*J. Sleep Res.* (2005) **14,** 393–400

# Task-dependent differences in subjective fatigue scores

S. RICHTER[1], K. MARSALEK[1], C. GLATZ[2] and A. GUNDEL[1]

[1]German Aerospace Centre, Institute of Aerospace Medicine, Department of Transport safety, Cologne and [2]University of Applied Sciences, Bremerhaven, Department of Medical Engineering, Germany

Accepted in revised form 9 June 2005; received 20 January 2005

**SUMMARY** The aim of the present study was to evaluate time-on-task effects on subjective fatigue in two different tasks of varying monotony during night-time testing (20:00 to 4:00 hours) in a sleep deprivation intervention. The experiment included eight test runs separated by breaks of approximately 20 min. Twenty healthy volunteers performed a driving simulator and the Mackworth clock vigilance task in four of the test runs each. Sequence of tasks was varied across subjects. Before and after each task, subjective sleepiness was assessed by means of the Karolinska sleepiness scale and subjective fatigue was rated on the Samn-Perelli checklist. Fatigue and sleepiness significantly increased over the course of the night. Both tasks led to an increase in fatigue and sleepiness across test runs. However, this time-on-task effect was larger in the vigilance than in the driving simulator task. It is important to note that fatigue and sleepiness in one test run were not influenced by the task performed in the preceding test run, that is there were no cross-over effects. The results suggest that time-on-task effects superimpose circadian and sleep-related factors affecting fatigue. They depend on the monotony of the task and can be quantified by means of a design including separate test runs divided by breaks.

KEYWORDS driving simulator, fatigue, model, time-on-task, vigilance

## INTRODUCTION

Driver's fatigue has often been cited as a cause of road accidents (Knipling and Wang, 1994; Maycock, 1997; Thomas and Attard, 1994; ten Thoren and Gundel, 2003; Williamson *et al*., 2001). Consequently, efforts have been made to develop strategies to warn a driver in case his driving performance is in danger of being deteriorated by fatigue. One example are non-intrusive, vision-based approaches, i.e. monitoring of the driver's eyes. If long periods of eye closure are detected, a warning signal is produced (Eriksson and Papanikolopoulos, 2001; Wierwille *et al*., 1994). While such systems bear the disadvantage of being expensive and not yet sufficiently reliable, an alternative approach is to predict phases of reduced alertness with the help of mathematical models simulating the most important factors causing fatigue (e.g. Akerstedt *et al*., 2004; Spencer and Gundel, 1998; Moore-Ede *et al*., 2004).

*Correspondence*: Dr Stefanie Richter, Institute of Aerospace Medicine/Transport Safety, German Aerospace Centre, Linder Hoehe, 51147 Cologne, Germany. Tel.: +49 2203 601-2616; fax: +49 2203 695211; e-mail: sr.@bunyip.de

The main factors are the time of day (i.e. the circadian rhythm), time since last sleep, sleep duration and sleep quality (Borbély, 1982; Lan *et al*., 2002; Moore-Ede *et al*., 2004). Most models typically assume that the circadian component has a sinusoidal and the sleep-related component an additive exponential effect on fatigue (Akerstedt *et al*., 2004; Borbély, 1982). As a third component, a time-on-task effect has been proposed (Akerstedt *et al*., 2004; Spencer and Gundel, 1998). That is, fatigue resulting from performing a task superimposes circadian and sleep-related fatigue and can be isolated by means of a design with several test runs separated by breaks. The break is supposed to 'set back' fatigue to the value which is caused by the combined influence of the time of the day and sleep-related factors.

Former studies investigated how the time of the day and time since last sleep affect performance in different kinds of tasks and a decrease in performance was taken to indicate increased fatigue (Graw *et al*., 2004; Williamson *et al*., 2001). Furthermore, results suggest that performance decrements across time depend on the type of task. More specifically, the degree of monotony in a task seems to be the crucial factor determining performance degradation (Bonnet, 1994; Kraemer

© 2005 European Sleep Research Society

*et al.*, 2000; Pilcher and Huffcutt, 1996; Porcu *et al.*, 1998; Williamson *et al.*, 2001). While some studies suggest that more complex tasks are more vulnerable to the effects of fatigue (Bonnet, 1994; Pilcher and Huffcutt, 1996), results of other studies show that more cognitively demanding tasks have a less fatiguing effect, i.e. they are more interesting and thus induce increased arousal. Kraemer *et al.* (2000) found time of the day effects for a calculation and a visual search test. No changes were found for a more complex reaction time task. In the study by Williamson *et al.* (2001) effects of sleep deprivation and time of day were found for simple reaction time, Mackworth clock vigilance and symbol digit coding tasks. No effects were found for a visual search and logical reasoning task. In a study by Porcu *et al.* (1998) the ability to perform visual-attentive tasks was substantially spared during a nighttime session. Performance on a monotonous letter cancellation task, however, degraded.

