THE DIMENSIONS OF DRIVER PERFORMANCE
DURING SECONDARY MANUAL TASKS

Richard A. Young
Linda S. Angell
General Motors Engineering
Warren, Michigan USA 48090-9055
E-mail: richard.a.young@gm.com, linda.s.angell@gm.com

Summary: This analysis identified the underlying dimensions of driver performance using data obtained from drivers engaged in secondary manual tasks. Randomly chosen subjects balanced for age and gender used one of five advanced navigation and communication systems while driving on a closed roadway. Fifteen driver performance variables were averaged and standardized across subjects for 79 tasks. There were high correlations between all variables. Principal Component Analysis (PCA) found that the vector of loadings defining the first principal component (PC1) was positive for all 15 variables, accounting for 61 percent of the total variation across all tasks. It is interpreted as *overall driver demand*. PC2 loaded with one sign on event detection and response variables, but opposing sign on visual-manual workload variables. It identified tasks making drivers more inattentive to outside events than expected, given a task's visual-manual workload, and accounted for 17 percent of total variation. It is interpreted as *low-workload-but-high-inattentiveness*. PC3 had loadings of opposing sign for peripheral vs. central event variables (5 percent of total variation). It is interpreted as *peripheral insensitivity*. The first three components together accounted for 83 percent of total variation, which is deemed substantial. Thus most of the information available through the 15 original variables can be summarized by only three PC variables. Because the vectors of loadings defining the components are orthogonal to each other as defined by PCA, no single variable by itself can capture all the important variations in driver performance during secondary manual tasks. A multivariate design and analysis is required.

INTRODUCTION

In today's vehicles, drivers manage multiple tasks, including *primary* driving tasks such as steering, accelerating, braking, speed and lane choice, navigation, and hazard monitoring—and *secondary* tasks that they may want or need to perform in the vehicle while *en route* to a destination. To minimize driver distraction, it is necessary to have a means of determining which secondary tasks have the potential to interfere with primary driving tasks. Such knowledge is also required to develop design and driver performance standards for secondary tasks. To accomplish these goals, it is important to know what key dimensions of driver performance need to be measured in order to provide a complete and valid assessment of effects on primary driving performance. Numerous measures of driver performance have been explored in the literature on driving behavior, and intercorrelations among measures have been described. However, few, if any, previous studies have attempted to examine the structure within these correlations as a

means of identifying a set of underlying dimensions that may capture most of the variance in driver performance data. This was the purpose of the study described here. Primary and secondary task performances were measured and the data were analyzed using Principal Components Analysis (PCA).

To examine how the additional multitasking from secondary tasks may affect primary driving performance, it was deemed important to use an on-road test scenario that included those elements of primary driving that impose a workload or demand on key resources (e.g., perceptual, attentional and response resources). For this study, consideration was given to the primary driving tasks that a driver manages in "real-world" driving: controlling the vehicle's movement, path and speed; maintaining awareness of the road and traffic situation; detecting and responding to roadway events; operating and monitoring vehicle systems; and making decisions about maneuvers. Additional primary tasks of navigating to a destination, maneuvering in traffic and communicating with road-users (horn, turn signals) are also normal behaviors in "real-world" driving, and were initially considered for inclusion in the test scenario. However, the on-road testing done here was performed on a two-lane roadway closed to other traffic, and as such, could not incorporate these additional primary tasks.

The test scenario used in this study was that of driving a segment of a roadway built according to interstate highway standards (with curves, hills and bridges in it, but no intersections).[1] Subjects were asked to maintain lane position, and control speed at 40 ± 5 mph.

In addition, we wanted to represent the processes associated with detecting and responding to roadway objects and events during driving. The *peripheral detection task* consists of a light presented at random times in the peripheral view of the driver.[2,3] We extended this method to illuminate a small, red light in the forward as well as side view (in an attempt to emulate, for example, having to respond to unexpected braking by a vehicle in traffic in the forward or side lanes). It was hypothesized that the inclusion of both lights would permit a more controlled examination of so-called *cognitive tunneling* effects under high driver workload.[2] We termed this task the *Visual Events Detection Task* to distinguish it from earlier versions.[2,3] Drivers responded to a light with a tap of the brake pedal,[4] which was intended to make the response modality as natural as possible and to provide consistency with stimulus-response mappings that are routinely used in driving for responding to events on or near the roadway.

In addition to the primary driving tasks, drivers in this study were asked to perform secondary tasks. Secondary tasks that are sometimes undertaken by drivers on the road today include conversing with other passengers, eating, drinking, grooming, listening to audio entertainment, using portable devices, placing phone calls and many others.[5] In the future, this set of secondary tasks might expand to potentially include such activities as access to advanced information sources, Internet and e-mail access, and others. For the purpose of this study, we focused on secondary tasks that involve the use of in-vehicle devices. The range of tasks associated with the selected secondary tasks was hypothesized to be suitable for determining the underlying dimensions of driver performance when drivers were concurrently managing a secondary task while driving.

OBJECTIVE AND SCOPE

The objective of the study was to determine the underlying dimensions of driver performance during the conduct of secondary manual tasks inside a moving vehicle. No other known study to date has done an explicit multivariate analysis of driver performance data while performing secondary tasks. The scope of the study was restricted to tasks that could be performed using the eyes and hands (visual-manual). This interface method is the most common in vehicles today, and so it was deemed important to study it first, compared to voice, haptic or other methods. We also restricted the scope to only one secondary task at a time, because we felt it necessary to understand the effect of single secondary tasks first, before studying them in combination.

