Behavior Research Methods
2005, 37 (1), 111-118

The Walter Reed palm-held psychomotor vigilance test

DAVID R. THORNE, DAGNY E. JOHNSON, DANIEL P. REDMOND, HELEN C. SING, and GREGORY BELENKY
Walter Reed Army Institute of Research, Silver Spring, Maryland

and

JORDAN M. SHAPIRO
Corware, Peoria, Arizona

This field-portable reaction time test and analysis software run on devices using the Palm operating system. It is designed to emulate a test and commercial device widely used in sleep deprivation, shift work, fatigue, and stimulant drug research but provides additional capabilities. Experimental comparisons with the standard commercial device in a 40-hour total sleep deprivation study show it to be comparably sensitive to selected experimental variables. A Pocket PC-compatible version is under developement.

We have developed software for a portable reaction time (RT) test that runs on Palm-OS-based personal data assistants (PDAs). The test is intended primarily for field studies where an experimenter collects data on multiple subjects under various conditions for later detailed analysis in the laboratory. It was designed to emulate test devices that have been widely used in sleep-deprivation research: the "unprepared simple reaction time test" of Wilkinson and Houghton (1982), which ran on a mixed digital/analog cassette recorder available in the U.K., and the fully digital "psychomotor vigilance test" (PVT) designed by Dinges and Powell (1985) available in the U.S. (Ambulatory Monitoring Inc., Model PVT-192). However, the software can be used as a generic RT test that provides additional user-selectable stimulus, feedback, control, and data options, which will be described later.

The main components of the system include (1) the PalmPVT.prc executable that runs on the hand-held device, (2) a Windows companion program that runs on the PC, and (3) HotSync software that controls bidirectional communication between the two. Test parameters are normally set in the Windows program and then transferred to the PDA but can also be set or changed manually on the PDA itself via a mechanism hidden from the subject.

This material has been reviewed by the Walter Reed Army Institute of Research, and there is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the position of the Department of the Army or the Department of Defense. Correspondence should be addressed to D. R. Thorne, Department of Behavioral Biology, Division of Neurosciences, WRAIR, 503 Robert Grant Ave., Silver Spring, MD 20910-7500 (e-mail: david.thorne@us.army.mil).

Once parameters are set and the application is started, instructions and a login screen are presented to the subject, who then initiates the test (Figure 1). If desired, multiple subjects can share the same device while being distinguished by name, number, or another experimenter-assigned alphanumeric code. Visual stimuli are presented on the liquid crystal display (LCD; Figure 2), and the subject responds by pressing a specified button on the device. Subjects can be allowed to declare their handedness and button preference on the first session, when this information is not known in advance. Operation does not require knowledge or use of a stylus, which can be dropped or lost in the field.

When a test session ends, a brief message is presented, and the device turns itself off. An experimenter option can prevent subjects from accessing the main menu and thus possibly launching other applications, running down the battery, or making undesirable changes or deletions. This apparent dedicated usage may also dissuade theft.

Results are stored by subject and session along with the date and time of administration. The raw data recorded for each test session consist of each sequential trial number, foreperiod, response time, response type (e.g., valid, anticipatory, wrong button), and elapsed time from the start of the session. Placing the PDA in its cradle and pressing the HotSync button transfers collected data to the Windows companion program, where it can be viewed and saved. Both raw data and user-selectable summary statistics (shown in Figure 3) can be exported for further processing in various file formats including Excel, Access, REACT,[1] txt, and csv. Most of these statistics can also be viewed on the device itself while still in the field to check compliance, progress, or results.

The PDA software has run successfully on several different models manufactured by Handspring, Palm, and

