Central Executive Functions Likely Mediate the Impact of Device Operation When Driving

Sachi Mizobuchi[1], Mark Chignell[1], Junko Suzuki[2], Ko Koga[2], Kazunari Nawa[2]

[1]University of Toronto/ Vocalage Inc.
5 King's College Rd., Toronto
Ontario, M5S 3G8, Canada
sachi.mizobuchi@utoronto.ca,
chignell@mie.utoronto.ca

[2]Toyota InfoTechnology Center Co., Ltd.
6-6-20 Akasaka, Minato-ku
Tokyo 107-0052, Japan
{ju-suzuki, ko-koga, nawa}@jp.toyota-itc.com

ABSTRACT
We measured multitasking performance across a range of device interfaces and investigated the relationship between task performance and three measures of cognitive capacity (assessing the executive processes of shifting, inhibition, and updating, respectively). In the first experiment, higher levels of ability on the three executive processes of shifting, updating and inhibition were associated with improved multitasking performance. However, the impact of cognitive demands was reduced when touch input and combined visual and audio output was used in the device interaction task. When a simulated driving task was added in the second experiment, the impact of cognitive demands increased, and the use of combined audio and video output no longer reduced the impact of Central Executive (CE) function ability on performance.

While inhibition, updating, and shifting all seemed to be involved when carrying out the device operation during simulated driving, the effects of shifting and updating ability were greater than the effect of inhibition. Detailed analysis of the data indicated that the impact of cognitive demand when using a device in a dual task setting was restricted to people with low levels of shifting, updating, and inhibition ability. It is suggested that future research on the impact of in-vehicle technology focus on drivers with low levels of ability on or more of a range of cognitive factors, including the CE functions of shifting, updating, and inhibition. Based on the present results it seems likely that less cognitively able drivers will be put at most risk by inappropriate or poorly designed device tasks and interfaces.

Keywords
Mental workload, Working memory, Central Executive, in-Vehicle information device

1. INTRODUCTION
Advances in information technology, telecommunications, and sensor design are transforming the types of tasks, and associated human-machine interactions (HMIs), that people deal with while driving. More tasks compete for the driver's attention, even as they provide benefits (e.g., navigation support). The many tasks that drivers routinely deal with other than driving include optional activities such as climate control, selection of radio or entertainment settings, and communication with passengers. During driving, different tasks generate additional physical, perceptual, and cognitive workload that may lead to overload or to inattention whilst driving. From the viewpoint of HMI design for in-car information systems, the challenge is thus to design interactions that are usable and efficient without unduly compromising driving safety (cf., Coughlin et al, 2011[2]). Numerous studies related to HMI design and driver workload and distraction have been reported [11] [7]) . The research reported in this paper) focuses on understanding the mechanisms by which HMI affects driver workload.

In the research study reported here, we examined participants' performance of a choice reaction task in a multi-tasking context with five different HMI conditions. Our goal was not to evaluate which HMI design is good or bad, nor was it to identify the absolute impact of each HMI condition. Instead, we sought to examine the extent to which performance differences were based on different cognitive functions. Our overall goal was to determine what aspects of HMI cause significant increases in cognitive workload for a driver and how that impact can be ameliorated.

2. RELATED RESEARCH
2.1 Working Memory and Central Executive Functions
Research on driving cognition has typically examined cognition from the perspective of the engineering psychology model of the mind summarized in [17]. That model emphasizes the issues of workload and attention. Divided attention and distraction are then considered in terms of the availability of attentional resources in the context of different combinations of perceptual, storage, and motor modalities. The deficiency in this approach is that it fails to consider considerable research evidence that cognitive performance is critically affected by control processes in the pre-frontal cortex (often referred to as CE, functions, [14], Chapter 7).

Researchers have sought to delineate the component CE functions that exist within human working memory. In [9], 35 teenage students performed a large set of working memory tasks including digit and word span, a modified memory-updating task, and five different complex working memory tasks. The resulting complex memory span measures and the memory-updating task showed high inter-correlations, however, none of the executive function tests exhibited any significant inter-correlations. The author

AutomotiveUI'11, November 29-December 2, 2011, Salzburg, Austria

interpreted this result as supporting fractionated CE, rather than a unitary system.

Miyake et al. [12] carried out confirmatory factor analysis (CFA) and structural equation modeling (SEM) with various types of cognitive test results performed by 137 participants. Their results suggested that inhibition, shifting, and updating are primary functions within the CE.

Holtzer et al. [6] conducted a variety of neuropsychological tests with 32 (16 young and 16 older) participants, and found evidence for 4 factors using principal-component factor analysis: Attention/Executive, Memory, Motor speed, and Cognitive status factors, respectively. Regression analyses revealed that these factors were significant predictors of performance on a delayed visual recognition task, but the relationship varied as a function of task condition. While the Memory and Motor Speed factors were the strongest predictors of single-task performance, the Attention/Executive factor was the most important predictor of dual-task performance. Their result implies that different CE functions are used, depending on the task type (in this case, single or dual task).