METHODS

Subjects

Subjects were licensed drivers, not previously familiar with, nor technically knowledgeable about, the systems under investigation. This restriction ensured that variation in prior experience would not bias the results. All subjects were recruited via newspaper advertisement from the greater Blacksburg/Roanoke area of Virginia, and screened by telephone survey. Equal numbers of subjects were assigned to two age groups: 18-34 and 45-65 years. Gender also was a controlled variable, with an equal number of males and females assigned to each experimental condition, within each age group. A total of 81 paid subjects participated in the on-road study.

Vehicles and Vehicle Systems

Of five vehicles delivered to the experimental site, four were sedans, and the other a sport-utility vehicle. One manufacturer made four of the vehicles, and another the fifth. Different in-vehicle information systems had been installed in four of the vehicles. One vehicle did not have an information system but had a prototype radio interface. Three of the systems were experimental prototypes for demonstration purposes, not scheduled for production then or now. The system without the information function is now in production, and the fifth was a commercial system.

Tasks

The tasks were selected to represent typical tasks that drivers would do in a vehicle with the functions available to them for a given system. The tasks chosen were selected from the categories of *communication* (e.g., phone dialing), *entertainment* (radio, CD, information), *navigation* (e.g., route destination entry), *Internet* tasks (e.g., get weather, sports stories), and *address book* (e.g. data entry). We also included conventional tasks that are commonly done in the vehicle today. Note that the task list was not identical for the five vehicles tested because the functions offered by the five vehicle systems were overlapping but not identical.

A given driver performed each of the tasks once during the test phase.[6] There were 13 to 18 task test trials undertaken by every driver in a given vehicle. The order of the tasks was individually randomized for each driver. The number of estimated steps required to complete the tasks ranged between two and 30 (for definition of estimated steps, see Ref. 7). The range of task difficulties was found in pilot work (not shown) to cover a wide range of task difficulties, from easy (e.g.,

"set cruise control") to quite hard (e.g., "enter 1100 Main St." using the manual speller). Some of the tasks were replicated across vehicles, particularly the conventional tasks, in a mixed design.

The sequence and number of task steps in all the tasks were optimized in an extensive task analysis before the main experiment began. The beginning state of the system for a given task was set the same way every time that task was performed in each system.

Experimental Design

The main independent variable manipulated between subjects was *Vehicle System*. Twelve to 18 drivers were stratified by age (younger, older) and gender (male, female) within each of the five vehicle systems. The main independent variable manipulated within subjects was *Tasks* (13 to 18 per vehicle system).

Dependent Variables

Variables collected for every task for every subject are listed in Table 1 (for details see Ref. 7).

Table 1. Dependent variables in current study.

#	Variable	Name
1	Task Completion Time	*tasktime*
2	Eyes-Off-Road Time	*eort*
3	Number of Glances to the In-Vehicle System	*glances*
4	Number of Lane Deviations	*lanedev*
5	Subjective Workload	*workload*
6	Subjective Situation Unawareness[8]	*sit_unaw*
7	Number of Speed Deviations	*speeddev*
8	Percent Unsuccessful Task Completion[9]	*per_unsu*
9	Percent of Total Visual Events Missed	*allmiss*
10	Percent of Forward Visual Events Missed	*hoodmiss*
11	Percent of Side Visual Events Missed	*sidemiss*
12	Mean Single Glance Time to System	*glncedur*
13	Time to Respond to Total Visual Events	*evnttime*
14	Time to Respond to Side Visual Events	*sidetime*
15	Time to Respond to Forward Visual Events	*hoodtime*

Facilities and Apparatus

Facilities. All on-road testing was conducted on the Virginia Tech Transportation Institute (VTTI) *Smart Road*. The completed portion of the Smart Road at the time of the study was a 1.7-mile, two-lane roadway with a banked turn-around at one end and a slower speed turn-around at the other end.[1] The road was closed to traffic other than the vehicles involved in the testing. The road has a center yellow line for determination of leftward lane violations as well as an uphill and downhill portion.

Vehicle Instrumentation. VTTI designed and programmed all vehicle instrumentation and data collection equipment (see Ref. 7 for details).

Visual Events Detection Task. The instrumentation for this task consisted of two red lights mounted on the vehicle (intended conceptually to emulate visual roadway events such as a vehicle braking), together with software to control their presentation. One light was mounted on the hood at the centerline of the driver (not the centerline of the vehicle). A second light was mounted above the left outside mirror (about 42 degrees to the left of the driver's forward line of sight). The side mirror light was 1.5 x 1 inches, and the hood light was 2.5 x 1 inches. One of the two lights randomly turned on three to six seconds after the brake pedal was pressed in response to a prior light event. Which light came on in that period (hood or side) was also randomly determined. A light remained on until the driver responded to it.[10] By providing a forward light as well as a side light, we provided additional control for head and eye movements. When the driver looked at the side light, the hood light acted as a "peripheral" light, and vice versa, controlling for a possible head or eye bias toward the side light (because then the front light would be in the periphery). The comparison to the forward event also allowed for a better measure of any potential true periphery effects, specific to peripheral event detection and not to event detection in general.

