Comparing Visual and Subjective Measures of Cognitive Workload

Tuomo Kujala
Department of Computer Science and
Information Technology
P.O. Box 35
FI-40014 University of Jyväskylä
+358 40 024 7392
tuomo.kujala@jyu.fi

ABSTRACT

As a part of two larger driving simulator experiments focusing on driver distraction, we analyzed the relationship of subjectively reported levels of mental demand (NASA-TLX) and the levels indicated by three visual measures of mental workload (saccadic peak velocity, percent road centre, pupil diameter). The results suggest that the visual metrics resembled the subjective ratings but the direction of the effects were opposite. It is suggested that the proposed visual metrics of mental workload might reflect in this case the influence of the visual demands of the secondary task on an in-car display instead of mental workload. Thus, we suggest that the effects of visual secondary tasks should be carefully considered before making assumptions on mental workload when using visual measures of cognitive workload in multitasking experiments with visual in-car user interfaces.

Categories and Subject Descriptors

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – Evaluation/methodology – Ergonomics.

General Terms

Human Factors; Measurement

Keywords

Driver distraction, cognitive workload, mental workload, visual measures, subjective measures, visual-manual in-car tasks.

1. INTRODUCTION

Saccadic peak velocity [3], percentage of fixations that fall within the road centre area (Percent Road Centre, PRC) [6], as well as pupil diameter [2] [5], have all been proposed to be sensitive for variations in the levels of mental workload in different task environments. As a part of two larger driving simulator experiments focusing on driver distraction, we analyzed the relationship of subjectively reported levels of mental workload (NASA-TLX [4]) and the levels indicated by the three visual measures of mental workload (saccadic peak velocity, PRC, pupil diameter).

2. EXPERIMENT 1 – VISUALLY GUIDED VOICE COMMANDS

In Experiment 1, we studied the comparative distraction effects of two verbally commanded in-car infotainment systems. The difference of interest between the two systems was the visual guidance provided by the other system, i.e., the available commands per application were shown for the participants in the other group. Even if the in-car system was commanded verbally, the participants had to glance the in-car display in both groups in order to navigate to the correct page in the menu showing the target application (by saying “next” or “back”). An application took voice commands only when visible at the in-car display. This was explained to minimize errors in command recognition.[1]

2.1 Method

The experimental design was mixed factorial design with 2 x 2 factors (group [Visual guidance vs. No visual guidance]) x trial [baseline driving vs. dual-tasking]). 24 volunteer university students participated (12 male, 12 female). They all had sufficient driving experience and normal or corrected vision. The participants were divided into two groups of 12 (Visual guidance vs. No visual guidance).

The experiments were conducted in a fixed-base medium-fidelity driving simulator in the University of Jyväskylä (Figure 1). Remote eye-tracking system SMI RED 500 Hz was calibrated to the front view of the driving scene tracking the participants’ gaze at the driving scene. A city environment was used in practicing the driving where as a rural two-lane road environment was used in the experimental trials. Wizard-of-Oz method was used for imitating the voice command controls of the infotainment system of which menu was displayed at 21.5” Dell touch screen display.

Figure 1: The driving simulator with HUD meters and the remote eye-tracking device above the steering wheel

---

[1] Due to confidentiality issues further details of the user interfaces are not described here.

The orders of the trials (baseline, dual-tasking) were varied and counterbalanced within groups. In total, a participant completed 12 different in-car tasks while driving in the dual-tasking condition. In the driving task, their task was to keep the right lane as well as maintain vehicle speed between 40 and 60 km/h.

Here, we analyze only the measures related to mental workload: the mental demand scale on NASA-TLX, saccadic peak velocities, percentage of fixations towards road centre (PRC), and pupil diameters. PRC was defined as "*the percentage of gaze data points labeled as fixations that fall within the road centre area, where the road centre area is a circle of 8° radius centred around the driver's most frequent gaze angle*" according to [1]. Saccadic peak velocity and pupil diameter data were provided directly by SMI's analysis software (BeGaze). It's important to note that these metrics were calculated only for the captured gaze data on the driving scene (i.e. it did not include in-car fixation data).

2.2 Results

No significant between-subject effects were found with the metrics reported here. The participants reported significantly higher mental demand for the dual-tasking trial than for the baseline driving trial (Figure 2), *F*(1,20) = 37.792, *p* < .001.

