Cite as: Petzoldt, T., Bellem, H., & Krems, J.F. (2014). The Critical Tracking Task: A potentially useful method to assess driver distraction? *Human Factors, 56*(4), 784-803. doi:10.1177/0018720813501864

The critical tracking task - A potentially useful method to assess driver distraction?

Tibor Petzoldt*, Hanna Bellem and Josef F. Krems

Chemnitz University of Technology, Chemnitz, Germany

E-mail: [tibor.petzoldt, hanna.bellem, josef.krems]@psychologie.tu-chemnitz.de

Tel: +49 (0) 371 / 531 36519; Fax: +49 (0) 371 / 531 836519

ABSTRACT

Objective: We report on four experiments that investigated the critical tracking task's (CTT) potential as a tool to measure distraction.

Background: Assessment of the potential of new in-vehicle information systems to be distracting has become an important issue. An easy-to-use method, which might be a candidate to assess this distraction, is the critical tracking task (CTT). The CTT requires an operator to stabilise a bar, which is displayed on a computer screen, such that it does not depart from a predefined target position. As the CTT reflects various basic aspects of the operational level of the driving task, we used it as a simple surrogate for driving to assess the CTT's capabilities.

Methods: We employed secondary tasks of varying demand, artificial tasks as well as tasks representative of secondary tasks while driving, and asked participants to perform them together with the CTT in parallel. CTT performance, secondary task performance and subjective ratings of load were recorded and analysed.

Results: Overall, the CTT was able to differentiate between different levels of demand elicited by the secondary tasks. The results obtained corresponded with our a-priori assumptions about the respective secondary tasks' potential to distract.

Conclusion: It appears that the CTT can be used to assess in-vehicle information systems with regard to their potential to distract drivers. Additional experiments are necessary to further clarify the relationship between driving and CTT performance.

Application: The CTT can provide a cost-effective solution as part of a battery of tests for early testing of new in-vehicle devices.

Keywords: *Inattention; In-vehicle information systems; Surrogate reference task; Evaluation methods*

INTRODUCTION

Driver distraction, with its causes as well as its effects, has received more and more attention in recent years. Webster's dictionary selected "distracted driving" as its word of the year for 2009 ("Distracted-driving campaign", 2010). The US Transportation Secretary announced an "administration wide effort to combat distracted driving" (US Department of Transportation, 2009). The term has risen to prominence for good reasons. Driver distraction has been shown to have detrimental effects on a variety of driving-related variables (see Regan, Lee, & Young, 2008 for an overview), and appears to be a relevant factor in traffic crashes (Kuratorium für Verkehrssicherheit, 2008; New Zealand Ministry of Transport, 2010; Stutts, Reinfurt, Staplin, & Rodgman, 2001). In the US, driver distraction was reported to have been involved in 16% of all fatal crashes in 2008 according to data from the Fatality Analysis Reporting System (FARS), and an estimated 21% of injury crashes were reported to have involved distracted driving, according to data from the General Estimates System (GES; Ascone, 2009). "Technology-based distraction" (Young, Regan, & Hammer, 2003) has become one of the major issues, as in-vehicle information systems (IVIS) and other related devices have become increasingly popular (Buettner, 2009; Starry, 2001). Pickrell and Ye (2011) used observational data to estimate that 660,000 vehicles were driven by people using hand-held cell phones at a typical moment, during daylight in the US in 2010. In addition, they report that about 0.9% of drivers were text-messaging or visibly manipulating other hand-held devices.