Finally, there are also examples for the effect of task complexity on driving performance. Previous studies show that cognitive overload caused by visual displays in navigations systems may increase driving errors (Liu, 2000; Parkes and Coleman, 1990). In the study by Liu (2000) all drivers showed a better control of their car when using multimodal and auditory displays compared with visual-only displays. On the contrary, studies with driving simulators using monotonous virtual environments also revealed performance degradations (Thiffault and Bergeron, 2003; Verwey and Zaidel, 2000). Thus, both cognitive overload and underload seem to impair driving performance. Regarding navigation systems, auditory wayfinding information may not only decrease visual interference and driving errors, but also have an alerting function paralleling the effect of social interaction, e.g. when driving with a passenger (Akerstedt and Landström, 1998). However, this has not been investigated so far.

Not only performance might be affected by the kind of task performed, but also subjective fatigue. Time-on-task effects on subjective fatigue for different kinds of tasks are not well investigated (Akerstedt *et al.*, 2004). However, as Johns (1998) noted, mathematical models that describe changes in sleepiness over time could be criticized for disregarding important influences on sleepiness. One of these is the time a subject spends on a task (i.e. time-on-task effects). A problem with the investigation of subjective fatigue is poorly defined concepts and terms, often used differently by different disciplines and investigators (Johns, 2000). In the present study, the psychological concept of fatigue as assessed by means of the Samn-Perelli checklist (Samn and Perelli, 1982) is distinguished from the concept of sleepiness. The latter can be considered as sleep propensity or the probability to fall asleep at a particular time (Johns, 2000) and might be assessed with the help of the Karolinska sleepiness scale (Akerstedt, 1990). An often-used indicator of sleepiness is the electroencephalogram (EEG). EEG measures are obtrusive and inconvenient, however, thus alternative measures are necessary from a practical point of view. A validation study has shown that subjective scales are one such alternative (Akerstedt and Gillberg, 1982).

The aim of the present study was to evaluate time-on-task effects on subjective fatigue in two different tasks of varying monotony, i.e. in the more monotonous Mackworth clock vigilance task and a more interesting driving simulation task. Results were supposed to provide evidence as to how monotony affects fatigue. Furthermore, results were considered relevant for the validation of the time-on-task component in mathematical fatigue models.

We firstly wanted to describe the combined influence of time of day and sleep-related factors, as well as the superimposed time-on-task effect on fatigue and sleepiness. Secondly, we wanted to validate the assumption that the more monotonous vigilance task induces higher fatigue and sleepiness than the driving simulation task. Thirdly, we were interested in changes in performance in the vigilance and driving simulation tasks over the course of the night. Finally, we investigated the effect of acoustic way finding information on fatigue, sleepiness and performance in the driving simulator task.

## MATERIAL AND METHODS

### Subjects

Twenty healthy volunteers participated in the study (10 female, 10 male; mean age 29.5 ± 7.5 years, range 19–42 years). All but one ambidextrous subject were right-handed according to the Oldfield handedness inventory (Oldfield, 1971). Mean score on the Pittsburgh sleep quality index (PSQI; Buysse *et al.*, 1989) was 4.6 ± 2.4. Apart from one subject with a PSQI score of 11, all subjects had a PSQI-score ≤7. Results of the subject with the higher score were not different from those of the remaining subjects, however. Participants were assigned to four subgroups (see below). These subgroups did not differ with respect to gender, age, handedness or PSQI (univariate analyses of variance, all $P > 0.5$). No subject reported chronic or current major medical illness or injury, medication or drug consumption, shift work or transmeridian travel within the last 3 months prior to the study. During the week preceding the study participants had to keep a sleep diary to assess sleep habits. They were instructed not to take daytime naps during that time, i.e. to go to sleep only once a day and to refrain from excessive physical activity, caffeine and alcohol consumption. Finally, subjects were told not to consume alcoholic or caffeine beverages during the 7 h before the experiment. All participants gave written informed consent and the study was approved by the local ethics committee.

### Experimental procedure

#### *Experimental tasks*

Two tasks had to be performed, a driving simulation and the Mackworth clock vigilance task. Both tasks took about half an hour, each. The driving simulation was performed by means of the '3D-driving school' (3D-Fahrschule; Besier 3D Edutainment) implemented on a Toshiba Satellite 3000 Notebook. The scene was projected on the wall (Epson LCD Projector Model EMP-75, Seiko Epson Corporation, Nagano, Japan) in front of the subject (67°), who sat in front of a steering wheel and pedals. On the steering wheel there were buttons used to go into the forward or reverse gear, for the turn signals, mirrors and for looking over the shoulder when turning left of right. The driving task involved a night drive through Berlin (Germany). The program measured driving errors like speeding, disregard of traffic signs or traffic lights, forgetting to look into the mirror or over the shoulder when turning left or right. Written information about the route the subjects had to follow was projected on the wall on the upper margin of the scene. In one half of the subjects, information about the route was also given verbally, i.e. a female voice was telling where to drive.

A handheld device (Palm m505, Palm One Inc. Milpitas, CA, USA), the display of which was 5.6 cm in width and height, was used to administer the vigilance task. In the Mackworth clock vigilance task (Mackworth, 1950) a small point (diameter: 2.7 mm) moves on a circular path (diameter: 41.0 mm) consisting of 32 successive dots of the same size as the target point. The interval between each movement of the target point has a duration of 2000 ms. A button has to be pressed if the point 'jumps' across one of the dots on its path, which happened, on the average, four times per minute (100 jumps). The number of jumps recognized (hits), as well as the number of false alarms were measured.