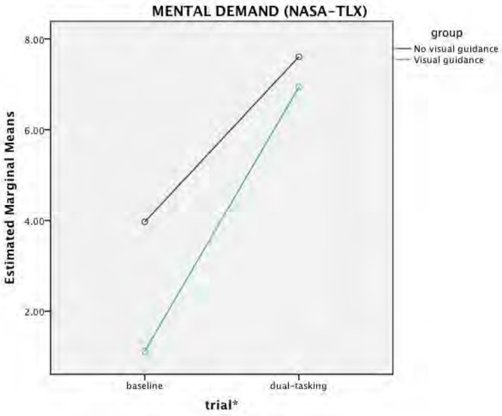

Figure 2: Subjectively reported mental workload (NASA-TLX, M, max 20) in Experiment 1 by trial and group

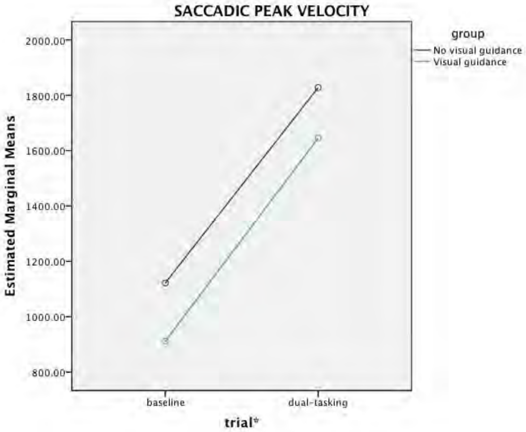

Figure 3: Saccadic peak velocities (M) by trial and group in Experiment 1

Saccadic peak velocities indicated also a significant effect of trial, *F*(1,19) = 38.154, *p* < .001, but the direction of the effect was the opposite than expected (Figure 3). According to theory [2], higher mental workload should be associated with lower saccadic peak velocities. PRC indicated also a significant effect of trial (Figure 4), *F*(1,22) = 11.158, *p* = .003. Again, PRCs were lower for the dual-tasking condition, indicating lower mental workload than in baseline driving, the opposite result to that of NASA-TLX.

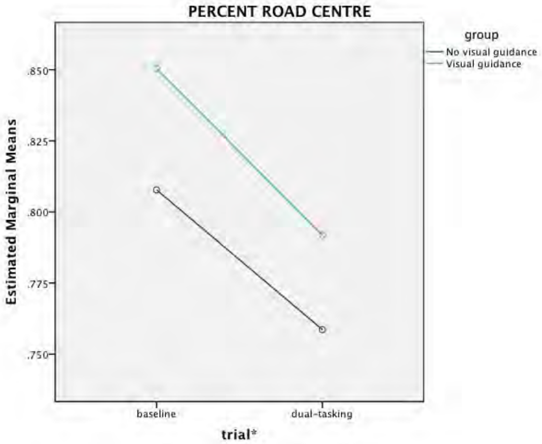

Figure 4: Percent road centre (M) by trial and group in Experiment 1

Pupil diameter did not indicate significant effects but the difference between baseline driving and dual-tasking approached significance (Figure 5), *F*(1,21) = 4.083, *p* = .056, again for the favor of dual-tasking (i.e. lower mental workload).

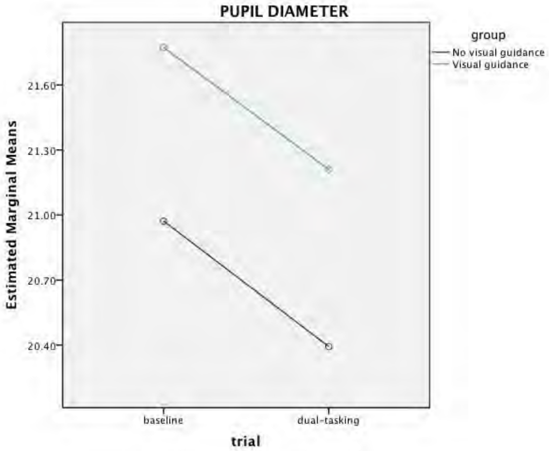

Figure 5: Pupil diameter (M) by trial and group in Experiment 1

3. EXPERIMENT 2 – VISUAL-MANUAL IN-CAR TASKS

In Experiment 2, two special Car Mode User Interfaces (UIs) providing access to a variety of smart phone applications were compared for their possible distraction effects. The differences between the UIs are not the essence here and will not be discussed.[2]

3.1 Method

Experiment 2 had almost identical setup as Experiment 1 with the exception that a smart phone placed at a dashboard holder was used for the in-car tasks. This time the in-car tasks required also much more visual-manual interaction.

