Ergonomics

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/terg20

Investigation of task performance variations according to task requirements and alertness across the 24-h day in shift workers

E. Galy [a] , C. Mélan [a] & M. Cariou [a]

[a] Université Toulouse, 5, allée Antonio Machado, 31058, Toulouse cedex 9, France

Version of record first published: 18 Sep 2008.

To cite this article: E. Galy , C. Mélan & M. Cariou (2008): Investigation of task performance variations according to task requirements and alertness across the 24-h day in shift workers, Ergonomics, 51:9, 1338-1351

To link to this article: http://dx.doi.org/10.1080/00140130802120218

Ergonomics
Vol. 51, No. 9, September 2008, 1338–1351

Investigation of task performance variations according to task requirements and alertness across the 24-h day in shift workers

E. Galy, C. Mélan* and M. Cariou

Université Toulouse, 5, allée Antonio Machado, 31058 Toulouse cedex 9, France

The study investigated time-of-day effects on task performance in shift workers in different tasks (reaction time, discrimination, probe recognition, free recall), by varying task-specific features. On each of six recordings, each programmed on a different day and in a randomised order, operators rated alertness and performed different tasks. Self-rated alertness varied according to a typical diurnal trend. Time of day also affected reaction time (slower responses at 03:00 hours), discrimination performance (lower accuracy at 03:00 hours in the most difficult condition) and recall (superior recall at 07:00 and 11:00 hours following deeper processing at encoding). The data demonstrated time-of-day effects on cognitive processes also involved in many real-job activities, despite the lack of control for a number of exogenous factors known to interfere with performance in work settings. Since in the cognitively more loaded tasks, time-of-day effects depended on task conditions, the findings are of operational concern in shift-work situations involving differential task requirements. In a real-job setting, performance variations were observed according to time of day and task requirements in a set of cognitively more or less demanding tasks. Task-specific research across the 24-h day enables a better understanding of operators' tasks and the development of supporting technology.

Keywords: shift-workers; alertness; reaction time; non-verbal discrimination; memory tasks

1. Introduction

Over the last few decades, a number of studies showed how performance measures are affected by sleep drive (homeostatic influence) and circadian phase. The demonstration of an interaction of circadian phase and time awake in the regulation of performance came from laboratory studies under controlled conditions and/or manipulation of the sleep/wake cycle (for a review, see Akerstedt 2007). Several theoretical models conceptualised this interaction by proposing an optimal position and duration for sleep, with any deviation impairing performance and alertness (Folkard and Akerstedt 1992, Achermann and Borbély 1994, Van Dongen *et al.* 2003). Systematic performance measures in relation to different simulated shift systems further confirmed a sharp performance drop in the early morning hours (Colquhoun *et al.* 1969). This was observed in most tasks that require sustained attention, such as reaction time performance or throughput in cognitive tasks (Colquhoun *et al.* 1968, Fröberg *et al.* 1975, Johnson *et al.* 1992). At the same time, physiological and subjective arousal was reported to strongly decrease (Akerstedt and

*Corresponding author. Email: cmelan@univ-tlse2.fr

ISSN 0014-0139 print/ISSN 1366-5847 online

DOI: 10.1080/00140130802120218
http://www.informaworld.com

Gillberg 1990). A simulation study revealed that at high fatigue levels errors involving a failure to act (errors of omission) were increased, whereas incorrect responses (errors of commission) were decreased in a simulated train-driving task (Dorrian *et al*. 2007). The differential effect of fatigue on error type was explained by a cognitive disengagement with the virtual train, which increased accident risk.

In real shift-work settings, the findings of higher occurrences of risk and accidents on night- compared to day-duty, further stressed the potential consequences of reduced arousal on safety and security (Folkard and Tucker 2003). Shift work would further aggravate the nocturnal decline in cognitive abilities observed under controlled conditions due to a chronic sleep debt that shift workers carry over to the night shift and also to day shifts following early morning shift start (Akerstedt 1991, Dinges *et al*. 1997, Pilcher *et al*. 2000, Fletcher and Dawson 2001). Although a nocturnal work performance drop has been reported in several studies (for a review, see Folkard and Tucker 2003), reliable performance measures may not be available in all occupational settings. In addition, masking effects of work performance changes across the 24-h day have been suggested for work activity itself (Andorre and Queinnec 1998) and for incentives (Blake 1971).

Several field studies collected, instead, interpolated performance measures by using simple and short neuropsychological tests. They reported reduced performance on the night shift for reaction time, mental arithmetic and search performance (Wojtczak-Jaroszowa *et al*. 1978, Tilley *et al*. 1982, Hossain *et al*. 2004). However, performance changes have not systematically been reported in field studies, for instance, for serial reaction time, digit symbol substitution and manual tracking (Costa *et al*. 1979, Lowden *et al*. 1998). Further, in memory tasks 24-h performance variations have been shown to depend on task requirements. Thus, process controllers' recall performance was higher at 18:00 hours compared to 02:00 hours for words associated with semantic encoding, while the contrary was observed for words processed according to physical attributes (Maury and Queinnec 1992). Another study revealed that air traffic controllers' recall performance depended on the modality of presentation and the recall procedure used (Galy *et al*. 2004, Mélan *et al*. 2007). In a free recall task they remembered systematically fewer words during the night and early morning hours. In a probe recognition task, this effect was only observed when they had to recall visual items but not auditory items.

