Sleep and Breathing
https://doi.org/10.1007/s11325-019-01859-7

METHODS • ORIGINAL ARTICLE

Auditory psychomotor vigilance testing in older and young adults: a revised threshold setting procedure

Virginie Gabel[1] · Mirelle Kass[1,3] · Daniel S. Joyce[1] · Manuel Spitschan[1,4] · Jamie M. Zeitzer[1,2]

Received: 10 January 2019 / Revised: 8 April 2019 / Accepted: 26 April 2019

Abstract

Background One of the most common ways to examine the daytime impact of sleep loss is the use of the psychomotor vigilance test (PVT). PVT metrics, including median reaction time (RT) and number of lapses, have been examined in a variety of studies in which both acute and chronic sleep times are manipulated. Most of these studies involve young, healthy individuals and use a visual stimulus. As light is a possible countermeasure to sleep loss, and sometimes incompatible with the use of visual PVT, PVT with auditory cues (aPVT) has been used. A threshold of 400 ms is commonly used to delineate lapses from normal RT in the aPVT. As aging can influence a variety of brain functions, we wanted to examine whether this lapse threshold was accurate for use in older adults.

Methods Twenty-eight young and 19 healthy older participants performed a 10-min auditory PVT approximately 90 min before habitual bedtime. The occurrence of lapses was determined by five objective RT thresholds: (1) 400 ms, (2) 500 ms, (3) 2 × median, (4) mean + 2 × SD, and (5) method 4 without outliers. Results of these methods were compared with a triplicate visual inspection of RT histograms to determine RT outside of the expected log normal distribution.

Results In both groups, methods 1, 4, and 5 performed poorly, while methods 2 and 3 were adequate, though method 3 was statistically superior.

Conclusion In both age groups, the use of twice the median as an objective threshold had the best concurrence with visual scoring.

Keywords Sleep · Reaction time · Aging · Alertness · Thresholds · Cognition

Introduction

Sleep deprivation and sleep restriction are widespread problems of modern society and can lead to, among other negative effects, a degradation of attention, especially vigilant attention [1]. Several tests can be used to measure these cognitive decrements in the laboratory, but the most widely used is the psychomotor vigilance test (PVT) [2], which measures a reaction time to a stimulus and can be used as an indicator of sustained attention. Performance on the PVT over time tracks cumulative time awake in a variety of settings and does not exhibit a learning effect; as such, it is a useful tool with which to study the objective impact of a variety of interventions on sustained attention [3–5].

The PVT itself is a response (button press) to an irregularly presented stimulus over what is typically a fixed 10 min. There are a variety of metrics that can be derived from a PVT, notably median reaction time, the number of lapses (failures of attention), and time-on-task decrement [6]. Most studies that use the PVT have a visual stimulus (typically, a millisecond counter). In studies in which the impact of lighting on sustained attention is being assessed, the use of a visual stimulus on the PVT is undesirable as it exposes the participant to additional light. As such, an auditory version of the PVT using a simple tone stimulus is often used in these settings. In the 10-min visual version of the PVT, a threshold of 500 ms is often used to define a lapse of attention [2]. That is, responses longer

✉ Virginie Gabel
vgabel@stanford.edu

[1] Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA 94305, USA

[2] Mental Illness Research Education and Clinical Center, VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304, USA

[3] Present address: Connecticut College, New London, CT 06320, USA

[4] Present address: Department of Experimental Psychology, University of Oxford, Oxford OX2 6GG, UK

Published online: 08 May 2019

than 500 ms are considered to be outside of the normal range of attentive responses and are defined as a period of inattentiveness. Indeed, the lapses and slower RTs might stem from changes in cortical sensory, attention, and motor control pathways [7, 8] and are increased under sleep deprivation. It is unknown, however, if the mechanism underlying lapses in visual attention during a PVT is the same as the mechanism underlying lapses in auditory attention during a PVT.

In studies that examine the auditory version of the PVT (aPVT), a fixed threshold of 400 ms is often used [9]. According to Jung et al., this fixed 400-ms threshold was adapted in an arbitrary fashion to better match the number of lapses observed in a visual PVT and an auditory PVT (cf., Figure 2 in that publication).

