Driver Sleepiness—Evaluation of Reaction Time Measurement as a Secondary Task

SD Baulk, LA Reyner, and JA Horne

Sleep Research Centre, Loughborough University, UK

Summary: The application of reaction time (RT) as a secondary task to determine sleepiness in drivers is of increasing interest, but is a problematic area. We assessed the extent to which RT reflected this sleepiness, and/or otherwise affected driving behaviour in sleep restricted, moderately sleepy people. They drove a real-car interactive simulator for two, two hour afternoon monotonous drives, with and without RT (counterbalanced). Simple auditory RT was used, with a semi-random inter-stimulus interval averaging 2½ minutes. Lane wandering (driving “incidents”), subjective and EEG measures of sleepiness were obtained. For both conditions all three indices changed significantly during the course of the afternoon circadian “dip”. However, this was not reflected in RT, which remained relatively stable. Nevertheless, RT provided more “stimulation” for the sleepy driver, and significantly reduced subjective sleepiness, with a trend for fewer incidents and a more alert EEG. Possible reasons for the disparity in sensitivity between RT and the other measures are discussed. Under this experimental protocol, RT did not provide a useful guide to driver sleepiness; it was merely a mechanism for increasing task load and reducing monotony. The drivers’ own insight into their sleepiness had more validity as a tool for assessing sleepiness.

Key words: Driver sleepiness; reaction time; driving performance; subjective sleepiness

INTRODUCTION

A DRIVER’S RESPONSE TIME IN APPLYING THE BRAKES IN AN EMERGENCY IS COMMONLY THOUGHT TO BE IMPAIRED BY SLEEPINESS. Thus, reaction time (RT) given as a task secondary to driving has been seen to be a valid method for monitoring sleepiness, measured by means of the driver pushing a steering wheel switch or a foot operated button in response to stimuli generated from within the vehicle. However, driving studies indicate that a sleepy driver will either respond almost normally to an emergency or not at all; hence a collision.1 That is, driver response time becomes disrupted by sleepiness rather than reflects a gradual decline.1 Research outside the field of driving has clearly demonstrated that the deterioration in RT performance with sleepiness is reflected in only a portion of responses being impaired, often through momentary lapses, as the majority of responses remain within normal parameters.2 But as these laboratory studies usually involve at least 24 hours without sleep, little is known about what might be expected with RT as a secondary task under less extreme but more typical conditions, for example, during the afternoon “dip” following a night of disturbed or restricted sleep, (i.e. in moderately sleepy drivers who otherwise show impaired driving [lane drifting] and declare themselves to be sleepy).

Another issue concerning the application of RT as a secondary task during driving, is that with monotonous roads the very act of responding to the RT stimuli could have an arousing effect, particularly if the stimuli are frequent. While this might seem to be beneficial, it masks underlying sleepiness. Moreover, too frequent a stimulus rate (e.g., averaging every 30 seconds) might even be counterproductive in distracting (even momentarily) the driver from attending to the road, which is why Mackie and Wylie3 advocated caution about such techniques. On the other hand, an infrequent stimulus (e.g., averaging 4—5 min) is unlikely to have such an effect, and be more sensitive to sleepiness. Unfortunately, the infrequency makes it of less use in detecting sleepiness, as a critical stimulus might come too late to avoid an accident. Ideally, one might undertake a frequency response study to determine an optimal stimulus presentation rate between these two extremes, but this would be a costly and time-consuming procedure.

With these issues in mind, we evaluated the application of simple RT using an auditory stimulus, given with inter-stimulus intervals averaging 2½ minutes, within the context of moderate and known levels of sleepiness, and using our standard driving protocol. In these respects we wished to see whether RT reflects this sleepiness, and/or otherwise affected driving behavior.

METHOD

Participants

Most sleep-related vehicle accidents occur in young adults4 and we targeted this age group. Participants selected for the study comprised 10 people, equal men and women, aged 20—29y, healthy (medication-free), experienced drivers (driving for >2y averaging >3 h per week), good sleepers, sleeping regularly 7—8 hours per night, were infrequent daytime nappers (< once a month), and were recruited by advertisement. None complained of daytime sleepiness, nor indicated potential sleep disorders in a screening questionnaire. Their Epworth Sleepiness Scores5 were within the normal range 0—10. They had the procedures fully explained to them, signed consent forms, and were paid to participate.