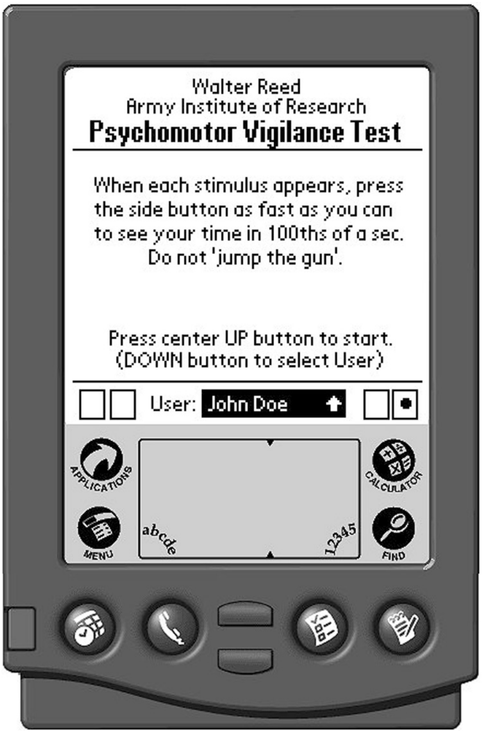

Figure 1. The startup instruction and login screen when multiple subjects share the same device. In the more typical use, the last instruction line is not displayed. The dot-filled rectangle indicates the response button a right-handed subject would normally press while holding the device with the left hand.

Sony. The Windows companion program has run successfully under Windows 95, 98, 2000, and XP.

Details and Differences Between the Palm Version and Previous Versions of the Test

Some of the differences between the PalmPVT, the commercial PVT-192, and the earlier British version were dictated by the characteristics of the supporting devices or the software that controlled them. Three categories of such differences are relatively straightforward and are described first. Other changes were deliberately introduced in this version to provide desirable features or to avoid undesirable ones. These differences and their reasons require more detailed explanation and are described subsequently.

Display. PalmPVT devices use an LCD screen, unlike the previous devices, which use a light emitting diode (LED) display. Two key differences between LCD and LED technology might be expected to influence reaction times: contrast ratio and rise time. LCDs have considerably lower contrast ratios than do LEDs. Reflective monochrome LCDs have contrast ratios as low as 4:1, which are relatively independent of ambient light, whereas transmissive backlit and color LCDs have higher ratios but are more affected by front surface reflection. LEDs in total darkness have a theoretical contrast ratio of infinity but in practice have ratios over 100:1, which vary with ambient illumination. Both LCDs and LEDs are directional, so that viewing them at an angle causes stimulus contrast and intensity to drop much more rapidly than the cosine rule. LCDs are considerably slower than LEDs, with typical rise times of 10s of msec versus less than 1 msec for LEDs. Combined, the lower contrast ratios and slower rise times of LCDs can be expected to affect absolute RT values. Whether this would increase or decrease sensitivity to experimental variables is an experimental question, but comparison of results obtained from different display technologies should be approached cautiously.

Timing resolution. Since the PVT-192 uses a 1-KHz timer whereas Palm-OS devices currently use a 100-Hz timer, the best-case measurement uncertainty for a single response occurrence would be 1 and 10 msec, respectively. If the timer cannot be reset at stimulus onset, this uncertainty doubles. The PalmPVT software is written to detect and adjust to faster clock rates if they become available and uses a “delay 0” command trick to effectively reset its otherwise unresetable timer by moving this second source of uncertainty into the prestimulus foreperiod, where the variation is inconsequential. Available documentation for the PVT-192 does not say whether its timer is resetable.

Actual measurement uncertainty of these devices is also limited by how rapidly the timer can be re-read. If the timer tick rate is 1 msec, but loop execution time can only read it every 50 msec, measurement uncertainty is 50 msec (averaging $+25 \pm 25$). PalmPVT execution

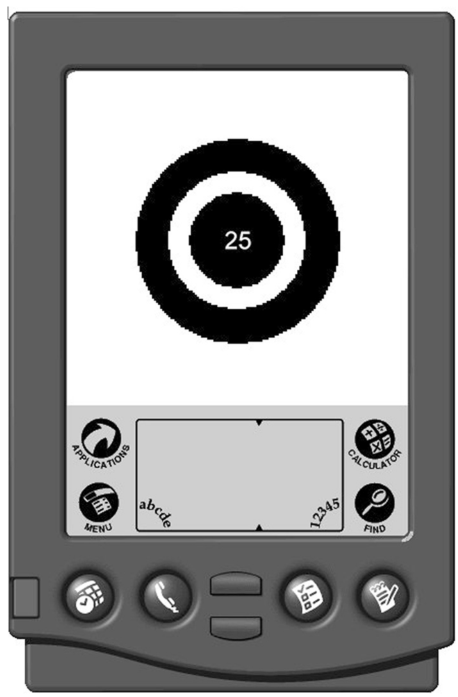

Figure 2. This shows the high-contrast graphic stimulus with superimposed numerals indicating an RT or elapsed time of .25 sec. Three stimulus modes and four feedback modes are available, including posttask feedback of the mean RT.