PROCEDURE

Two training areas were designated at opposite ends of the Smart Road. When leaving either training area, the vehicle would first pass a pair of large standard orange traffic barrels on opposite sides of the road. These barrels marked the *Start Gate* at which drivers reached a speed of 40 mph, and began responding to the event lights. The vehicle then passed a small standard traffic cone on the right side of the road, 0.1-mile from the Start Gate. This cone marked the approximate point at which the experimenter told the driver "begin" the task. Drivers were instructed in training to say "done" whey they felt they had completed a task. A pair of barrels at the end of the 1.0-mile run marked the *Finish Gate*, where the driver decelerated into the training area in the turn-around. Drivers therefore had a maximum of 0.9 miles, driving at the nominal speed of 40 mph, to complete the task (about 81 seconds). If a driver finished sooner, he or she drove at will to the Finish Gate. After being trained on the next task in the training area, the driver then went in the opposite direction on the road for the next test trial, and so forth.

All subjects were given a demo trial, then a training trial, for every task while in the training area. The training for each task was done immediately before the road test trial. This procedure, as per previous investigations with this method (not shown), typically produces a level of training that is likely at or near the maximum of the short-term learning curve for that particular task for that subject. Further details of the experimental procedure are given in Ref. 7.

Preparation of Data for Analysis and Experimental Controls

The measures reported below include both successful and unsuccessful task trials (but not trials on which a vehicle system malfunctioned because of software or hardware issues).[11] Preliminary analyses examined the statistical distributions of the data. The data in this experiment exhibited heterogeneity of variance for many of the variables.[12] As a control, statistics designed for use with Poisson distributions were initially applied and compared with

standard statistics suitable for use with normal distributions. No significant differences in results were found. Therefore, findings reported here are from the more familiar normal distribution statistics, which were robust to the non-normality exhibited by the underlying data distributions.

In preparing data for analysis, some possible limitations of the data for the Visual Events Detection Task were discovered. Because there was a slight grade to the roadway, there was some noise in the data that resulted from braking responses to control speed on the downhill grade. Careful processing of the data was successful in reducing this noise.[4] Also, even though attempts were made after the preliminary experiments to block sunlight glare, it was possible that sunlight at certain angles behind the vehicle reflected off the lens of the hood light back into the subject's eyes. Because all task orders were randomized, these limitations could not have affected one task more than another. Furthermore, for tasks with completion times of less than six seconds, sometimes by chance no light would be on during the task, producing no visual event response data for that trial. This fact necessarily leads to less robust measures of percent misses and response times than for the other variables for such tasks. In fact in a preliminary report on a portion of these data,[7] it was thought necessary to combine data from front and side visual events to have a sufficient number of observations to compute the measures reliably. However, the more powerful methods of multivariate data analysis in this study found that the side vs central event detection results were quite meaningful (see Results).

In the analyses, mean values for each task were computed by averaging across subjects performing that task. The data were initially examined for younger and older groups separated. In general, correlations between dependent variables were slightly higher for the younger group than for the older. However, all major findings reported below were the same in both groups. Therefore, the results reported here were calculated across the two age groups combined. The same held true for gender. Preliminary data analyses using Analysis of Variance techniques found no significant differences because of Vehicle System, so all 79 tasks were pooled into one data set.

Each variable for each task was averaged across all subjects tested for that task, yielding a 79 x 15 input data matrix. A 15 x 15 correlation matrix between the variables was then calculated across the 79 tasks. Principal Component Analysis (PCA) was then used to reduce the dimensionality of the standardized data set (i.e. each variable is normalized to a mean of zero and a standard deviation of one) to the major orthogonal dimensions.[13] PCA has been widely used in many scientific fields so the mathematical details are well known and will not be presented here.[14,15,27] Suffice it to say that the results were initially run in a common statistical package (Minitab), and checked by running the data in SPSS. Two outside statistical experts were contracted to independently check and run the same data in SAS or Minitab, and all methods concurred in the computations. There are occasional sign differences in the component loadings and scores that arise in the different packages. These have no effect on any of the results or conclusions reported here. The PCA methods in Ref. 15 were used for the reported analyses.

RESULTS

Table 2 shows the correlation coefficient matrix between all 15 dependent variables across all tasks. Every correlation coefficient in Table 2 is positive. A total of 101 out of 105 possible unique paired correlations are statistically significant at the 0.05 level or better ($r \geq 0.221$, two-

tailed, df = 77). All values in the upper left quadrant (variables one through eight) are statistically significant at the 10^{-6} level or better ($r \geq 0.522$). Most *r*-values in the lower right quadrant (variables nine through 15) are likewise significant at 10^{-6} or better.[16]

Using Principal Component Analysis, all the redundancy between variables in Table 2 was removed, and the information was simplified and represented in just a few principal components without loss of significant information (Fig. 1 and Table 3).

