Article

Pre Print

This article is a version of the paper before peer review.

Author(s)

Young, M S and Stanton, N A

Title

Miles away. Determining the extent of secondary task interference on simulated driving

Original Citation

Theoretical Issues in Ergonomics Science, Volume 8, Issue 3, 2007, Pages 233 – 253

This version is available at:

http://dx.doi.org/10.1080/14639220601129228

Miles away. Determining the extent of secondary task interference on simulated driving.

Mark S. Young* and Neville A. Stanton

School of Engineering and Design, Brunel University

Uxbridge, Middlesex UB8 3PH, UK

* - author for correspondence

Tel: +44 (0)1895 266527

Fax: +44 (0)1895 269763

email: m.young@brunel.ac.uk

Miles away. Determining the extent of secondary task interference on simulated driving.

Abstract

There is a seemingly perennial debate in the literature about the relative merits of using a secondary task as a measure of spare attentional capacity. One of the main drawbacks is that it could adversely affect the primary task, or other measures of mental workload. The present experiment therefore addressed an important methodological issue for the dual-task experimental approach – that of secondary task interference. The current experiment recorded data in both single- and dual-task scenarios to ascertain the level of secondary task interference in the Southampton Driving Simulator. The results indicated that a spatial secondary task did not have a detrimental effect on driving performance, although it consistently inflated subjective mental workload ratings. However, the latter effect was so consistent across all conditions that it was not considered to pose a problem. General issues of experimental design, as well as wider implications of the findings for multiple resources theory, are discussed.

Keywords: attentional resources, driving simulator, experimental design, mental workload, secondary task

1. INTRODUCTION

Driver multi-tasking has been of increasing concern lately, with the proliferation of in-car devices competing for attention--particularly in the visual channel (Lansdown, 2001, Sodhi, Reimer and Llamazares, 2002). Whilst drivers may have up to 50% spare visual capacity (Hughes and Cole, 1986), it seems that in-vehicle tasks have a detrimental effect on driving performance regardless of their specific modality demands, and multiple secondary tasks compound these detrimental effects (Lansdown, Brook-Carter and Kersloot, 2002; 2004).

Nowhere has this problem attracted more popular attention than with the use of mobile phones while driving. Although there is good evidence that manual dialling can affect steering ability (Jenness, Lattanzio, O'Toole, Taylor and Pax, 2002, Salvucci, 2001), the case for an advantage of hands-free units is weakening (Kubose, Bock, Dell, Garnsey, Kramer and Mayhugh, in press; Patten, Kircher, Östlund and Nilsson, 2004), implying that the cognitive effort of holding a conversation can interfere with visual and other resources (cf. Sodhi et al., 2002).

At a theoretical level, this is a problem of divided attention, and how efficiently drivers can share their limited processing resources between tasks. Attentional resources can come from a single pool (cf. Kahneman, 1973), or there may be multiple pools dedicated to, for instance, verbal or spatial tasks (e.g., Baddeley, 2003; Wickens, 2002; see also Young and Stanton, 2002a, for a discussion of working memory versus attentional resource explanations of performance). By implication, tasks that are very similar in terms of demand are likely to interfere with each other more than tasks that are different.

This presents us with a methodological quandary when it comes to measuring driver attention, and in particular, mental workload (MWL). Previous experiments in the Southampton Driving Simulator (SDS; Young and Stanton, 2002, 2004) relied upon a secondary task as a measure of spare attentional capacity and hence an indirect measure of MWL. However, there are limitations in using the secondary task technique. The main argument against it is the problem of intrusiveness, particularly at low workload levels (Wierwille and Gutmann, 1978). The basic assumption of the technique is that only spare capacity is directed to the secondary task. Whilst there is evidence that intentional prioritisation of the primary task can attenuate interference (Temprado, Zanone, Monno and Laurent, 2001), Kantowitz (2000) has criticised experiments in driving on the basis that this assumption may not hold true for that particular domain. Despite heavily emphasised instructions to maintain priority on the primary task, then, the possibility remains that participants in previous studies may have allowed the demand characteristics of the experiment to interfere with their driving performance. To validate the results of previous experiments, and to guide the design of future studies, it is necessary to find out whether the type of secondary task has any effect on driving performance.

1.1. The debate

Secondary task measures have been used to discriminate MWL levels on the flight deck (Ephrath and Young, 1981, Thornton, Braun, Bowers and Morgan, 1992, Wickens, Gempler and Morphew, 2000) and across varying driving demands (Harms, 1991, Verwey and Veltman, 1996). However, the type of secondary task used seems to differ with every researcher in the field. Various authors have used visual tasks (Brouwer, Waterink, van Wolffelaar and Rothengatter, 1991, Ephrath and Young, 1981, Verwey and Veltman, 1996, Wickens et al., 2000), forced-choice tasks

(Thornton et al., 1992), and mental calculation (Harms, 1991, Recarte and Nunes, 2002).

The choice of secondary task is critical to ensure construct validity as a measure of MWL (Kantowitz, 200), but this decision largely seems to depend on the researcher's opinion about attentional resource theories (cf. Wickens, 1984, 2002, Wickens and Hollands, 2000). Baber (1991) favoured a multiple resource approach. In an investigation of automatic speech recognition (ASR) systems, participants were required to give verbal commands to a process control system, while performing either a verbal or a spatial secondary task. It was found that participants' responses to the secondary task were quicker if it was spatial than if it was verbal. This was thought to be consistent with multiple resources theory, in that the verbal secondary task was most disrupted by a verbal primary task:

> '…in line with the predictions of multiple resource theory, the fact that ASR use is a verbal activity means that it can be paired with a spatial reasoning task without detriment to either task.' (Baber, 1991; pp. 61-62)

The use of a secondary task designed to access the same processing code as the primary task is therefore a valid MWL metric under a multiple resources assumption. However, it also raises the problem of interference between the two tasks (Wickens and Liu, 1988). For instance, Verwey and Veltman (1996) found that while a visual secondary task was a good measure of short-term MWL peaks in driving, it also increased the frequency of steering corrections. Brouwer et al. (1991) used a similar set-up to the SDS, with manual responses (via buttons on the steering wheel) being made to a visual secondary task. Older participants (aged between 63 and 65 years) in particular found that manual secondary task responses interfered with their driving performance. However, if responses were made vocally, performance was better. These results support the multiple resources prediction that performance on concurrent tasks will be maximised if input modes, response devices, and tasks are as dissimilar as possible (Wickens and Hollands, 2000).

As an alternative to a multiple resource model of MWL, then, some authors adopt a unitary capacity assumption. Zeitlin (1995) used an auditory secondary task to assess MWL in real-world driving conditions. The use of an alternate modality for the secondary task was intended to minimise interference. This strategy appeared to work – secondary task performance was degraded as traffic density increased, with no intrusion to driving performance. The authors qualified this by stating that in the real-world setting, it was likely that participants were more concerned about maintaining driving performance, making the secondary task a true measure of spare capacity in this case.

However, using a secondary task that draws upon separate resources may not accurately reflect spare capacity, as multiple resources theory predicts a separate pool for verbal and spatial processing. Secondary tasks which make demands on the same attentional resources as the primary task appear to be more sensitive to changes in demand (e.g. Baber, 1991, Liu, 1996). Moreover, it is possible for a dual task experiment to compete for the same resource pool without adversely affecting the primary task (Baber, 1991); conversely, some studies have found that mental secondary tasks can affect vehicle control (e.g., Patten et al., 2004, Recarte and Nunes, 2002, Sodhi et al., 2002). Furthermore, there is evidence that there are other qualitative aspects of secondary tasks which can determine the extent of their interference effects, such as whether they are forced-pace or interruptible (Lansdown, Brook-Carter and Kersloot, 2004, Noy, Lemoine, Klachon and Burns, 2004), or whether they are actually perceived as a subset of the primary task (Cnossen, Meijman and Rothengatter, 2004).