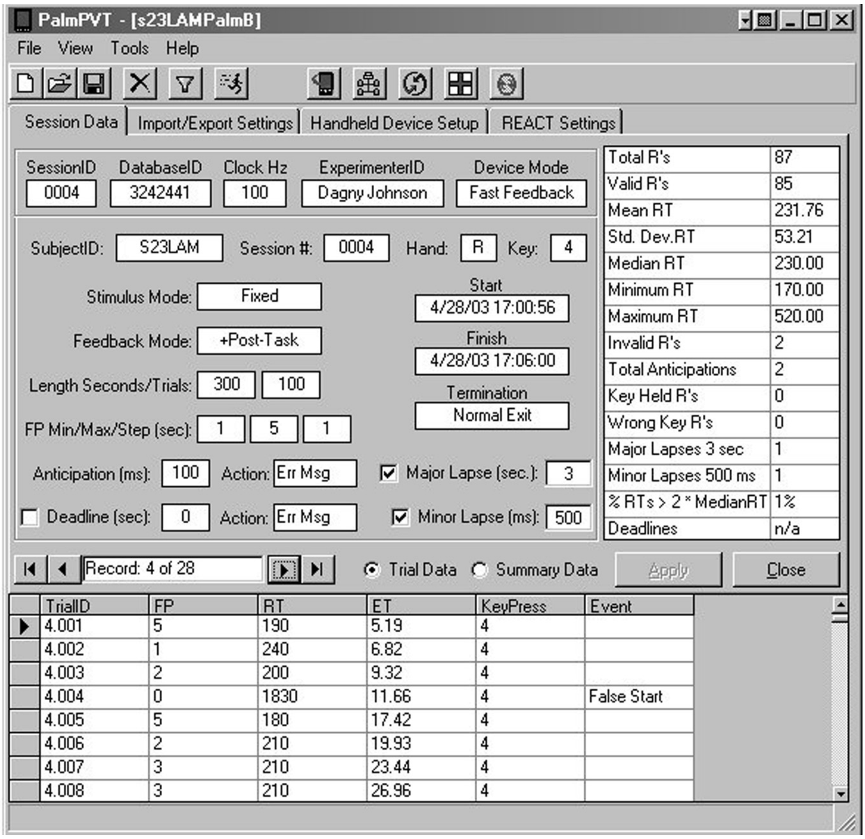

Figure 3. One screen from the Windows companion program showing raw data along the bottom and summary statistics in the right column for 1 of 28 test sessions uploaded from the hand-held device. Time and date of administration, parameter settings, and identifier variables are also shown. Any or all can be selectively exported in various file formats.

times were optimized by the use of in-line code and by selecting among alternative commands those commands that were measurably the fastest. Execution times using the traditional PVT incrementing stimulus (described below) can sacrifice temporal resolution by an amount that depends on the particular microprocessor and clock speed used. For example, measurements on an early 16-MHz Palm M-100 and a more recent 16-MHz Handspring Visor gave resolution estimates of 30.6 and 26.4 msec, respectively—worse if RT deadlines were imposed. However, when using the suggested default stimulus (described more fully below), execution times are less than or close to the tick rate, so single RT uncertainty currently remains at or near 10 msec. Since available documentation for the 12-MHz PVT-192 firmware does not give loop execution time, true resolution is unknown. Note that these uncertainties are worst-case values for single RT occurrences and that mean resolution improves with averaging and an increasing *N*. However, variance does not improve with averaging.[2]

Size, weight, and cost. The brick-sized PVT-192 is 21 × 12 × 58 cm and weighs 658g. Its current price is about $2,500. Most PDAs are about the size of a 1–2-cm deck of 3 × 5 cards or smaller and weigh 100–200g. Current prices start around $100. More expensive PDAs with cameras, phones, MP3, or wireless capabilities, or QWERTY keyboards are unnecessary or undesirable. These extras increase battery drain, which can be a particular problem in field studies where the PC, cradle, and recharger may be miles or days away. However, the higher contrast and speed of color displays may be desirable, even though color is not used.

