2011/806

ISSN 1797-3457 (verkkojulkaisu)
ISBN (PDF) 978-951-25-2289-7

TIIVISTELMÄRAPORTTI (SUMMARY REPORT)

Quantitative sleepiness and performance testing in drivers: Markers, metrics, and models of fatigue

Research in the Sleep and Performance Research Center at Washington State University Spokane

Pia M. Forsman[*] and Hans P. A. Van Dongen
Sleep and Performance Research Center
Washington State University Spokane
702 South Campus Facility, P.O. Box 1495,
Spokane, WA 99210-1495, USA
[*]phone: +1-509-358-7756, email: pia.forsman@wsu.edu

Abstract

Lane deviation has frequently been used as an index of drowsy driving in cars and trucks. Research has tried to quantify driving performance on the basis of several other signals and metrics, with the aim to warn the driver before his/her performance is critically degraded. However, a lack of published largescale comparisons between available signals and metrics precludes informed decisions on which signals and metrics may best predict driver drowsiness. This research aimed to fill this gap. We show that metrics derived from steering performance constitute an additional dimension of driving performance, which when combined with lane deviation may non-invasively enhance the ability to detect driver drowsiness.

A total of 41 healthy adults (ages 22-39y; 12 females) participated in a 2-week in-residence laboratory study. The study included a practice day, five simulated workdays, a one-day or two-day break, another five simulated workdays, and a recovery day. Eleven participants were randomized to a day shift condition (daily time in bed 22:00-08:00) and the other 30 participatns were assigned to a night shift condition (daily time in bed 10:00-20:00). Each workday included four 30-minute driving sessions administered at fixed intervals through the day (the sessions in the day shift condition started at 09:00, 12:00, 15:00, and 18:00, and the sessions in the night shift condition started at 21:00, 00:00, 03:00, and 06:00). The participants drove a simulated Ford Taurus in a standardized scenario of rural highways on a high-fidelity driving simulator (PatrolSim IV, L-3 Communications). The standardized scenario included ten straight, uneventful road segments with speed limit 88.5 km/h and length 0.8 km. From these segments we extracted 87 metrics of driving performance.

The variance in the data set (41 subjects, 40 driving sessions, 87 metrics) was examined with principal component analysis. This yielded two dominant dimensions of driving performance, which were represented by metrics reflecting steering variability and lane variability. The steering dimension explained 33% of the variance and the lane dimension explained another 14% of the variance. Mixed-effects ANOVA revealed significant interactions of condition (day shift or night shift) by time of day (the first, second, third, or fourth driving session of the workday) for the steering dimension ($F=3.11$, $P=0.026$) and lane dimension ($F=4.51$, $P=0.004$).

Our results suggest that steering-related metrics may provide more information about driving performance than lane-related metrics do. Real-time driver drowsiness detection may substantially improve with technologies incorporating steering-related metrics.

Postiosoite	MATINE Puolustusministeriö PL 31 00131 HELSINKI	Sähköposti	matine@defmin.fi
Käyntiosoite	Eteläinen Makasiinikatu 8 00130 HELSINKI	WWW-sivut	www.defmin.fi/matine
Puhelinvaihde	(09) 16001	Y-tunnus	FI01460105
Pääsihteeri	(09) 160 88310	OVT-tunnus/verkkolaskuosoite	003701460105
		Itellan operaattorivälittäjätunnus	003710948874
Suunnittelusihteeri	(09) 160 88314	Verkkolaskuoperaattori	Itella Information Oy
Toimistosihteeri	050 5555 837	Yhteyshenkilö/Itella	helpdesk@itella.net
Faksi kirjaamo	(09) 160 88244		

1. Johdanto / Introduction

Sleepiness behind the wheel ("drowsy driving") is one of the main contributors to road crashes (NTSB, 1999). In Europe up to 20% of all traffic accidents are believed to be due to driver sleepiness (AWAKE, 2002). In the U.S., falling asleep while driving causes at least 100000 crashes annually: 40000 lead to nonfatal injuries and over 1500 lead to fatal injuries (Royal, 2002). These crashes happen primarily between midnight and 06:00 and are recognizable because they involve a single car with a sober driver who is alone and does not attempt to prevent the car from drifting off the road. In this context, finding countermeasures against sleep-related traffic accidents has increasingly become a priority over the last decade.

Prior research has focused on developing objective systems that monitor the driver and warn him/her when sleepiness or degraded driving performance compromise safety. The research trends towards systems that integrate both sleepiness detection and poor driving performance (Vadeby et al., 2010). Thus far, though, no commercially available system provides a sufficiently reliable warning system (Anund and Kircher, 2009).