Design and Procedure

On an initial day following a normal night’s sleep, participants
had an afternoon 2 hour baseline practice drive on the simulator (see below). A week later they underwent two treatment conditions, a week apart, with both conditions having sleepiness enhanced by sleep restricted to 5 hours (delayed bedtime) the night before. Driving was between 14:00 and 17:00, when daytime sleepiness tends to increase due to the circadian “afternoon dip” and sleep-related accidents increase.6,7 Participants slept at home, with sleep monitored by wristactimeters. Both conditions comprised an initial “warm-up” 30-minute drive followed by a 30-minute break and then a two-hour continuous drive (see below). In the RT condition participants also had to respond to an audible stimulus by pressing, as soon as possible, a thumb-activated microswitch on the steering wheel. Stimuli were randomized with an inter-stimulus range of 50 to 300 seconds, averaging 150 seconds. RT began during the initial 30 min period. We used this simple form of RT in order to minimize practice effects and distraction from driving. The second condition was “normal,” without RT. The order of these two conditions was counterbalanced.

**Simulator**

This has been described in detail elsewhere8 and comprises an immobile car with an interactive full-size computer-generated bending, dull, and monotonous roadway projected on to a 2.0m x 1.5m, screen located 2.3m from the windshield. There were two “up” and two “down” lanes, hard shoulder and simulated auditory “rumble strips.” Participants sit in the driving seat and drive at their normal cruising speed within white lane markings. Lane drifting is the usual manifestation of sleepy driving, which was automatically detected by the computer data-logger from continuously recorded steering data. A car-wheel crossing a lateral lane marking is the criterion for this detection, and identified as an “incident.” An unobtrusive infrared camera films the driver’s face, which is recorded with the roadway using a split-screen video display. The video data are further analyzed by a skilled assistant “blind” to the experimental conditions, whereby all the automatically identified incidents are checked to see whether: (1) these are due to driver distraction (looking elsewhere), which are discounted, or (2) to episodes associated with sleepiness (i.e., eye “rolling” or vacant staring ahead). The EEG (see below) during (1) shows little alpha and theta activity, whereas during (2) either or both activities are present, often with slow “eye rolling” on the EOG (see below). Additional quality checks on these video, EEG and EOG data are undertaken “blind,” by one of us (LAR).

**Subjective Sleepiness**

Every 200s, participants were asked to respond verbally with a number from the nine-point Karolinska Sleepiness Scale (9): 1=extremely alert, 2=very alert, 3=alert, 4=rather alert, 5=neither alert nor sleepy, 6=some signs of sleepiness, 7=sleepy, no effort to stay awake, 8=sleepy, some effort to stay awake, 9=very sleepy, great effort to keep awake, fighting sleep). The scale and descriptors were printed on the car’s dashboard, within easy view of the driver.

**EEG**

Electrodes were attached for one channel of EEG (C3-A1). To identify “eye-rolling” two channels of EOG were also recorded (electrodes 1cm lateral to and 1 cm above left outer canthus, and 1 cm lateral to and 1cm below right outer canthus; both referred to A2). The EEG was digitally recorded using the “EMBLA” system (Flaga, Iceland), and spectrally analysed in four seconds periods. High and low band-pass filtering of the EEG at >15 Hz and <4 Hz removed slow eye movements and muscle artifact. There was some unavoidable eye blink contamination on the EEG, which was mostly filtered out, and does not bias the EEG outcomes.10 Increases in EEG power in the alpha (8—11 Hz) and theta (4—7 Hz) ranges indicate increasing sleepiness (e.g.,10,11). We10 have found that under these conditions summation of alpha and theta (i.e., slower EEG activity in the range 4—11 Hz) showed more consistent EEG findings in relation to driver sleepiness than either alone. For example, some sleepy subjects display alpha surges, others show little alpha but much theta (c.f. 10,11). EEG power in this frequency range, for the four-second periods, was averaged in one-minute epochs. It was then standardised (to remove individual differences in mean EEG power levels) for each participant by the following transformation,10 before averaging across participants:

\[
\text{transformed epoch} = \frac{\text{Difference from mean of first 30 minutes}}{\text{Standard deviation of first 30 minutes}}
\]

**Statistical Analyses**

For the two-hour continuous drive, incidents, EEG, and subjective sleepiness data were averaged into four, 30-minute periods per subject and condition, and two-way (condition x time) repeated measures ANOVAs were applied (using the Huynh-Feldt epsilon adjustment to degrees of freedom). In the case of the RT data over the two-hour period, these were analyzed in 15-
RESULTS

Figure 1 presents mean KSS scores over the two hours of driving for both groups and sleep conditions. When assessed in 30-minute periods, subjective sleepiness was significantly less with RT (F=7.89 [1.9] p<0.02; Huynh-Feldt e=1). The effect of time was also significant (F=12.61 [1.5,13.5] p<0.02; d.f. corrected by e=0.51). This time effect also showed a significantly quadratic ("inverted U") trend, with an increase in sleepiness during the mid-afternoon circadian dip, followed by an improvement in alertness. There was no significant interaction between condition and time. Figure 2 gives the mean numbers of driving incidents. Although there was a trend for these to be lower with RT, this failed to reach significance. Again, there was a significant effect of time (f=6.31 [2.5; 22.2] p<0.004; d.f. corrected by e=0.82), that was also significantly quadratic. There was no significant interaction. With regard to the EEG, Figure 3 shows both conditions to have similar trends for the first hour or so, with RT leading to a less sleepy EEG thereafter. However there were no overall significant effects for either condition or time.