Logie et al. [10] examined the neuropsychological basis for a multiplicity of CE functions, showing that the function for multitasking could be selectively impaired. They conducted three experiments comparing groups of Alzheimer's disease (AD) patients and healthy older and younger participants on visuo-spatial tracking and digit sequence recall, as both single tasks, and when performed concurrently. They found that AD patients showed a clear dual task deficit but were no more sensitive than control groups to varying demand. Based on their results they suggested the existence of an identifiable cognitive resource for dual task coordination within a multiple component working memory system.

2.2 Impact of Cognitive Load on Driving

Previous research on driver cognition has tended to focus on the impact of workload and stress on performance, often assuming an inverted-U shaped relationship between physiological arousal and performance that moderates that relationship (e.g., Kahneman, 1973 [8]; Hancock et al., 1987 [5]). Following the influential research of Wickens (e.g., 2002 [16]), research on driving has also examined the role of multiple attentional resources in causing or mitigating driver distraction. However, the simplistic notion that one can avoid resource conflicts by assigning different attentional sources to different tasks does not seem to work well in practice.. for instance the elevation of accident risk in hands-free cell phone use has been found to be just as high as when the cell phone is held in the hand while driving. (add reference)

Bauman et al. [1] investigated the effect of CE load on driving performance (as assessed by time to collision and driving speed). In their experiment, 48 participants drove in a driving simulator while either driving with no secondary task, or while also performing a working memory task (auditory monitoring). Two types of auditory monitoring task were used, one designed not to interfere with situation awareness, and the other (memory updating) that was expected to interfere with the comprehension and prediction function of situation awareness. The results generally supported the researchers' hypothesis that participants received less benefit from being provided with a warning signal when they had to perform the memory updating task. The researchers concluded that the CE component (updating) of working memory is strongly involved in situation awareness processes.

Engström et al. [4] examined the effect of working memory load on emergency braking performance. Participants in a driving simulator followed a lead vehicle. One group of participants performed a mental arithmetic task (to count down in increments of 7 from a given 3-digit number), and another group performed a baseline task (repeat a single digit immediately after it was presented). The driving scenario of interest involved a lead vehicle that suddenly reduced its speed so that participants had to brake to avoid crashing. The results showed that participants who had their working memories loaded with the more demanding task responded significantly slower than those who simply repeated back a single number.

Thus prior research demonstrated that placing a load on CE impairs driving performance, including aspects associated with visual attention and manual response. In the present research we sought to demonstrate that impacts of CE loading on driving could be further differentiated in terms of the relative impact of the shifting, updating, and inhibition functions, respectively.

3. Experiment 1

In the first experiment reported here, we examined the effect of device operation condition on visual attention (in terms of choice reaction task performance), and also investigated if performance was moderated by the cognitive ability of the participants. We hypothesized that if a particular device operation condition requires a particular executive function, we should see correlations between cognitive test results (i.e., ability) associated with that executive function and performance in the visual attention task.

3.1 Method

3.1.1 Participants

Thirty people participated in the experiment. The participants consisted of 16 university/college students (7 males and 9 females, aged from 21 to 28, Mean=23.7, SD=1.8) and 14 working people (8 males and 6 females, aged from 36 to 64, Mean=45.9, SD=9.5). All of the participants were English speakers with normal vision and hearing. The working people for the study were selected by a recruiting firm, based on screening criteria provided by the researchers.

3.1.2 Tasks

3.1.2.1 Cognitive tests

We selected three cognitive tests to measure each participant's ability:

Choice Response test (Inhibition test): In this task, a visual object was presented on a screen either to the left or right of centre. Participants were asked to press the "/" key on the qwerty keyboard immediately after object A (gray circle) appeared in the right (same side), and press the "z" key when object B (striped circle) appeared in the right (opposite side). This task consisted of 3 blocks of trials with 16 trials in each block. In the first two blocks, only object A or B was presented within a block. In the third block, both objects A and B were presented, in different orders for different participants. Since participants had to refrain from making the natural response on some trials *Inhibition* was deemed to be the characteristic function of this task. This task is closely related to the Simon Task, which involves a strong component of inhibition on incompatible trials (e.g., Simon, 1990 [13]).

Visual pattern order memory test (Updating test): In this task, a series of geometric outline figures were presented to participants in sequence, and then a pair of those figures was presented on a display. Participants were required to answer which of the two figures appeared earlier in the previous sequence. The sequence length started from 2, and went up to a length of 6 with each length being used in 5 trials (for a total of 25 trials). Since participants had to maintain the order information in their head until the choice pair was presented (similar to the N-back task),, *Updating* was deemed to be the most relevant function of this task.

Wisconsin Card Sort Test (WCST; Shifting test): In this task, four stimulus cards were presented to participants. The objects on the cards could differ in color, quantity, and shape. The participants were then given an additional card and were asked to choose which one of the four original cards conformed to the same category as the additional card. As the classification rule was not provided to the participants, they had to guess the rule. They did this based on the pattern of feedback provided to them ("correct" or "incorrect"), after they chose one of the four cards to match with the additional card, over a series of trials. In this experiment, the classification rule changed after 10 correct responses under the rule. The task was finished when a participant completed 6 different rules or 128 trials, whichever came earlier. Previous research [12] has shown this test to involve *Shifting* predominantly.