Individual sleepiness was assessed by means of the Karolinska sleepiness scale. In the original scale (Akerstedt, 1990), sleepiness scores range from 1 to 9, and only odd scores involve verbal descriptions. In the present study, a modified version of the scale was used. Scores ranged from 1 (extremely alert) to 10 (extremely sleepy, cannot keep awake) and verbal descriptions were given for both even and odd scores. Subjects were asked to indicate which state description applied to them in the last 10 min. Fatigue was measured with the help of the Samn-Perelli checklist (Samn and Perelli, 1982). In the Samn-Perelli checklist, descriptions of 10 different states of fatigue or wakefulness are given in random order, i.e. not in increasing order of fatigue. The subject is asked to assess for each of the 10 states if it fits the individual state (score = 1), if the individual state is better (score = 0), or worse than the one described in the scale (score = 2). Thus, a total score of 20 (maximum fatigue) can be reached. Also the fatigue and sleepiness ratings were performed with the help of the handheld device.

### *Experimental design*

Subjects had to come to the lab twice. In the first session, which took about 2 h, subjects were given information about the course of the experiment and individual data about each subject was assessed. The Pittsburgh Sleep Quality Index was administered and subjects had to give their written consent to take part in the study. Afterwards subjects were familiarized with the driving simulation task. First, the use of the buttons on the steering wheel and the pedals were explained. Then, participants had to perform nine test drives with increasing complexity provided by the software. Subjects were finally asked to fill-in a standard sleep diary during the following 7 days preceding the actual experiment.

The actual experiment was performed in a second session about 1 week after familiarization. It started at 20:00 hours and finished at 4:00 hours. There were eight experimental test runs with a duration of 60 min and starting at the beginning of each full hour. In four test runs the driving simulation had to be performed, in the remaining four test runs, the Mackworth clock vigilance task was administered. The order of tasks was varied across subjects. Ten subjects (five male and female, each) started with the driving simulation (DS), the other 10 with the Mackworth clock vigilance task (MCVT). In the first group the sequence of tasks was: DS–MCVT–MCVT–DS–DS–MCVT–MCVT–DS. In the second group the sequence mirrored that in the first group, i.e. was MCVT–DS–DS–MCVT–MCVT–DS–DS–MCVT. Five randomly chosen subjects of each group performed the driving simulation with and without additional acoustic wayfinding instruction. The design was not balanced, because during the vigilance task acoustic information was never given. That is, the presentation of wayfinding instructions was only varied in the simulator task.

### *Statistical analyses*

Initial analysis revealed that the presentation of acoustic wayfinding information in the driving simulation task had no transfer effect on performance in the vigilance task or on fatigue and sleepiness scores. Therefore, in the statistical analyses concerning vigilance, fatigue and sleepiness, groups with and without acoustic wayfinding information were merged. In contrast, in the driving simulation task, acoustic wayfinding information affected driving errors. Thus, groups with and without such information were considered separately when analyzing driving errors.

The main analysis regarding sleepiness and fatigue scores was performed with SAS® software (SAS Institute Gmbtt, Heidelberg, Germany) (procedure MIXED) and involved a mixed model with TIME (eight test runs), TASK (vigilance vs. driving simulation), and PREPOST (before vs. after task performance) as fixed and SUBJECT and SUBJECT by TIME as random factors. This model turned out to be the most appropriate regarding Akaike's information criterion (AIC; Akaike, 1974).

Two additional analyses were performed on subjective scores. First, to illustrate 'break-effects', i.e. the reduction of fatigue and sleepiness due to the break, we computed mixed models with the fixed factors TIME (break 1 to break 7), TASK (vigilance versus driving simulation), and BREAK (before break versus after break) and the random factors SUBJECT and SUBJECT by TIME. Secondly, to account for cross-over effects, i.e. effects of the task performed preceding the task of interest, we computed a mixed model with TASK (vigilance versus driving simulation), SEQUENCE (vigilance preceding versus driving simulation preceding), and

PREPOST (before versus after task performance) as fixed factors and SUBJECT as random factor.

The number of hits in the vigilance task was analyzed by means of a mixed model with TIME as fixed factor and SUBJECT as random factor. To account for the fact that data are not normally distributed, number of hits was transformed with the formula ln[(101 − number of hits)/number of hits]. Number of hits was subtracted from a reference value of 101 instead of 100, because there were subjects showing the maximum number of 100 hits which would have led to a difference of zero. For the analysis of false reactions, this transformation could not be performed as there were several subjects showing no false reactions at all and division by zero is not defined. A nonparametric statistical analysis was therefore performed, i.e. a Friedman test with the factor TIME on the number of false reactions. Misses in the vigilance task were not analyzed because the number of misses is dependent on the number of hits and its analysis does not add important information regarding vigilance changes. Errors in the driving simulation task were analyzed by means of a mixed model with TIME and ACOUSTIC WAYFINDING INFORMATION as fixed factors and SUBJECT as random factor.

Finally, to analyze if changes in fatigue and sleepiness went along with performance changes during the night, cross-correlations were calculated. Across the eight test runs, we considered mean fatigue and sleepiness scores both before and after task performance (i.e. PRE and POST) on the one hand and mean transformed hits and false alarms in the vigilance task, as well as errors in the driving simulation task on the other. That is, mean subjective fatigue and sleepiness at the beginning of each test run was cross-correlated with the performance measures in the test runs. The same was carried out for mean fatigue and sleepiness at the end of each test run.