Most of the studies in which the aPVT has been administered have examined young individuals. Older individuals can experience both a decrease in overall cognitive function [10, 11] and a reduction in speed of auditory processing [12, 13]. As such, the normal range of aPVT reaction times in an older individual could be shifted to longer RTs (i.e., slower response). The use of the 400-ms threshold in older adults taking an aPVT, consequently, might be inappropriate.

In this study, therefore, we examined whether the fixed 400-ms aPVT lapse threshold is an appropriate cutoff for both young and older individuals or whether an alternate value should be used.

Methods

Study participants

Two separate cohorts, young (18–35 years old) and older (55–85 years old), were recruited for two different studies (ClinicalTrials.gov: NCT01119365, NCT02632318). Participants were recruited through online or poster advertisements. Participants were without any sleep problem and in stable self-reported health. Further exclusion criteria included smoking, extreme chronotypes [14], use of illegal drugs or drugs that could impact light sensitivity or sleep, excessive alcohol intake [15], and a history of shift work. The protocol was approved by the Stanford University Institutional Review Board and all procedures adhered to the principles outlined in the Declaration of Helsinki. While the experiment was registered in ClinicalTrials.gov, the current analyses were exploratory in nature.

Twenty-eight young participants (13 women, mean age ± SD: 28.1 ± 3.91 years old) and 19 older participants (11 women, mean age ± SD: 69.0 ± 8.4 years old) were enrolled (N = 47). Both groups were required to keep at least one week of a regular sleep-wake cycle prior to coming to the laboratory. This schedule helped to ensure a predictable phase angle between the circadian system and the timing of sleep, as well as a regular amount of sleep (7–8 h per night) to avoid sleep deprivation. Compliance to this schedule was verified using wrist actigraphs (Motionlogger, Ambulatory Monitoring Inc., Ardsley, NY) and self-reported sleep logs.

Procedures

Upon entry to the laboratory, participants were screened for proximal use of alcohol and urine; use of either was exclusionary. In the first study, young individuals came to the laboratory 9 h before their habitual bedtime, while in the second study, older individuals came to the laboratory 3 h before their habitual bedtime. From the time of entry into the laboratory, participants remained in a windowless, light-controlled, sound-attenuated bedroom, without any information concerning time of day. Following a practice aPVT, older participants had a PVT 2 h before sleep time. This aPVT was conducted under a white fluorescent broad-spectrum light (~ 150 lx in the horizontal angle of gaze). Following a practice aPVT, young individuals were exposed to the same fluorescent lighting as the older individuals, except that their lighting was dim (< 15 lx in the horizontal angle of gaze). Young individuals took five additional aPVT in the evening; the aPVT administered 90 min before habitual bedtime was examined herein to align with the only aPVT done by the older group in the evening.

aPVT

The participant was asked to press a button as quickly as possible in response to a 1000-Hz tone. Tones were delivered at a constant volume that the individual participant could hear comfortably. Each tone was separated by a random interval of 1 to 6 s [16], and each tone continued until the response occurred. No RT feedback was provided. Each session comprised approximately a hundred trials and lasted 10 min.

The test was administered with a custom-built aPVT based on an Arduino Uno microcontroller board and a single hardware button which stops the stimulus and recorder when pressed. Data was logged by the Arduino and interfaced via Python (Python 2.7.12) with a PC running Linux (Linux 4.8.0).

Analysis

We evaluated the reaction times for the single aPVT session in each of the 47 participants. The first three reaction times recorded were discarded as an "adaptation" response to performing the test. We also removed "false start" responses (reaction times < 150 ms), as reaction times this rapid would likely represent responding in anticipation of the stimulus. There were no false start responses in the older group, and there were four false starts (0.14%) removed in the young

group. We examined the histogram distributions of the remaining reaction times in each participant (see Fig. 1). Log normal curves were fitted to each of the histograms, and these graphs were examined by three observers (EK, VG, JMZ), each blind to the age of the participant, who independently determined whether a reaction time fell outside of the expected distribution (i.e., was a "lapse"). A consensus was obtained for each reaction time being within a normal range or being a lapse. We did not use a confidence interval to determine whether a value was outside of a log normal distribution as the presence of outliers drove the log normal distribution to include a greater number of values that would have otherwise been outside of the distribution. The consensus among the three observers constitutes the "standard" to which objective determinations of lapses were compared. Five different thresholds were used to objectively demark the occurrence of a lapse: (1) fixed threshold of 400 ms (typically used in the literature for aPVT [9]), (2) fixed threshold of 500 ms (typically used in the literature for visual PVT [17]), (3) twice the median value, (4) mean value plus twice the standard deviation, and (5) mean value plus twice the standard deviation after extreme values (defined as all the values over twice the median) were removed. Distributions were fitted using Origin (OriginPro 2017, OriginLab Corporation, Northampton, MA) and analyzed with R (RStudio 3.5.0 [18]). Data are presented with standard deviations. Individual statistical tests are noted below. The sum of the square of the error (SSE) was calculated as follows: $\sum_{i=1}^{n}(VL_i - TL_i)^2$, such that VL is the number of visual lapses, TL is the number of threshold-based lapses, and n is the number of participants.