The mean RT data for the RT condition can be seen in Figure 4. These levels remained fairly stable throughout the 2 h drive, with the fluctuations across the 15-minute epochs being small and non-significant. Figure 5 is compiled from Figures 1 and 4 and compares both KSS and RT trends over time. The correlation between these values over the two hours was non-significant at r=0.08 (d.f.=9). RT data were checked for possible "lapses" (i.e., response times in excess of two seconds). These occurred for only three out of a total of 480 RT responses (48 responses for each participant). Although each was also associated with a driving incident, the reverse was not true. That is, for the great majority driving incidents, which were indicative of lapses, there was no apparent RT lapse to be seen in the nearest RT response, either preceding or following the incident.

DISCUSSION

Under these driving conditions and level of sleepiness, the auditory RT method we adopted was not sufficiently sensitive to sleepiness and lapses in driving ability, even though there were manifest signs of sleepiness in the form of increased lane weaving (incidents). However, inasmuch that the RT provided more activity and "stimulation" for the sleepy driver during a monotonous drive, it reduced subjective sleepiness and somewhat improved driving ability as measured by incidents. It could be argued that had we used more participants, then RT might have shown small but statistically significant effects. However, in terms of practical use, such statistical criteria are not appropriate.

There are various reasons for the disparity between driving incidents and RT performance. The RT stimuli only occurred on average every 2½ minutes, whereas the average interval for the incidents was approximately seven minutes. As each event (incident or RT response) only occupied a second or so, it is unlikely that both would coincide, especially as the RT task was an externally paced event, whereas the incidents were caused by the driver. Also, both types of event could alert the driver for a while, and facilitate normal behavior with respect to the next RT stimulus or potential for lane drifting.

On the other hand, lapsing detected by these RT methods may be too late for a speeding driver to avoid a collision, as the vehicle may already be running off the road during the lapse. A further issue is that even during stage 1 sleep a person has about a 50% chance of responding normally (with a microswitch) to an auditory stimulus without fully awakening. Such findings suggest that even during a lapse, a sleepy person could still respond to an auditory stimulus (i.e., not all lapses are detected by RT).

We have shown previously that sleepy drivers have good insight into their level of sleepiness. The present findings further confirm this, as shown by increasing sleepiness reflected by the KSS data. The failure of the RT data to reflect this, given that sleep-related incidents also increased, as did EEG determined sleepiness, indicates that subjective sleepiness provides a better gauge to sleepiness than does RT.

What must seem to be ill-conceived commercial RT devices...
aimed at sleepy drivers are coming on to the market. We examined one of these, that was claimed by the manufacturers to help maintain or improve driver alertness. Although based on auditory RT, it was more demanding of the driver’s attention than was the case here. Alert drivers could cope adequately with both driving and the RT system, but when they became sleepy the device distracted them from the immediate driving task, and led to more incidents. This made them realize that they were having difficulty coping, which in turn heightened their perception of sleepiness. However, the RT responses showed no clear or consistent changes, indicating that the sleepy drivers put more compensatory effort into the RT device (perhaps in order to avoid triggering the loud and rather aggravating warning tone), at the expense of driving ability. In sum, a RT system demanding effort and attention by the sleepy driver may lead to unwanted distraction from driving, whereas a less demanding RT system of the type we used here, may be beneficial in improving overall alertness (i.e., the “inverted U” concept for arousal and performance applies here). Nevertheless, neither system was effective in detecting sleepiness.

Together, all these factors illustrate some of the difficulties with RT as a secondary task to detect driver sleepiness. Maycock in his recent unpublished, comprehensive review of this topic concluded that many aspects of this area require further investigation, and it can not be said with any confidence that RT as fatigue monitoring method could be made both effective and reliable. Above all, and as Maycock pointed out, we do not really know what is being monitored by RT within the context of driving.

In summary, under this experimental protocol, RT did not provide a useful guide to driver sleepiness. Seemingly, it was merely a mechanism for increasing task load and reducing monotony. The drivers’ own insight into their sleepiness had more validity as a tool for assessing sleepiness.

ACKNOWLEDGMENTS

This work was supported by the UK Department for Transport, Local Government and the Regions. We would like to thank Pauline Barrett for her assistance.

REFERENCES

15. Reyner LA, Barrett PR, Horne JA. The effectiveness of a commercial reaction time device to identify the onset of driver sleepiness. Sleep Res Online, 1999;2(suppl 1):240