Taken together, these findings indicate that performance trends depend both on the task used and on the task components under investigation. This interpretation is in agreement with the results obtained by more systematic investigations of performance variations in controlled laboratory studies. In perceptual-motor tasks, such as simple reaction time tasks restricted to immediate throughput of information, performance was usually shown to peak in the evening (Monk *et al*. 1985, Casagrande *et al*. 1997, Monk and Carrier 1997, Heuer *et al*. 1998, Owens *et al*. 2000, de Gennaro *et al*. 2001, Wright *et al*. 2002). A reduction in response speed during the night, coincident with an increase in response accuracy, suggested a strategy change rather than a global nocturnal performance drop (Monk and Carrier 1997). When the emphasis was on short-term memory processing, superior performance was reported in the morning, while for long-term memory tasks performance peaks were observed in the late afternoon (Baddeley *et al*. 1970, Folkard *et al*. 1976, Folkard and Monk 1980, Oakhill and Davies 1989).

These findings have been interpreted as indicating that low arousal in the morning would favour more automatic processes, such as maintenance rehearsal in short-term memory. Progressive increment of arousal as the day progresses would tend to favour more demanding processing, involving, in particular, semantic attributes from long-term memory sources (Folkard 1979, Folkard and Monk 1980). This interpretation may also account for the differences of recall performance across the 24-h day according to the recall procedure, as reported by a previously mentioned field study (Mélan *et al*. 2007), but also in a controlled laboratory study (Lorenzetti and Natale 1996). The latter study indeed revealed superior recall in the afternoon when subjects had been instructed to use a free recall procedure, and in the morning when they had been informed to use a recognition procedure. Greater degrees of elaboration would be necessary for free recall (relying on subjects' ability to actively retrieve memory traces) compared to recognition (requiring subjects to simply discriminate memory traces; Brébion *et al*. 2005, Prince *et al*. 2005). A similar hypothesis has been proposed to account for recall performance differences over the 24-h day according to item modality. Auditory presented items would benefit from a longer-lasting sensory trace and/or automatic phonological encoding compared to visually presented items (Baddeley 1986, Penney 1989, Cowan *et al*. 2002). This would explain that only recall for visually presented items was reduced when alertness was low (Mélan *et al*. 2007).

The present study tested these hypotheses in a real shift-work setting by investigating operators' task performance across the 24-h day in a set of experimental tasks. In a repeated measures design, operators (in charge of in-door security of a nuclear power plant) provided alertness self-ratings every 4 h and performed tasks involving cognitive functions ranging from psychomotor reaction time to elaborate processing of meaningful words. While a nocturnal performance decrement was expected for all tasks, diurnal performance peaks were expected to vary between tasks. Performance would peak in the morning for visual discrimination and probe recognition, involving more automatic processing, and later during the day for free recall, requiring more elaborate processing. In each task, the overall trend should be modulated by an interaction with the task component that was manipulated, i.e. difficulty in the discrimination task (parallel bars shifted by 0, 1, 2 or 3 pixels), item modality in the probe recognition task (visual and auditory list and probe presentation) and item attributes in the free recall task (typographic, structural and semantic encoding conditions). In agreement with the variations observed in laboratory studies, it was expected that later performance peaks or larger performance variations in the more demanding task conditions would be found.

2. Method

2.1. Subjects

The study was carried out with 23 volunteer shift workers supervising, on various control monitors, the in-door security of a nuclear plant in the south-west of France. They came from six teams of operators (total of 36) and were all men, aged 41.5 years (range 35–53 years), working in shift-work systems for 11.8 years (range 4–20 years). They worked a continuous rapidly forward rotating shift system, with three mornings, two afternoons, two nights followed by 5 d off. The 8-h shifts started at 06:00, 14:00 and 22:00 hours respectively.

2.2. Material and procedure

Satellite controllers completed Thayer's check list and performed four different neuropsychological tasks on each of six testing times that were distributed over the 24-h day. Testing took place in a quiet room, during duty pauses introduced for the sake of the study, by avoiding meal times and shift changeover times, i.e. at 03:00, 07:00, 11:00, 15:00, 19:00 and 23:00 hours. The subject first completed the check list, then performed a discrimination task, followed by two memory tasks (starting randomly with either) and a reaction-time task. To ensure no confounding effects of time of day (TOD) and practice, testing times were randomised across participants, with a single testing time per day for a given subject. Prior to the beginning of the study, operators completed Horne and Östberg's (1976) questionnaire. The results indicated that no participant displayed extreme morningness or eveningness.

2.2.1. *Alertness*

Participants completed a French paper-and-pencil version of Thayer's activation–deactivation adjective checklist (Thayer 1978). They quoted, for each of 20 adjectives, one of the following responses: 'not at all'; 'don't' know'; 'little'; 'much'; having an assigned weight respectively of 1, 2, 3 and 4. Two adjectives, however, have a negative weight assigned to them. The scores were grouped into four factors, general activation (GA), deactivation sleep (DS), high activation (HA) and general deactivation (GD). The ratio GA/DS provided an index of subjects' alertness.

2.2.2. *Visual discrimination*

Each of 24 trials started with the display of two horizontal parallel bars, 3 cm long and 0.5 cm apart. Depending on the trial, the bars were either shifted horizontally to each other by 1, 2 or 3 pixels (in each case, 1/4 of trials), or they were not shifted (1/4 of the trial). Items were displayed in the centre of a computer screen while the following instruction appeared on the bottom of the screen: 'The lines are shifted (press key M)/The lines are not shifted (press key Q)'. The participant's response initiated a 2-s inter-trial interval, before the next trial started. Software, specifically designed for this study, presented the four discrimination conditions randomly and recorded the subject's response accuracy and latency.

