work may be required to extend our findings to older or more sleep deprived populations. This study also informed work validating the *touch down* form of PVT-Touch against an existing PVT implementation using a physical button in a sleep deprivation study [10]. We are excited by the prospect of integrating PVT-Touch into existing mobile health assessment platforms, such as ohmage [4], and using PVT-Touch in experience-sampling studies for in situ alertness and sleepiness assessment. PVT-Touch can reach a larger number of users in a low-cost, low-overhead, and more ecologically valid manner than prior PVT implementations. Further examination of *finger lift* in sleep deprivation studies may be valuable to see if it is more sensitive than button-based versions. The *goal crossing* technique may also be useful in situations like games, where comparison to external standards is not relevant but some users may prefer it. Finally, other non-touch input modalities may be explored, such as sound or accelerometer-based movement.

REFERENCES

[1] D. F. Dinges and J. W. Powell, "Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations," *Behavior Research Methods, Instruments, & Computers*, vol. 17, no. 6, pp. 652–655, Nov. 1985.

[2] M. Basner and D. F. Dinges, "Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss.," *Sleep*, vol. 34, no. 5, pp. 581–91, May 2011.

[3] D. R. Thorne, D. E. Johnson, D. P. Redmond, H. C. Sing, G. Belenky, and J. M. Shapiro, "The Walter Reed palm-held psychomotor vigilance test.," *Behavior Research Methods*, vol. 37, no. 1, pp. 111–118, 2005.

[4] N. Ramanathan, F. Alquaddoomi, H. Falaki, D. George, C. Hsieh, J. Jenkins, B. Longstaff, J. Ooms, J. Selsky, H. Tangmunarunkit, and D. Estrin, "ohmage: An open Mobile System for Activity and Experience Sampling," *Pervasive Health '12*, pp. 203–204, 2012.

[5] R. L. Potter, L. J. Weldon, and B. Shneiderman, "Improving the accuracy of touch screens: an experimental evaluation of three strategies," in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '88*, 1988, pp. 27–32.

[6] J. Accot and S. Zhai, "Beyond Fitts' law: models for trajectory-based HCI tasks," in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '97*, 1997, pp. 295–302.

[7] S. Loh, N. Lamond, J. Dorrian, G. Roach, and D. Dawson, "The validity of psychomotor vigilance tasks of less than 10-minute duration," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 2, pp. 339–346, May 2004.

[8] M. Johns, "A new method for measuring daytime sleepiness: the Epworth sleepiness scale," *sleep*, 1991.

[9] M. W. Johns, "Reliability and factor analysis of the Epworth Sleepiness Scale.," *Sleep*, vol. 15, no. 4, pp. 376–81, Aug. 1992.

[10] M. Kay, M. A. Grandner, B. Jared, R. A. Lang, W. N. F., and J. Kientz, "Initial Validation of an Android-Based Psychomotor Vigilance Task," *Sleep '13 (in press)*, 2013.