Given the growing distribution of in-vehicle devices, it is vital to assess the distraction-related demands of each of these systems. Large field operational tests (e.g. Karlsson et al., 2009, Sanchez et al., 2012), which are "[...] undertaken to evaluate a function, or functions, under normal operating conditions in environments typically encountered by the participants [...]" (FESTA consortium, 2011; p. 1) are able to contribute to this assessment with a high ecological validity. However, the cost associated with these tests is immense, which makes them impossible to implement on a regular basis. Such projects can help gather information about potential effects of these systems, but they are limited in use to differentiating between specific systems, brands, or functions, as they cannot establish actual causality. Studies in a driving simulator are somewhat more suitable in this regard, but still, the efforts that must be invested are rather high. Therefore, researchers have come up with a number of easy-to-use methods to coarsely measure the distraction elicited by IVIS. The occlusion method (Senders, Kristofferson, Levison, Dietrich, & Ward, 1967; see also Baumann, Keinath, Krems, & Bengler, 2004; Keinath, Baumann, Gelau, Bengler, & Krems, 2001) targets visual distraction, especially focusing on (non-)interruptability of secondary tasks. This method has become an ISO-standardised procedure (ISO 16673, 2007). As the ISO standard notes, however, its primary purpose is to allow for statements about time spent with eyes-off-the-road. The method itself has no motor component that would allow for further assessment of potential driving performance decrements.

Also, as testing is typically carried out by having participants orient themselves to the task, occlusion does not provide a measure of how the position of a secondary task display might affect performance. SAE 2364 (Society of Automotive Engineers, 2004) picked up the method (calling it "Interrupted Vision Method"), and developed compliance criteria that have to be met by a navigation function that shall be accessible while the vehicle is moving. The guideline also introduced the so called 15 second rule, which states that "Any navigation function that is accessible by the driver while a vehicle is in motion shall have a static total task time of less than 15 seconds." (Society of Automotive Engineers, 2004, p. 7). Similar guidelines have been put forward by the Alliance of Automobile Manufacturers (2006) and the National Highway Traffic Safety Administration (2012). The peripheral detection task (PDT; Jahn, Oehme, Krems, & Gelau, 2005) tries to assess cognitive and visual distraction by making use of the fact that visual and cognitive load can narrow the driver's functional field of view (Miura, 1986). However, while the task itself is fairly simple, it is usually employed as an additional task in more elaborate simulated (Martens & van Winsum, 1999; Törnros & Bolling, 2006) or real-world (Patten, Kircher, Östlund, & Nilsson, 2004; Patten, Kircher, Östlund, Nilsson, & Svenson, 2006) driving settings. Another dual task method, the lane change test/task (LCT; Mattes, 2003), has also been developed to address the issue of visual distraction. Here, the task is to control a vehicle on a three-lane road at a constant speed, while repeatedly performing lane changes as instructed by signs at the side of the road. The degradation of driving performance when operating an additional secondary task serves as a measure of distraction (Mattes & Hallén, 2009). Unfortunately, while unevenly spaced, the lane changes are rather predictable, as the signs are visible throughout the drive. It can be argued that this is a problem especially with self paced secondary tasks, as this allows for strategies of attention allocation that might distort the assessment of the secondary task's true distraction potential.

A task that, in the context of driver distraction, only has been used as an IVIS surrogate is the critical tracking task (CTT; Jex, McDonnell, & Phatak, 1966). In this function, the task has been subject to some criticism, as the continuous monitoring and input required are rather uncommon in most IVIS (e.g. Petzoldt, Bär, Ihle, & Krems, 2011). However, it appeared to us that the task's characteristics bear some similarity to central aspects of the driving task (see section "The CTT and driving"). Whether the CTT holds potential as a method to measure the demands imposed by distraction of in-vehicle tasks has to be investigated.

The CTT

The CTT was first described by Jex et al. (1966). In their rather technical account, they explained the CTT as a task "in which a human operator is required to stabilize an increasingly unstable first-order controlled element up to the critical point of loss of control" (Jex et al., 1966, p. 138). Others have likened the task to “balancing a stick on one’s fingertip, with the stick’s length decreasing with time” (Burns & Moskowitz, 1980, p. 261). Basically, most of the CTT’s formulations use some image of a virtual bar on a computer screen that tends to depart from a predefined target position. Participants use joystick or keys to bring the bar back to the target position; a task that becomes increasingly difficult over time. The level of instability is represented by a λ-value (with $\lambda = 1/T$, T being the divergent time constant measured in *s*; for details on CTT dynamics see e.g. Jex et al., 1966), which increases constantly. The value at which control is lost is recorded as a performance measure. The task has been found to correlate substantially with subjective ratings of workload (Rehman, Stein, & Rosenberg, 1983; Rosenberg, Rehman, & Stein, 1982), allowing for the assumption that the CTT is a valid method to assess demand on a general level.