Stimulus characteristics. Stimulus intensity (more accurately, the on/off difference or ratio) has a large effect on RT value (Cattell, 1886). Unless the effects of stimulus intensity itself are being measured, the preferred visual stimulus in most psychophysical RT experiments is a constant-intensity supraliminal light change with a rapid rise time (e.g., using a shutter or gas discharge bulb). Wilkinson and Houghton's (1982) original device used a 4-digit, 7-segment LED display to provide numeric feedback of the subject's RT, whereas the PVT-192 uses an LED dot-matrix display. In both devices, the LED display is also used as the RT *initiating* stimulus, by starting a "running timer" that counts up from zero in milliseconds and halts with the response. This unconventional stimulus was presumably a design compromise that saved additional equipment, fabrication steps, and battery power. Although this was clever in some ways and may even have some unmeasurable incentive characteristics, it has a number of undesirable consequences. First, stimulus intensity changes within the RT itself in a very nonlinear fashion related to the number of segments (or matrix dots) illuminated for a given digit as the digits and number of digits change. Second, both the relatively low intensity of LEDs and the still lower contrast of an LCD result in operation at the steepest portion of the psychophysical function where small variations in intensity have large effects on RT. Both effects can be expected to increase variance. Third, for durations less than approximately 100 msec, the Bunsen-Roscoe law holds, and stimulus intensity and duration interact (Wells, 1913). The consequence of these nonlinear intensity changes and jumps at 1-, 10-, and 100-msec intervals is unknown but can also be expected to increase variance in ways that may be prudent to avoid. Fourth, although LED rise times are very rapid, LCD rise times are very slow. The maximum on/off change for a typical monochrome LCD used in inexpensive PDAs is from approximately 20% to 80% reflectance, with a rise time on the order of 50 msec. Driving the display faster than this means that the actual change is even smaller, further exacerbating the above effects.

Two additional problems with a running-timer stimulus are unrelated to intensity or duration. First, a running-timer stimulus necessarily provides feedback, whether it is desired or not. Immediate feedback may allow subjects to detect their own performance changes while a test session is still in progress and to compensate by increased effort or attention, thereby attenuating experimental effects. Feedback has, in fact, been shown to reduce sensitivity to sleep deprivation (Wilkinson, 1961). Second, the temporal resolution of an RT measurement is also a function of loop-execution time, as mentioned above, and displaying a running timer is computationally more time intensive than might first be thought. For example, it can require reading multiple one-byte timer values and converting them to a single binary long integer, to decimal, and then to ASCII, which involves multiplications, additions, shifts, storing/retrieving intermediate results, and lookup operations, and then printing to particular locations in screen memory or the display driver, checking for a response occurrence, and conditionally looping back to repeat the entire operation. Whether the PVT-192 uses outboard components and/or internal commands that relieve the CPU of some of this overhead is undocumented, but with inexpensive PDAs running as slow as 16 MHz, these loop-execution times are slower than the timer tick rate and therefore reduce timing resolution.

For all the reasons elaborated above, the traditional running-timer stimulus was considered undesirable for the PalmPVT. One is provided as a menu option for historical purposes and apparent backward compatibility, but it is not the recommended default. The small-character-sized numeric running timer is replaced by a larger graphic whose intensity does not change with the RT and which does not require CPU-intensive looping operations. An easily discriminable high-contrast stimulus was desired, such as the black-and-white checkerboard used in photic driving of the EEG. Instead, a similar but arguably more attractive stimulus of alternating black-and-white circular annuli is used, in the form of a bull's-eye or target. Its ≈33-mm diameter is an arbitrary compromise between large size and short drawing time. It is possible to superimpose the running timer on the bull's-eye for a high-contrast stimulus, but timing resolution is compromised.

Feedback options. Feedback of the subject's RT value after each response is unavoidable with the running-timer stimulus but is at the experimenter's discretion with the PalmPVT's graphic stimulus. Feedback of the subject's mean RT can also be provided at the end of each session for either stimulus mode, instead of, or in addition to, immediate feedback. This delayed, rather than immediate, feedback is less likely to attenuate experimental effects and may be preferred over no feedback (also optionally available) in order to maintain motivation in a repeated measures design.