**Table 2. Correlation matrix between all 15 variables across all 79 tasks**

	1 tasktime	2 eort	3 glances	4 lanedev	5 workload	6 sit_unaw	7 speeddev	8 per_unsu	9 allmiss	10 hoodmiss	11 sidemiss	12 glncedur	13 evnttime	14 sidetime	15 hoodtime
1 tasktime	1.000	.984	.989	.923	.931	.910	.914	.876	**.576**	.521	.462	.521	.393	.312	.281
2 eort	.984	1.000	.989	.920	.917	.895	.875	.882	**.564**	.508	.464	**.583**	.369	.289	.273
3 glances	.989	.989	1.000	.922	.935	.911	.885	.866	**.566**	.499	.462	**.535**	.352	.276	.247
4 lanedev	.923	.920	.922	1.000	.882	.846	.820	.797	**.578**	.517	.477	.504	.400	.336	.269
5 workload	.931	.917	.935	.882	1.000	.985	.815	.775	**.540**	.459	.451	.479	.374	.299	.211
6 sit_unaw	.910	.895	.911	.846	.985	1.000	.799	.742	.509	.439	.417	.470	.367	.294	.200
7 speeddev	.914	.875	.885	.820	.815	.799	1.000	.823	**.527**	.500	.397	.420	.309	.229	.224
8 per_unsu	.876	.882	.866	.797	.775	.742	.823	1.000	.431	.431	.358	.407	.240	.210	.209
9 allmiss	**.576**	**.564**	**.566**	**.578**	**.540**	.509	**.527**	.431	1.000	.775	.849	**.544**	**.673**	**.534**	**.541**
10 hoodmiss	.521	.508	.499	.517	.459	.439	.500	.431	**.775**	1.000	.460	.394	.512	.346	**.568**
11 sidemiss	.462	.464	.462	.477	.451	.417	.397	.358	**.849**	.460	1.000	.487	**.587**	**.605**	.376
12 glncedur	.521	**.583**	**.535**	.504	.479	.470	.420	.407	**.544**	.394	.487	1.000	.412	.307	.316
13 evnttime	.393	.369	.352	.400	.374	.367	.309	.240	**.673**	.512	**.587**	.412	1.000	**.812**	**.734**
14 sidetime	.312	.289	.276	.336	.299	.294	.229	.210	**.534**	.346	**.605**	.307	**.812**	1.000	.383
15 hoodtime	.281	.273	.247	.269	.211	.200	.224	.209	**.541**	**.568**	.376	.316	**.734**	.383	1.000

Note: Bolded *r*-values are statistically significant at *p* < 0.000001 (*r* > 0.522). See Table 1 for definition of variable labels.

Fig. 1 (top) shows the vector of loadings defining the first principal component (PC1). It has positive loadings for all variables, decreasing in size for variables one through 15.[17] Although PC1 is a weighted average of all the dependent variables, it is most highly loaded by variables one to eight. It accounts for 61 percent of the total variation in the standardized task data.

Fig. 1 (middle) shows the vector of loadings defining the second principal component (PC2). PC2 is a contrast between variables one through eight (negative loadings) and variables nine through 15 (positive loadings). The loadings in the positive group are larger in absolute value than those in the negative group with the exception of variable 12 (glance duration). The loadings in the positive group for PC2 are also larger than the loadings on PC1 for those same variables nine through 15, again with the exception of glance duration. PC2 accounts for 17 percent of the total variation in the standardized task data. This variation is completely separate from the variation explained by PC1, because of the guaranteed orthogonality of the components.

Fig. 1 (bottom) shows the vector of loadings defining the third principal component (PC3). It had high positive loadings for forward event variables 10 and 15, and high negative loadings for side event variables 11 and 14. Loadings on all other variables for PC3 were negligible. PC3 accounted for 5 percent of the total variation. This variation is completely separate from the variation in the data explained by PC1 or PC2, because of the orthogonality of the components.
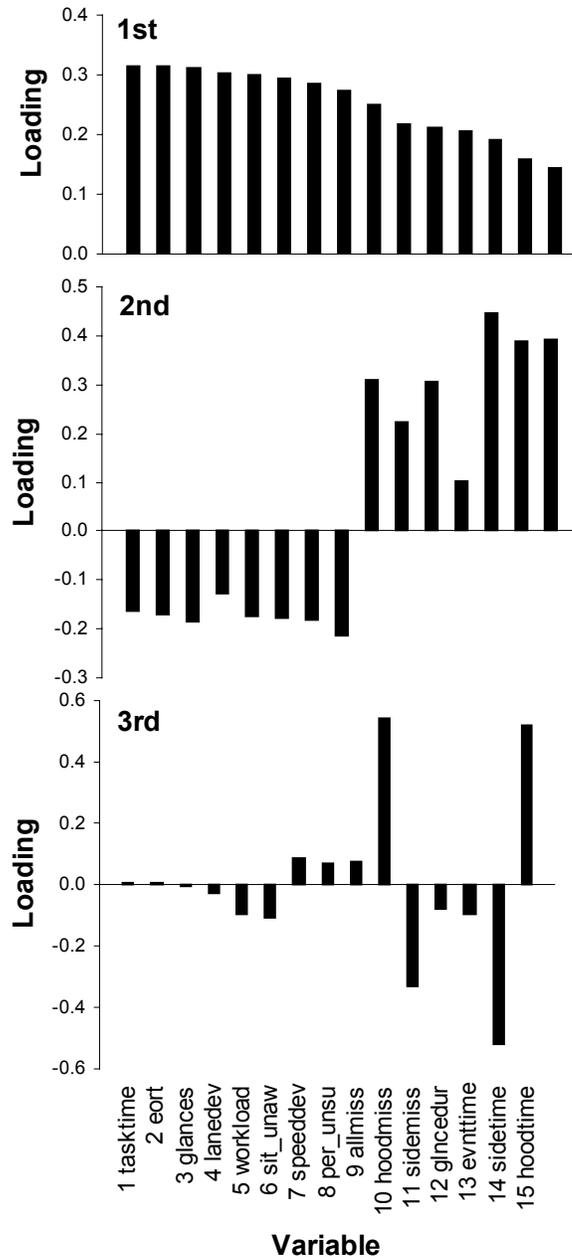
Fig. 1. Loadings of first three components