The CTT has been frequently used in research on various kinds of cognitive and psychomotor impairment. Most of these studies have been concerned with medication-, drug- and alcohol-related performance decrements (Barnett, Licko, & Thompson, 1985; Burns & Moskowitz, 1980; Klein & Jex, 1975; Ramaekers, Uiterwijk, & O’Hanlon, 1992; Ramaekers et al., 2006a; Ramaekers et al, 2006b; Ramaekers, Muntjewerff, van Veggel, Uiterwijk, & O’Hanlon, 1998). Elmenhorst et al. (2009) assessed the impairment caused by sleep deprivation. Damos et al. (1984) even studied the effects of “exotic” environments (e.g. environments with low frequency motions which cause seasickness) on human performance, concluding that the CTT might be a useful method in that regard.

The use of the task is, however, not limited to the single task setup. Jex (1967) explained the CTT’s use as a primary task in conjunction with other secondary tasks to assess the secondary task’s workload. He argued that higher workload levels should result in earlier loss of control, represented by a smaller critical λ. In addition, he also proposed the option of employing the CTT as a secondary task. The instability of the task should be set at a constant subcritical level, demanding frequent, but not continuous attention. With this setup, he argued, the CTT might serve as a source of continuous workload that is reflected by the performance on any primary task. Following this approach, Burke, Gilson and Jagacinski (1980) studied the parallel use of two CTTs, as well as a two-dimensional implementation of the task, i.e. one that is unstable in two directions. In a similar account, Derrick (1988) tried to assess the validity of different measures of workload by requiring participants to operate various tasks (among them the CTT) in parallel. In a study reported by Wickens and Kessel (1980), the CTT was used as a source of distraction. It was found that the detection of dynamic system failures was impaired by the secondary task. Wickens, Braune and Stokes (1987) assessed age differences in the speed and capacity of information processing by using the CTT in conjunction with different versions of the Sternberg memory search task. It occurred that processing speed decreased with age, whereas time-sharing abilities were unaffected.

Although Jex (1967) clearly differentiated between primary and secondary tasks, this practice is not always evident in the measurement of driver distraction (still, for simplicity, in this paper we use the terms “primary task” for the CTT and “secondary task” for any concurrent task). For the LCT, for example, it is explicitly required that no instructions shall be offered to participants on how to prioritize between primary and secondary tasks. The same approach can be used with the CTT. Employing the “constant subcritical instability” setup, and using the bar’s deviation from the target position as a performance measure, it is possible to achieve a continuous measurement of load. Wickens and Kessel (1980, see also Wickens et al., 1987) reported on the calculation of a tracking error as performance measure. Such a metric appears to have an advantage over the previously described critical λ. When using the critical λ as performance criterion, a single trial lasts as long as the participant has control over the task, and ends once he loses it. This may vary dramatically, resulting in differences in task duration and number of secondary task trials. As the secondary task is usually not a continuous one, this may give rise to a variety of problems. With the continuous measurement, on the other hand, a stop criterion can be set through the secondary task (i.e. operation of both tasks until the secondary task is finished) as well as the CTT (i.e. operation of both tasks for a predefined duration). Thus, the length of trials is more independent from participants’ individual skills in controlling the CTT (noting the fact that secondary task duration may also depend, at least partially, on individual skills).