The sample consisted of 20 volunteer university students (10 male, 10 female) with sufficient driving experience and normal or corrected vision. They were divided into two groups of 10, corresponding to the user interfaces UI1 and UI2. As such, the experimental design was mixed factorial with 2 x 2 factors (group [UI1 vs. UI2]) x trial [baseline driving vs. dual-tasking]). A participant completed 5 different in-car tasks while driving in the dual-tasking condition.

3.2 Results

None of the metrics reported here indicated any significant between-subject effects.

The mental demand metric of NASA-TLX revealed a significant effect of trial (Figure 6), *F*(1,18) = 43.891, $p < .001$. Similarly to Experiment 1, the participants reported the mental demand in the dual-tasking condition higher than in the baseline driving.

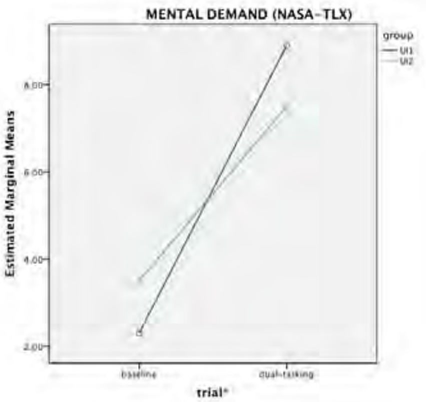

Figure 6: Subjectively reported mental workload (NASA-TLX, M, max 20) in Experiment 2 by trial and group

The saccadic peak velocities indicated no significant effect of trial but the difference between baseline and dual-tasking trials approached significance (Figure 7), *F*(1,18) = 3.261, *p* = .088. It seemed that the baseline driving would have had again the higher levels of mental workload.

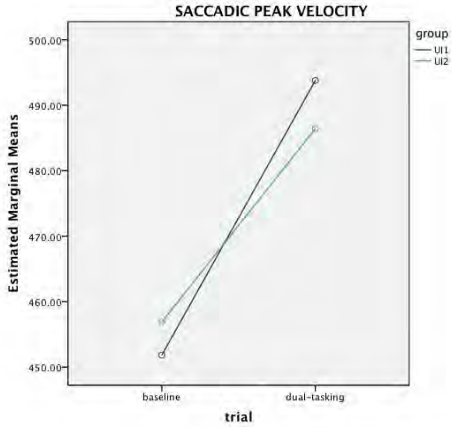

Figure 7: Saccadic peak velocities (M) by trial and group in Experiment 2

PRCs indicated significant effect of trial (Figure 8), *F*(1,18) = 6.735, *p* = .018. Again, dual-tasking seemed to lead to lower levels of mental workload. Also pupil diameters indicated significant effect of trial (Figure 9), *F*(1,18) = 11.394, *p* = .003. Pupils were again less dilated in the dual-tasking condition.

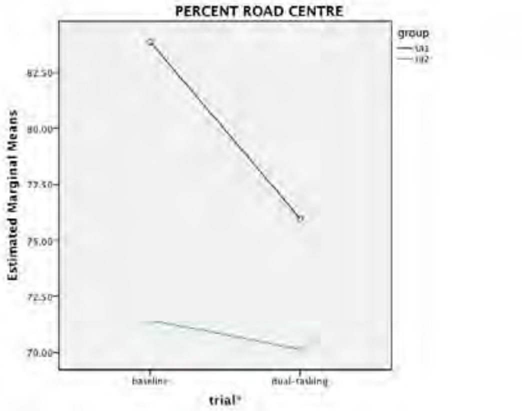

Figure 8: Percent road centre (M) by trial and group in Experiment 2

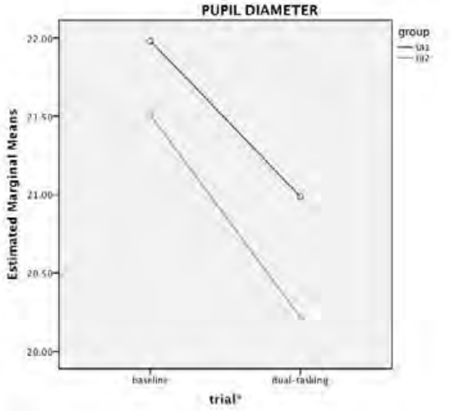

Figure 9: Pupil diameter (M) by trial and group in Experiment 2

---

[2] Again, confidentiality issues prevent descriptions on further details of the user interfaces.