Figure 2 is a two-dimensional plot of loadings for the first two PCs. The tight cluster of variables in the lower right of Figure 2 is reflected in the high correlations between variables one through eight in Table 2. Likewise, the loose cluster of variables in the upper left of Figure 2 is reflected in the high correlations between variables nine through 15 in Table 2.[18] Thus, the PCs are accurately representing the pattern of the correlations between the variables, as expected.
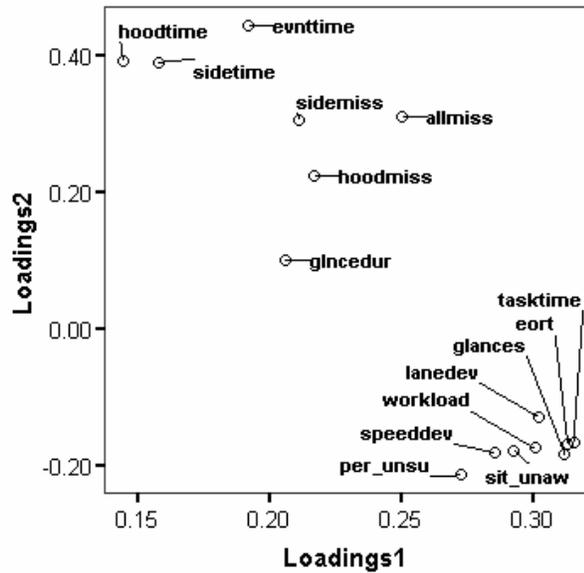
Figure 2. Two-dimensional plot of loadings for first two principal components

Table 3. Loadings of first three components

	PC1	PC2	PC3
1 tasktime	0.315711	-0.164919	0.007909
2 eort	0.313484	-0.169237	0.004130
3 glances	0.312389	-0.183500	-0.007459
4 lanedev	0.302439	-0.128792	-0.027367
5 workload	0.300917	-0.174431	-0.095792
6 sit_unaw	0.293140	-0.177664	-0.106193
7 speeddev	0.286123	-0.180773	0.086272
8 per_unsu	0.272728	-0.213594	0.068821
9 allmiss	0.250694	0.310538	0.072896
10 hoodmiss	0.217295	0.223664	0.541177
11 sidemiss	0.211500	0.305251	-0.333838
12 glncedur	0.206086	0.101393	-0.079801
13 evnttime	0.192501	0.444665	-0.096737
14 sidetime	0.158071	0.390114	-0.520020
15 hoodtime	0.144768	0.392142	0.520291

DISCUSSION

The effect on primary driving of a total of 79 in-vehicle secondary tasks was studied with 81 drivers. The first three principal components found by analyzing the correlation matrix between the 15 dependent variables accounted for 83 percent of the total variation across all 79 tasks, which is deemed substantial. Thus most of the information available through the 15 original variables can be summarized by just three Principal Component (PC) variables. The PCs reflected the underlying dimensions of driver performance, at least for the variables and tasks measured in this study.

PC1 is interpreted as *overall driver demand*. All variables tested loaded positively on PC1, particularly driver workload variables (task time, eyes-off-road time, glances, lane and speed deviations, subjective workload, subjective situation unawareness, and unsuccessful task completions). Event detection and response variables loaded less strongly but in the same direction as the visual-manual workload variables one through eight. For example, tasks scoring high on PC1 have long task times, high subjective workloads, and involve many glances off the road, along with high miss rates and long reaction times to outside events.[19] Because PC1 encompasses both driver visual-manual workload variables as well as event detection variables, it is referred to as *overall driver demand*. It explains most of the variance in all the tasks studied.

PC2 is a new driver performance dimension not previously identified in the published literature to our knowledge. It specifies tasks that make drivers less attentive or more insensitive to outside events than expected, given their visual-manual workload, and contrasts them with tasks for which attentiveness is greater than expected for their given visual-manual workload. For example, tasks with high positive scores on PC2 have few glances off the road, few lane deviations, low subjective workload and high subjective situation awareness, yet paradoxically have high miss rates and long reaction times to outside events.[20, 21] We interpret PC2 as *low-workload-but-high-inattentiveness*.[22,23] PC2 is hypothesized to be associated with a larger class of driver attentional phenomena we term *mind-off-road*.[24] PC2 is orthogonal to overall driver demand, as defined by PC1, and explains about one-sixth of the overall variance.

PC3 is interpreted as *peripheral insensitivity*, in particular opposing responsiveness for side vs. forward events.[25] It is hypothesized to be a possible narrowing of visual attention to the central visual field. Eye movement variables are not likely to explain the peripheral effect as observed in this study.[26] Peripheral insensitivity identifies variance in the driver performance data that is unique and separate from (orthogonal to) the variance attributable to PC1 and PC2. Peripheral insensitivity explains about one-twentieth of the overall variance.