Results

In the course of a single 10-min aPVT, there were 105 ± 4.49 responses obtained in each young participant and 101 ± 4.71 responses obtained in each older participant. Responses in both the young and older groups were non-normally distributed, with a log normal distribution fitting each individual reasonably well (Fig. 1; average ± SD adjusted R^2 of the log normal fits = 0.91 ± 0.079, range = 0.62–1.00). The average of the individual mean reaction times is 290 ± 44.7 ms and 365 ± 154 ms in the young and old, respectively ($p < 0.05$, t test). The average of the individual median reaction times is 272 ± 41.9 ms and 294 ± 37.4 ms in the young and old, respectively ($p = 0.072$, t test).

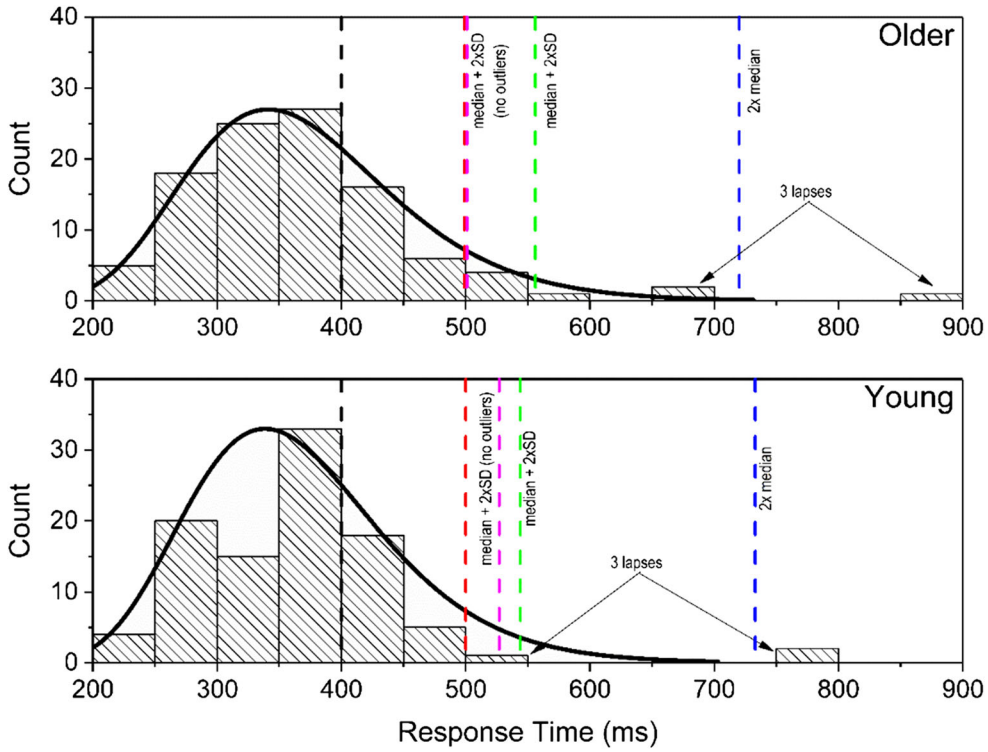

Fig. 1 Example histograms of reaction times from an older (top) and a young (bottom) participant completing a single aPVT trial. The different thresholds examined in this manuscript are provided as follows: (1) 400 ms (black dashed line): 26 lapses (older), 29 lapses (young); (2) 500 ms (red dashed line): 3 lapses (older), 7 lapses (young); (3) 2 × median (blue dashed line): 2 lapses at 720 ms (older), 1 lapse at 733 ms (young); (4) median + 2 × SD (green dashed line): 2 lapses at 556 ms (older), 3 lapses at 544 ms (young); and (5) median + 2 × SD (no outliers) (magenta dashed line): 3 lapses at 500 ms (older), 5 lapses at 527 ms (young). The arrows indicate the reaction times that were identified by three independent viewers and scored as "lapses," three in both the young and older groups