2.2.3. *Probe recognition*

On each of 32 trials, six common French nouns were presented, one by one (750 ms), at a rate of one word per second. A probe item was presented 4 s after presentation of the last list item, together with the following response instruction: 'The word figured in the list (press M key)/The word did not figure in the list (press Q key)'. The subject's response initiated a 5-s intertrial interval, followed by a brief sound signalling the start of the next trial. The list items were randomly selected from a 24-item set and on half the trials the probe had occurred in the list. Item lists and probes could be presented either in the visual modality (centre of a computer screen) or in the auditory modality (loud speakers). The four trial types resulting from presentation modality of lists and probes (auditory–visual, auditory–auditory, visual–visual, visual–auditory) were randomised across trials by the aforementioned tailor-made software. To rule out practice effects across the six testing times, six different item sets were used. All items of a set belonged to the same word category (fruit, vegetables, music instruments, furniture and animals).

2.2.4. *Free recall*

This protocol was previously described by Maury and Quéinnec (1992). Briefly, on each of 18 trials, a common French noun was presented in the middle of a computer screen

(300 ms), immediately preceded by a question that induced processing of the word's typographical attributes ('Is the word written in capitals', 1/3 trials), structural attributes ('Do the vowels of the word occur in alphabetical order', 1/3 trials) or semantic aspects (1/3 trials). In the latter case, the question varied according to the item presented, such as: 'Is it used to open a door', followed by the item 'handle'. Simultaneously with each word, the following instruction occurred at the bottom of the screen: 'Response 'Yes', press key M; Response 'No', press key Q'. The response initiated an interpolated reaction time task (see below) before the next trial started. Following the last trial, subjects were asked to write on a white sheet all the words they could remember from the 18-item list.

2.2.5. Reaction time

This task was interpolated after each trial in the free recall task, in order to prevent maintenance rehearsal. The response after each question/word sequence in the free recall task automatically generated a 1-pixel spot that moved randomly on the screen for several seconds and finally 'exploded'. Following this event, the subject had to press the 'space bar' as rapidly as possible, which initiated a second, and then a third trial. Thereafter, the program shifted back to the recall task, with the presentation of a question starting the second trial, and so on.

2.3. Statistics

All measures recorded were analysed by parametric tests. ANOVA with repeated measures across TOD were used to analyse subjects' self-reported alertness (one-way ANOVA), and factorial ANOVA investigated TOD effects together with task-specific components in the cognitive tasks. Tukey's test was used for post hoc analyses. The study further explored whether task performance variations observed across TOD were associated with variations in subjects' alertness. This hypothesis was tested by using Pearson's correlation test as follows. Correlations were computed within participants ($n = 6$ for each person separately), by using alertness/test scores at the six TOD values. Then, the 23 r values were tested against the null hypothesis by using a one-sample t test. Although this procedure violates the assumption of independence of scores, it provides an indication whether, in general, alertness was correlated with test performance.

3. Results

Operators' alertness curve obtained by concurrent ratings on six recordings across the 24-h day is represented by Figure 1. A one-way ANOVA with repeated measures revealed a significant TOD effect ($F(5,5) = 3.73$, $p < 0.01$), with significant lower alertness at 03:00 hours than at 15:00 and 19:00 hours.

The effects of TOD on operators' performance in the cognitive tasks were explored together with task-specific factors by using factorial ANOVA. In the visual discrimination task, a two-way ANOVA of TOD (repeated measures) and discrimination condition revealed a significant effect of discrimination condition on response accuracy ($F(3,23) = 199.13$, $p < 0.001$; Figure 2) and an interaction between response accuracy and TOD ($F(5,23) = 1{,}5639$, $p < 0.05$). Post-hoc analyses indicated that the latter effect resulted from less accurate responding at 03:00 compared to 11:00 hours in the 1 pixel shift condition ($p < 0.01$). Furthermore, subjects responded overall more accurately in the most obvious discrimination conditions (3-pixel shift, no-shift) compared to the 1- and

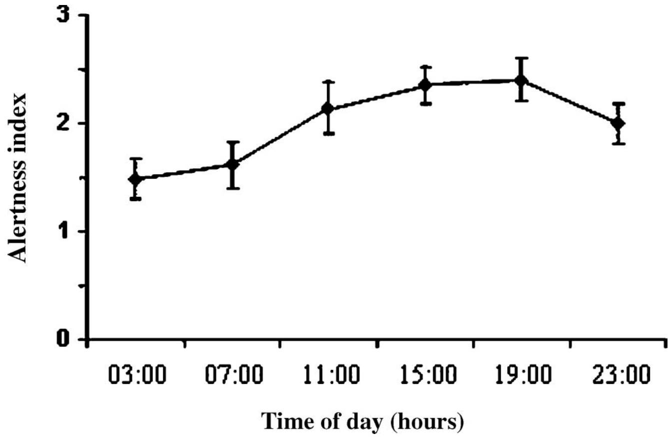

Figure 1. Mean (±SE) self-ratings of alertness on a check list provided every 4 h across the 24-h day.

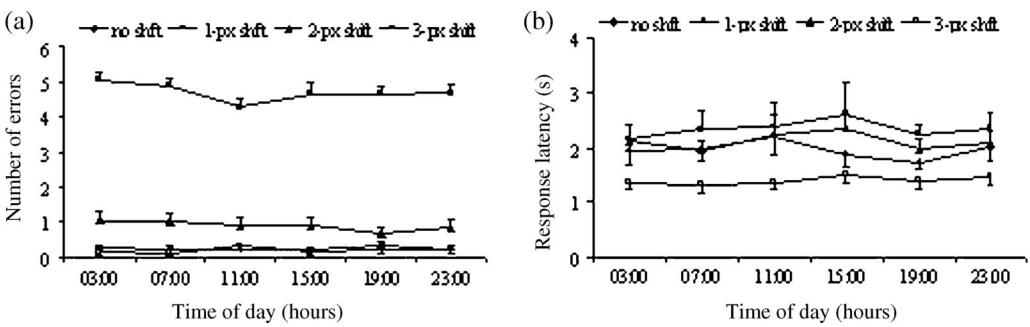

Figure 2. a) Mean (±SE) number of errors and b) mean (±SE) response latency in four visual discrimination conditions, at six times across the 24-h day.