The CTT and driving

While research supports the assumption that the CTT is suitable to assess the load imposed by a secondary task in general, its use as a surrogate for the driving task is, at first glance, probably less obvious. However, theoretical models of driving suggest that the CTT shares various characteristics of the driving task. More specifically, the task is closely related to what Michon (1979) called the operational level of driving, which mainly covers the basic skills of steering, accelerating and braking. With regard to steering, Godthelp (1986) noted that “most of the available steering-control models are based on the fundamental assumption that the automobile driver acts as an error-correcting mechanism with permanent attention allocated to the steering task” (p. 211). Although he went on to argue that attention allocation does not have to be permanent to drive safely, and that a certain margin of error is regularly accepted by the driver (Godthelp, 1984), the more general assumption that steering is some form of error correction that requires a substantial amount of attention is still valid. Longitudinal aspects of the driving task, like car following, can broadly be described in a similar way. Many models in traffic simulation assume that drivers aim for a desired time headway or time distance separation from the lead vehicle (Short, Pont, & Huang, 2004), resulting in velocity adjustments when the lead vehicle changes speed. Accordingly, controlling the CTT might be likened to keeping a safe distance to a vehicle in front that repeatedly accelerates and decelerates. Following these examples, the CTT is a direct reflection of the driving task at the operational level; a view that is supported by O'Donnell, Moise and Schmidt (2005), who argued that the successful operation of the CTT especially requires focussed attention and visual motor control. Just like lateral and longitudinal control in driving, the CTT requires the user to continuously correct an error, or, in Jex et al.'s (1966) terminology, the stabilisation of an unstable element, which can only be achieved through the regular devotion of a sufficient level of attention towards the task. In accordance, the operational level of driving has been shown to be sensitive to driver distraction (e.g. Angell et al., 2006; Carsten et al., 2005). Hurwitz and Wheatley (2002) reported on a study in which drivers had to perform either an auditory or a visual secondary task while driving. Although both secondary tasks only required monitoring, the authors found deteriorated lane keeping performance when performing the visual secondary task. In an analysis of naturalistic driving data, Peng, Boyle and Hallmark (2013) found that inattention with eyes-off-road while engaged in a secondary task lead to an increase in the standard deviation of lane position. Longitudinal control is affected as well. Jamson and Merat (2005) reported detrimental effects of concurrent auditory and visual secondary tasks on time-to-contact, brake reaction time and minimum time and distance headway in a simulated environment. In an on-road study, Lamble, Laakso and Summala (1999) found increased detection thresholds in car following situations for visually demanding in-car displays. Depending on eccentricity of the display, TTC decreased from 8s down to 4s. Similarly, as the CTT requires the regular devotion of sufficient attention to be operated properly, it can be assumed that the withdrawal especially of visual attention has profound effects on CTT performance, which should manifest themselves in higher deviations from the centre position.

Overall, the findings suggest that the CTT, as a single task, places demands on the operator that are fairly similar to various aspects of driving. However, evidence supporting the claim that the task is a promising candidate to assess driver distraction is still missing. The four experiments described in this paper aim to provide deeper insights into the advantages and limitations of this method. The goal of Experiments I and II was to investigate the CTT's potential as a method to assess driver distraction, using valid, but simple secondary tasks. If such evidence was found, then a more differentiated approach (i.e. artificial tasks tapping into specific aspects of the CTT), as well as a more applied approach (real in-vehicle information systems), could be pursued in later experiments to verify results. Thus, Experiment III followed this more differentiated approach, employing standardised artificial secondary tasks previously described in studies that assessed the LCT. Finally, the goal of Experiment IV was the practical application of the CTT by comparing the amount of distraction caused by different navigation system tasks.

EXPERIMENT I – LEGIBILITY

In a first experiment, we sought to find general evidence that the CTT could serve as a method to assess distraction. To this end, it was necessary to find a secondary task that could claim face validity for being representative of in-vehicle tasks, but at the same time being simple to employ and manipulate.