3.1.2.2 Baseline task

We used a visual attention task as a "baseline" to measure cognitive workload caused by the device operation tasks. In this task, either a red, down-pointing triangle (r=20 pixels) or a gray circle was presented on the main display for 1500 ms. Participants were instructed to tap a right foot pedal immediately after (and only when) they saw a red triangle. After 10 practice trials, they then performed 20 trials of this task. The study used six different inter-trial time intervals (750, 1250, 1750, 2250, 2750, 3250 ms) between the end of one trial (when the participant pushed a foot pedal to make his or her response) and the display of the stimulus to be responded to in the next trial The position that the object appeared in was varied randomly across trials, as were the inter-trial intervals. There was a ½ chance on each trial of getting each of the two types of stimuli (red triangles and gray circles) and a 1/6 chance of getting each of the six inter-trial intervals after each trial.

3.1.2.3 Device operation task

Participants performed target selection tasks, selecting a music title from a 27-title list. We developed five different device conditions as described below (summarized in Table 1). In all conditions, the instructions indicating which target to select were given in audio[1] (e.g., "Please choose title 6 Bad Romance").

L1: This condition provided a typical touch screen interface; i.e., a visual music title list (Figure 1) was presented on the touch screen and participants could scroll the list by sliding their finger on the screen and selecting a target title by touching on it. A confirmation window popped up, asking for confirmation of the selection, upon which the participants selected a title.

L2: This condition used the "Earpod" interface [18]; i.e., participants operated the touch screen without visual information. As participants scrolled the list by sliding their finger, the system "spoke" the title corresponding to the finger position using a synthesized voice. The selection point on the list moved one title at a time when participants slid one finger on the screen, and three titles at a time were skipped when a participant used two fingers to make the sliding gesture. Participants could then select a target by single tapping, and confirm the selection by double-tapping on the screen.

[1] We used the voice synthesizer provided by AT&T http://www2.research.att.com/~ttsweb/tts/demo.php

L3: This condition provided the L1 interface described above, plus audio feedback; i.e., the visual title list was displayed on the touch screen. As participants slid their finger on the screen, the system read out the selected title in a synthesized voice.

L4: This condition used an audio-only interaction. Participants could select one from a set of eight voice commands: up (move to one earlier number of title), up fast (moving up while skipping three titles), down, down fast (also skipping three titles), select, confirm, repeat selection (the system re-read the title at the current selection point) and repeat instruction (the system repeated which title the person was instructed to look for on the current trial). As participants moved through the list, the system would read out the titles at the current selection point in the synthesized voice.

L5: This condition provided the same interface as for L4 plus visual feedback; i.e., a visual title list was provided on the secondary display in addition to the audio information.

In all conditions except L4, participants interacted with the secondary display. In L4 and L5, voice commands were processed with the Wizard of Oz method; that is, an experimenter, using a remote computer, operated the device manually, based on the participant's voice commands.

Table 1. Device Operation Task Conditions.

	Input	Output
L1	touch	visual
L2	touch	audio
L3	touch	visual + audio
L4	voice	audio
L5	voice	visual + audio

Figure 1. An example of visual output of the list menu.

3.1.2.4 Subjective evaluation of workload

After each condition of the device operation task, participants were asked to evaluate the level of workload that they experienced during the task, using the NASA TLX scale.

3.1.3 Apparatus

The cognitive tests[2] and the baseline task programs were run on desktop computers (Dell dimension 515 with Windows XP SP3 and Apple Mac Pro 2006 with OS X 10.6, respectively), and they were shown on an Apple LED Cinema display (30 inches, 2560 x 1600 pixel in resolution; referred to below as "the main display").

The device operation task programs were run on a Mac Pro (2006) desktop computer. Audio information was provided on the computer, and visual information was presented on an Apple iPad (the "secondary display"). The main display was set on a desk in front of a participant at an approximately 80cm distance, while the secondary display was set below the main display at an angle of approximately 45 degrees. For the L4, and L5, conditions the remote controller program for the experimenter ran on an Apple

[2] The three cognitive test programs were developed by Ken Seergobin at the Centre for Biological Timing & Cognition, University of Toronto.

Macbook (laptop), which was wirelessly connected to the Mac Pro.

3.1.4 *Procedure*
All experimental sessions were conducted on the campus of the University of Toronto in December 2010. Participants participated in the experiment individually.

Upon signing on a consent form for participation to the experiment, participants first performed the three cognitive tests. The order of the tests varied among participants in order to avoid order effects. Each participant was exposed to all five device operation conditions, with block of 15 trials per device interface condition. Ordering of conditions was counterbalanced between participants to the extent possible given that there were 5! (120) possible orderings of the conditions and only 30 participants. Participants were instructed to respond as quickly and accurately as they could, and to allocate their attention in such a manner as to perform as well as they could on both of the tasks. Participants were paid for their participation and signed a consent form before participating, in accordance with a research protocol that was approved by the University of Toronto Ethics Review Board.