Foreperiods. In both the current and previous devices, the intervals between a response and the following stimulus can either be fixed or variable with experimenter-selected minimum and maximum values (e.g., 2–10 sec). In previous implementations, foreperiod values varied continuously between these limits and were randomized without further constraints, frequently resulting in undesirable runs of similar (short or long) values that could not be controlled across subjects, sessions, or time. In the long term, this will converge on the desired mean, but short-term variations can be considerable. This becomes of concern in one commonly used metric involving 1-min means, where variations in the standard error and an already small N become large. The PalmPVT uses N discrete foreperiods determined by a user-specified step size and then randomizes without replacement in blocks of $2N$. This equalizes frequencies in both the short and long term, assures that each foreperiod can precede or follow any other, including itself, only once per block, and also holds the number of trials per minute more nearly constant. This is because the number of trials per minute is determined more by the average foreperiod (e.g., 6 sec) than by variations in the average RT (e.g., 250–500 msec).

If a subject responds during a foreperiod before stimulus presentation, the same foreperiod is repeated. This is to counteract the tendency for anticipations during longer foreperiods to be rewarded, on average, by shorter foreperiods, thereby encouraging more frequent anticipations. This discourages both deliberate cheating and learning without awareness.

Session length. In the previous implementations of this test, sessions were terminated after a fixed elapsed time (typically 10 min). In the present version, sessions are terminated after either a fixed time or a fixed number of valid responses, whichever occurs first. Dual termination criteria are sometimes useful as a safeguard to prevent excessive values arising from certain forms of cheating. However, the experimenter can select between a fixed time or a fixed count by setting the "other" parameter to an unrealistically large value. A fixed count has the advantage of not forcing an increase in the standard error of the mean when subjects slow down. It can also reduce the portion of the slow-down effect due to learning rather than to sleepiness when subjects are exposed to repeated timed tests.[3] Even when fixed time is selected, task termination is still response contingent—it will not end during an on-going trial and will not allow (and thereby encourage) a subject to simply "wait it out."

Other differences. The PVT-192 can present an auditory stimulus[4] instead of or in addition to its running-timer visual stimulus, using auxiliary earphones. This capability was not incorporated into the PalmPVT because of nonavailability on many PDAs, nonstandard connectors on PDAs providing it, poor volume control, the need for additional equipment, and increased battery drain. The PVT-192 can also present a pre- and post-test single-item mood scale using an experimenter-selected word (e.g., *sleepy*), which the subject self-rates on a 10-point scale.

The PalmPVT allows the user to set the criterion cutoff values for defining poststimulus anticipations, major and minor lapses, deadlines, and whether such responses are separately tabulated as invalid or are included in the computed mean, median, standard deviation, and *N*. Lapse criteria and inclusion/exclusion can be changed after the fact in the Windows companion program, with statistics recomputed for purposes of exploratory analysis.

Software operation. When started, the software presents the instruction and login screen, waits for the subject to press the "begin" button, reads the desired parameter settings, jumps to the appropriate 1 of 12 subprograms,[5] initializes the database, prerandomizes the foreperiods as described above, then times each sequential foreperiod, presents the designated stimulus, and restarts the timer. If the stimulus is the running timer, it (re)reads the timer, converts to decimal and ASCII, erases the old value and displays the new, and repeats this as rapidly as execution times allow. When a response occurs, the software encodes and stores its value(s), presents an error message or feedback if appropriate, and repeats the sequence until done. It then computes and stores various totals and summary statistics, displays an "end of session" message and the mean RT if so selected, pauses, and turns itself off.

When the login screen is present, a secret entry by the experimenter can cause the program to present a number of screens for examining recorded data, viewing help files, gaining access to the main menu if this has been denied to the subject, and setting or changing parameter values, preferences, and identifiers. The software then checks any changed parameter values for validity and returns to the login screen. Once begun, another secret entry by the experimenter can abort the ongoing session.

The operation of the Windows companion program is beyond the intended scope of this paper but is explained in extensive help files available in the program itself.