Other criticisms of secondary task methods centre around their sensitivity, and their influence on other MWL measures. Wierwille's research (Wierwille, Gutmann, Hicks and Muto, 1977, Wierwille and Gutmann, 1978) found that a visual secondary task was only sensitive to gross changes in driving performance, and was less informative than examining primary task measures. However, the secondary task data were not redundant, as they did reflect spare attentional capacity. Finally, there is some evidence that the introduction of a secondary task can affect responses on subjective instruments, such as the NASA-TLX (Meshkati, Hancock and Rahimi, 1990, Liu, 1996). If subjective ratings are susceptible to secondary task interference, they may not be a reliable guide to primary task MWL.

1.2. The experiment

There are a number of issues which need resolving here. First and foremost, there is the possible problem of secondary task interference on both primary task and subjective measures of MWL. Since many experiments apparently lack a single-task control condition (Kantowitz, 2000), the present study sought to elucidate any interference effects by including such a baseline. Furthermore, there is also a question surrounding whether the secondary task should draw upon the same attentional resources as the primary task. If it is possible to use, say, a verbal secondary task in an experiment about driving, this may have practical benefits of reduced interference. It would also have major theoretical ramifications for the multiple resources approach. The possibility of a general reservoir underlying the separate resource pools (cf. Matthews, Sparkes and Bygrave, 1996, Tsang and Velazquez, 1996, Brown, 1997) may have to be reconsidered. This also has implications for other studies in the SDS (Young and Stanton, 2002, 2004), as a multiple resources approach has been throughout our research.

The present experiment used the SDS to answer two questions. On the one hand, does a concurrent spatial secondary task interfere with performance on the primary driving task? Also, a different type of secondary task was introduced, designed to draw upon verbal processing resources. The verbal secondary task was used to assess whether single- or multiple-resource models are better suited to model driver MWL. Driving was assessed alone and with each secondary task (spatial and verbal), and baseline secondary task performance was also recorded. Finally, the NASA-TLX (Hart and Staveland, 1988) was administered in all conditions to determine if there was any effect of dual task methods on subjective MWL.

2. METHOD

2. 1. Design

This experiment was designed to assess the level of interference (if any) between the primary driving task and the secondary task measure of spare capacity. It also provided an opportunity to evaluate the multiple resources view of attention described by Wickens (1984, 2002, Wickens and Hollands, 2000). The whole experiment took place in the Southampton Driving Simulator (SDS).

A within-subjects design was adopted. Six conditions, each lasting 10 minutes, covered all combinations of driving with and without the secondary task, as well as baseline secondary task performance without the driving task. Automatic transmission was employed in the driving conditions, with participants required to control steering, accelerator and brake only. Participants were instructed to catch up with and then maintain a consistent speed and distance headway from a lead vehicle, which was travelling at a constant 70mph. The choice of headway was left with the participant. The main advantage to this approach was that following a car motivated participants to drive at a relatively constant speed, thereby controlling objective demand across conditions. Otherwise, participants may have compensated for increased workload by reducing speed, which might contaminate results. Furthermore, a constant speed implied that participants all drove approximately equal distances, again controlling for workload differences which may otherwise have been incurred.

A single-carriageway track consisting of a mixture of curved and straight sections was used, with no hills or wind gusts to disturb control. There were no other vehicles in the participants' lane except for the lead vehicle, so no overtaking was necessary. However, there were oncoming vehicles, so participants were encouraged to remain in their own lane. The NASA-TLX was completed after each run.

Two different secondary tasks were used. The spatial task consisted of a rotated figures task (as used by Baber, 1991), presented in the lower left corner of the screen (figure 1). Each stimulus was a pair of stick figures (one upright; the other rotated through 0°, 90°, 180° or 270°) holding one or two flags. The flags were simple geometrical shapes, either squares or diamonds. The task was to make a judgement as to whether the figures were the same or different, based on the flags they were holding (see figure 2 for an example). The verbal task presented participants with a premise (e.g. 'A is before B') alongside a conclusion (e.g. 'AB'), and the task was to decide whether the conclusion was true or false (in this example, the conclusion is true). This task is similar to one developed by Johnson-Laird (1989), for research in mental models. Although the verbal nature of the task could be questioned, it is the contention of these authors that its presentation necessitates verbal processing. Premises and conclusions were presented simultaneously in the lower left corner of the screen, the same location as the spatial secondary task. Each task was self-paced, with responses being made via buttons attached to the steering column stalks, and brief visual feedback was provided before presentation of the next stimulus.

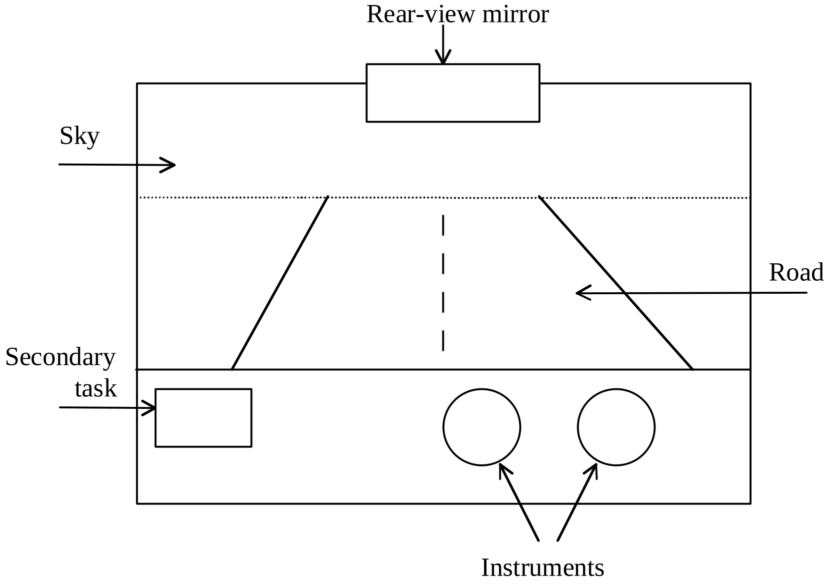

Figure 1. Schematic representation of the simulated environment

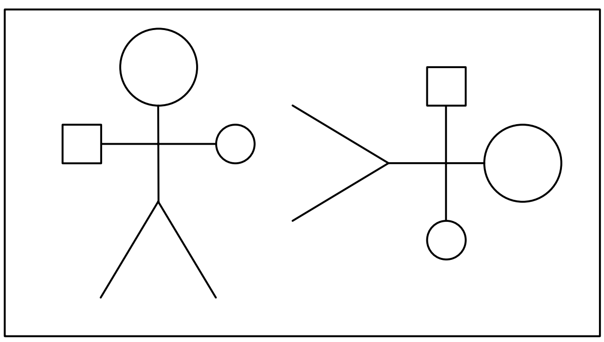

Figure 2. Example secondary task stimuli. In this case, the correct answer is ‘same’

There were three generic conditions: driving alone, driving with secondary task, and secondary task alone. A 'drive alone' trial was included at the beginning and end of the design, in order to obtain two baseline levels of driving performance. The number of participants did not allow for complete counterbalancing of the conditions, so the design in figure 3 was used.