2-pixel shift conditions (in each case, $p < 0.01$) and in the 2-pixel compared to the 1-pixel discrimination condition ($p < 0.001$). Discrimination condition also affected subjects' response latency ($F(3,23) = 4.783$, $p < 0.01$), with more rapid responding in the 3-pixel shift condition compared to the remaining conditions (in each case, $p < 0.01$).

Figure 3 illustrates the mean number of correct responses in a verbal probe recognition task according to TOD and presentation modality of the verbal items. A two-way ANOVA with repeated measures revealed that task performance was affected by presentation modality of the item lists ($F(1,23) = 22.11$, $p < 0.001$) and of the probes ($F(1,23) = 9.66$, $p < 0.01$), indicating better performance following auditory rather than visual item presentation, both at encoding and recall. Further, TOD interacted with presentation modality ($F(5,23) = 2.166$, $p < 0.05$), due to significant higher recall at 03:00 hours following auditory compared to visually encoded words ($p < 0.001$).

In a free recall task, the mean number of words recalled according to TOD was analysed together with the type of question that was presented prior to each word at encoding (Figure 4) and the position in the list (Table 1). The figure indicates more marked TOD effects following questions inducing processing of semantic or typographic aspects rather than of structural aspects. Furthermore, whether items were issued from the initial (positions 1 to 6), intermediary (positions 7–12) or end (positions 3–18) portion of the list, recall performance was overall higher on daytime compared to night-time recordings. A three-way ANOVA of these factors revealed significant effects for encoding question type ($F(2,23) = 22.30, p < 0.001$) and list portion ($F(2,23) = 3.9, p < 0.05$) and an interaction

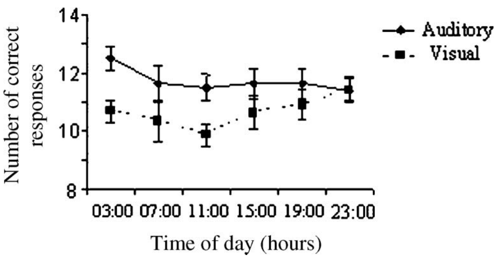

Figure 3. Mean (±SE) number of correct responses in a probe recognition task according to presentation modality of item lists, at six times across the 24-h day.

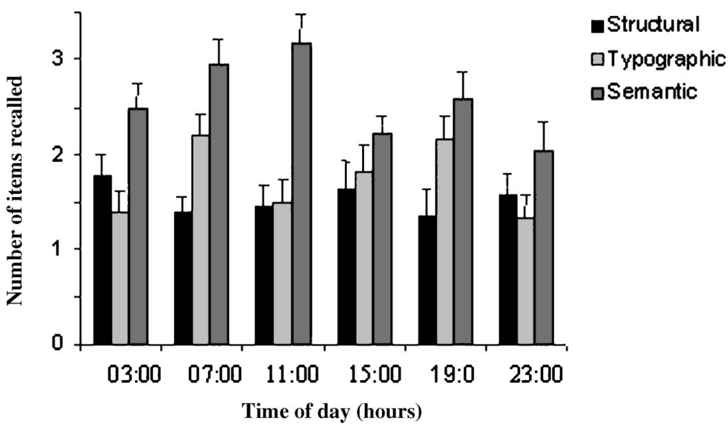

Figure 4. Mean (±SE) number of words recalled according to the type of inducing question preceding each word, at each of six times across the 24-h day.

Table 1. Mean number of words recalled (±SE) according to their position in the list.

	Time of day (hours)					
List portion	03:00	07:00	11:00	15:00	19:00	23:00
Initial	1.83 (0.30)	2.57 (0.27)	2.45 (0.26)	2.00 (0.27)	1.75 (0.29)	2.14 (0.27)
Intermediary	1.91 (0.31)	1.39 (0.26)	1.77 (0.27)	1.77 (0.27)	1.95 (0.27)	1.33 (0.24)
Terminal	1.87 (0.33)	1.61 (0.23)	1.91 (0.35)	1.95 (0.26)	2.20 (0.32)	1.62 (0.28)

Note: initial list portion: positions 1 to 6; intermediary list portion: positions 7 to 12; terminal list portion: positions 13 to 18.

between each of these factors and TOD (TOD × question type $F(10,23) = 1.99$, $p < 0.05$; TOD × list portion $F(10,23) = 2.01$, $p < 0.05$); ($F(4,23) = 5.97$, $p < 0.001$). Post-hoc analyses revealed superior recall for words preceded by a semantic question rather than by a structural question at 07:00 and 11:00 hours (respectively, $p < 0.05$, and $p < 0.01$) or by a typographic question at 11:00 hours ($p < 0.01$). In addition, items issued from the initial list portion were better recalled at 07:00 than at 19:00 hours ($p < 0.05$). Moreover, a significant interaction between question type and list portion indicated significant higher recall performance when words issued from the initial list portion were preceded by a semantic question rather than by a structural or typographical question ($p < 0.001$ in each case). A similar effect was observed when items issued from the end portion of the list were preceded by a semantic rather than by a typographical question ($p < 0.01$).