3.2 Results

3.2.1 *Baseline task performance*
Participant responses were each classified into one of four mutually exclusive events; a Hit (pedal pressed for the right target), a Miss (no pedal pressed even though the red triangle target was presented), a False alarm (pedal pressed in response to the non-target grey circle), or a Correct rejection (no pedal pressed in response to the non-target). The accuracy was then calculated as (fHit + fCR) /(fHit + fMiss + fFA + fCR), where the f-prefix indicates the frequency of each of the four response categories within a block. We also calculated means of median response time (RT) of Hit responses by device condition (L1～L5) and for the baseline task only). In order to have a single measure of accuracy that did not reflect different positions on a speed-accuracy tradeoff, we developed a measure (see the appendix for more details on its derivation) called accuracy increment (AccInc) to assign combined performance scores to participants based on deviation in accuracy from the expected accuracy of that person (based on their speed and accuracy relative to the speed-accuracy trend observed across all the participants in the experiment). Figure 2 reports the AccInc of the baseline task (AccInc_Baseline) by each condition. An ANOVA showed a significant main effect of conditions (F=4.54, p<.01). Post hoc analysis showed that AccInc was significantly higher in baseline only condition than it was in conditions 2 (p<.005), 3 (p<.005), and 5 (p<.05), whilst AccInc was significantly higher in condition 4 than it was in conditions 2 and 3. Aside from the baseline condition (6), the participants showed the highest performance in condition 4 and the lowest performance in conditions 2 and 3.

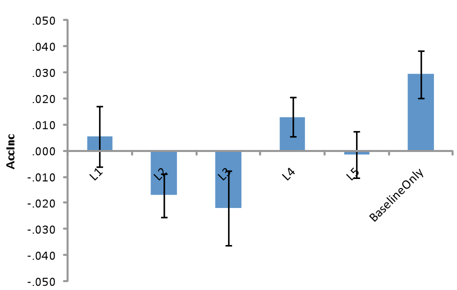

Figure 2. Baseline task performance by condition.

3.2.2 *Cognitive test performance and baseline task performance*
Prior to the analysis, the cognitive test data was processed as follows.

- Choice Response test (Inhibition test): AccInc_Inhibition was calculated based on the accuracy and RT of correct responses of the same side trial in block 3.
- Visual pattern order memory test (Updating test): AccInc_Updating was calculated based on the accuracy and RT of correct responses.
- WCST (Shifting test): Corrected Perseverative Responses were calculated by subtracting the number of ambiguous Perseverative responses from the overall number of Perseverative responses.
- Modified AccInc_Baseline (M-AccInc_Base): This measure was calculated by subtracting AccInc_Baseline in each device operation condition from AccInc_Baseline in the baseline task only condition (control) . This measure was intended to show the incremental impact of the device operation tasks.

Pearson product-moment correlations were then calculated between level of performance on the cognitive tests and the impact of device operation calculated by subtracting the AccInc in each condition from the AccInc in baseline only condition. Table 2 shows the resulting matrix of correlations (only correlations greater than r= 0.30 are shown). Greater levels of inhibition and Updating ability reduce the negative impact of adding the device condition to the baseline task. Note that only the correlations with shifting are positive, because the shifting measure counted number of perseverative results, which is an indicator of lack of shifting ability.

Shifting test performance significantly (p<.05) correlated with the baseline task performance in most conditions whereas Inhibition test performance correlated significantly with the baseline task performance only in condition 1 (one asterisk in the table indicates significance at the .05 level and two asterisks indicates significance at the .01 level).

Table 2. Correlation table among cognitive test performance and baseline task performance

	Inhibition	Updating	Shifting
L1	-.375*	-.399*	.495**
L2			.427*
L3			
L4		-.398*	.664**
L5		-.501**	.564**

Next, for each of the three CE functions examined, we divided the participants into three or four groups ordered according to measured ability on that function, and we then compared the baseline task performance among those groups. For each function the groupings were creating based on visual inspection of the frequency histogram for that function.

3.2.2.1 *Inhibition test*
The impact of inhibition ability on baseline task performance while also doing a device interaction task was assessed using AccInc. The AccInc_Baseline data were entered into a 4 (group) x 6 (device condition: L1-L5 plus baseline only) ANOVA. The results showed significant main effects of group (F[3, 26]=4.19, p<.05) and device(F[5, 130]=5.63, p<.001) Post hoc analysis (Sheffé test) showed that AccInc_Baseline, in group 1, was significantly lower than other groups in the device conditions (as shown in Figure 2) but was comparable in the baseline only

condition (all post hoc analyses in this paper used an alpha level of .05). There were no significant differences amongst group 2, 3, and 4.

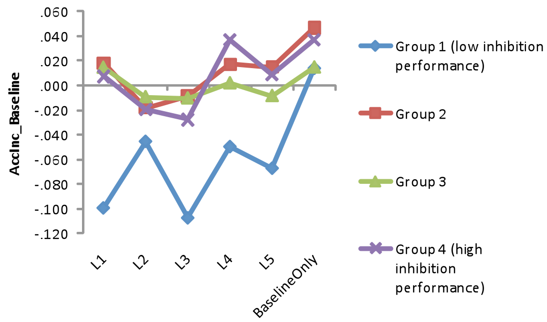

Figure 3. Mean AccInc_Baseline by Inhibition test performance group and by the device condition

3.2.2.2 Updating test

For the data on the grouping of updating ability, the AccInc_Baseline data were entered into a 4 (group) x 6 (device condition: L1-L5 plus baseline only) ANOVA. The results showed a significant main effect of group ($F[3, 26]=5.89$, $p<.01$), device condition ($F[5, 130]=6.00$, $p<.001$) and an interaction between group and device conditions ($F[15, 130]=1.50$, $p<.01$). Post hoc analysis showed that AccInc_Baseline in group 1 and 2 were significantly lower than 3 and 4. In contrast to the device condition, level of updating ability did not affect performance in the no device (baseline only) condition.