Using the literature-derived fixed threshold of 400 ms, 25 of the 28 young participants (89%) had at least one lapse and 13 of the 19 older participants (84%) had at least one lapse (Table 1). Of the participants who had at least one lapse, the average number of lapses was 7.56 ± 6.86 in the young and 10.8 ± 10.4 in the older participants. It was evident from visual observation of the distribution curves of reaction times, however, that many of the "lapses" defined by a 400-ms threshold still fell within a log normal distribution.

According to the visual scoring of lapses, 68% of the older participants had at least one lapse and there were 5.31 ± 5.20 lapses per older participant with a lapse. In the young participants, visual scoring identified 75% of the individuals having at least one lapse and 4.38 ± 3.14 lapses per young individual with a lapse. The standard 400-ms threshold overestimated lapses by more than twofold in older adults and 1.7-fold in young adults (Table 1). The 500-ms and twice the median thresholds performed much better, with slight overestimation (older cohort) or underestimation (young cohort) of lapses. The mean plus twice the standard deviation threshold performed poorly as all individuals were determined to have lapses, but the average number of lapses per individual was underestimated. Similarly, the threshold based on the mean plus twice the standard deviation after extreme values were excluded determined all individuals had lapses and overestimated the number of lapses. Using the sum of the square of the error as a marker of goodness of fit to the standard, the threshold set by twice the median performed the best in both age groups.

Discussion

Our data indicate that the use of a single, fixed threshold to determine the occurrence of lapses on an aPVT, in either healthy young or older adults, is insufficient. The standard fixed threshold of 400 ms yields nearly twice the rate of lapses during an evening aPVT session as would be determined through visual inspection of the response time distribution. The use of the fixed 500-ms threshold, typically used on a standard visual PVT, performs better, though the use of a variable threshold set at twice the median value has the best performance.

We determined the standard of the presence or absence of a lapse through visual inspection and agreement among three independent viewers. Overlaid atop the histogram distributions (see Fig. 1) was a log normal distribution. Data conformed well to a log normal distribution, not unexpectedly as the shortest response times are limited by a physiologic speed at which faster times are not possible. While the use of standard deviations is often helpful in determining outliers in normally distributed data, their inclusion in two different variable threshold methods was only somewhat helpful and did not reach the performance of the variable threshold calculated as twice the median.

Our data were obtained during a baseline session in two separate protocols, one involving healthy young adults and the other involving healthy older adults. There were notable differences in the protocols that make problematic the direct comparison of the reaction times in the two groups. First, the older group entered the lab later in the day than the young group and was exposed to normal room lighting prior to and during the aPVT. The young group was exposed to very dim light for 6.5 h prior to and during their aPVT. Even the normal room lighting to which the older adults were exposed has the potential for changing performance on the PVT [19]. Second, the aPVT that the older participants performed was their first session of the test, while young participants had five sessions of the aPVT prior to taking the aPVT for which the results are reported. We chose to align participants by approximate time relative to normal sleep rather than test number as the PVT has a relatively shallow learning curve [1, 4], and we anticipated similar results between the cohorts. Thus, our data are insufficient to compare the difference in reaction times or lapses between the two cohorts, but are sufficient to examine the different thresholds for determination of lapses in the two

Table 1 Determination of lapses by various threshold methods

		Standard	Threshold-based				
		Visual	Fixed (400 ms)	Fixed (500 ms)	2 × median	Median + 2 × SD	Median + 2 × SD (no outliers)
Older	% with a lapse	68%	84%	68%	58%	100%	100%
	No. of lapses (non-zero)	5.31 ± 5.20	10.8 ± 10.4	6.85 ± 6.95	6.18 ± 5.93	3.11 ± 1.66	7.16 ± 5.80
	SSE	n/a	1464	134	63	502	315
Young	% with a lapse	75%	89%	75%	75%	100%	100%
	No. of lapses (non-zero)	4.38 ± 3.14	7.56 ± 6.86	3.81 ± 3.13	2.90 ± 2.35	3.86 ± 1.90	6.43 ± 2.82
	SSE	n/a	1913	526	383	426	730

SSE sum of the square of the errors, *SD* standard deviation, *n/a* not applicable

cohorts. The most parsimonious interpretation of the data is that the use of a variable threshold (twice the median from a baseline, non-sleep-deprived session) would be the most appropriate to use when analyzing the aPVT in either young or older healthy individuals.