## RESULTS

### Sleepiness and fatigue

*Changes across the test night*

Figure 1 shows that sleepiness significantly increased across time (significant effect of TIME: $F_{7,125} = 47.15$, $P < 0.0001$) and that this increase was linear (*posthoc* linear contrasts, $P < 0.0001$). Moreover, subjective ratings rose over the course of the task, i.e. there was a significant PREPOST effect ($F_{1,143} = 66.34$, $P = 0.0001$). These PREPOST effects differed for the factor TIME (significant interaction between PREPOST and TIME: $F_{1,143} = 3.7$, $P = 0.0010$). Variation of the PREPOST effect with TIME was evident in the fact that pre–post differences in sleepiness were not significant in the first and sixth test run in *posthoc* analyses of simple main effects with a Bonferroni adjusted $\alpha$ of 0.00625. The interaction between TIME and TASK ($F_{7,125} = 0.40$, $P = 0.8982$) and the triple-interaction between TIME, TASK, and PREPOST were not significant ($F_{7,143} = 0.69$, $P = 0.6798$).

Figure 2 shows that fatigue increased across time (significant effect of TIME: $F_{7,118} = 30.45$, $P < 0.0001$). *Posthoc*

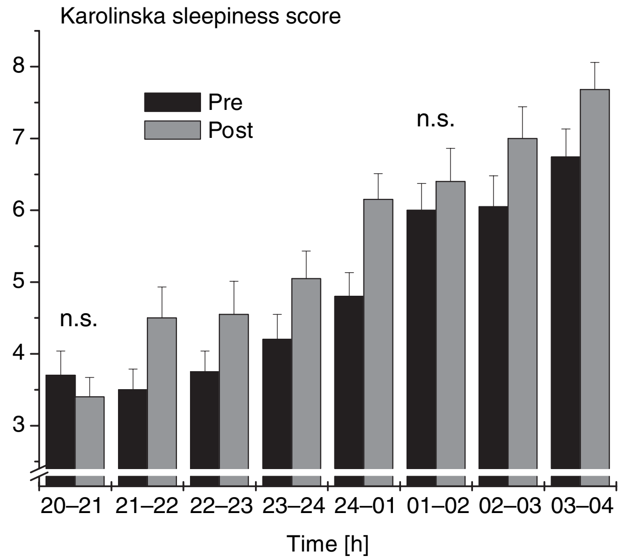


**Figure 1.** Mean values and standard errors of sleepiness at the beginning (pre) and end (post) of each task performance, averaged across the two tasks. Sleepiness significantly increased across TIME ($P = 0.0001$) and across test runs (PREPOST effect, $P = 0.0001$). PREPOST effects differed for the factor TIME (PREPOST × TIME: $P = 0.0010$). Differences in sleepiness before and after task performance were significant in all but the first and sixth test run. The TIME by TASK ($P = 0.8982$) and TIME by TASK by PREPOST interaction ($P = 0.6798$) were not significant. n. s. = not significant.

contrasts revealed significant linear ($P < 0.0001$) and quadratic ($P = 0.024$) changes. Moreover, subjective ratings significantly rose across test runs, i.e. there was a significant PREPOST effect ($F_{1,135} = 73.44$, $P = 0.0001$). These PREPOST effects varied with TIME (significant interaction between PREPOST and TIME: $F_{7,135} = 2.71$, $P = 0.0116$). *Posthoc* computations of simple main effects with a Bonferroni adjusted $\alpha$ of 0.00625 revealed that pre–post differences in fatigue were not significant in the first and last test run. The interaction between TIME and TASK was not significant ($F_{7,135} = 1.08$, $P = 0.3798$), but the triple interaction between TIME, TASK, and PREPOST ($F_{7,135} = 3.66$, $P = 0.0012$). To more specifically describe this effect, for each test run *posthoc* analyses of variance on fatigue scores were calculated with TASK as between-subjects factor and PREPOST as within-subject factor. Considering a Bonferroni-adjusted $\alpha$ of 0.00625, significant TASK by PREPOST interactions were found in the second ($P = 0.001$), seventh ($P = 0.005$) and eighth test ($P = 0.002$) run.

The analysis of break effects on sleepiness and fatigue similarly revealed significant effects of TIME (sleepiness, fatigue: $P = 0.0001$) and TASK (sleepiness: $P = 0.0001$, fatigue: 0.0014). More important, significant effects of BREAK (sleepiness: $P = 0.0121$, fatigue: 0.0001), and BREAK by TASK (sleepiness: $P = 0.0001$, fatigue: $P = 0.0003$) were found showing that the degree of recovery due to the break was dependent on the task. The TIME by TASK effect was significant for fatigue ($P = 0.0013$), but not