The most accurate techniques to monitor sleepiness employ physiological measures (Johns et al., 2008, Vadeby et al., 2010). However, these tend not to be practicable for drivers, because the detector systems – involving electrodes or glasses – on which they rely are obtrusive to the driver. The most promising technique currently available is perhaps the video-based or infrared beam-based measurement of percent eye closure (Johns et al., 2007, Bergasa et al., 2006), although this technique requires high sampling frequency and is prone to data loss in sunny conditions or when the driver looks away.

Monitoring driving performance is unobtrusive because performance is measured indirectly from the behavior of the car – via signals of, e.g., its lane position, steering wheel angle, and speed. From these signals ("embedded performance metrics"), researchers usually compute lane deviation and other metrics derived from lane position, but metrics derived from steering and speed are also used (Kircher et al., 2002, Berglund, 2007, Mattson, 2007). In fact, in a literature search we found 87 metrics that have been used for research purposes. A few studies also made comparisons between a handful of metrics (Vadeby et al., 2010, Berglund, 2007, Mattson, 2007). However, it is unclear which metrics provide the best proxy for driver performance, because to our knowledge there is no extensive data-driven comparison between and evaluation of the signals and metrics. The present research contributes such a comparison and evaluation.

2. Research objectives and accomplishment plan

The overall aim of this 2-year project was to develop sensitive and specific metrics of driver drowsiness that are based on on-road, real-time signals.

For this purpose we used a high-fidelity driving simulator. Most research on driving performance employs simulators because they provide safety for the study participants, high repeatability, and ecological validity when the interest is in relative changes of performance (Philip et al., 2005).

Our dataset is drawn from two different simulator studies. *N*=41 participants participated in the two-week in-laboratory studies, in which they were assigned to either a day shift or a night shift schedule. A total of 40 driving sessions were administered to each participant at 09:00, 12:00, 15:00 and 18:00 in the day shift schedule, or at 21:00, 00:00, 03:00 and 06:00 in the night shift schedule. We quantified each driving session with 87 selected metrics and explored the resulting dataset with principal component analysis (PCA). PCA identifies the underlying dimensions of a dataset and ranks them according to the amount of total variance that they can explain (Hatcher, 1994, Davies and Fearn, 2005). The dataset does not only reflect the variance in the original signals drawn from the simulator. It also reflects the variance that the shift conditions induce, as well as the variance among all the metrics. Metrics that correlate strongly with the dominant dimensions, or the factor scores, should provide potent predictors of driver drowsiness.

3. Materials and methods

Our dataset is drawn from two different simulator studies, Study 1 and Study 2.

3.1 Participants

Twenty-five participants (ages 22 to 39, mean(±SD) 27.3(±5.5), males 50%) completed a 14-day in-laboratory study (Study 1), and 16 participants (ages 22 to 39, mean (±SD) 27.5(±5.65), males 100%) completed a 16-day in-laboratory study (Study 2). The participants were recruited with advertisements in local newspapers and on the internet. They received compensation ($2235 in Study 1 and $2485 in Study 2). The inclusion criteria were as follows: valid driver's license, good health (questionnaires and physical examination), absence of sleep disturbances (questionnaires, sleep diary, at-home actigraphy and baseline polysomnography), no shift work or travel across time-zones within one month of entering the study (questionnaires), no medications, no smoking, and no susceptibility to simulator adaptation sickness. The participants received detailed information about the study and signed an informed consent form before inclusion in the study. The study was approved by the Institutional Review Board of Washington State University.

3.2 Protocol

In Study 1, the experiment had two different shift conditions: day shifts (time in bed between 22:00 and 08:00) and night shifts (time in bed between 10:00 and 20:00). We randomized 12 participants to the day shift and 13 to the night shift. After five shifts, each participant had a 34-hour restart period (a total of 20 hours of sleep and 14 hours awake) before commencing the next five shifts. Each condition commenced with two baseline days and ended with a recovery day.

In Study 2, the experiment had only a night shift condition (time in bed between 10:00 and 20:00). After five shifts each participant had a 58-hour restart period (a total of 30 hours of sleep and 28 hours awake) before commencing the next five shifts. The schedule of baseline and recovery days was identical to that of the night shifters in Study 1.

Each participant took a total of 40 test sessions during the shifts in the experiment. The day shifters had scheduled test sessions at 09:00, 12:00, 15:00, and 18:00. The night shifters had scheduled test sessions at 21:00, 00:00, 03:00, and 06:00.