Drive alone	Drive+Spatial	Drive+Verbal	Verbal alone	Spatial alone	Drive alone
Drive alone	Drive+Spatial	Drive+Verbal	Spatial alone	Verbal alone	Drive alone
Drive alone	Drive+Verbal	Drive+Spatial	Verbal alone	Spatial alone	Drive alone
Drive alone	Drive+Verbal	Drive+Spatial	Spatial alone	Verbal alone	Drive alone
Drive alone	Verbal alone	Spatial alone	Drive+Spatial	Drive+Verbal	Drive alone
Drive alone	Verbal alone	Spatial alone	Drive+Verbal	Drive+Spatial	Drive alone
Drive alone	Spatial alone	Verbal alone	Drive+Spatial	Drive+Verbal	Drive alone
Drive alone	Spatial alone	Verbal alone	Drive+Verbal	Drive+Spatial	Drive alone

Figure 3. Design of experiment

Dependent variables included evaluative measures of driving performance on longitudinal and lateral control (see below), total number of correct responses on the secondary task, and the subjective responses for the NASA-TLX. For the TLX, the Overall Workload score (OWL) was subject to analysis. This was calculated as the arithmetic mean of the raw scores on each of the six TLX subscales.

2.2. Participants

16 Expert driver participants (eight male) took part in this experiment. This allowed for two participants (one of each gender) to be run on each line of the design as specified above. Average age of participants was 20.8 (SD = 1.00), average annual mileage was 2438 (SD = 1289), and they had held their driving licences for an average of 2.47 years (SD = 0.99). The mileage statistics were particularly low for the participants in this study, due to sample being comprised exclusively of students who only drove during university vacations. As they were all qualified drivers with some degree of experience, though, it was deemed appropriate to classify the participants as Expert drivers.

Participants for this experiment were recruited mainly via the participant pool of the Department of Psychology. The experiment was designed according to the ethical guidelines of the British Psychological Society, and approved by the ethical committee of the Department of Psychology.

2.3. The Southampton Driving Simulator (SDS)

The SDS is a medium-fidelity, fixed-base driving simulator. The simulator consists of the front half of a Ford Orion. The steering wheel, accelerator and brake pedal produce analogue voltages. Appropriate hardware reads these voltages and converts them into digital signals to be fed into the simulation computer. An Acorn Archimedes computer runs the simulation and generates the display image. A medium-resolution colour monitor displays a view of the road and a simulated instrument panel. The resolution of the display limits the visibility range to 200 metres, at which distance another vehicle is one pixel wide. The refresh rate is 25 frames per second. The area of the screen occupied by road view is approximately 2m wide by 1.1m tall, and approximately 2.9m from the participant's eyes. The visual angle subtended at the eyepoint is therefore approximately 40º horizontal by 20º vertical. The display shows: the single-carriageway road, in solid colour with a central broken white line; other traffic in both directions; and simple roadside objects such as speed limit signs. Collisions with other vehicles or the edge of the road are detected and lead to simulated crashes. Other vehicles follow a fixed path with scripted speed changes.

The SDS software records data at a rate of 2Hz. The following data are logged: speed, lateral position on the road, distance from the vehicle in front, distance from oncoming vehicle, steering wheel and pedal positions, and collisions. The simulator was set up to run with automatic transmission at all times.

2.4. Procedure

Participants were invited into the simulator laboratory and given a 10-minute practice run, in order to familiarise themselves with the control of the driving simulator. After the practice, the nature of the secondary tasks to be used was explained to participants, with an emphasis on it being a subsidiary task when performed while driving (i.e. 'attend to it only when you feel you have time to do so').

Such instructions have been given in previous experiments using a secondary task (e.g. Young and Stanton, 2001a, b), so the present study provides the opportunity to assess their efficacy. Examples of secondary task stimuli were presented to ensure that participants understood the task. The remaining experimental instructions, including those for the primary driving task, were also given at this point.

Participants then performed the six experimental trials in the order predetermined by the design. For the conditions in which the secondary task was performed in the absence of driving, participants were simply instructed not to drive, and the secondary task was presented in the lower left corner of the simulator screen as normal. By remaining in the simulator and using the same interface to respond to the secondary task, any experimental confounds were minimised. At the end of every trial, the NASA-TLX was completed. The instruction to only rate the driving task, not the secondary task, was emphasised (except in non-driving conditions, when participants were asked to rate the secondary task). Given that this experiment was searching for evidence of secondary task interference on subjective ratings, the effectiveness of this instruction was particularly under scrutiny. At the end of the experiment, participants were debriefed as to the nature of the study.

2.5. Data reduction

For an assessment of driving performance, evaluative measures of longitudinal and lateral control were needed. Longitudinal control measures involve speed and headway. However, simple measures of location (i.e. mean, median) do not necessarily provide evaluative information about how well participants are performing. Given the instructions to participants (maintain constant speed and headway), it would be logical to adopt a measure of consistency (or rather, inconsistency) for these variables. Fortunately, Bloomfield and Carroll (1996) described such a measure, in their derivation of instability. 'A linear equation that is the line of best fit for a series of points on the track of a vehicle can be used to describe the position of the vehicle relative to the center of the lane' (Bloomfield and Carroll, 1996; p. 336). A similar line can be calculated for vehicle speed. The sampling rate of the SDS allows such equations to be calculated for the 1200 data points on each of the speed and headway variables. The standard error around this line represents the driver's ability to maintain stability in the measure. This is a better measure of driving performance than standard deviation, as it reflects the drivers' consistency in their own performance, rather than deviation from an absolute measure (J. R. Bloomfield, personal communication, December 15 1999).

For lateral control, it was considered that instability measures would not be an appropriate reflection of driving performance on a road which involves both curved and straight sections. Popular measures of lateral control (such as instability, RMS error, or time-to-line-crossing) assume that 'good' driving performance is characterised by the vehicle remaining consistently in the centre of the lane. However, modern driving techniques (e.g. Coyne, 1994) advocate a shallow trajectory when negotiating curves (i.e. approach on the outside of the curve, aim for the apex, then drift out on exit). This strategy has the effect of 'straightening' the curve, improving stability of the car as well as driver's vision. Good driving is therefore not necessarily characterised by maintaining a constant lane position, so the usual measures of lateral control will be confounded. Instead, then, simple measures of lane excursions were used to evaluate lateral control, with the assumption then being that good driving performance is rewarded with fewer lane excursions. Total number of lane excursions, and time spent out of lane, were the dependent variables for lateral control. All of the driving performance measures were filtered for outliers and extreme values, and these data points were removed prior to analysis.

3. RESULTS

3.1. Primary task data

The driving task variables were entered into a repeated measures ANOVA, with experimental condition (i.e. first drive, drive+secondary task, final drive etc.) as the independent variable. Only four of the six trials involved a driving task, so ‘condition’ was a within-subjects variable with four levels. As the present study was investigating potential interference effects, the first drive was again used to establish baseline performance. Therefore, simple contrasts with the first drive as the reference category were deemed to be most appropriate. Furthermore, a post-hoc test was used for each variable to determine if the type of secondary task (i.e. Drive+Spatial vs. Drive+Verbal) had an influence on the driving task.

There was a significant effect of driving task on the mean number of lane excursions ($F_{3,45} = 18.2$, $p < 0.001$). Compared to the first drive, lane excursions increased when performing each of the spatial ($F_{1,15} = 6.77$, $p < 0.05$) and verbal secondary tasks ($F_{1,15} = 9.75$, $p < 0.01$), but a decrease was observed in the final drive ($F_{1,15} = 9.60$, $p < 0.01$). A post-hoc test found no difference in number of lane excursions between the two secondary task conditions. The descriptive data are presented in figure 4.