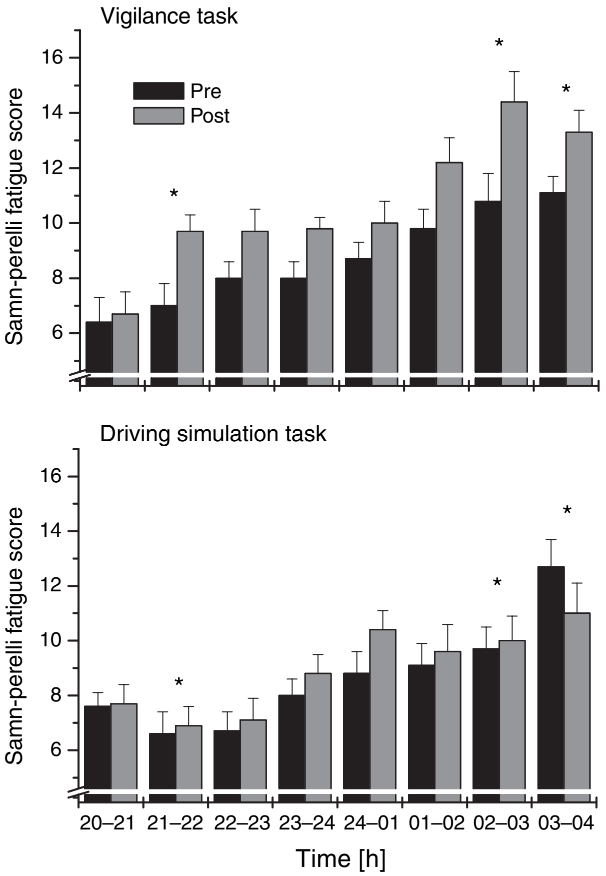


**Figure 2.** Mean values and standard errors of fatigue at the beginning (pre) and end (post) of each test run, separated for the vigilance (top) and driving simulation (bottom) task. Fatigue significantly increased across TIME ($P = 0.0001$) and across test runs (PREPOST effect: $P = 0.0001$). PREPOST effects differed for the factor TIME (PREPOST × TIME: $P = 0.0116$), but the TIME by TASK interaction was not significant ($P = 0.3798$). Finally, the TIME by TASK by PREPOST interaction was significant ($P = 0.0012$) and *posthoc* tests showed that this was because of the second, seventh and eighth test run (*significant with $P < 0.00625$).

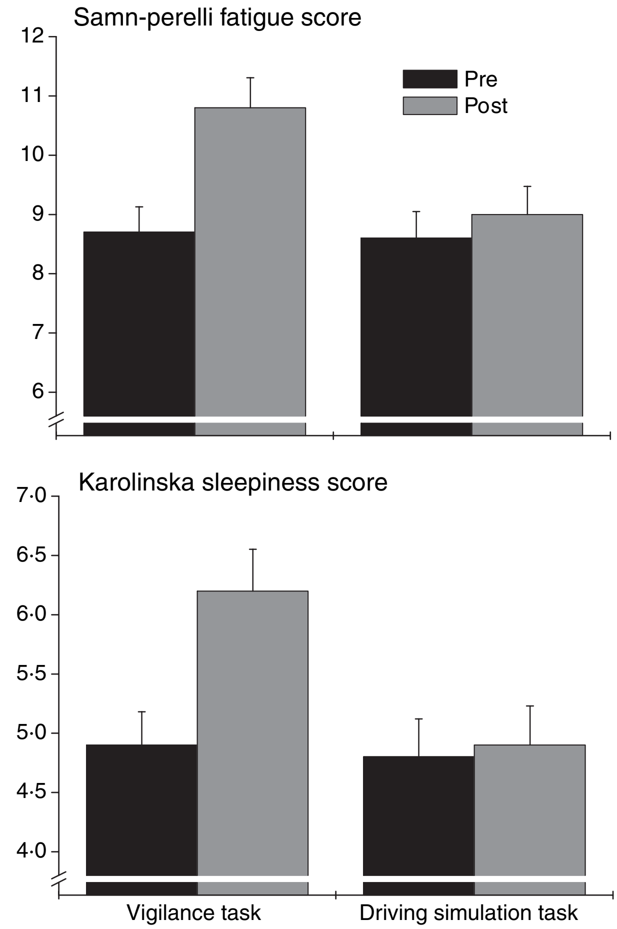


**Figure 3.** Mean values and standard errors of fatigue and sleepiness at the beginning (pre) and end (post) of the vigilance (left side) and driving simulation task (right side), averaged across test runs. Fatigue was larger in the vigilance than in the driving simulation task (TASK-effect: $P = 0.0001$) and increased more strongly across task performance in the vigilance as compared with the driving simulation task (TASK by PREPOST interaction: $P = 0.0001$).

for sleepiness ($P = 0.5732$). All other effects were not significant ($P > 0.36$).

### *Time-on-task effects*

Averaged across the assessments before and after a test run, subjective fatigue and sleepiness were larger in the vigilance task than in the driving simulation task (significant effect of TASK; sleepiness: $F_{1,143} = 26.97$, $P = 0.0001$; fatigue: $F_{1,135} = 16.26$, $P = 0.0001$). However, Fig. 3 shows that this was because of the fact that the vigilance task induced larger fatigue and sleepiness across task performance than the driving simulation task (significant interaction between TASK and PREPOST; sleepiness: $F_{1,143} = 40.17$, $P = 0.0001$; fatigue: $F_{1,135} = 40.17$, $P = 0.0001$). *Posthoc* computations of simple main effects with a Bonferroni adjusted $\alpha$ of 0.025 revealed that the pre–post difference was significant for the vigilance (sleepiness, fatigue: $P = 0.0001$), but not the driving simulation task (sleepiness: $P = 0.25$; fatigue: $P = 0.19$).