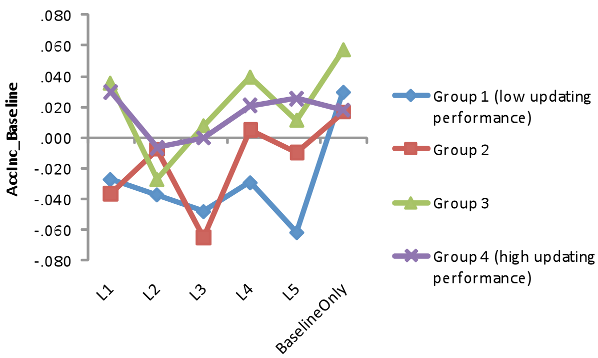

Figure 4. Mean AccInc_Baseline by Updating test performance group and by the device condition

3.2.2.3 Shifting test

The participants were classified into 3 levels based on their WCST_CPR scores. The AccInc_Baseline data were entered into a 3 (group) x 6 (device condition: L1-L5 plus baseline only) ANOVA. The results showed a significant main effect and device condition ($F[5, 135]=6.09$, $p<.001$). Post hoc analysis on the effect of condition showed that AccInc_Baseline in condition 6 (baseline only) was significantly higher than other groups, but there was no significant difference among conditions L1–L5. Thus, we combined conditions L1-L5 as a "device operation condition", and carried out a new 3 (level, i.e., people with the lowest level of shifting ability, medium and the highest) x 2 (condition) ANOVA. The result showed a significant main effect of condition ($F[1, 26]=14.8$, $p<.01$), and an interaction between group and device condition ($F[2, 26]=3.5$, $p<.05$). The performance of participants with relatively low shifting ability was significantly affected by adding the device interaction (dual) task, while the corresponding performance for people with the highest level of shifting ability was not (

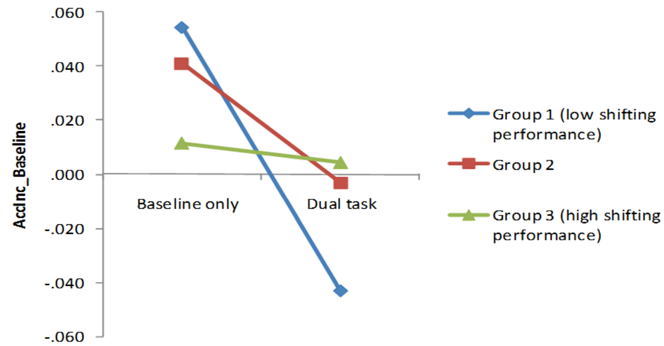

Figure 6).

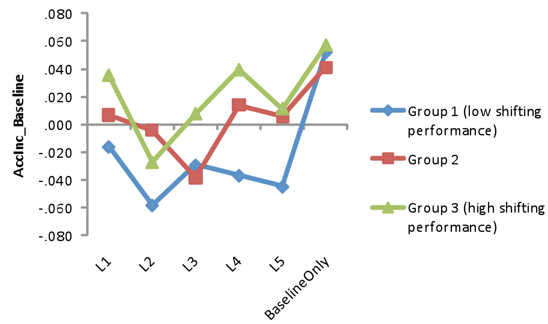

Figure 5. Mean AccInc_Baseline by Shifting test performance group and by the device condition.

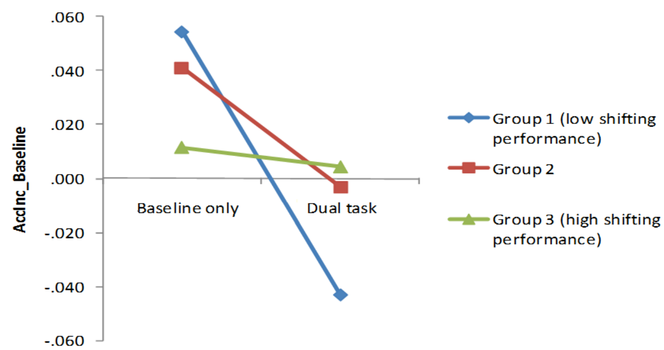

Figure 6. Mean AccInc_Baseline by Shifting test performance group and by the (new) device condition.

3.2.3 Subjective Workload

Figure 7 shows the mean TLX scores of overall workload by task type (single/dual) and device condition (L1-L5). In the single task condition, the overall workloads in L2 and L4 were relatively high whereas those in L1, L3, and L5 are relatively low. Workload data was not collected in the baseline only condition.