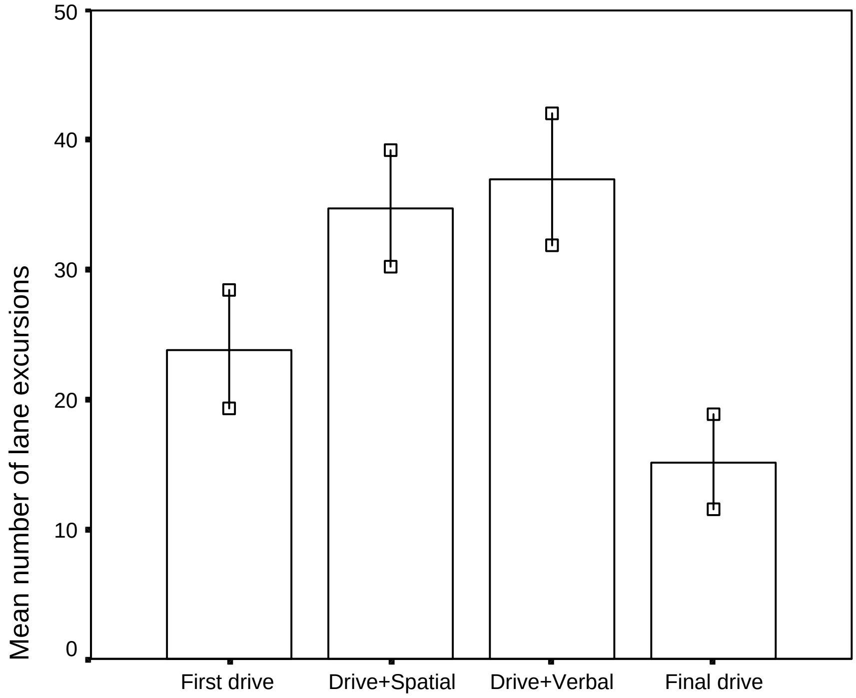

Figure 4. Mean number of lane excursions across experimental conditions. Error bars represent one standard error

A significant effect of driving task was found for time spent out of lane ($F_{3,36}$ = 7.18, $p < 0.005$), although none of the specified contrasts reached significance. Post-hoc testing revealed that the source of the main effect was a significant difference between driving with a spatial secondary task and the final drive (t_{15} = 6.02, $p < 0.001$). Since there was no significant difference with the initial baseline drive, it is likely that this result simply represented a practice effect. The descriptive statistics are presented in figure 5.

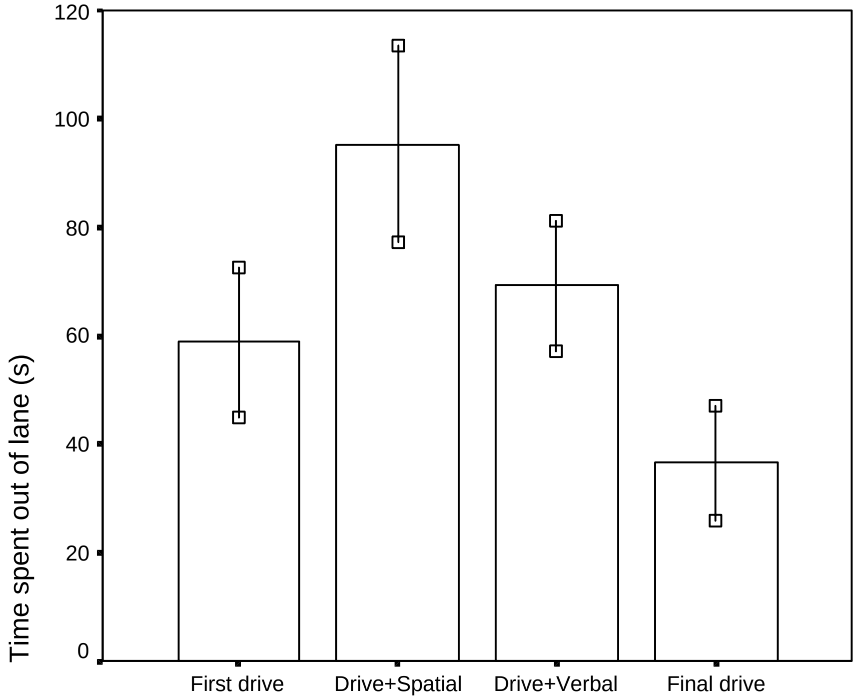

Figure 5. Time spent out of lane (s) across experimental conditions. Error bars represent one standard error

There was a significant driving task effect on speed instability ($F_{3,39} = 4.38$, $p < 0.01$), although none of the specified contrasts were significant. A visual inspection of the data suggested that the source of the main effect lay in a significant decrease in speed instability from driving with the verbal secondary task to the final drive. A post-hoc test confirmed this assumption ($t_{14} = 3.08$, $p < 0.01$). The descriptive data are presented in figure 6.

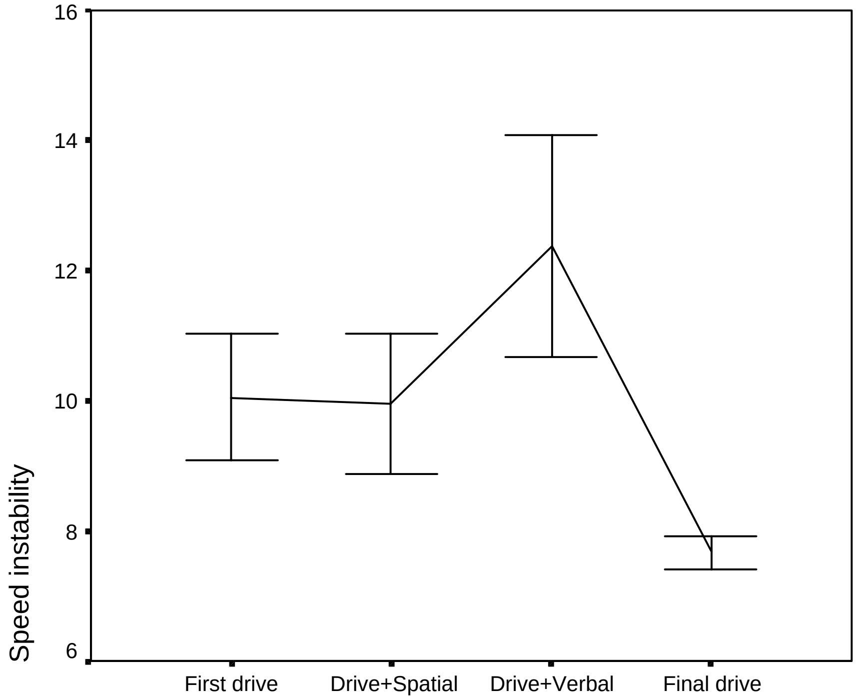

Figure 6. Speed instability across experimental conditions. Error bars represent one standard error

A significant effect of driving task was observed for headway instability ($F_{3,39} = 7.50$, $p < 0.001$). Headway instability increased significantly from the first drive to driving with the verbal secondary task ($F_{1,13} = 6.26$, $p < 0.05$). Further post-hoc testing revealed that headway instability also differed significantly depending on the type of secondary task ($t_{13} = -2.24$, $p < 0.05$). The descriptive data are illustrated in figure 7.