### *Cross-over effects*

The analysis revealed that fatigue and sleepiness in a test run did not depend on the task performed before, no matter what task was actually performed in the considered test run. Moreover, the changes in fatigue and sleepiness across test runs (PREPOST effect) did not differ depending on the preceding task. The latter effect was again not modulated by the task actually performed. Thus, there was neither a significant main effect of SEQUENCE, nor any significant interaction between two or all of the factors SEQUENCE, TASK, and PREPOST (all $F$-values $<1.0$ and all $P$-values $>0.4$). As in the main analysis, the TASK effect, PREPOST effect and TASK by PREPOST interaction were significant also in the presence of the SEQUENCE factor (all P-values $\leq 0.0016$).

### Performance in the vigilance and driving simulation tasks

*Vigilance task*

Figure 4 reveals that the transformed number of hits in the Mackworth clock vigilance task changed significantly across time (main effect of TIME; $F_{7,54} = 9.48$, $P = 0.0001$). *Posthoc* comparisons with a Bonferroni correction for multiple tests ($\alpha = 0.025$) further showed that the TIME effect was significant both in the group in which the vigilance task was performed in the second, third, sixth and seventh test run ($P = 0.0001$) and in the group in which the vigilance task was performed in the first, fourth, fifth and eighth test run ($P = 0.0002$).

Regarding false reactions, the overall TIME effect across the eight test runs was not significant ($\chi^2 = 7.374$, $P = 0.391$). *Posthoc* analysis (Friedman-test, $\alpha = 0.025$) within the four test runs of each experimental group revealed no significant effects, too. That is, the TIME effect on the number of false reactions was not significant both in the group in which the vigilance task was performed in the second, third, sixth and seventh test run ($P = 0.445$), and in the group in which the vigilance task was performed in the first, fourth, fifth and eighth test run ($P = 0.392$).

*Driving simulation*

Number of errors in the driving simulation task was not significantly affected by the TIME factor ($F_{7,47} = 1.93$, $P = 0.0853$). The respective means and standard errors for the eight test runs were: test run 1: 39.4 ± 5.2, test run 2: 34.9 ± 4.0, test run 3: 32.1 ± 5.5, test run 4: 33.2 ± 5.1, test run 5: 42.5 ± 6.7, test run 6: 28.7 ± 4.5, test run 7: 27.3 ± 3.7, test run 8: 32.3 ± 6.7. However, on the average more errors were made in the condition without acoustic wayfinding information (mean: 41.1 ± 3.4) than in the group with acoustic wayfinding information (mean: 26.8 ± 3.6; main effect of ACOUSTIC WAYFINDING INFORMATION $F_{1,47} = 15.54$, $P = 0.0003$). The interaction between TIME and ACOUSTIC WAYFINDING INFORMATION was not significant ($F_{7,47} = 1.39$, $P = 0.2326$).

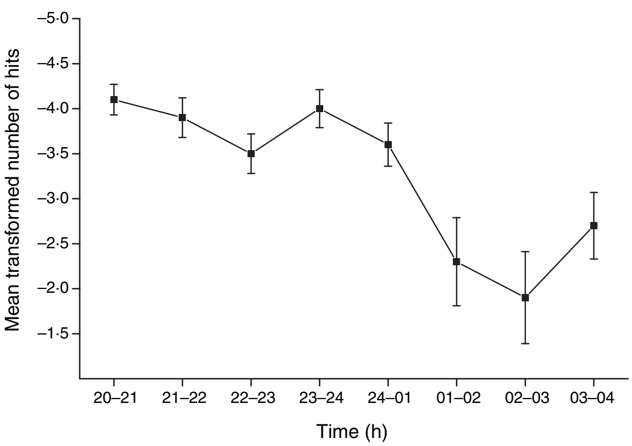


**Figure 4.** Mean values and standard errors of transformed number of hits in the Mackworth clock vigilance task in the eight test runs. The transformation formula was [ln(101 – number of hits)/number of hits)]. Transformed number of hits significantly decreased across time ($P = 0.0001$).

### Correlations between fatigue and performance

Cross-correlations between PRE and POST sleepiness and fatigue scores on the one hand and errors in the driving simulation task on the other were not significant ($r = -0.255$ to $-0.410$; $P > 0.05$). Cross-correlations between fatigue scores and false reactions were larger ($r = 0.473$–$0.637$), but also not significant ($P > 0.05$). Finally, there were high cross-correlations between fatigue scores and transformed number of hits ($r = 0.851$–$0.911$; $P < 0.05$).

## DISCUSSION

The main result of the present study was that the vigilance task induced larger time-on-task effects on sleepiness and fatigue than the driving simulation task comparing changes over the period of task performance. These effects were not obscured by cross-over effects from the task performed before. The analysis of break effects further showed that fatigue and sleepiness which had developed during the test run reduced in the subsequent break. This reduction was the larger, the larger the preceding increase during the test run. These results support the assumption that there are changes in fatigue and sleepiness that can be ascribed to task performance. These time-on-task effects are task-dependent and can be quantified by means of a design including separate test runs divided by breaks.