3.3 Measurements

We used a high-fidelity driving simulator (PatrolSim IV, L-3 Communications) that we adapted for research purposes by installing additional hardware and software (Moore et al., 2009). We developed a standardized driving scenario that involved rural highway driving with 10 straight and uneventful road segments ("straightaways") and 5 to 7 randomly located encounters (with pedestrians or dogs crossing the road). The drive totaled 30 minutes if the driver abided by the speed limit of 88.5 km/h throughout the scenario. The simulator continuously sampled lane position, steering wheel angle, driving speed, accelerator usage, car yaw angle, and engine torque at 72 Hz. From these signals we extracted the 10 straight segments and concatenated them so as to emulate a single continuous straightaway. Since each segment was about 30 seconds (depending on the driving speed), the concatenated straightaway that we used for analyses (see section 3.4) was about 5 minutes.

Each driving session was paired with a control battery of performance tests. Immediately prior to a driving session we administered the 10-minute psychomotor vigilance task (PVT) test (Van Dongen et al., 2003). The outcome metric was the number of lapses, defined as reaction times longer than 500 ms. Immediately after a driving session we administered computerized versions of the Karolinska Sleepiness Scale (KSS) and the 3-minute digit-symbol substitution task (DSST). In the KSS participants rated their feeling of sleepiness from 1 (very alert) to 9 (very sleepy). In the DSST participants were shown a key where digits (1 to 9) were associated with nine symbols. During the test, symbols were shown one at a time and participants typed the corresponding number. The outcome metric was the total number of correct responses.

3.4 Data analyses

We searched the literature for metrics that researchers have used to characterize driver performance. We found the metrics from papers on driving and papers on analysis of physiological signals. We also developed new metrics when we identified a performance aspect that was not reported in the literature on driving. From the 5-minute straightaway signals (see section 3.3) we extracted, in total, 87 metrics. We used Matlab 7.5.0 for the computations.

Study 1 yielded a 1000×87 matrix (from 25 participants, 40 bouts per participant, and 87 driving metrics). Study 2, which we used as a validation study, yielded a 640×87 matrix. To reduce the dimensionality of each matrix we performed principal component analyses (PCA) with orthogonal varimax rotation (SAS 9.2; SAS Institute, Inc.). We inspected the scree plots of eigenvalues to determine how many principal components to retain before rotation in order to explain most of the variance in the data. Given the relatively large number of metrics, we only interpreted principal component loadings with an absolute value of 0.5 or greater (Hatcher 1994).

To evaluate the effect of circadian timing on the retained principal components in study 1, we performed a mixed-effects analysis of variance (ANOVA) of shift type (day, night) by time of day. A significant interaction effect ($P \leq 0.05$) would indicate that the retained principal components could be indicative of drowsiness. To evaluate whether changes in the outcome of the PVT, KSS, and DSST in the control battery could explain changes in the driving performance, we also performed mixed-effects ANOVA of shift type by time of day with PVT, KSS, and DSST as covariates.

4. Results and discussion

The scree plot of eigenvalues indicated that there were two principal components, which together explained 47% of the total variance in the driving data (Fig. 1). Metrics capturing steering variability clustered on the fist principal component (*steering variability*), and metrics capturing lane variability clustered on the second principal component (*lane variability*).

Above we found that the principal component *steering variability* explained more variance than the principal component *lane variability* did. To verify the finding we performed a post hoc PCA involving four steering-related, four lane-related, and four speed-related metrics (we applied the metrics standard deviation, variance, root mean square, and average on each of the steering-, lane-, and speed signals). The scree plot revealed that there were still two principal components, which together explained 47% of the total variance in the data. Moreover, the steering-related metrics still clustered on the first principal component; the lane-related metrics clustered on the second principal component.

Mixed-effects ANOVA showed that with *steering variability*, the interaction of shift type by time of day was significant (F=3.11, P=0.026, see Fig. 2A). With *lane variability*, the interaction of shift type by time of day was significant (F=4.51, P=0.004, see Fig. 2B). The finding that circadian timing affected *steering variability* and *lane variability* suggested that they could be indicative of drowsiness. The mixed-effects ANOVA of shift type by time of day with PVT, KSS, and DSST as covariates showed that *steering variability* covaried with PVT (F=13.8, p<0.001, compare Figs. 2A and C) and KSS (F=16.6, p<0.001), but not with DSST (F=0.97, p=0.33). *Lane variability* covaried with PVT (F=14.5, p<0.001), KSS (F=37.5, p<0.001), and DSST (F=56.8, p<0.001, compare Figs. 2B and D). This means that for both *steering variability* and *lane variability* the PVT and KSS were significant covariates, but only for *lane variability* the DSST was a significant covariate. Thus, *lane variability* had a learning curve not seen in *steering variability*. It indicates that the PCA teased apart metrics that quantify different aspects of drowsiness.