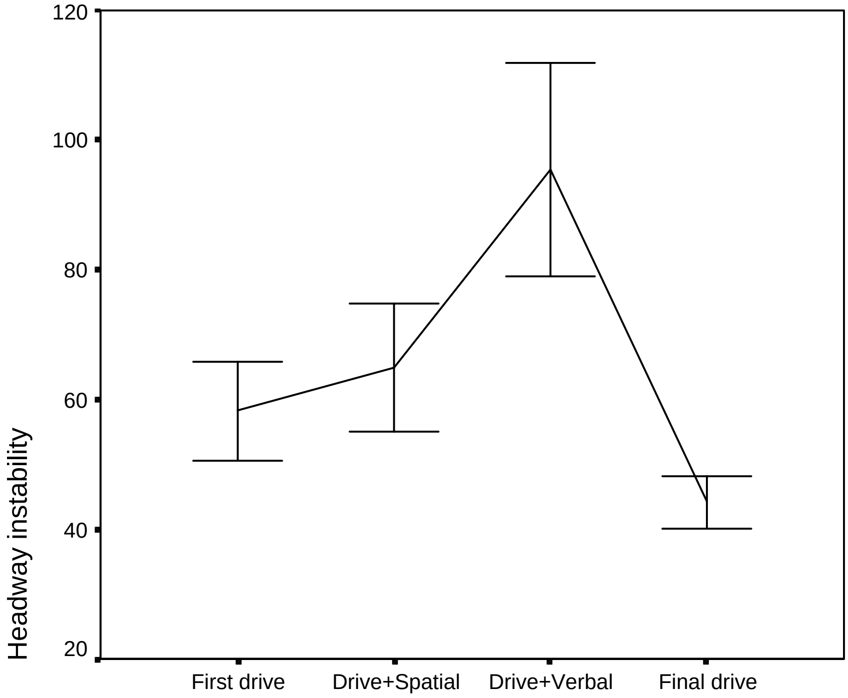

Figure 7. Headway instability across experimental conditions. Error bars represent one standard error

3.2. Secondary task data

The secondary task was performed in four out of the six conditions. The ANOVA design for secondary task data, therefore, involved four levels: Drive+Spatial, Drive+Verbal, Spatial alone, and Verbal alone. A visual inspection of the two types of task revealed that error rate was slightly higher for the verbal task than for the spatial task. Whilst this difference was significant ($F_{3,45} = 10.5$, $p < 0.001$), the data patterns for number of correct responses and total number of responses were equivalent. Therefore, the dependent variable for the secondary tasks continued to be number of correct responses.

A significant main effect of task type was observed ($F_{3,45} = 96.7$, $p < 0.001$). There was no single set of orthogonal contrasts which efficiently dealt with the comparisons of interest, so post-hoc tests were used to determine the source of the main effect. The statistics are presented in table 1. In sum, more correct responses were made if the secondary task was spatial than if it was verbal, and also if the secondary task was performed alone than if there was a concurrent driving task. The latter effect was expected, as the instructions required participants to assign priority to the driving task. However, the effect of task type implies that the verbal task was more difficult than the spatial task. The descriptive statistics are presented in figure 8.

TABLE 1

Test statistics for secondary task scores

Comparison	t	df	$p <$
Drive+Spatial vs. Drive+Verbal	6.25	15	0.001
Spatial alone vs. Verbal alone	-15.9	15	0.001
Drive+Spatial vs. Spatial alone	-7.18	15	0.001
Drive+Verbal vs. Verbal alone	-11.4	15	0.001

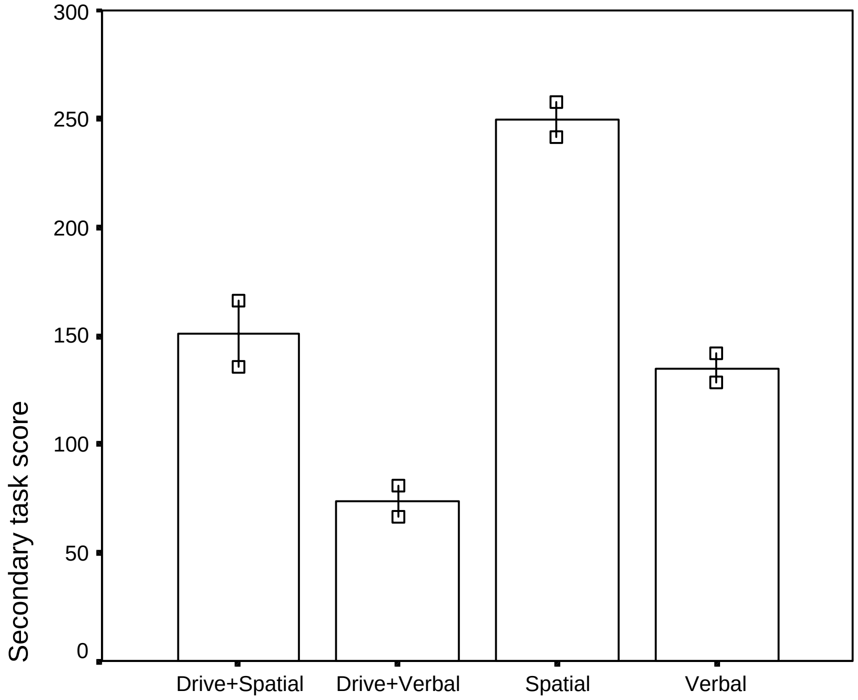

Figure 8. Secondary task scores across experimental conditions. Error bars represent one standard error

3.3. Subjective data

The NASA-TLX was completed after every single condition. Overall Workload (OWL) was calculated as the arithmetic mean of the raw scores on each of the six TLX subscales. Rather than analyse these data with six levels of the independent variable, though, the analysis was split to address difference aspects of the experiment. On the one hand, there is the problem of primary task interference, whether from the presence of a secondary task or from time-on-task. To resolve this issue, the OWL data were entered into a similar analysis as the primary task variables. That is, a repeated measures ANOVA with four levels (First drive, Drive+Spatial, Drive+Verbal, Final drive), with simple contrasts using the first condition as a reference category. The second aspect is whether subjective ratings differ according to type of secondary task, and for this another ANOVA was performed, resembling that for the interference analysis. In other words, the four secondary task conditions (Drive+Spatial, Drive+Verbal, Spatial alone, Verbal alone) were entered into a repeated measures ANOVA. Two sets of post-hoc tests were used to determine the effect of the type of secondary task, and the effect of dual- vs. single-task conditions.

Overall Workload was significantly affected by the presence of a secondary task ($F_{3,45} = 17.7$, $p < 0.001$). Compared to the first drive, there were significant increases in perceived workload in the Drive+Spatial ($F_{1,15} = 11.2$, $p < 0.005$) and Drive+Verbal conditions ($F_{1,15} = 46.9$, $p < 0.001$).

Furthermore, the type of secondary task influenced subjective ratings of Overall Workload ($F_{3,45} = 16.2$, $p < 0.001$). The statistics are presented in figure 9 and table 2, but in general, the verbal secondary task was rated as being of higher workload than the spatial task. Also, performing both primary and secondary tasks together was given higher OWL ratings than performing the secondary task alone.