Sleepiness and fatigue revealed mostly similar results, which is in accordance with the study by Williamson *et al.* (2001). One possible explanation is that subjects were not able to differentiate between both concepts. This seems less likely, however, as in the introductory phase of the experiment participants were told about the difference between sleepiness and fatigue, and subjects had to deal with the two concepts already in the week before the main experiment when working on their sleep diaries.

Moreover, sleepiness and fatigue seemed to produce different results under specific circumstances. We found a significant three-way interaction between TIME, TASK, and PREPOST for fatigue showing that larger time-on-task effects for the vigilance than for the driving simulation task were found in the second, seventh and eighth test run. In contrast, there was only a significant TASK by PREPOST interaction effect on sleepiness. The latter revealed that – independent of time – the vigilance tasks generally induced larger fatigue across test runs than the driving simulation task. End-of-experiment effects which were stronger for fatigue than for sleepiness may be a possible explanation. These effects are a reflection of the fact that subjects behave differently in the last test runs compared with earlier test runs just because they know the experiment will end soon.

In the more monotonous vigilance task, performance decreased more strongly in the four test runs of each group than in the driving simulation task. Performance degradation was especially pronounced in the third and fourth test run. This is in agreement with the study by Williamson *et al.* (2001) in which performance effects were more pronounced especially in later test runs in the Mackworth clock vigilance task in comparison with less monotonous visual search or logical reasoning tasks.

In the driving simulation task, no time-dependent performance changes were found. However, several points have to be considered when interpreting this result. First, there was a slight trend for errors in the driving simulation to decrease across test runs and type-2 errors cannot be excluded given a small *P*-value of 0.08. Secondly, it is unclear if error-enhancing effects of fatigue and error-reducing effects of learning canceled each other out. We could have used different routes in the four experimental test runs. However, these should have been comparable with respect to the number of turns etc., which was not possible because of software restrictions. Moreover, the complexity of the scenery, e.g. the frequency of trees and houses, could not have been controlled. We tried to control learning effects by means of an extended, standardized training session before the actual experiment, however, this might not suffice. Thus, learning effects should be more adequately controlled in driving simulation tasks in future studies.

One might further criticize a circular definition of monotony. That is, a task is classified as being more monotonous, because performance decrements over time are larger and subjective fatigue scores are higher. In the present study, however, the classification of a task as being more or less monotonous was set beforehand. In the vigilance task, the stimulus material consisted of a circle of dots only. Moreover, the only reaction required was to press a button if a moving dot jumped across one of the dots of the circular path. The driving simulation task of the present study was more varied because subjects had to drive through Berlin and not on a motorway. That is both the scenery and the motor reactions required were far more diversified. Therefore, it appears we were able to validate our assumption of differences in monotony in the two tasks applied by means of subjective fatigue and sleepiness ratings.

In the driving task, there were fewer errors in the condition with acoustic wayfinding information than in the group where subjects were distracted by reading route information. This result has important implications for navigation systems and is in agreement with previous studies showing that distraction caused by visual displays increase driving errors (Liu, 2000). In Liu's study, subjects were instructed to perform a push button and a navigation task while using displayed traffic information to drive through a scenario. The information was presented visually only, aurally only, or by multimodal display. All drivers made significantly fewer errors in responding to hazard warnings and showed a better control of car direction and speed when using multimodal and auditory displays compared with visual-only displays.

Interestingly, and in contradiction to our expectation, acoustic wayfinding information had no significant influence on subjective fatigue and sleepiness ratings. This may be due to the fact that the driving simulation task was alerting in itself, so that the acoustic wayfinding information had no additional effect. Subjective fatigue and sleepiness might be diminished by acoustic information when driving a less diversified or demanding route.

In summary, we found task-dependent time-on-task effects on subjective sleepiness and fatigue. That is, the more monotonous vigilance task led to higher fatigue and sleepiness scores than the more interesting driving simulation task. These task-dependent differences in subjective scores argue in favor of the validity of the scales used here. Moreover, it appears that time-on-task effects should be considered in mathematical models predicting fatigue, which are currently based on circadian and sleep-related factors only. In addition, performance seems to degrade differently in tasks of different monotony paralleling the effects on fatigue scores. However, these results have to be taken with caution and require further investigation.

## ACKNOWLEDGEMENT

We like to thank the participating subjects for their time and effort.

## REFERENCES


Akaike, H. A new look at the statistical model identification. IEEE T. *Automat. Contr.*, 1974, AC-19: 716–723.

Akerstedt, T. Subjective and objective sleepiness in the active individual. *Int. J. Neurosci.*, 1990, 52: 29–37.

Akerstedt, T. and Gillberg, M. Experimentally displaced sleep: effects on sleepiness. *Electroen. Clin. Neuro.*, 1982, 54: 220–226.

Akerstedt, T. and Landström, U. Work place countermeasures of night shift fatigue. *Int. J. Ind. Ergonom.*, 1998, 21: 167–178.

Akerstedt, T., Folkard, S. and Portin, C. Predictions from the three process model of alertness. *Aviat. Space Environ. Med.*, 2004, 75: A75–A83.