Experimental Comparison of the Palm and PVT-192 Versions of the Test

In order to validate the PalmPVT device for field use, the palm-based version and the commercial version were compared during a total sleep-deprivation study using PalmPVT parameter values designed to improve subject acceptance and compliance under field conditions. The purpose of the study was to compare the devices, with respect to their response to progressive sleep loss and recovery in the sleep laboratory, while using shortened foreperiods and test durations, a fixed-intensity stimulus, and delayed versus immediate feedback. A secondary purpose was to examine time-on-task order effects.

Method. Twelve subjects (6 males, 6 females) in their twenties and thirties performed a series of 30 PVT test sessions over 4 days. The sessions were repeated every 2 h during waking hours on a schedule consisting of one baseline day followed by 8 h in bed, 40 h of continuous wakefulness, 10 h in bed, and one recovery day. Each PVT test session consisted of a conventional 10-min PVT-192 task, preceded and followed by a 5-min PalmPVT task. Foreperiods for the PVT-192 varied randomly from 2–10 sec (the values typically used in previous laboratory studies). Foreperiods for the Palm varied randomly from 1–5 sec (values deemed more acceptable in the field), thus yielding about the same number of responses per task. The PVT-192 used its numeric running-timer stimulus, whereas the PalmPVT used a fixed-intensity graphic bull's-eye. The PVT-192 and one PalmPVT (Palm-I) displayed immediate feedback after each response, whereas the other PalmPVT (Palm-D) provided only delayed post-task feedback of the mean RT. The order of these two PalmPVT tasks was alternated from session to session.

Results. Absolute RTs for the PVT-192 tended to be faster than the pretest/posttest averaged values for the PalmPVTs but with frequent reversals. Differences in one direction or the other were expected, as discussed in the previous section, due to the combined mechanical, electronic, optical, and operational differences between the devices, so this finding is not particularly meaningful or informative. For evaluating application validity, the devices were compared in terms of response speed (1/RT) relative to their Day-1 baseline means using a 2-point running average (because of the alternation), as is illustrated in Figure 4. All three devices showed the same pattern, with speed declining during sleep deprivation to a minimum around 0730 on Day 3 and then rising throughout the day.

An order effect (overall session time-on-task) was apparent, with speed on the third PVT in each test block significantly slower than on the first ($F = 12.314$, $p = .002$). Figure 5 shows this for Palm-I. A similar pattern was seen for Palm-D.

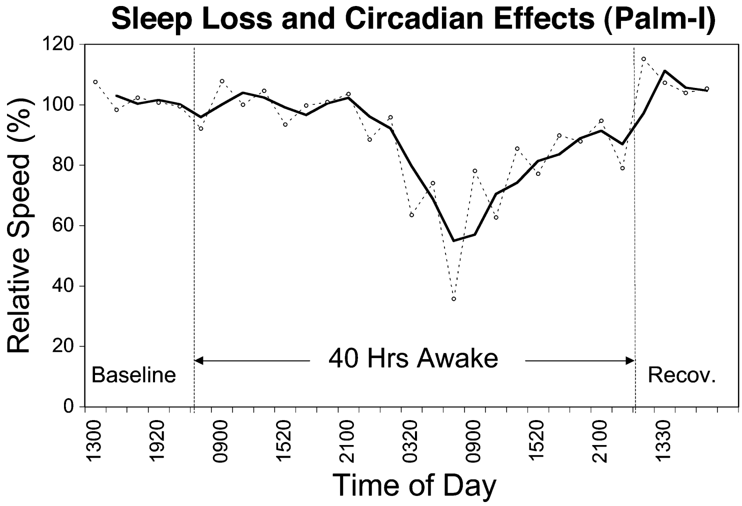

Figure 4. Response speed (1/RT) relative to the baseline-day average versus time of day, for a 5-min Palm PVT using immediate feedback. Vertical dashed lines denote nights in bed. Solid line is the 2-point running average used in Figures 6 and 7. N = 12 subjects.

As can be seen in Figure 6, delayed feedback appeared to increase sensitivity, yielding greater decrements at the circadian trough (−52% vs. −45%) and for the following period of increasing sleep loss, but these differences were not statistically significant. The subjects *were* getting immediate feedback on an otherwise identical PalmPVT within minutes of each other, so it is not known what effects delayed feedback might have if given exclusively.