TABLE 2

Test statistics for Overall Workload ratings

Comparison	t	df	p <
Drive+Spatial vs. Drive+Verbal	-3.92	15	0.001
Spatial alone vs. Verbal alone	4.40	15	0.001
Drive+Spatial vs. Spatial alone	2.91	15	0.05
Drive+Verbal vs. Verbal alone	4.67	15	0.001

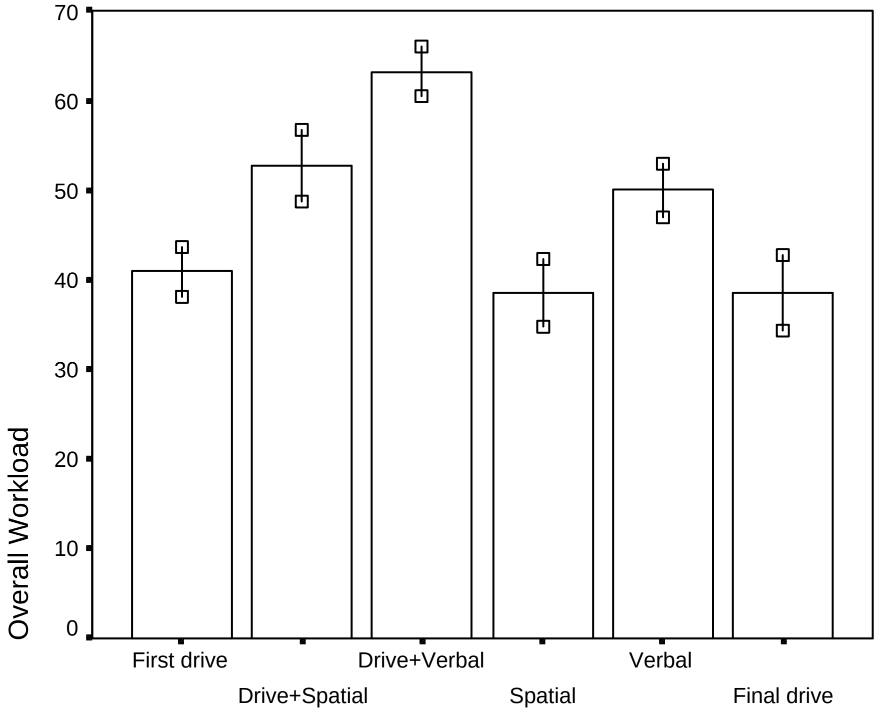

Figure 9. Overall Workload ratings across experimental conditions. Error bars represent one standard error

Given that these results were similar to the secondary task data, a correlation between the two variables was performed across the Drive+Spatial and Drive+Verbal conditions. The analysis yielded a modest, albeit significant result (r_{32} = -0.483, p < 0.01). Therefore, from a MWL point of view, these two measures can be said to measure different aspects of the same underlying construct (cf. Young and Stanton, 2001a). Although the shared variance is only slightly less than 25% here, the significant association suggests that the variables are tapping a common factor.

4. DISCUSSION

4.1. Summary of results

In consideration of the primary task performance data, several results are notable. There is something of an interference effect for lateral driving control, in that more lane excursions occurred (compared to baseline) when concurrently performing a secondary task, regardless of whether it was verbal or spatial. Informal observations suggested that participants were on average closer to the centreline in secondary task conditions, so it is reasonable to assume that they were mostly drifting into the opposite lane. The effect on lateral control may simply have represented interference between manual responses, rather than attentional resources, since steering and responding to the secondary task could potentially conflict with each other. Longitudinal control, on the other hand, mostly revealed an interference effect from the verbal secondary task. Multiple resources theory might predict this on the basis of stimulus-response compatibility (Wickens, 1984, 2002, Wickens and Hollands, 2000). The fact that the secondary task consisted of visual input, verbal processing and manual response could have caused conflict within the dimension of processing codes. However, if attentional resources are completely separate, there is no reason that this should interfere with the spatial processing of driving at all. Furthermore, there is no clear explanation from multiple resources theory as to why the verbal task interfered with both longitudinal and lateral control, while the spatial task only affected number of lane excursions.

As far as the secondary task results are concerned, fewer correct responses were made if the task was verbal than if it was spatial. This result held whether the secondary task was performed alone, or if it was performed concurrently with the driving task. The logical conclusion from this is that the verbal task was more difficult than the spatial task. In addition, performing either secondary task alone allowed participants to make more responses than when also required to maintain driving performance. This result indicates that participants were indeed allocating some priority to the driving task, as requested.

Finally, subjective reports on the TLX revealed a distinct interference effect of a concurrent secondary task. Both the spatial task and the verbal task increased perceived MWL when compared to driving without a secondary task. Furthermore, the verbal task consistently produced higher OWL ratings than the spatial task, whether combined with driving or performed alone. Again, this supports the conclusion that the verbal task was more difficult than the spatial task.

4.2. Implications: Interference

Contrary to all predictions from multiple resources theory, and from previous experimental research, driving performance in this study suffered a greater effect of interference from the verbal secondary task. The spatial secondary task only affected number of lane excursions, whereas the verbal task additionally interfered with headway instability. These results are consistent with previous findings that manual dialing affects steering (Jenness, Lattanzio, O'Toole, Taylor and Pax, 2002, Salvucci, 2001), and that verbal processing can affect longitudinal control (Kubose et al., in press). However, on the basis of the present study, we are unlikely to upset theories of attentional resource pools, as it seems from the secondary task and subjective data that the verbal task was plainly more difficult than the spatial task. It is not surprising, then, that greater interference was observed.

The fact that interference was observed at all, though, suggests that participants were not entirely obeying the instructions. Participants were clearly required to maintain performance on the driving task at all times, and only attempt the secondary task when they felt they were able. If this had been strictly followed, there would not be any interference at all from either secondary task. Therefore, the use of any subsidiary task has to be called into question. Demand characteristics seem to incite participants into wanting to do well on each task, in spite of the experimental instructions. Unless there can be a way of ensuring that participants maintain their primary task performance, researchers will have to be aware of this limitation with the secondary task technique.

A further consideration involves response competition. Lateral control was partially affected by both secondary tasks, probably due to the fact that the secondary task buttons were located on the steering column. Again, strictly speaking, participants should not have attempted to respond if they thought their driving performance would have been affected. Instead, they should have postponed their response until such a time when steering would not be affected (although this would then distort the secondary task as a pure measure of spare attentional capacity). Ideally, then, the method of responding to the secondary task should have been more harmonious with the manual demands of steering. One solution would be to locate the buttons actually on the steering wheel itself, rather than on the main column stalks.

Interference on the subjective responses is a more serious issue. There was a widespread effect of performing both tasks, as many TLX scores were inflated when compared to single-task driving. Again, this was in spite of instructions to only rate the driving task when completing the TLX. Nonetheless, the fact that the effect was so consistent across most of the TLX subscales can actually be turned to advantage. Wierwille and Gutmann (1978) found that a visual secondary task interfered with multiple dependent variables on the primary driving task, but it affected them all in similar ways. The authors argued that this made it a predictable and therefore controllable effect. The same could be said for the TLX data in this study. Indeed, the interference effect could even be quantified using the data from this experiment. Whether or not this assumption could be extrapolated to conditions outside this study is another question.

4.3. Implications: Multiple resources

Due to the finding that the verbal and spatial tasks were not matched for difficulty, one cannot be conclusive about the multiple resources issue. One tentative conclusion though, is that the evidence hints at a general resource reservoir underlying the separate multiple pools, commonly feeding both spatial and verbal resources. This would allow for a difficult verbal task interfering with other spatial tasks, since it will be draining the general pool and restricting available resources for spatial processing. If the processing codes for verbal and spatial tasks were entirely separate, there should be no interference at all between two such tasks, irrespective of their difficulty. This experiment demonstrated that they could not be perfectly timeshared, implying at least some sharing of resources between two ostensibly different tasks.