A similar analysis was performed for operators' performance in a detection task that was interpolated following each question/word presentation in the free recall task. TOD effects on subjects' mean response latencies on each of three successive trials (Table 2) were investigated by including two additional factors (Table 3), i.e. the question type preceding the detection task (structural, typographic and semantic) and the list portion (initial, intermediary and end portions). A four-way ANOVA of these factors revealed a significant effect of time of recording ($F(5,161) = 2.94$, $p < 0.05$), indicating slower responses at 03:00 compared to 07:00 hours ($p < 0.05$) and 19:00 hours ($p < 0.01$). Furthermore, highly significant differences between trials ($F(2,161) = 40.86$, $p < 0.001$) and list portions ($F(2,161) = 10.24$, $p < 0.001$) resulted from slower responses on the first compared to the second and third trials ($p < 0.001$, in each case), and on trials inserted in the intermediary ($p < 0.001$) and end portions ($p < 0.001$) of the list compared to the initial list portion. No interaction was found between task factors and TOD.

Table 2. Mean response latencies (±SE) in a visual detection task according to time of day and trial.

	Time of day (hours)					
	03:00	07:00	11:00	15:00	19:00	23:00
Trial 1	0.28 (0.03)	0.29 (0.03)	0.42 (0.07)	0.35 (0.06)	0.29 (0.03)	0.30 (0.04)
Trial 2	0.28 (0.04)	0.31 (0.05)	0.46 (0.09)	0.47 (0.14)	0.41 (0.08)	0.23 (0.07)
Trial 3	0.53 (0.17)	0.40 (0.08)	0.31 (0.06)	0.35 (0.08)	0.46 (0.08)	0.40 (0.09)

Table 3. Mean response latencies (±SE) across trials in the detection task according to time of day, type of inducing question (upper panel) and list portion (lower panel).

		Time of day (hours)					
		03:00	07:00	11:00	15:00	19:00	23:00
Type of inducing question	Structural	0.41 (0.14)	0.34 (0.10)	0.41 (0.10)	0.37 (0.04)	0.35 (0.05)	0.36 (0.03)
	Typographic	0.37 (0.09)	0.36 (0.08)	0.36 (0.06)	0.45 (0.10)	0.36 (0.04)	0.38 (0.05)
	Semantic	0.40 (0.12)	0.36 (0.05)	0.40 (0.08)	0.39 (0.05)	0.37 (0.06)	0.36 (0.04)
List portion	Beginning	0.24 (0.02)	0.35 (0.06)	0.38 (0.07)	0.43 (0.09)	0.42 (0.07)	0.36 (0.07)
	Intermediary	0.55 (0.22)	0.39 (0.09)	0.40 (0.06)	0.47 (0.13)	0.37 (0.06)	0.32 (0.06)
	End	0.38 (0.07)	0.35 (0.05)	0.40 (0.09)	0.33 (0.06)	0.41 (0.08)	0.41 (0.07)

To investigate whether trends of task performance and self-reported alertness trends were associated, correlations across TOD were calculated between alertness scores and either performance measure for each participant. Table 4 summarises ranges of individual correlation values for each performance measure and the corresponding mean correlation. Testing the 23 r values against the null hypothesis revealed overall significant correlations across TOD between alertness on one hand and discrimination latency, visual recognition performance and free recall performance on the other hand (Table 4).

4. Discussion

In real-job conditions, operators displayed reliable alertness variations and task performance variations across the 24-h day. A nocturnal drop was found for task performance and self-reported alertness, despite a lack of control for a number of factors that have been shown to affect psychological measures in controlled laboratory conditions, including light exposure, noise and motivation (Smith *et al*. 2002, Wright *et al*. 2002). In shift workers, the nocturnal decline of psychological measures is generally aggravated by chronic sleep deprivation (Akerstedt 1991, Dinges *et al*. 1997, Fletcher and Dawson 2001, Pilcher *et al*. 2000). The disruptive effects of shift work have been shown to depend on various shift-scheduling features. Somewhat smaller effects have been shown for rapid compared to slow rotations, for systems including a limited number of successive night shifts and for late vs. early morning-shift start (Knauth 1996, Tucker *et al*. 1998, Della Rocco and Nesthus 2005). As in the present job situation, operators worked according to a rapid forward rotating shift-scheduling system that included a maximum of two or three successive night shifts and late night-to-morning shift changeover, the disruptive effects on the nocturnal decline of operators' psychological measures may have been less pronounced than may have been expected in other shift-scheduling systems.

Operators' self-rated alertness displayed a typical day-oriented trend that peaked at 19:00 and reached its minimum at 03:00 hours. The finding of a trend similar to those reported in other real work settings (Akerstedt 1991, Dinges *et al*. 1997, Tucker *et al*. 1998) supports the hypothesis of a strong dependency of alertness on the endogenous systems underlying circadian rhythms (Folkard and Akerstedt 1992, Achermann and Borbély 1994). The marked differences between day- and night-time reported in the present study contrasted, however, with the restricted range of alertness ratings reported in another shift-work study (Tucker *et al*. 1998). This difference may be related to methodological

Table 4. Ranges and mean ($\pm$SD) values for Pearson's correlation test computed across time of day between alertness scores and each performance measure in the different tasks.