Bonnet, M. H. Sleep deprivation. In: M. H. Kryger, T. Roth, W. C. Dement (Eds) *Principles and Practice of Sleep Medicine*, 2nd edn. W.B. Saunders, Philadelphia, PA, 1994: 50–67.

Borbély, A. A. A two-process model of sleep regulation. *Hum. Neurobiol.*, 1982, 1: 195–204.

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R. and Kupfer, D. J. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res.*, 1989, 28: 193–213.

Eriksson, M. and Papanikolopoulos, N. P. Driver fatigue: a vision based approach to automatic diagnosis. *Transport. Res. C-Emer 9*, 2001, 9, 399–413.

Graw, P., Kräuchi, K., Knoblauch, V., Wirz-Justice, A. and Cajochen, C. Circadian and wake-dependent modulation of fastest and slowest reaction times during the psychomotor vigilance test. *Physiol. Behav.*, 2004, 80: 695–701.

Johns, M. W. Rethinking the assessment of sleepiness. *Sleep Med. Rev.*, 1998, 2: 3–15.

Johns, M. W. A sleep physiologist's view of the drowsy driver. *Transport. Res. F-Traf*, 2000, 3: 241–249.

Knipling, R. R. and Wang, J. S. *Crashes and Fatalities Related to Driver Drowsiness/fatigue (NHTSA Research Notes, November)*. National Highway Traffic Safety Administration, US Department of Transport, Washington DC, 1994.

Kraemer, S., Danker-Hopfe, H., Dorn, H., Schmidt, A., Ehlert, I. and Herrmannn, W. M. Time-of-day variations of indicators of attention: performance, physiologic parameters, and self-assessment of sleepiness. *Biol. Psychol*., 2000, 48: 1069–1080.

Lan, P., Ji, Q. and Looney, C. G. Information fusion with Bayesian networks for monitoring human fatigue. *ISIF*, 2002, 1: 535–542.

Liu, Y. C. Effect of advanced traveler information system displays on younger and older drivers' performance. *Displays*, 2000, 21: 161–168.

Mackworth, N. H. *Researches on the Measurement of Human Performance*. Medical Research Council, Report Series, 268, 1950.

Maycock, G. Sleepiness and driving: the experience of U.K. drivers. *Accident. Anal. Prev*., 1997, 29: 453–462.

Moore-Ede, M., Heitmann, A., Guttkuhn, R., Trutschel, U., Aguirre, A. and Croke, D. Circadian alertness simulator for fatigue risk assessment in transportation: application to reduce frequency and severity of truck accidents. *Aviat. Space Environ. Med*., 2004, 75: A107–A118.

Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 1971, 9: 97–113.

Parkes, A. M. and Coleman, N. Route guidance systems: a comparison of methods of presenting directional information to the driver. In: E. J. Lovesey (Ed.) *Contemporary Ergonomics*. Taylor & Francis, London, 1990: 480–485.

Pilcher, J. J. and Huffcutt, A. I. Effects of sleep deprivation on performance: a meta-analysis. *Sleep*, 1996, 19: 318–326.

Porcu, S., Bellatrecchia, A., Ferrara, M. and Casagrande, M. Sleepiness, alertness and performance during a laboratory simulation of an acute shift of the wake-sleep cycle. *Ergonomics*, 1998, 41: 1192–1202.

Samn, S. W. and Perelli, L. P. *Estimating Aircrew Fatigue: A Technique with Implications to Airlift Operations* (Technical Report No SAM-TR-82–21). USAF School of Aerospace Medicine, Brooks AFB, TX, 1982.

Spencer, M. B. and Gundel, A. *A PC-based Program for the Assessment of Duty Schedules in Civil Aviation: The Way Forward* (Report DERA/CHS/PP5/CR/980069/1.0). DERA, Farnborough, UK, 1998.

Thiffault, P. and Bergeron, J. Monotony of road environment and driver fatigue: a simulator study. *Accident Anal. Prev*., 2003, 35: 381–391.

Thomas, C. and Attard, S. *Drowsiness and Fatal Accidents*. Centre Européan d'Etudes Socio-economiques et Accidentologiques des Risques, Nanterre, France, 1994.

ten Thoren, C. and Gundel, A. Müdigkeit als Unfallursache im Stadtbereich - eine Befragung von Unfallbeteiligten. *Somnologie*, 2003, 7: 125–133.

Verwey, W. B. and Zaidel, D. M. Predicting drowsiness accidents from personal attributes, eye blinks and ongoing driving behaviour. *Pers. Individ. Dif*., 2000, 28: 123–142.

Wierwille, W. W., Ellsworth, L. A., Wreggit, S. S., Fairbanks, R. J. and Kirn, C. L. *Research on vehicle-based driver status/performance monitoring: development, validation, and refinement of algorithms for detection of driver drowsiness*. National Highway Traffic Safety Administration Final Report: DOT HS 808 247, 1994.

Williamson, A. M., Feyer, A.-M., Mattick, R. P., Friswell, R. and Finlay-Brown, S. Developing measures of fatigue using an alcohol coparison to validate the effects of fatigue on performance. *Accid. Anal. Prev*., 2001, 33: 313–326.

Copyright of Journal of Sleep Research is the property of Blackwell Publishing Limited. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.