In Figure 7, the PVT-192 is compared with the PalmPVT when both presented immediate feedback. Although the Palm-I showed the greater maximum decrement (−45% vs. −37%), the difference was not statistically significant, nor were the differences during recovery.

Conclusions

RTs as measured by the PalmPVT are responsive to sleep deprivation and circadian effects, and to time-on-task, fatigue, and order effects. In terms of relative response speed, shorter foreperiods and task durations yield results comparable to the longer values of the PVT-192 in spite of different stimulus characteristics.

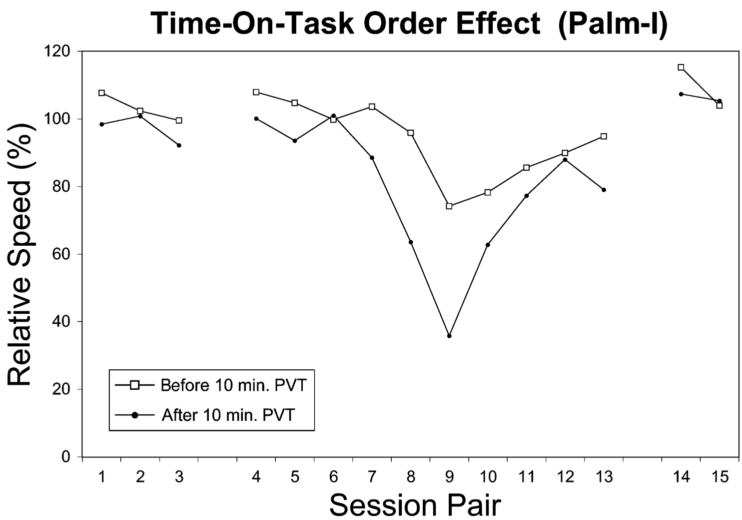

Figure 5. Relative response speed for the Palm PVT using immediate feedback when it alternately preceded and followed another 5-min Palm PVT and 10-min PVT-192.

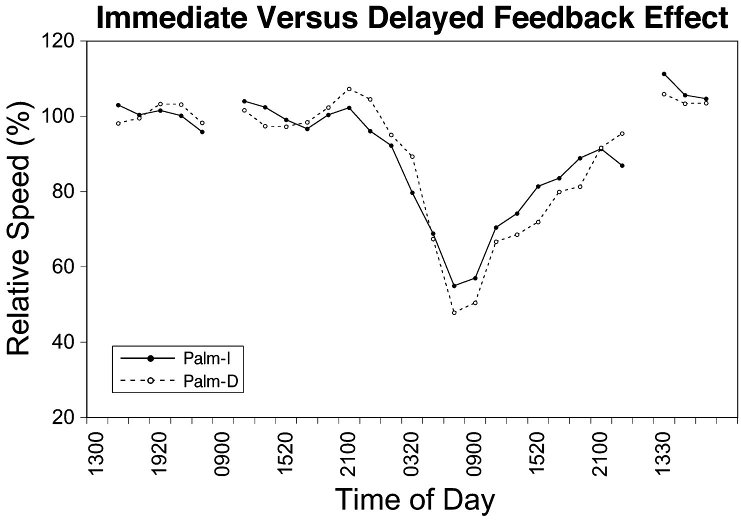

Figure 6. Averaged relative response speed for two 5-min Palm PVTs using either immediate (I) feedback of each RT or post-task delayed (D) feedback of the mean RT.