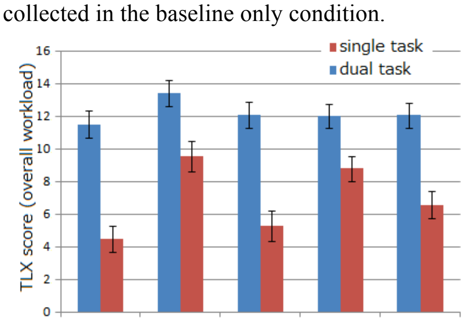

Figure 7. Mean scores of overall workload by task type and device condition and standard errors.

3.3 Discussion

In a series of preliminary analyses the effect of age was considered. However in every case the effect of CE functions proved to be a stronger determinant of performance and thus the following reporting of results focuses on the CE functions as explanatory variables. One other feature of this study was that we contrasted a group of students with a group of people outside the university community. This allowed us to explore a wider range of cognitive ability than would be possible when participants were all associated with the university community. Analysis of the differences between the university and non-university groups were also carried out, with significant differences between the groups being observed, but the results were consistent with the age differences and differences in cognitive abilities observed, and are thus not reported in this paper. In our samples, differences in cognitive ability were found to be the strongest predictor of performance and were likely to be the main driver of the effects of age and sample (university vs. non-university) that were also observed.

In experiment 1 the baseline choice reaction time task was carried out while also performing a menu selection task on one of five different device interfaces (in addition to a control, no-device condition). The 30 participants in the study had a range of test scores on the three tasks which were chosen to assess level of ability with respect to each of the executive processes of interest (shifting, updating, and inhibition).

There was no correlation between scores on the three executive abilities and performance on the baseline only task, suggesting that the demands of the baseline task alone were not great enough to require superior ability on any of the three executive processes we examined.

When the baseline task was performed along with use of the device interfaces, the only significant correlation between inhibition ability and the baseline task was in the presence of the L1 interface. In contrast, shifting and updating ability were correlated with baseline task performance in the presence of the L1, L4, and L5 device interfaces. Performance on the baseline task in the presence of the L2 and L3 device interfaces was not affected by ability on the three CE functions, except in the case of the L2 device interface whether shifting ability was implicated. Where correlations with the CE functions were observed in Table 2, it suggests that either these processes were involved in the device use task, or they were involved in managing the two tasks together.

Workload was reported as moderate in the conditions where the devices were used in a single task setting, but high in all the dual task conditions. This demonstrates the limitations of the NASA TLX workload measure for evaluating interfaces in vehicles since it is prone to ceiling effects that make it unable to differentiate between interfaces used in multi-task situations where workload is likely to be high.

People with the lowest level of ability, on all three CE tests, performed at roughly the same level as others in the baseline only condition, but were significantly worse than the others in the dual task condition.

We hypothesized earlier that if a particular device operation condition requires a particular executive function, we should see correlations between cognitive test results (i.e., ability) associated with that executive function and performance in a probe task performed at the same time as the device operation. Moderate correlations were in fact found in Experiment 1 between the the CE functions and some of the device operation conditions. Significant correlations with shifting ability were found for all but one of the device conditions (L3). Correlations with Updating were found for all conditions except L2 and L3. The L3 condition, with touch input, and both visual and audio output, was the only device condition where superior levels of CE function did not appear to be beneficial to performance of the baseline task. In contrast, the

Voice input conditions appeared to be cognitively demanding, benefiting from higher levels of updating and shifting ability. Overall, these results suggest that voice input increases cognitive demands, while providing both visual and auditory output tends to reduce them.

4. Experiment 2

In the second experiment, we extended the scope of the first experiment to the driving context: we investigated the effect of a device operation task on choice reaction task performance while simulated driving, in relation to individual cognitive capacity. Would our findings about relative cognitive demand of different device operation interfaces transfer to a simulated driving context?

4.1 Method

4.1.1 Participants

Eleven people who participated in experiment 1 also participated in Experiment 2. They were chosen so as to have different levels of the three executive functions.

4.1.2 Apparatus

Figure 8 shows the setting used in experiment 2. A Virage VS5000M driving simulator was used for the driving task. The driving course was an oval road of approximately 4 km without any traffic rights and turns.

The baseline task program ran on the Apple Mac Book and the target objects were projected onto the top margin (a black surface) of the driving simulator screen using an Epson EX5200 projector . Instead of the foot pedal used in experiment 1, participants were asked to press a button, (Apple magic mouse) mounted in the center of the steering wheel, as soon as a red triangle was presented.

The device operation task was the same as for experiment 1 except that the secondary display was mounted in front, and to the lower right, of the driver's seat.

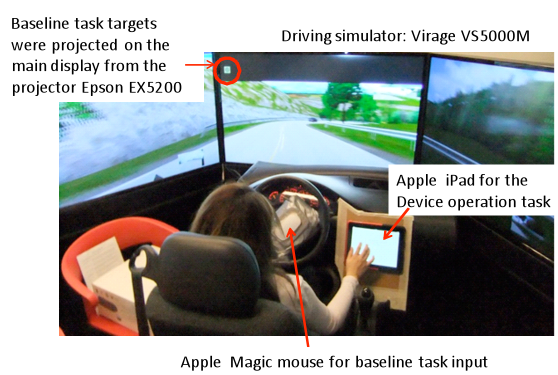

Figure 8. Experimental setting in Experiment 2.