Future work may find it possible to replace the large set of variables used here by a limited subset of variables. To retain the same amount of the original variation as p PCs one may only need (p+1) or (p+2) variables, and this option may be much easier to interpret than the PCs – see Section 6.2 of Ref. 27. Software has been written that implements the relevant methodology.[28]

Future work may also find that the PCA method may reduce errors in classifying tasks as inside or outside control limits for acceptable driving performance. Type I errors (false positives) mean that a task is classified as outside control limits when it is really inside control limits, leading to unnecessary redesign or lockout of the task while driving. Overall Type I errors can be reduced because one is looking at fewer PCs than variables and hence there is less scope for such errors.[29] Type II errors (false negatives) mean that a task is classified as inside control limits when it is really outside control limits, allowing a task to be used while driving that should be redesigned or locked out. Type II errors can be reduced because the PCA method may find different types of outliers (tasks outside of control limits) not seen by the original variables.[30] Finally, additional reduction of errors may be obtained by a multivariate generalization of the Student *t*-test called T^2 giving an overall measure of variability across all the PCs (see Ref. 15, p. 21).

CONCLUSION

Overall driver demand as defined by PCA is a separate driver performance dimension from *low-workload-but-high-inattentiveness*, and vice versa. *Peripheral insensitivity* is an additional dimension, separate from the first two. It follows that no single variable by itself can capture all the important variations in driver performance during the conduct of secondary in-vehicle manual tasks. A multivariate design and analysis is required.

ACKNOWLEDGMENTS

---

[1] Roadside ravines and/or hillsides on both sides of the roadway, as well as bridges and bridge abutments, meant that the driving experience was more challenging than on a closed-course test track or straightaway on which lane wander might have little practical consequence.

[2] Martens, M. H., and van Winsum, W. (2000). "Measuring distraction: The peripheral detection task." Posted at http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/Welcome.htm.

[3] Olsson, S. and Burns, P. C. (2000). "Measuring driver visual distraction with a peripheral detection task." Posted at http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/Welcome.htm.

[4] Our current experiments use a left foot switch on the floor of the vehicle to the left of the brake, like a headlamp dimmer switch on older vehicles, to more easily discriminate brake responses for reducing speed when going downhill, from light event responses.

[5] Stutts, J. C., Reinfurt, D. W., Staplin, L., and Rodgman, E. (2001). "The Role of Driver Distraction in Traffic Crashes," Chapel Hill, N.C.: Univ. of North Carolina Highway Safety Research Center, prepared for AAA Foundation for Traffic Safety, May.

[6] A pilot study (not shown) found that three trials of the same task, with each task set being run in one block, randomizing task order within each block, led to smaller subject variances around the task means, as would be expected by basic statistics. Running more than one trial forces a trade-off given fixed resources in that fewer subjects and/or fewer tasks can be run. Although not done here, some efficiency ensues if tasks less than about 20 seconds are run three trials in a row on one drive down the 0.9 mile course, using what would otherwise be time required just to drive down to the next training area after the task is done. Conversely, one could study tasks that were many minutes long (e.g. phone conversations) by pausing and resuming the task to allow the driver to turn around at the ends of the roadway.

[7] Angell, L. S., Young, R. A., Hankey, J. M. and Dingus, T. A. (2002) "An Evaluation of Alternative Methods for Assessing Driver Workload in the Early Development of In-Vehicle Information Systems," SAE Proceedings, 2002-01-1981, Government/Industry Meeting, May 15, Washington, DC, USA.

[88] Situation awareness rather than unawareness was the original experimental variable, measured on a scale from 1 to 100 with 1 being low situation awareness and 100 being high. For purposes of this report, the measured values were subtracted from 100 for simplification of analysis and clarity of presentation. The final variable was thus a measure of situation *unawareness* rather than situation *awareness*. This change ensured that this variable was "pointing in the same direction" as the other variables as a function of task difficulty. Control analyses (not shown) found that this simple reversal in scale in the data analysis for this variable did not affect any of the results or conclusions of the study.

[9] An *unsuccessful trial* was defined as the subject not reaching the goal of the task that he or she was asked to perform. "Success" was based solely on this criterion, and not on keeping within a lane, or detecting visual events. Success or non-success also did not require that the subject use the exact path or steps to the goal on which they were trained. For example, if the task were "increase HVAC fan speed two notches," and when the subject said "done" the fan speed was set at the second higher notch, that would be considered a success whatever actual steps were taken or however he or she got there. Unsuccessful task completions because of software bugs, or mechanical/electrical breakdown of a prototype system under test, were not counted as driver failures and are not included in the data or analyses reported here.

[10] In our later studies, the light is turned off after 3.5 seconds if the driver has not responded to it, and recorded as a "miss." This change protects against the possibility that in a hypothetical situation where a driver might fail to respond to a light at all for an extremely long period (e.g., the entire 81-second test trial), that only a single missed light would be recorded, leading to a possible false negative for missed events. This change would not have materially affected the results reported here.

[11] Preliminary control analyses (not shown) found little difference in the main results from excluding or including trials with unsuccessful task completions in the main data. Therefore, to provide a more complete overview of performance, the successful and unsuccessful trials were combined into one large data set for the tabulation of the dependent variables.