This conclusion is not definitive, though, as both primary and secondary tasks shared visual input and manual response, so there could be potential competition on these resource dimensions, or even between input modality (cf. Wickens and Liu, 1988). Given the nature of the tasks, it is possible that participants were switching their attention between driving and the secondary task, rather than processing them in a purely parallel manner (see Townsend, 1990, for a review of the debate between serial and parallel processing). In that case, the observed interference may simply have been due to inappropriate sampling of the secondary task – with the verbal task demanding longer glance durations than the spatial task. On the present findings, it is difficult to distinguish between a unitary resource model and an attention switching explanation. In all likelihood, though, drivers were probably employing a mixture of strategies – serially processing each task, while manual responses could be carried out in parallel (e.g., responding to the secondary task while perceptual/cognitive resources were returned to the driving task).

Therefore, although the evidence favours a single-resource argument, it by no means rules out the multiple resource model of MWL. It is the opinion here that a hybrid theory consisting of the general reservoir model, is the most likely candidate for the structure of attentional resources. Whilst this proposition cannot be confirmed on the basis of the results obtained so far, even Wickens notes the complementarity between single- and multiple-resource models in his textbook (Wickens and Hollands, 2000), and goes so far as to suggest the possibility of a common resource in some of his recent research:

> '…the added perceptual activity imposed by processing the visual [task] demands the same resources as the cognitive activity of the mental [task]…' (Wickens et al., 2000; p. 100)

4.4. Practical implications

In combination with the results of previous studies (Kubose et al., in press; Patten et al., 2004; Sodhi et al., 2002), the present experiment suggests that human attentional resource pools may not be totally mutually exclusive after all. Indeed, Liao & Moray (1993) posited that a single channel model is of more use in real world situations, which generally have more than two tasks anyway. The most obvious application of these conclusions is in the perennial controversy about the use of mobile phones while driving. The legislation against handheld phones assumes that any detrimental effect on driving is due to manual responses interfering with proper control of the vehicle. However, we can now be reasonably certain that the cognitive processing involved in holding a conversation can absorb attentional resources otherwise engaged in driving. Moreover, even a handsfree system requires some visual glances and manual interaction, particularly in making and answering calls, which can affect vehicle control (Jenness, Lattanzio, O'Toole, Taylor and Pax, 2002, Salvucci, 2001). Such a task is akin to the visual input, verbal processing and manual responding used for the verbal secondary task in this study. Whilst we have acknowledged that the theoretical explanations have been confounded by the experimental design, the results would support the case against all mobile phone use while on the move.

With a proliferation of driver information and entertainment systems being introduced into cars, it seems that vehicle designers must be ever more careful not to distract the driver on any attentional channel. However, there is some hope for accommodating technology while maintaining performance, in the shape of multisensory displays (Lansdown, 2001, Sarter, 2000) or integrated interfaces (e.g., Michon, 1993). If, as we have suggested, the source of the interference is structural (i.e., visual scanning or manual response competition), then the use of auditory or even tactile displays, coupled with vocal responses, could improve performance. Such technology would also provide an interesting new perspective on the theoretical debate – whether haptic processing has its own set of attentional resources.

5. CONCLUSIONS AND FUTURE DIRECTIONS

On the basis of the results in this study, there is little to suggest that the spatial secondary task used in our previous experiments (Young and Stanton, 2002, 2004) might adversely impact on driving performance. The only performance variable affected was number of lane excursions, and it is probably fair to assume that this was due to manual response competition. The other lateral control measure, time out of lane, did not suffer, nor did any of the longitudinal instability measures. Furthermore, although the introduction of a secondary task does inflate subjective MWL responses, this effect seems to be consistent across all trials. Indeed, a visual inspection of the data could even lead an optimistic researcher to suggest that the relation is a simple additive one.

One of the reasons for conducting the present experiment was to guide the design of future experiments into driver MWL. The most pertinent piece of advice would clearly be that if a secondary task were to be used, ensure that the method of responding does not interfere with the task of driving itself. One suggestion here has been to locate the response buttons onto the steering wheel itself. Furthermore, whilst it is evident that subjective responses are influenced by the presence of a secondary task, the consistency of this effect presents little concern to the overall conclusions.

There are a few exceptions to this which were not specifically addressed in the present study; the first concerns the issue of mental underload – for instance, while driving with automation. If the performance of a secondary task contributes to mental workload, it may not be possible to induce an underload state, regardless of the primary task demands (Liu, 2003, found that a mobile phone task actually improved driving performance in otherwise low workload situations). In that case, differences in performance with automation (if any) might not be attributable to mental underload. Now it is known that the secondary task can inflate overall workload, it would probably be sensible not to use the secondary task in experiments investigating the effects of underload on performance. There may also be individual differences with interference effects, such as gender (Lansdown, 2002, Lansdown et al., 2004, Lesch & Hancock, 2004) and particularly level of expertise. Since skilled performers have more spare capacity, they are less susceptible to interference from secondary tasks than novices (e.g., Beilock, Wierenga and Carr, 2002, Lansdown, 2002). Future research could investigate how underload and expertise interact with secondary task interference. In the meantime, the correlation between the secondary task and the TLX data suggests that although some information may be lost by omitting the secondary task, the underlying construct of MWL can still be accessed with the subjective measure.

Finally, there is the choice of secondary task type. Despite the concerns about the pure multiple resource model, it seems that best practice is still to use a spatial secondary task in experiments on driving, a predominantly spatial task itself. In the present experiment, the spatial task had the least interference effect on the primary task, and this could probably be rectified with appropriate placement of the response buttons.

The theoretical implications of this experiment were perhaps restricted by the fact that the verbal task did not match the spatial task for difficulty. It was therefore not possible to be entirely conclusive as to whether the results were best explained by attention switching (i.e., structural interference), a unitary resource model, or multiple resources theory. It is the opinion here, though, that support is growing for a hybrid model of attentional resources, incorporating aspects of both single- and multiple-resource models. The idea of a general reservoir supplying separate pools of resources is becoming accepted in the literature (e.g. Matthews et al., 1996, Tsang and Velazquez, 1996, Brown, 1997). Future work could be dedicated to investigating the validity of the multiple resource model of MWL, by using matched verbal and spatial secondary tasks, in addition to exploiting different input and response modalities. Exactly how to match the tasks for difficulty would present quite a challenge, and the different combinations of resource dimensions would make for quite a large study, certainly beyond the scope of the present paper. It would, nevertheless, be a valuable contribution to attention research. Notwithstanding such a study, in practical terms the conclusions from this experiment are clear: even if it has been designed to draw upon different resources, any in-vehicle task can potentially interfere with driving.

6. REFERENCES

Baber, C. (1991). Speech technology in control room systems: a human factors perspective. Chichester: Ellis Horwood.

Baddeley, A. (2003). Working memory: Looking back and looking forward. Nature Reviews Neuroscience, 4(10), 829-839.

Beilock, S. L., Wierenga, S. A., and Carr, T. H. (2002). Expertise, attention, and memory in sensorimotor skill execution: Impact of novel task constraints on dual task performance and episodic memory. Quarterly Journal of Experimental Psychology, 55A(4), 1211-1240.

Bloomfield, J. R., and Carroll, S. A. (1996). New measures of driving performance. In S. A. Robertson (Ed.), Contemporary ergonomics 1996. (pp. 335-340). London: Taylor and Francis.