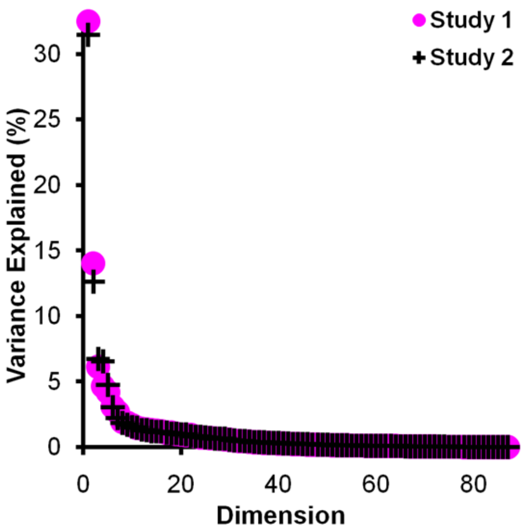

Fig. 1. Scree plot of eigenvalues expressed as variance explained. The variance explained is the eigenvalue divided by the number of components *M* (*M*=87). Component 1 explained 33% of the variance and component 2 explained 14% of the variance.

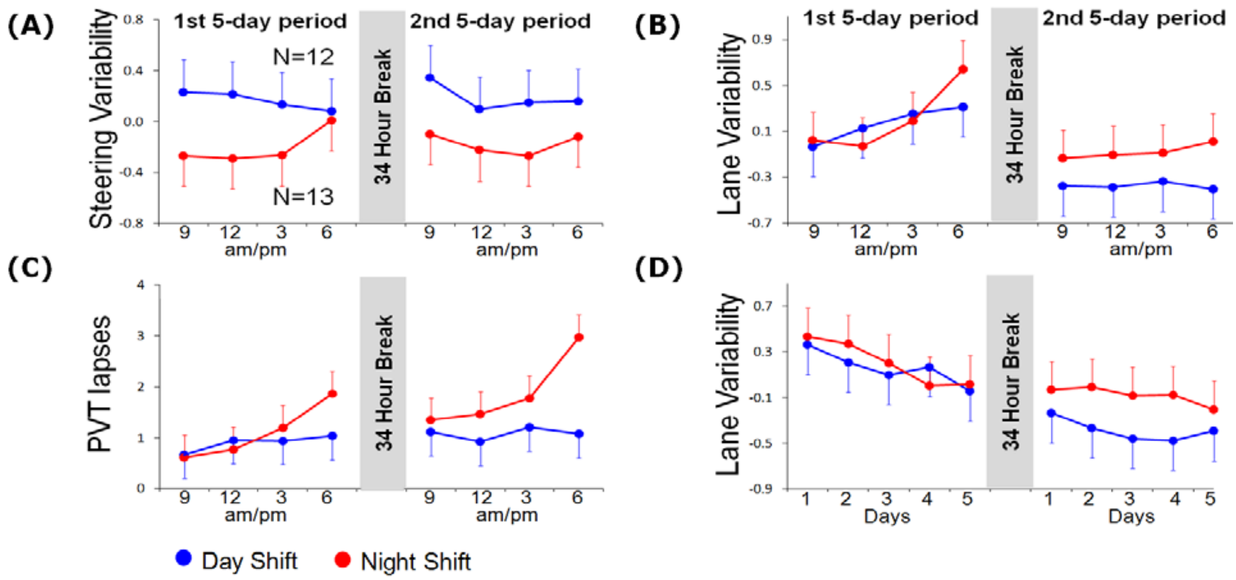

Fig. 2. Driving performance and PVT performance in the day-shift (blue) and night-shift (red) conditions of study 1. In panels A, B, C: steering variability, lane variability, and number of PVT lapses as a function of time of day, collapsed over days within each 5-day period. In panels D: lane variability as a function of days, collapsed over time of day for each 5-day period. The panels show group means (with standard errors). Times of day shown on the x-axes are through the day (from 9 am until 6 pm) for the day-shift condition and through the night (from 9 pm until 6 am) for the night-shift condition.