4.1.3 Procedure

The experiment was carried out with each participant individually. After general instructions for the experiment were presented, each participant was given some practice trials so that he or she could get used to the driving simulator and the course. Participants then performed the baseline task while driving, followed by subjective rating of the workload they experienced. The participants were instructed to follow a lead car that was programmed to run at a speed of 60km/h. It was also programmed to slow down when the interval between two cars became larger than a certain distance, so that participants did not lose contact with the lead car during test sessions.

4.2 Results

4.2.1 The Effect of Condition

As in experiment 1 there appeared to be a ceiling effect in reported workload for the dual task condition with workload being uniformly high across the device interfaces. There was also no significant difference between reported workload in the dual and triple (driving, plus baseline task, plus device interface condition) task conditions, likely also because of a ceiling effect.

The error rate in experiment 2 was slightly over 5 percent (104 errors out of a total of 1767 trials). Over half of the errors were made by two of the participants. Figure 9 shows the median RT results across the Device interface conditions (as is typical with RT data, the RT distribution was positively skewed, thus the median RT provides a better characterization of the typical RT).

It can be seen that not only are response times for carrying out the menu selection task slower in the second experiment, but also that response times vary between device interface conditions much more in experiment two than they did in experiment 1. ANOVA found a significant main effect of condition ($F[5]=4.79$, $p<.01$), and a post-hoc test showed significant differences between L1 and L5, L2 and L4, L2 and L5, L3and L4, L3 and L5, and L3 and the Baseline only condition in the second experiment.

4.2.2 The Effect of Condition and CE Performance

We then looked at what might be causing the slower response times when simulated driving was added in the second experiment. The median response time of the baseline task for a participant using a device interface was subtracted from the corresponding median response time in the first experiment. This created a way to assess the relative slowing in menu selections for different interfaces that was caused by adding the simulated driving task.

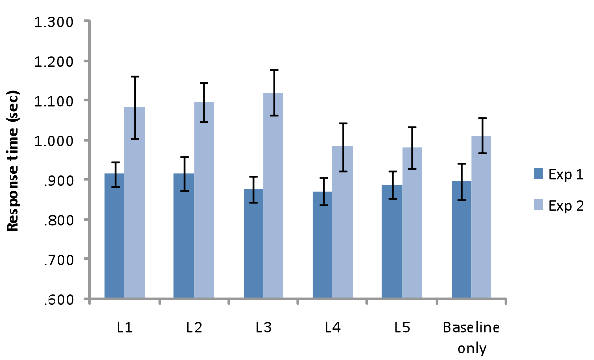

Figure 9. Comparison of Device Interface median RTs and standard errors across the Two Experiments.

Table 3. Correlations between Executive abilities and the Slowing Caused by Adding the Simulated Driving Task.

	L1	L2	L3	L4	L5
Inhibition	0.428	- 0.4	- 0.391	- 0.223	- 0.36
Updating	- 0.248	- 0.529	- 0.656*	- 0.444	- 0.673*
Shifting	0.304	0.556	0.777**	0.617	0.578

Table 3 shows the correlation between the three executive abilities and the degree of slowing caused by the simulated driving task. Due to the small sample size only relatively large correlations are statistically significant. Higher updating ability and shifting ability were related to less slowing in performance when using the L3 device interface. Shifting and updating ability showed a tendency to reduce slowing of performance on the L2, L4, and L5 interfaces although not to a statistically significant extent in most cases. There were moderate, but not significant, correlations involving inhibition ability, but in every case, except the L1 interface (where the correlations were generally lower), the correlation with inhibition was lower than the correlations with the other executive abilities.

4.3 Discussion

Analysis of slowing in device interface tasks attributable to addition of the driving task (by subtracting device condition response times in the absence of driving - Experiment 1 - from the equivalent device response times when doing simulated driving in Experiment 2) showed moderate to strong correlations (and in the expected directions) for all three CE functions assessed in this research. Thus the results of Experiment 2 suggest that all three of the CE functions (shifting, updating, and inhibition) assessed in this research have an impact on menu selection performance across a range of interface types, when that performance occurs whilst the person is also engaged in simulated driving. The benefit of using both visual and audio output in terms of apparent reduction in cognitive demand that was observed in Experiment 1 no longer applied in Experiment 2. This suggests that the addition of the simulated driving task, and the additional task switched that it required, may have added sufficient cognitive demand to outweigh the benefit of using output in both visual and audio modalities.

5. CONCLUSIONS

The pattern of results in these two experiments suggests that all three of the executive processes (shifting, updating, inhibition) are likely to be involved when people try to do menu selections while driving. Of the three executive functions, shifting and updating were found to be more strongly involved. In contrast, workload was not found to be sensitive to differences in interface design in dual and triple task contexts. Overall, the role of executive functions in device performance appeared to be stronger when the devices are used in the context of simulated driving.

While this report examined the impact of driving on a menu selection task carried out in a simulated vehicle, future research should examine how actual driving performance is affected by addition of device interactions, and how the impact of device interactions is mediated by levels of CE functioning.