Brouwer, W. H., Waterink, W., van Wolffelaar, P. C., and Rothengatter, T. (1991). Divided attention in experienced young and older drivers: lane tracking and visual analysis in a dynamic driving simulator. Human Factors, 33(5), 573-582.

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. Perception & Psychophysics, 59(7), 1118-1140.

Cnossen, F., Meijman, T., and Rothengatter, T. (2004). Adaptive strategy changes as a function of task demands: a study of car drivers. Ergonomics, 47(2), 218-236.

Coyne, P. (1994). Roadcraft: The police driver's handbook. London: HMSO.

Ephrath, A. R. and Young, L. R. (1981). Monitoring vs. man-in-the-loop detection of aircraft control failures. In J. Rasmussen and W. B. Rouse (Eds.), Human detection and diagnosis of system failures (pp. 143-154). New York: Plenum Press.

Harms, L. (1991). Variation in drivers' cognitive load. Effects of driving through village areas and rural junctions. Ergonomics, 34(2), 151-160.

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), Human mental workload (pp. 138-183). Amsterdam: North-Holland.

Hughes, P. K., and Cole, B. L. (1986). What attracts attention when driving? Ergonomics, 29(3), 377-391.

Jenness, J. W., Lattanzio, R. J., O'Toole, M., Taylor, N., and Pax, C. (2002). Effects of manual versus voice-activated dialing during simulated driving. Perceptual and Motor Skills, 94, 363-379.

Johnson-Laird, P. N. (1989). Mental models. In M. I. Posner (Ed.), Foundations of cognitive science. (pp. 469-499). Cambridge, MA. MIT Press.

Kantowitz, B. H. (2000). Attention and mental workload. Ergonomics for the new millennium: Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and the 44th Annual Meeting of the Human Factors and Ergonomics Society, San Diego, CA, July 29 – August 4 2000. Volume 3: Complex Systems and Performance (pp. 456-459). Santa Monica, CA: HFES.

Kubose, T. T., Bock, K., Dell, G. S., Garnsey, S. M., Kramer, A. F. and Mayhugh, J. (in press). The effects of speech production and speech comprehension on simulated driving performance. Applied Cognitive Psychology.

Lansdown, T. C. (2001). Conflicting demands in the driving task. In D. Harris (Ed.), Engineering Psychology and Cognitive Ergonomics. Volume Five: Aerospace and Transportation Systems (pp. 347-356). Aldershot: Ashgate.

Lansdown, T. C. (2002). Individual differences during driver secondary task performance: verbal protocol and visual allocation findings. Accident Analysis and Prevention, 34, 655-662.

Lansdown, T. C., Brook-Carter, N., and Kersloot, T. (2002). Primary task disruption from multiple in-vehicle systems. Intelligent Transportation Systems Journal, 7(2), 151-168.

Lansdown, T. C., Brook-Carter, N., and Kersloot, T. (2004). Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. Ergonomics, 47(1), 91-104.

Lesch, M. F., and Hancock, P. A. (2004). Driving performance during concurrent cell-phone use: are drivers aware of their performance decrements? Accident Analysis and Prevention, 36, 471-480.

Liao, J., and Moray, N. (1993). A simulation study of human performance deterioration and mental workload. Le Travail Humain, 56(4), 321-344.

Liu, Y. (1996). Quantitative assessment of effects of visual scanning on concurrent task performance. Ergonomics, 39(3), 382-399.

Liu, Y. (2003). Effects of Taiwan in-vehicle cellular audio phone system on driving performance. Safety Science, 41, 531-542.

Matthews, G., Sparkes, T. J., and Bygrave, H. M. (1996). Attentional overload, stress, and simulated driving performance. Human Performance, 9(1), 77-101.

Meshkati, N., Hancock, P. A., and Rahimi, M. (1990). Techniques in mental workload assessment. In J. R. Wilson and E. N. Corlett (Eds.), Evaluation of human work: a practical ergonomics methodology. (pp. 605-627). London: Taylor and Francis.

Michon, S. A. (1993, Ed.). Generic Intelligent Driver Support. London: Taylor & Francis.

Moss, S. A., & Triggs, T. J. (1997). Attention switching time: A comparison between young and experienced drivers. In Y. I. Noy (Ed.), Ergonomics and safety of intelligent driver interfaces. (pp. 381-392). Mahwah, NJ: Lawrence Erlbaum Associates.

Patten, C. J. D., Kircher, A., Östlund, J., and Nilsson, L. (2004). Using mobile telephones: cognitive workload and attention resource allocation. Accident Analysis and Prevention, 36, 341-350.

Recarte, M. A., and Nunes, L. (2002). Mental load and loss of control over speed in real driving. Towards a theory of attentional speed control. Transportation Research Part F, 5, 111-122.

Salvucci, D. D. (2001). Predicting the effects of in-car interface use on driver performance: an integrated model approach. International Journal of Human-Computer Studies, 55, 85-107.

Sarter, N. B. (2000). The need for multisensory interfaces in support of effective attention allocation in highly dynamic event-driven domains: The case of cockpit automation. International Journal of Aviation Psychology, 10(3), 231-245.

Sodhi, M., Reimer, B., and Llamazares, I. (2002). Glance analysis of driver eye movements to evaluate distraction. Behavior Research Methods, Instruments, and Computers, 34(4), 529-538.

Thornton, C., Braun, C., Bowers, C., and Morgan, B. B. (1992). Automation effects in the cockpit: A low-fidelity investigation. Proceedings of the Human Factors Society 36th Annual Meeting: Vol. 1 (pp. 30-34).

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee but theny can (and should) be distinguished. Psychological Science, 1(1), 46-54.

Tsang, P. S., and Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. Ergonomics, 39(3), 358-381.

Verwey, W. B., and Veltman, H. A. (1996). Detecting short periods of elevated workload: a comparison of nine workload assessment techniques. Journal of Experimental Psychology: Applied, 2(3), 270-285.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman and R. Davies (Eds.), Varieties of attention (pp. 63-101). New York: Academic Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science, 3(2), 159-177.

Wickens, C. D., Gempler, K., and Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. Transportation Human Factors, 2(2), 99-126.

Wickens, C. D., and Hollands, J. G. (2000). Engineering Psychology and Human Performance (3rd edition). Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D. and Liu, Y. (1988). Codes and modalities in multiple resources: a success and a qualification. Human Factors, 30(5), 599-616.

Wierwille, W. W., Gutmann, J. C., Hicks, T. G., and Muto, W. H. (1977). Secondary task measurement of workload as a function of simulated vehicle dynamics and driving conditions. Human Factors, 19, 557-565.

Wierwille, W. W., and Gutmann, J. C. (1978). Comparison of primary and secondary task measures as a function of simulated vehicle dynamics and driving conditions. Human Factors, 20, 233-244.

Young, M. S. & Stanton, N. A. (2002a). Attention and automation: new perspectives on mental underload and performance. Theoretical Issues in Ergonomics Science, 3(2), 178-194.

Young, M. S. and Stanton, N. A. (2002b). Malleable Attentional Resources Theory: A new explanation for the effects of mental underload on performance. Human Factors, 44(3), 365-375.

Young, M. S. and Stanton, N. A. (2004). Taking the load off: investigations of how adaptive cruise control affects mental workload. Ergonomics, 47(9), 1014-1035.

Zeitlin, L. R. (1995). Estimates of driver mental workload: A long-term field trial of two subsidiary tasks. Human Factors, 37(3), 611-621.