5. Conclusions

This study involved repeated driving sessions throughout a 2-week period of simulated day-shift and night-shift conditions. We showed that among the 87 selected metrics of driving performance there are two dominant dimensions, *steering variability* and *lane variability*, which together explain 47% of the total variance in driving performance. We also showed that whereas both *steering variability* and *lane variability* covary with PVT and KSS, only *lane variability* covaries with DSST, which indicates that metrics of *steering variability* and *lane variability* capture different aspects of driving performance.

The finding that *steering variability* conveys more information about driving performance agrees with the finding of Sandberg et al. (2011), who in a real-road study found no effects of circadian timing on the standard deviation of lane position. If steering is more informative about driving performance and driver drowsiness, this would be a positive finding in the sense that it is more difficult to monitor the lane position of the car – usually this is done by videoing lane markers and by applying picture recognition software.

In the current study the control battery PVT, which is a well-validated test of performance impairment from sleep loss and circadian misalignment, showed that the performance impairment in the current night-shift participants was modest (Fig. 2C) compared to the documented impairment from a night of total sleep loss or a week of sustained sleep restriction with 6 hours sleep per night (Van Dongen et al., 2003). Therefore, the next step would be to repeat the study with higher levels of sleep deprivation.

6. Scientific publishing and other reports produced by the research project

Abstract:

Forsman P, Mott C, Vila B, Van Dongen H. Combining lane deviation with steering metrics of simulated driving to detect driver drowsiness. *SLEEP* 2011; Volume 34, Abstract Supplement, A113-A114. (http://www.journalsleep.org/Resources/Documents/2011abstractsupplement.pdf, haettu 1.11.2011)

Presentation:

Forsman P. New metrics for detecting drowsy driving in real time as derived from high-fidelity driving simulators. Associated Professional Sleep Societies, Minneapolis, June 2011.

Manuscript:

We are finalizing a manuscript on the current findings for submission to an international peer-reviewed journal.

7. References

Anund A, Kircher K (2009). Advantages and disadvantages of different methods to evaluate sleepiness warning systems. VTI rapport 664A. Available: www.vti.se/publications

Awake Consortium (IST 2000-28062). System for effective assessment of driver vigilance and warning according to traffic risk estimation (AWAKE), Sep. 2001–2004. [Online]. Available: http://www.awake-eu.org

Bergasa L, Nuevo J, Sotelo M, Barea R, Lopez M (2006) Real-Time System for Monitoring Driver Vigilance. IEEE Transactions on intelligent transportation systems 7(1):63-77.

Berglund J (2007). In-vehicle prediction of truck driver sleepiness-steering related variables. Linköping University, Report number LiTH-ISY-EX-07/3942—SE

Davies A and Fearn T (2005). Back to basics: the principles of principal component analysis. Spectroscopy Europe 16(6):20–23.

Hatcher L (1994). A step-by-step approach to using SAS for factor analysis and structural equation modeling. Cary, NC: SAS Institute Inc.

Mattsson K (2007). In-vehicle prediction of truck driver sleepiness. Master's thesis. Luleå University of Technology 107 CIV.

Moore J, Van Dongen H, Belenky G, Mott C, Huang L, Vila B (2009). Use of a driving simulator to assess fuel inefficiency as a downstream effect of driver sleepiness in controlled laboratory experiments. Sleep 32:A387-A388.

Johns M, Tucker A, Chapman R, Crowley K, Michael N (2008). Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers. Somnologie 11:234–242.

Johns M, Chapman R, Crowley K, Tucker A (2008). A new method for assessing the risks of drowsiness while driving. Somnologie 12:66–74.

NTSB (1999) Evaluation of U.S. Department of Transportation Efforts in the 1990s to Address Operator Fatigue. Safety Report NTSB/SR-99/01. Washington, DC.: National Transportation Safety Board.

Philip P, Sagaspe P, Taillard J, Valtat C, Moore N, et al. (2005). Fatigue, sleepiness, and performance in simulated versus real driving conditions. Sleep 28(12):1511-1516.

Royal D (2002). Volume I—Findings report; national survey on distracted and driving attitudes and behaviours. The Gallup Organization, NHTSA.

Sandberg D, Anund A, Fors C, Kecklund G, Karlsson J, Wahde M, Åkerstedt T (2011). The characteristics of sleepiness during real driving at night – a study of driving performance, physiology and subjective experience. Sleep 34(10): 1317-1325.

Vadeby A, Forsman A, Kecklund G, Åkerstedt T, Sandberg D, et al. (2010). Sleepiness and prediction of driver impairment in simulator studies using a Cox proportional hazard approach. Accident Analysis and Prevention 42:835–841.

Van Dongen H, Maislin G, Mullington J, Dinges D (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 2:117-126.