The impact of the dual task and device interface conditions was found to mainly affect people with the lowest level of cognitive ability in our sample. This is in keeping with Spearman's "law of diminishing returns" ([15], cited in [3]) which implies that any two measures of cognitive ability will tend to exhibit a nonlinear relationship, with a stronger relationship at lower levels of ability. Based on the present results we propose that new in-vehicle information technologies should be tested not with university students or other convenient samples, but with segments of the driving population who have lower cognitive abilities.

Finally, the present research demonstrates that device interactions do differentially load the different CE functions, and that different device interfaces may modify the cognitive demand. The addition of a simulated driving task to device operation further increases load on CE functions, and on shifting and updating in particular.

6. ACKNOWLEDGMENTS

We would like to thank the team of Stephen Shi, Ryan Kealey, Anita Ko, Shadi Ghajar-Khosravi, Phil Lam, Aleksandra Jovicic, and Pierre Duez who helped to set up and run the experiments, and who also helped in the hardware and software development required in this study. We would also like to thank Ken Seergobin for his advice concerning the selection and measurement of CE functions used in this study.

7. REFERENCES

[1] Baumann, M. R. K, Petzoldt, T., Hogema, J. and Krems, J. F. 2008. The effect of cognitive tasks on predicting events in traffic. *In Proceedings of the European Conference on Human Centred Design for Intelligent Transport Systems.* Lyon, France, 2008, pp. 3-11.

[2] Coughlin, J.F., Reimer, B. & Mehler, B. (2011). Monitoring, Managing and Motivating Driver Safety and Well-Being IEEE Pervasive Computing, 10(3), pp. 14-21.

[3] Der, G., & Deary, I. J. (2003). IQ, reaction time and the differentiation hypothesis. *Intelligence, 31,* 491–503.

[4] Engström, J., Mikael Ljung, Aust. and Matias, V. 2010. Effects of Working Memory Load and Repeated Scenario Exposure on Emergency Braking Performance, Human Factors: *The Journal of the Human Factors and Ergonomics Society*. 52, 5, October 2010 , pp. 551-559(9)

[5] Hancock, P.A., & Chignell, M.H. (1987). Adaptive control in human-machine systems. In P.A. Hancock (Ed.), *Human factors psychology* (pp. 305-345). North Holland: Elsevier Science Publishers.

[6] Holtzer R, Stern Y, Rakitin B. C. 2004. Age-related differences in executive control of working memory. *Memory & Cognition*. 32, pp. 1333–1345.

[7] Hurts, K., Angell, L.S., and Perez, M.A. (2011). The Distracted Driver: Mechanisms, Models, and Measurement. In P. De Lucia (Ed.), Reviews of Human Factors and Ergonomics, Volume 7, 3-57.

[8] Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

[9] Lehto, J. (1996). Are Executive Function Tests Dependent on Working Memory Capacity? The Quarterly Journal of Experimental Psychology Section A, 49:1, 29-50

[10] Logie, R. H., Cocchini, G., Della Sala, S., & Baddeley, A. (2004). Is there a specific executive capacity for dual task co-ordination? Evidence from Alzheimer's disease. Neuropsychology, 18, 504–513.

[11] Matthews R., Legg S., Charlton S., (2003), Distraction, driving, multiple-task, workload, performance, Accid Anal Prev, 35 (4), 451-7.

[12] Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H. & Howerter A. 2000. The unity and diversity of executive functions and their contributions to complex ''Frontal Lobe'' tasks: A latent variable analysis. C*ognitive Psychology*, 41, pp. 49–100.

[13] Simon, J.R. (1990). The effects of an irrelevant directional cue on human information processing. In R.W. Proctor & T.G. Reeve (Eds.), Stimulus-response compatibility: An integrated perspective (pp. 31-86). Amsterdam: Elsevier.

[14] Smith, E.E. & Kosslyn, S.M. (2007). Cognitive Psychology: Mind and Brain.Upper Saddle River, NJ: Prentice Hall.

[15] Spearman, C. (1926). Some issues in the theory of ''G'' (including the law of diminishing returns). Paper presented at the British Association Section J-Psychology, Southampton

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

[17] Wickens, C. D., & Hollands, J. G. (2000). Engineering psychology and human performance (3rd ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

[18] Zhao, S., Dragicevic, P., Chignell, M.H., Balakrishnan, R., and Baudisch , P. (2007). earPod: Eyes-free Menu Selection with Touch Input and Reactive Audio Feedback. *Proceedings of CHI2007,* 1395-1404. NY: ACM Press.

8. APPENDIX: Calculating AccInc

The accuracy increment was calculated for each participant in the following steps: 1) standardized RT and accuracy scores (z-scores) were calculated; 2) regression analysis with standardized response times as the predictor and standardized accuracy as the criterion was carried out to develop predicted z-accuracy scores for each level of standardized speed (response time); 3) For each participant and combination of conditions that were of interest, predicted z-accuracy was subtracted from actual z-accuracy to obtain the AccInc measure for that participant and combination of conditions. Thus AccInc estimates the increment (or decrement) in accuracy for each participant (in standard deviation units) over that which would be expected based on their response time and the speed accuracy tradeoff observed in the sample. We carried out equivalent analyses of variance for both untransformed accuracy and AccInc as the dependent measure and in every analysis AccInc resulted in a higher F-value and a more sensitive test.