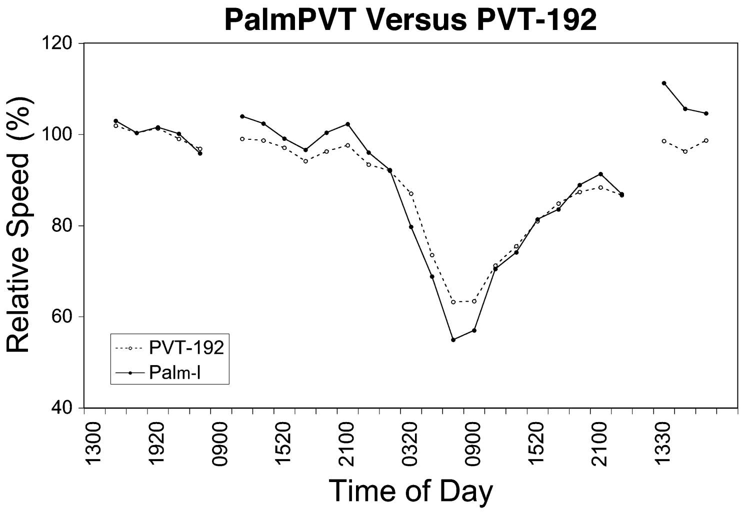

Figure 7. Averaged relative response speed for a 5-min Palm PVT with 1- to 5-sec foreperiods and a 10-min PVT-192 with 2–10-sec foreperiods, with both using immediate feedback.

Software Availability

The PalmPVT software is available without cost to researchers and practitioners in the life sciences with the agreement that (1) copies will not be distributed to others, (2) it will not be used for profit, and (3) any presentations or publications of findings using it will cite the Walter Reed Army Institute of Research as its source. At this writing, we cannot distribute it electronically, but the software developer is making it available for download on www.corware.com as a public service. Software registration requires agreeing to the above terms, submitting an e-mail address for download and notice of any future bug reports or fixes, and optionally answering a survey on the experimental variables and population of intended use. The registration process typically requires one business day to clear before download is enabled. A small (≈100K) stand-alone program is available for installation on the PDA, but most users will want to download the (≈10MB) full-release package, which contains both it and the Windows companion and HotSync software.

A Pocket PC-compatible version of the software should be available by the time of this publication.

REFERENCES

CATTELL, J. M. (1886). The influence of the intensity of the stimulus on the length of the reaction time. *Brain*, **8**, 512-515. Reprinted (1947). **70**, 103-106.

DINGES, D. F., & POWELL, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers*, **17**, 652-655.

WELLS, G. R. (1913). The influence of stimulus duration on reaction time. *Psychological Monographs*, **15**(Suppl.), 1-69.

WILKINSON, R. T. (1961). Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. *Journal of Experimental Psychology*, **26**, 263-271.

WILKINSON, R. T., & HOUGHTON, D. (1982). Field test of arousal: A portable reaction timer with data storage. *Human Factors*, **24**, 487-493.

NOTES

1. REACT is a commercial data-analysis program for the PVT-192 that provides many but not all of the same statistics as the PalmPVT system (and vice versa). It generates individual-session frequency distributions and per-minute regression plots that are labor intensive in Excel. Price and information are available at www.ambulatory-monitoring.com/reaction.html.

2. Since the distribution is uniform, the standard deviation will equal the single-case uncertainty divided by the square root of 12.

3. A repeated measures design is by nature a learning paradigm, and all of the principles of learning will apply, whether desired or not. When fixed-time tasks are administered repeatedly, subjects learn to slow down. This is indirectly rewarded by requiring fewer responses per session but results in less reliable means and increased variability. This slow-down effect is shown by both animals and humans and operates independently of motivation or awareness. It can be viewed as an example of "the law of least effort" (after Guthrie, Hull, and Tolman) or as the difference between interval and ratio contingencies of reinforcement (after Skinner). The effect increases with continued exposure and is amplified if the task is effortful, aversive, or boring. Such an effect is of negligible concern with a few repetitions but is a potential quantitative confound in longer studies. Fixed-count tasks do not exhibit this effect, where slowing is indirectly punished by increased test duration and any speedups (beyond the normal practice effect) tend to be self-limited by increased effort and by the "irreducible minimum" time required to perform the test.

4. This stimulus is a 1000-Hz tone with screwdriver-adjusted volume. The auditory and visual stimuli differ functionally as well as in modality since the auditory stimulus has a fixed intensity and does not rapidly say "One, two, three, four, . . . two hundred-fifty . . . ."

5. The possible combinations of parameter settings and options result in many different program "paths." Writing a separate program for each, although highly redundant, avoided the execution time penalty of many if–then tests and conditional jumps that a single program would have to perform with every loop and trial of a session, by trading computer memory for speed. The memory available in even 2-MB PDAs is sufficient for storing the program and data from scores or hundreds of tests.

(Manuscript received May 18, 2004;
revision accepted for publication September 24, 2004.)