	Performance measure	Range of r values	Mean r values	t(22)
Discrimination	Latency	− 0.91 to 0.75	− 0.24 (0.11)	− 2.21*
	Errors	− 0.68 to 0.44	− 0.11 (0.09	− 1.18
Probe Recognition	Auditory	− 0.77 to 0.72	0.13 (0.11)	1.25
	Visual	− 0.58 to 0.86	0.25 (0.09)	2.68*
Free Recall	Physical	− 0.64 to 0.83	0.22 (0.10)	2.23*
	Structural	− 0.72 to 0.96	21 (0.09)	2.38*
	Semantic	− 0.42 to 0.86	0.19 (0.08)	2.36*

*Indicates that mean r values differed from the null hypothesis, with $p < 0.05$.

differences, given that in the present study operators rated alertness concurrently on 4-hourly intervals, rather than retrospectively on a single session. In addition, the alertness score was computed from responses to a list of adjectives rather than by quoting a single-item scale and could thus provide a more precise description of operators' functional state. Although other factors may explain the finding of more or less shaded alertness curves in shift-work studies, the idea that the rating method used contributes to this difference may deserve further investigation. On the other hand, the possibility cannot be excluded that in real-job conditions, operators rate alertness simply according to what they should feel like while being at work or what they expect to feel like given the TOD.

This latter interpretation may, however, not account for the task performance variations observed over the 24-h day in the present study. The role of practice effects may also be excluded, given that performance trends were obtained by randomising the order of the six recording sessions across participants and by performing a single session on a given day. Similarly, in the memory tasks, learning effects were, in principle, ruled out by randomising different task versions across participants and recordings. In light of these considerations, task performance variations according to TOD most probably resulted from operators' functional state at the time of testing. Thus, with increasing alertness during daytime, operators displayed significant superior performance in several cognitively more or less demanding tasks. More specifically, they responded more rapidly in a reaction time task, performed more accurate responses in a non-verbal visual discrimination task and recalled more verbal items in the memory tasks. The finding that TOD affected different performance measures may indicate that operators' nocturnal performance drop resulted from the impairment of some general, non-specific capacity involved during information processing. This is in agreement with the hypothesis that sleep deprivation reduces the non-specific arousal level of the body, but has no specific effects (see, for instance, Wilkinson 1992). According to this point of view, sleep deprivation effects on reaction time and trough-put in cognitively more loaded tasks would result from the monotony and lack of novelty that characterise these tasks. In contrast, tasks that are 'too complex, too interesting, too variable and, above all, too short' would intrinsically encourage sleep-deprived people to apply compensatory efforts and to perform normally (Wilkinson 1992, pp. 254–256). Likewise, Balkin *et al.* (2004) found that monotonous tests that demanded constant attention, such as psychomotor vigilance tests, were most sensitive to partial sleep deprivation compared to tests involving higher cognitive load, such as logical reasoning, problem solving and decision making, which were least sensitive (Balkin *et al.* 2004).

Accordingly, the present results may reflect a differential sensitivity of the tasks to sleep deprivation as shift work is associated with systematic partial sleep loss (Pilcher *et al.* 2000). The present results only partly support this assumption. On the one hand, task performance decrements in early morning hours have been observed for reaction time. This finding is also in agreement with a number of studies performed under controlled conditions in the laboratory (Monk *et al.* 1985, Casagrande *et al.* 1997, Monk and Carrier 1997, Heuer *et al.* 1998, Owens *et al.* 2000, de Gennaro *et al.* 2001, Wright *et al.* 2002), in simulated shift-work conditions (Colquhoun *et al.* 1968, Fröberg *et al.* 1975, Johnson *et al.* 1992) and in real shift-work settings (Tilley *et al.* 1982, Hossain *et al.* 2004). On the other hand, in the more demanding cognitive tasks, TOD effects interacted with the task components that were manipulated. Thus, TOD impaired response accuracy in the discrimination task only in two out of four task conditions, i.e. when the difference between items to be discriminated was least obvious, while no such effect was found in those conditions involving a more marked to-be-discriminated difference. Likewise, recognition performance was superior on early morning hours for auditory compared to visually encoded words. In a free recall task, operators displayed superior recall during daytime, but only when elaborate processing was favoured during encoding, rather than processing of structural or physical word attributes.

Harrison and Horne (2000) proposed to differentiate those cognitive demanding tasks involving divergent skills, such as innovating and updating strategies, from those based on convergent rule-based skills. The authors reported a number of studies demonstrating sleep deprivation effects in the former but not in the latter tasks. Although the tasks used in the present study were rule based, it may be that the changing task conditions involved strategy changes that may have increased the task-sensitivity to TOD effects. Strategy changes in relation to arousal changes have also been proposed to account for performance variations observed in laboratory studies. In the morning, when arousal was low, performance was shown to peak in those tasks involving more automatic processing, while increasing arousal across the day would favour more elaborate processing, such as processing of semantic aspects from long-term memory (Baddeley *et al.* 1970, Folkard *et al.* 1976, Folkard 1979, Folkard and Monk 1980, Oakhill and Davies 1989, Lorenzetti and Natale 1996). This hypothesis was favoured by the finding of significant correlations between operators' self-rated alertness and their task performance. More especially, when they rated alertness at a low level, their response speed in the discrimination task and their recall and recognition performance were also reduced, and vice versa. In addition, as reported above, significant TOD effects were observed in all the neuropsychological tests used, with overall higher performance during the day. However, no clear performance peak was observed in either task in the present study. This may be related to the marked inter-individual task performance differences, evidenced by large standard errors, in particular in the cognitively more loaded tasks. Inter-individual task performance differences may be related to differences in operators' sleepiness, workload, adjustment to shift work and work conditions (Akerstedt 2007, Bustamante *et al.* 2007, Costa and Sartori 2007), in addition to the afore-mentioned lack of control for external factors.