A major contributor to distraction caused by IVIS is the legibility of information. The European statement of principles on human–machine interface (Commission of the European Communities [CEC], 2008) refers to legibility as an important design aspect of IVIS, pointing to international standards. Specifically, the ISO standard (ISO 15008, 2009) lists numerous features of information systems that have to be considered when developing such systems, including different aspects of legibility. Stevens, Quimby, Board, Kersloot and Burns (2002) regarded brightness, contrast, resolution, character/character spacing, font and case as relevant variables for display legibility. For navigation systems, character size, combinations of colours, background luminance, map orientation and amount of information have been identified as crucial factors (Kimura, Marunaka, & Sugiura, 1997). Pauzie (2002) reported on the influence that the size of displayed information has on glance frequency and duration. A small character size appears to force especially older users to look at the screen significantly longer in order to successfully read the text. This age-related difference disappears with larger characters. Text–background colour combinations affect legibility, but also whether a display has a pleasant appearance or not (Greco, Stucchi, Zavagno, & Marino, 2008). For our experiment, we chose to use character size and character–background colour combination as factors. If the CTT is useful for the assessment of driver distraction, the differences in legibility of the information should be reflected in CTT performance. More specifically, we hypothesised that larger characters and higher character–background contrast lead to better CTT performance than smaller characters and lower contrast.

Method

Participants. Twenty-four students from Chemnitz University of Technology took part in this experiment. Eighteen participants were female and 6 were male, with a mean age of 22.6 years (*SD* = 4.8), and all possessed a valid driving licence. Each participant completed all experimental conditions.

Material.

CTT. The CTT used in this experiment (and all follow-ups; Wagner & Weir, 2006) employed the constant subcritical instability setup described previously. The level of instability was set at a medium level of difficulty (as found in pre-tests, λ = 1.5, i.e. rate of divergence 1.5 rad/s or divergent time constant of 0.67 s). This level of difficulty ensured that the CTT was sufficiently demanding to be sensitive to smaller differences in distraction, while simultaneously ensuring that every participant could complete the task. Our implementation used a horizontal bar that continuously left its target position at the centre of the screen (Figure 1). It was presented on a 19” screen, positioned centred in front of the participants (Figure 2). The task was located at the centre of the screen and covered approximately 50% (215 mm × 200 mm) of it. Positions at the boundaries equalled deviation values of ±100 mm, the centre position was defined as 0. Participants were able to control the bar by pressing the up and down keys on the keyboard, which gave discrete and incremental commands to the target, i.e. the down arrow key pulled the target line down an incremental amount, and vice versa. When the bar’s deviation from the centre became too large (values > 40 mm or < –40 mm), it changed its colour from black to red, indicating the need for immediate action. Such a position was reached when not attending to the task for 4 s (assuming that the original position was at the centre of the screen). The position of the bar relative to the centre was recorded, and, based on these data, the mean absolute deviation was computed as a performance metric.

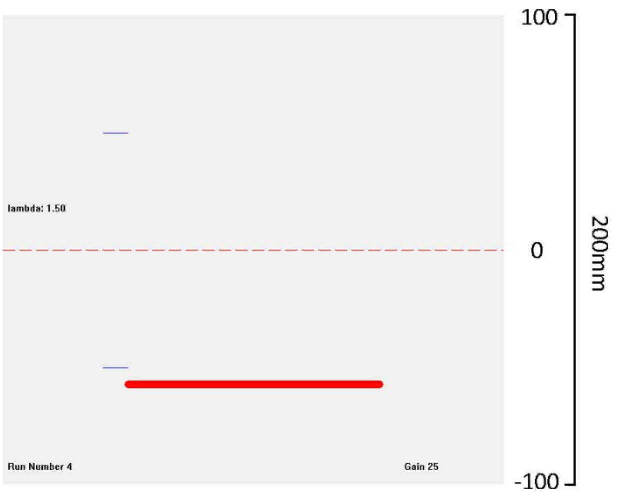

Figure 1. Critical tracking task (CTT); example screen.