[12] In particular, as the mean for a task became larger, the variances across subjects for that task also increased. Such a relationship is typical for variables based on counts, which typically have a Poisson rather than normal distribution. Number of glances, lane deviations, and speed deviations were explicit count variables in this study. Even the temporal variables often exhibited such a skewed distribution, with a long tail into the high end of the histograms.

[13] In PCA, the vectors of loadings are orthogonal and the PCs themselves (the PC scores of the individual tasks) are uncorrelated. PCA is the only transformation of the variables that has both of these properties. Rotation of the vectors can preserve orthogonality of the loadings, but the PC scores for the tasks then become correlated, so no rotation was done here.

[14] Young, R. A. (1986). "Principal-component analysis of macaque lateral geniculate nucleus chromatic data," J. Opt. Soc. Amer. A, **3**, pp. 1735-1742. Posted at http://hometown.aol.com/ryoungweb, Ref. 15.

[15] Jackson, J. E. (1991). *A User's Guide to Principal Components*, Wiley & Sons, pp. 4-25.

[16] Nine of the 105 possible unique paired correlations in Table 2 may have been high because of certain aspects of the methods. Eyes-off-road time is the product of number of eye glances and the mean glance duration on any individual test trial. Therefore when the task means across subjects are correlated across all tasks, eyes-off-road time (var. 2) would be expected to have a high correlation with *glances* (var. 3, $r = 0.989$) and *glncedur* (var. 12, $r = 0.583$). Likewise on any given task trial, the total missed events are the sum of the missed hood lights and the missed side lights. Therefore, the percent *allmiss* (var. 9) would be expected to have a high correlation with the percent *hoodmiss* (var. 10, $r = 0.775$) and percent *sidemiss* (var. 11, $r = 0.849$). Also on any given trial, the mean response time is the average of the totals of the hood light response times and side light response times. Therefore, *evnttime* (var. 13) would be expected

to have a high correlation with *sidetime* (var. 14, $r = 0.812$) and *hoodtime* (var. 15, $r = 0.734$). Finally, if there were a missed event, the reaction time for that event was counted as 10 seconds (see Ref. 7), leading to increased correlations between each of the three missed event variables nine through 11 and their corresponding reaction time variables 13 to 15. These correlations cannot simply be left out of the correlation table because standard PCA statistical packages require an invertible matrix. Entire variables cannot be dropped out *a priori* because variables 1-15 have each been previously identified as important in evaluations of driver performance (e.g. Ref. 7). PCA considers all correlations in the correlation matrix, and guarantees that the resultant components are uncorrelated. Therefore it seemed plausible that the components would be robust to these nine artificially high correlations, so the PCA was done with the correlation table as shown in Table 2. Arguing in favor of the inclusive approach adopted in this paper is that if any of these variables had an exact linear relationship, they would have automatically been discovered as PCs with zero variance. That is, the original correlation matrix would not have been of "full rank" as a 15 x 15 matrix. This didn't happen in the analysis, and it can be seen from the original task data (not shown) and the correlation matrix (Table 2) that these exact relationships do not hold. This result may occur because each variable mean was calculated across subjects for a given task, after averaging all the individual glances and glance durations for that task trial for that subject. If there had been exact linear relationships, then we could have dropped one variable in each such relationship from the analysis, but this was not necessary as it turned out.

[17] In fact, the data variables were pre-ordered in this way before submission to the PCA, for clarity of presentation. Such re-ordering has no effect on the actual PCs computed.

[18] Note that the same Figure 2 would result, independent of the order of the variables in the correlation matrix. That is, if the variables were permuted in the correlation matrix, Figure 2 would remain the same. As stated in Endnote 17, the order of the variables in the correlation matrix does not affect the components.

[19] For tasks scoring high on PC1, many glances to the system could be a putative proximal cause of low event performance, since if the driver looked at the system rather than the road, he or she would necessarily tend to miss or respond slowly to the lights.

[20] The opposite loading of glances (and eyes-off-road time) and event detection variables for PC2 indicates that missed events in PC2, unlike PC1, cannot simply be attributed to eye glance behavior. That is, tasks scoring highly positive on PC2 have few glances and low eyes-off-road time, yet show many missed events, the opposite effect from PC1. This observation further lends support to the hypothesis that PC2 may be associated with attentional or cognitive effects, possibly related to *mind-off-road*.

[21] Could PC2 not be associated with attentional effects at all, but simply arise from a single long-duration eye glance during short tasks? Short tasks might have only one glance, and one light event (because events are given only every three to six seconds). If that single glance is an "orienting glance" to a new location, it might be relatively longer in duration than other glances to the same location. Hence there might be a higher percentage of missed events and slowly responded events for a short task simply due to the long glance duration off the road. Indeed, glance duration (*glancedur*, var. 12) loads positively (but weakly) on PC2 as shown in Figure 1 and Table 3 (loading = 0.101393). However, this hypothesis would imply a negative correlation between glance duration and task time, but Table 2 shows in fact there is a high positive correlation ($r = 0.521$, $p < 0.00001$). This hypothesis would also imply a negative correlation between task time and missed events, and task time and event response time, but again Table 2 shows statistically significant positive correlations ($r = 0.576$, $p < 0.000001$; $r = 0.393$, $p < 0.001$). Hence the current data cannot rule out the hypothesis that PC2 is associated with attentional or cognitive effects related to "mind-off-road," that are independent of eye movement behavior (see also Ref. 20).