Funding This study was supported by Velux Stiftung (NNo) and the U.S. Department of Defense (NNo).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The protocol was approved by the Stanford University Institutional Review Board. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Lim J, Dinges DF (2008) Sleep deprivation and vigilant attention. Ann N Y Acad Sci 1129:305–322. https://doi.org/10.1196/annals.1417.002
2. Dinges DF, Pack F, Williams K, Gillen KA, Powell JW, Ott GE, Aptowicz C, Pack AI (1997) Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep 20(4):267–277
3. Basner M, Dinges DF (2011) Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. Sleep 34(5):581–591
4. Doran SM, Van Dongen HP, Dinges DF (2001) Sustained attention performance during sleep deprivation: evidence of state instability. Arch Ital Biol 139(3):253–267
5. Lisper HO, Kjellberg A (1972) Effects of 24-hour sleep deprivation on rate of decrement in a 10-minute auditory reaction time task. J Exp Psychol 96(2):287–290
6. Dinges DF, Powell JW (1985) Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behav Res Meth Instr 17(6):625–655. https://doi.org/10.3758/BF03200977
7. Drummond SP, Bischoff-Grethe A, Dinges DF, Ayalon L, Mednick SC, Meloy MJ (2005) The neural basis of the psychomotor vigilance task. Sleep 28(9):1059–1068
8. Naito E, Kinomura S, Geyer S, Kawashima R, Roland PE, Zilles K (2000) Fast reaction to different sensory modalities activates common fields in the motor areas, but the anterior cingulate cortex is involved in the speed of reaction. J Neurophysiol 83(3):1701–1709. https://doi.org/10.1152/jn.2000.83.3.1701
9. Jung CM, Ronda JM, Czeisler CA, Wright KP Jr (2011) Comparison of sustained attention assessed by auditory and visual psychomotor vigilance tasks prior to and during sleep deprivation. J Sleep Res 20(2):348–355. https://doi.org/10.1111/j.1365-2869.2010.00877.x
10. Salthouse TA (2010) Selective review of cognitive aging. J Int Neuropsychol Soc 16(5):754–760. https://doi.org/10.1017/S1355617710000706
11. Harada CN, Natelson Love MC, Triebel KL (2013) Normal cognitive aging. Clin Geriatr Med 29(4):737–752. https://doi.org/10.1016/j.cger.2013.07.002
12. Tun PA, Williams VA, Small BJ, Hafter ER (2012) The effects of aging on auditory processing and cognition. Am J Audiol 21(2):344–350. https://doi.org/10.1044/1059-0889(2012/12-0030)
13. Deary IJ, Corley J, Gow AJ, Harris SE, Houlihan LM, Marioni RE, Penke L, Rafnsson SB, Starr JM (2009) Age-associated cognitive decline. Br Med Bull 92:135–152. https://doi.org/10.1093/bmb/ldp033
14. Horne JA, Ostberg O (1976) A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int J Chronobiol 4(2):97–110
15. World-Health-Organization (2001) The alcohol use disorders identification test: guidelines for use in primary care, 2nd edn. World Health Organization, Geneva, p 2
16. Basner M, Mollicone D, Dinges DF (2011) Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. Acta Astronaut 69(11–12):949–959. https://doi.org/10.1016/j.actaastro.2011.07.015
17. Dinges DF, Kribbs NB (1991) Performing while sleepy: effects of experimentally-induced sleepiness. Human performance and cognition. Sleep, sleepiness and performance 1991. John Wiley & Sons, Oxford, pp 97–128
18. R Foundation for Statistical Computing V, Austria (2014) R: a language and environment for statistical computing. R Development Core Team. https://www.r-project.org/
19. Cajochen C, Zeitzer JM, Czeisler CA, Dijk DJ (2000) Dose-response relationship for light intensity and ocular and electroencephalographic correlates of human alertness. Behav Brain Res 115(1):75–83