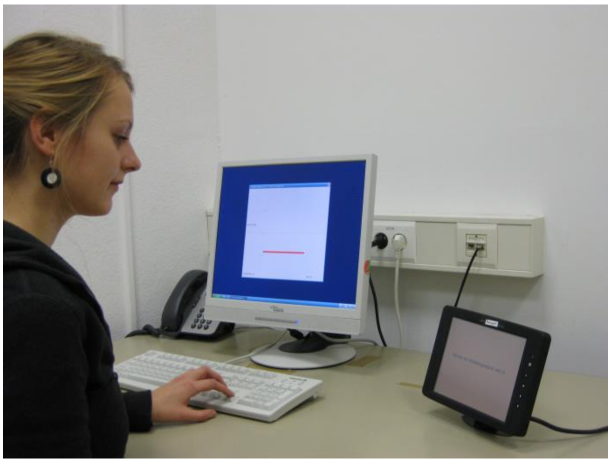

Figure 2. Experimental setup.

Reading task. The reading task employed short questions that were read and answered by participants. The questions were very easy (e.g. “Name a country beginning with D” – with “Deutschland” being an obvious answer for German participants), as their main purpose was to assure that participants read the presented information. Questions were screened in a pretest to eliminate those that participants considered too difficult. The structure of questions always followed the same pattern – “Name [something] beginning with [letter]”, to ensure comparability of reading times between items. Four conditions were generated from different foreground-background colour contrasts (blue and yellow vs. red and purple; contrast ratio 17:1 vs. 1.8:1) and different character sizes (5 mm vs. 2 mm; visual angle approx. 0.57° vs. 0.23°): (i) high foreground-background colour contrast with large characters, (ii) high foreground-background colour contrast with small characters, (iii) low foreground-background colour contrast with large characters and (iv) low foreground-background colour contrast with small characters. Colour combinations were chosen based on the work of Shie and Lin (2000), who reported that blue text on a yellow background went with the highest rate of identification and the highest preference ratings (with preference including aspects like clearness and visual comfort), whereas red text on a purple background was identified significantly less often, and was rated second to least preferable. The number of questions read was recorded as a reading performance metric. Questions were presented on an 8.37″ screen that was placed to the right of participants, where an aftermarket navigation system would have typically been positioned. The presentation was controlled by the experimenter and participants were required to answer questions as quickly as possible. As soon as a question was answered, the next one was displayed.

Procedure. First, participants received instructions on the CTT and reading tasks, followed by a brief CTT training session. After recording a baseline trial with the CTT as a single task, participants had to perform the CTT and the reading task simultaneously in four trials of 3 min duration each. Following the LCT ISO standard (ISO 26022, 2010), participants were told to perform the tasks with equal emphasis on their performance for both (similar for all follow-ups). The order of the four different legibility conditions was balanced over participants. The experiment was completed in about 25 min.

Results

In Figure 3, CTT standard deviation values are displayed for the five conditions. As can be seen, all four dual-task conditions produced higher deviations than the baseline. A two-factor repeated measures ANOVA revealed that character size had a significant impact, $F(1, 23) = 32.87$, $p < .001$, $\eta_p^2 = .59$, with smaller characters causing higher deviations. There was no significant main effect for colour contrast, $F(1, 23) = 1.33$, $p = .260$, $\eta_p^2 = .06$, and no interaction, $F(1, 23) = .56$, $p = .463$, $\eta_p^2 = .02$. For completeness, we also compared each experimental condition to the baseline with a paired t-test (Bonferroni-corrected for multiple comparisons), resulting in highly significant differences for all four comparisons (each $p < .001$).

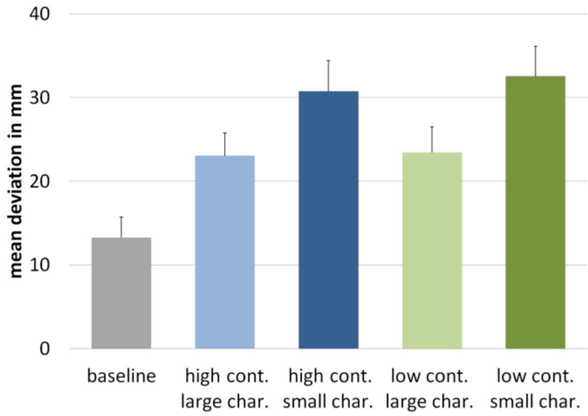

Figure 3. Mean deviation from the centre position.