[22] PC2 may be an *opponent* dimension because of the opposite signs of the loadings of the workload variables and visual event detection variables. That is, tasks scoring highly positive on PC2 have short task times, few glances, low perceived workload, and low situation unawareness, but high percentages of missed events and high event response times. The term *opponent* as used here does not necessarily imply an active inhibition of visual event sensitivity by physiological processes arising from tasks with low

visual-manual workload or vice versa, even though there is some evidence from neurophysiological studies that opposite signs on component loadings may be associated with neural inhibition (see Ref. 14). However, the pattern of loadings on PC2 demonstrates at least a *contrast* between visual-manual workload and inattentiveness to visual events, for tasks scoring highly on PC2.

[23] The term *inattentiveness* is used here in a general sense to mean insensitivity or lack of responsiveness to outside visual events. The data here do not allow for a direct determination of the *cause* of the inattentiveness. Indeed, the data do not preclude any combination of physical, physiological, behavioral, or cognitive causes. For example, tasks with low visual-manual workload might have high event insensitivity arising from (1) high mental task demand, (2) low mental task demand giving rise to "day-dreaming" or pre-occupation with internal thoughts; and/or (3) reduced attentional "supply" because of fatigue, emotional stress, or other confounding co-factors during the task. Such co-factors were common during OnStar embedded cell phone call tasks followed within ten minutes by a crash serious enough to cause airbag deployment –Young, R. A., "Association between embedded cellular phone calls and vehicle crashes involving airbag deployment," Proc. of Driving Assessment 2001: The First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Aspen CO, Aug. 2001, http://ppc.uiowa.edu/driving-assessment/2001/Summaries/Downloads/download.html#speaker3.

[24] In tasks scoring positively on PC2, the driver's subjective rating of outside road unawareness (variable six) loads in opposing fashion to objective measures of event inattentiveness (vars. 9-11 and 13-15). That is, during such tasks drivers tend to rate themselves as aware of outside events, even though they miss or respond slowly to them, lending further evidence that PC2 may be associated with *mind-off-road*. PC3 loads negligibly with situation awareness, also suggestive of lack of conscious awareness of peripheral insensitivity. See Levin, D., Momen, N., Drivdahl, S., and Simons, D. (2000) "Change blindness blindness: The metacognitive error of overestimating change-detection ability," Visual Cognition, **7** (1/2/3), 397–412; Levin, D., Drivdahl, S., Momen, N. and Rock, M. (2002) "False predictions about the detectability of visual changes: The roles of beliefs," Consciousness and Cognition **11**, pp. 507-527.

[25] Note that the periphery insensitivity effect associated with PC3 could not be uniquely specified if only responses to a peripheral light probe had been measured, without a central light probe and other dependent variables (as in Refs. 2 and 3). In such a case, there is an obscuration (i.e., confounding) of the separate and distinct contributions of PC1, PC2, and PC3 to driver performance. Tasks giving rise to high miss rates to a single peripheral light would score on all three components, and it could not be determined if the effect were due to overall task demand (PC1), general inattentiveness associated with low workload (PC2), or true peripheral insensitivity (PC3). On the other hand, even with both peripheral and central lights, these data cannot rule out the possibility that PC3 may just be an "opposing" event sensitivity engendered by having to respond to any two lights rather than one, with the peripheral location being superfluous. Future work could control for this possibility by a two-light control condition in which both lights are placed in the central visual field (fovea and parafovea areas). If the same opposing sensitivity to the two lights occurred, the opponent effects in dimension three would more likely be attributable to having to attend and respond to any two lights, rather than a peripheral vs. central insensitivity effect.

[26] Glances, glance duration and eyes-off-road variables barely load on this component, and the central light controls for bias in glances or head rotation to the side (see Methods).

[27] Jolliffe, I. T. (2002). *Principal Component Analysis*, Springer Verlag, 2nd Ed.

[28] Cadima, J. and Jolliffe, I. T. (2001). "Variable selection and the interpretation of principal subspaces," J. Agricultural, Biological and Environmental Statistics, **6**, 62-79.

[29] Assume the probability that any one of the 15 original variables is in control limits is 0.95 *(i.e.*, $\alpha = 0.05$). The probability that *all* of them are in control limits assuming they are all uncorrelated (which they are not) is $(0.95)^{15} = 0.46$, so the effective Type I error is roughly $\alpha = 0.50$, not 0.05. Thus if one were attempting to control 15 uncorrelated variables, at least one or more of these variables would falsely indicate an out-of-control condition over one-half of the time. If the variables were perfectly correlated (which again they are not here), the Type I error would remain at 0.05. It is difficult to calculate the Type

I error when the variables are correlated, as they are here, but it must be between 0.05 and 0.5. However, because PCs are uncorrelated, the Type 1 error may be easily computed for 3 PCs: $1 - (0.95)^3 = 0.14$. Setting the limits on each individual PC at a conservative significance level of $0.05/3 = 0.017$ guarantees the desired overall type I error of at most $\alpha = 0.05$ (see Ref. 15, pages 20-21). Of course, it must be emphasized that this reduction in overall Type I error occurs not because PCs are somehow better than variables, but simply because one is looking at "fewer" PCs and hence there is less scope for such errors. [30] That is, the PCA method can also help reduce false negatives (for example, see Ref. 15, p. 23, point C).