Accordingly, future field studies of the present kind should include objective sleep measures, in addition to subjective alertness or sleepiness ratings in order to explore the effects of sleepiness and adjustment to shift work on operators' task performance (Tucker *et al.* 1998). The tests used in the present study may be of limited ecological value as far as job-task complexity, real-job performance and consequences of performance impairments on safety and security are concerned. In this respect it may be useful to include in future studies a workload index, as discussed in several recent studies (Roth *et al.* 2006, Noyes and Bruneau 2007, Pretorius and Cilliers 2007). Thus, although no direct relationship may exist between the present results and real-job performance, the neuropsychological tests enabled suitable outcome measures reflecting cognitive processes also involved in real-job operations. Furthermore, data provided evidence of subtle changes in operators' information-processing strategies in relation to TOD. This further confirmed that TOD effects on memory depend on specific task-related factors (Mélan *et al.* 2007). Moreover, the data emphasise that organising shift work and work-break systems to maintain adequate concentration and avoid fatigue and boredom is still a major concern, in particular for safety-related occupations. Practical implications also include often-overlooked recommendations based on laboratory studies, and in particular on the findings of systematic performance variations according to presentation modality and depth of processing of verbal material (see section 1). The demonstration of superior recall for auditory compared to visual material is of particular relevance given the rather limited use of auditory compared to visual interfaces in many job situations, even for signalling emergency conditions.

References

Achermann, P. and Borbély, A.A., 1994. Simulation of daytime vigilance by additive interaction of homeostatic and a circadian process. *Biological Cybernetics*, 71, 115–121.

Akerstedt, T., 1991. Sleepiness at work: Effect of irregular work hours. *In*: T. Monk, ed. *Sleep, sleepiness and performance*. Chichester: John Wiley & Sons Ltd, 129–152.

Akerstedt, T., 2007. Altered sleep/wake patterns and mental performance. *Physiology and Behavior*, 90 (2–3), 209–218.

Akerstedt, T. and Gillberg, M., 1990. Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52, 29–37.

Andorre, V. and Quéinnec, Y., 1998. Changes in supervisory activity of a continuous process during night and day shifts. *International Journal of Industrial Ergonomics*, 21, 179–186.

Baddeley, A.D., 1986. *Working memory*. Oxford: Oxford University Press.

Baddeley, A.D., *et al.*, 1970. Memory and time of day. *Quarterly Journal of Experimental Psychology*, 22, 605–609.

Balkin, T.J., *et al.*, 2004. Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *Journal of Sleep Research*, 13, 219–227.

Blake, M.J.F., 1971. Temperament and time of day. *In*: W.P. Colquhoun, ed. *Biological rhythms and human performance*. New-York: Academic Press, 109–148.

Brébion, G., *et al.*, 2005. Word frequency effects on free recall and recognition in patients with schizophrenia. *Journal of Psychiatry Research*, 39, 215–222.

Bustamante, E.A., Bliss, J.P., and Anderson, B.L., 2007. Effects of varying the threshold of alarm systems and workload on human performance. *Ergonomics*, 50, 1127–1147.

Casagrande, M., *et al.*, 1997. Assessing vigilance through a brief pencil and paper letter cancellation task (LCT): Effects of one night of sleep deprivation and of the time of day. *Ergonomics*, 40, 613–630.

Colquhoun, W.P., Blake, M.J., and Edwards, R.S., 1968. Experimental studies of shift-work I: A comparison of 'rotating' and 'stabilized' 4-hour shift systems. *Ergonomics*, 11 (5), 437–453.

Colquhoun, W.P., Blake, M.J., and Edwards, R.S., 1969. Experimental studies of shift-work. 3. Stabilized 12-hour shift system. *Ergonomics*, 12 (6), 865–882.

Costa, G., *et al.*, 1979. Re-entrainment of diurnal variation of psychological and physiological performance at the end of a slowly rotated shift system in hospital workers. *International Archives Occupational Health*, 44, 165–175.

Costa, G. and Sartori, S., 2007. Ageing, working hours and work ability. *Ergonomics*, 50 (11), 1914–1930.

Cowan, N., *et al.*, 2002. Deconfounding serial recall. *Journal of Memory and Language*, 46, 153–177.

Cowan, N., Saults, J.S., and Brown, G.D., 2004. On the auditory modality superiority effect in serial recall: Separating input and output factors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 639–644.

de Gennaro, L., *et al.*, 2001. Visual search performance across 40 h of continuous wakefulness: Measures of speed and accuracy and relation with oculomotor performance. *Physiology and Behavior*, 74, 197–204.

Della Rocco, P.S. and Nesthus, T.E., 2005. Shift-work and air traffic control: Transitioning research results to the workforce. *In*: B. Kirwan, M. Rogers, and D. Schäfer, eds. *Human factors impacts in air traffic management*. Aldershot, UK: Ashgate, 243–278.

Dinges, L.S., *et al.*, 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.

Dorrian, J., *et al.*, 2007. Simulated train driving: Fatigue, self-awareness and cognitive disengagement. *Applied Ergonomics*, 38, 155–166.

Fletcher, A. and Dawson, D., 2001. Field-based validations of a work-related fatigue model based on hours of work. *Transportation Research*, 4 (Part F), 75–88.

Folkard, S., 1979. Times of day and level of processing. *Memory and Cognition*, 7, 247–252.

Folkard, S., *et al.*, 1976. The effect of memory load on the circadian variation in performance efficiency under a rapidly rotating shift system. *Ergonomics*, 19, 479–488.

Folkard, S. and Akerstedt, T., 1992. A three-process model of the regulation of alertness-sleepiness. *In*: R.J. Broughton and R.D. Ogilvie, eds. *Sleep, arousal, and performance*. Boston: Birkhaüser, 11–26.