4. DISCUSSION

In both experiments, the reported mental demand on NASA-TLX [4] indicated that the participants felt the dual-tasking condition significantly more demanding than the baseline condition. The reported average levels of mental demand were quite low in both experiments (max 20) but nevertheless the effect of dual-tasking was significant. Saccadic peak velocities indicated also significant effect of dual-tasking. However, the direction of the effect was the opposite than for NASA-TLX: the saccadic peak velocities were slower in the baseline condition, indicating greater mental workload compared to dual-tasking according to the theory behind the metric [3]. The metric of PRC indicated similar effects as saccadic peak velocities. Percentage of fixations towards road centre decreased significantly in the dual-tasking conditions. This would not be a surprise if we had analyzed all the eye-movements, including the fixations towards the in-car display. In that case, with increasing in-car visual demands the percentage is actually expected to decrease [4], but then we would not be measuring mental workload but visual load of the secondary task. However, we included into the analyses only the gazes towards the driving scene. Also pupil diameter indicated similar effects than the other visual measures of mental workload, to opposite direction than NASA-TLX. Pupil diameters seemed to decrease significantly in the dual-tasking conditions compared to baseline driving. Again, this was unexpected finding because pupil diameters should increase with higher mental workload [2].

There are a couple of features in the current experimental designs that could explain the unexpected results. The simplest explanation is that the participants actually had to put more mental effort to the baseline driving than in dual-tasking but were not able to report this. For some reason, they actually reported the opposite. However, this is an unlikely explanation because 1 (driving task) + 1 (in-car tasks) is typically more than 1 (driving task). NASA-TLX as well as other performance data not reported here seems to indicate this was the case also here. Technical problems with the eye-tracking system could provide another explanation but the calibrations were done carefully and it does not seem likely that there would have been significant effects with three different metrics even if there were some systematic technical error in the measurement.

Instead, it is possible that the interplay of eye-tracking only the gazes at the driving scene and the visual demands of the in-car tasks caused the observed effects, but because of different reasons for the different metrics. For the saccadic peak velocities, one can speculate that the saccades to and from the in-car display are included in the analysis for the dual-tasking condition, and because the participants tried to minimize the eyes-off-road time, the saccadic peak velocities are very high for these saccades. For the PRC metric the explanation could be the HUD speedometer located outside of the road centre but still on the driving scene. It is possible the participants had to make more glances at the speedometer in the dual-tasking condition to check the speed after each in-car glance. The larger drop for the UI1 than UI2 visible in Figure 7 could provide evidence for this explanation because of the greater visual demands of UI1 (not reported here), even if the difference did not become significant with these sample sizes (n=10). The more glances to the in-car display, the more inspection of the speedometer. The larger size of the average pupil diameters in the baseline driving could be explained, not by increased mental workload, but instead by the increase in illumination and/or closer viewing distance when fixating at the in-car display. Even if the in-car glances were not included in the analyses, these factors can reduce pupil size, and after fixating back at the driving scene it could take some time for the pupils to adjust. In other words, the dual-tasking data could include fixations with decreased pupil size due to delays in adjustment for changed brightness or viewing distance.

In all the three highly speculative explanations suggested above, the main source of measurement error is the visual demand of the in-car tasks and the corresponding eye-movements between the driving scene and the in-car display (as well as the HUD speedometer). Overall, the results suggest that the visual metrics indicated effects of dual-tasking likewise the subjective ratings but the direction of the effects were opposite. It is suggested that the proposed visual metrics of mental workload might reflect in this case the influence of the visual demands of the secondary task on an in-car display instead of mental workload. Thus, we suggest that the effects of visual secondary tasks should be carefully considered before making assumptions on mental workload when using visual measures of cognitive workload in multitasking experiments with visual in-car user interfaces.

5. ACKNOWLEDGMENTS

Our thanks to the OPTIMUM (Optimal Multitasking with Ubiquitous Media) research project funded by TEKES (Finnish Funding Agency for Technology and Innovation).

6. REFERENCES

[1] Ahlström, C., Kircher, K., and Kircher, A. (2009). Considerations when calculating percent road centre from eye movement data in driver distraction monitoring. *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 132-139.

[2] Beatty, J. (1982). Task Evoked Pupillary Responses, Processing Load and Structure of Processing Resources. *Psychological Bulletin, 91*, 276-292.

[3] Di Stasi, L.L., Marchitto, M., Adoracion, A., Baccino, T., and Canãs, J.J. (2010). Approximation of on-line mental workload index in ATC simulated multitasks. *Journal of Air Transport Management, 16*, 330-333.

[4] Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX: Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183), Amsterdam: Elsevier.

[5] Palinko, O., Kun, A.L., Shyrokov, A., and Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, 141-144.

[6] Victor, T.W., Harbluk, J.L., and Engström, J.A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F, 8,* 167-190.