Folkard, S. and Monk, T.H., 1980. Circadian rhythms in human memory. *British Journal of Psychology*, 71, 295–307.
Folkard, S. and Tucker, P., 2003. Shift work, safety and productivity. *Occupational Medicine (London)*, 53, 95–101.
Fröberg, J.E., *et al.*, 1975. Circadian rhythms of catecholamine excretion, shooting range performance, and self self-ratings of fatigue during sleep deprivation. *Biological Psychology*, 2, 175–188.
Galy, E., Camps, J.F., and Mélan, C., 2004. Influence of the perceptual mode on serial probe recognition across the 24-hr day in experienced shift-workers. *Cognitive Systems*, 6, 249–254.
Harrison, Y. and Horne, J.A., 2000. The impact of sleep deprivation on decision making: A review. *Journal of Experimental Psychology: Applied*, 6, 236–249.
Heuer, H., *et al.*, 1998. Effects of sleep loss, time of day, and extended mental work on implicit and explicit learning of sequences. *Journal of Experimental Psychology: Applied*, 4, 139–162.
Horne, J.A. and Ostberg, O., 1976. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, 4, 97–110.
Hossain, J.L., *et al.*, 2004. Subjective and objective evaluation of sleep and performance in daytime versus nighttime sleep in extended hours shift-workers at an underground mine. *Journal of Occupational and Environmental Medicine*, 46 (3), 212–226.
Johnson, M.P., *et al.*, 1992. Short-term memory, alertness and performance: a reappraisal of their relationship to body temperature. *Journal of Sleep Research*, 1 (1), 24–29.
Knauth, P., 1996. Design of shiftwork systems. *In*: W.P. Colquhoun, G. Costa, and S. Folkard, eds. *Shiftwork: Problems and solutions*. Frankfurt am Main: Peter Lang, 155–173.
Lorenzetti, R. and Natale, V., 1996. Time of day and processing strategies in narrative comprehension. *British Journal of Psychology*, 87, 209–221.
Lowden, A., *et al.*, 1998. Change from an 8-hour shift to a 12-hour shift, attitudes, sleep, sleepiness and performance. *Scandinavian Journal of Work and Environmental Health*, 24 (Suppl. 3), 69–75.
Maury, P. and Quéinnec, Y., 1992. Effects of 24-hour day on depth of processing in recall memory. *British Journal of Psychology*, 83, 249–260.
Mélan, C., Galy, E., and Cariou, M., 2007. Mnemonic processing in air traffic controllers (ATCs): Effects of task parameters and work organization. *International Journal of Aviation Psychology*, 17 (4), 391–409.
Monk, T.H., *et al.*, 1985. Diurnal variation in mood and performance in a time isolated environment. *Chronobiology International*, 2, 185–193.
Monk, T.H. and Carrier, J., 1997. Speed of mental processing in the middle of the night. *Sleep*, 20, 399–401.
Noyes, J.M. and Bruneau, D.P.J., 2007. A self-analysis of the NASA-TLX workload measure. *Ergonomics*, 50, 514–519.
Oakhill, J. and Davies, A.M., 1989. The effects of time of day and subjects' test expectations on recall and recognition of prose materials. *Acta Psychologica*, 72, 145–157.
Owens, D.S., *et al.*, 2000. Diurnal variations in the mood and performance of highly practiced young women living under strictly controlled conditions. *British Journal of Psychology*, 91, 41–60.
Penney, C.G., 1989. Modality effects and structure of short-term verbal memory. *Memory and Cognition*, 17, 398–422.
Pretorius, A. and Cilliers, P.J., 2007. Development of a mental workload index: A systems approach. *Ergonomics*, 50, 1503–1515.
Pilcher, J.J., Lambert, B.J., and Huffcutt, A.I., 2000. Differential effects of permanent and rotating shifts on self-report sleep length: A meta-analytic review. *Sleep*, 23 (2), 155–163.
Prince, S.E., Daselaar, S.M., and Cabeza, R., 2005. Neural of relational memory: successful encoding and retrieval of semantic and perceptual associations. *Journal of Neurosciences*, 25, 1203–1210.
Roth, E., *et al.*, 2006. Evolvable work-centred support systems for command and control: creating systems users can adapt to meet changing demands. *Ergonomics*, 49, 688–705.
Smith, M.E., McEvoy, L.K., and Gevins, A., 2002. The impact of moderate sleep loss on neurophysiologic signals during working-memory task performance. *Sleep*, 25, 784–794.
Thayer, R.E., 1978. Toward a psychological theory of multidimensional activation (arousal). *Motivation and Emotion*, 2, 1–34.
Tilley, A.J., *et al.*, 1982. The sleep and performance of shift workers. *Human Factors*, 24, 629–641.

Tucker, P., *et al.*, 1998. Shift length as a determinant of retrospective on-shift alertness. *Scandinavian Journal of Work, Environment and Health*, 24 Suppl 3, 49–54.
Van Dongen, H.P.A., *et al.*, 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*, 26, 117–126.
Wilkinson, R.T., 1992. The measurement of sleepiness. *In*: R.J. Broughton and R. Ogivie, eds. *Sleep, arousal and performance*. Boston: Birkhauser, 254–265.
Wojtczak-Jaroszowa, J., *et al.*, 1978. Changes in psychomotor and mental task performance following physical work in standard conditions, and in a shift-working situation. *Ergonomics*, 21, 801–809.
Wright, K.P., Hull, J.T., and Czeisler, C.A., 2002. Relationship between alertness, performance, and body temperature in humans. *American Journal of Physiology*, 253, R1370–R1377.