Secondary task performance was measured as the number of questions read during the 3 min run. Performance was better for the large characters, with 58.6 questions read (*SE* = 1.7) in the high contrast condition and 52.8 (*SE* = 1.8) with low contrast, compared to 49.3 (*SE* = 1.8) and 47.8 (*SE* = 2.0) when characters where small. A two-factor repeated measures ANOVA confirmed that similar to the CTT, character size had a significant impact, $F(1, 23) = 46.27$, $p < .001$, $\eta_p^2 = .67$. Colour contrast had a significant effect as well, $F(1, 23) = 36.37$, $p < .001$, $\eta_p^2 = .61$. There also was a significant interaction between character size and colour contrast, $F(1, 23) = 5.45$, $p = .029$, $\eta_p^2 = .19$. As the values show, colour contrast appeared to have an influence on performance mainly for the large character condition.

Since the 3 min duration of a single trial was chosen rather arbitrarily (following the LCT procedure; Mattes, 2003), we also tried to assess whether a shorter duration (e.g. half of the original 3 min) might produce the same results. To visualise the performance over time, Figure 4 shows CTT performance segmented in ten 18s portions. As can be seen for the dual task conditions, performance decreased slowly but steadily over time. We also calculated CTT standard deviation values for the first and second half (= 90 s each) separately. For all dual task conditions, performance in the second half was worse than in the first (differences in mean values between first and second half from -0.9 to -4.6 mm), whereas baseline performance slightly increased (difference of 1.8 mm). When comparing the two segments and the overall performance, we found highly significant correlations between all three scores: first and second half: $r = .94$; first half and overall: $r = .98$; second half and overall: $r = .99$.

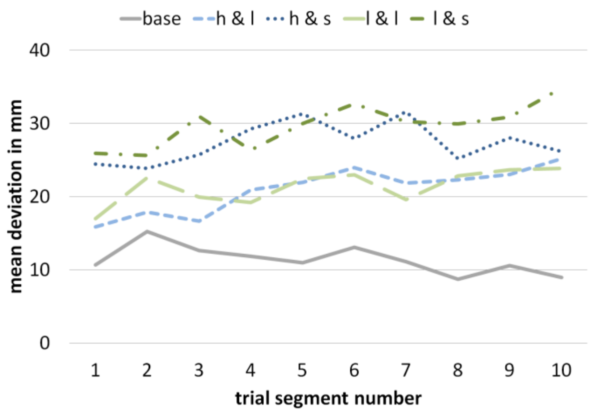

Figure 4. Mean deviation from the centre position over time. Each trial segment covers 18 s. (Conditions: base = baseline, h & l = high contrast & large characters, h & s = high contrast & small characters, l & l = low contrast & large characters, l & s = low contrast & small characters)

Discussion

Our first experiment provides first support for the assumption that the CTT might be a useful method to assess driver distraction evoked by IVIS. The task was able to reliably differentiate between small and large characters. However, it failed to directly capture the expected effect of colour combination. Statements of the participants after the experiment suggest that they did not experience substantial differences between the two colour conditions during the trials. However, secondary task results imply that the different colour combinations did have an effect on secondary task performance. Just as several authors have (e.g. Burns, Harbluk, Foley, & Angell, 2010; Harbluk, Mitroi, & Burns, 2009) argued, the inclusion of secondary task data can be vital to assess the complete picture of driver distraction.

Participant's statements after completion of the experiment provided further valuable information. Many complained about the length of the trials, arguing that the dual-task situation was rather demanding. Especially at the end of trials, attention and concentration levels decreased. These subjective reports are supported by the data. The 3 min trial duration was chosen following the LCT procedure (Mattes, 2003). However, the LCT requires slightly different subtasks (lane keeping, lane change from middle to left lane, from middle to right lane, etc.) to be completed within the task, so this duration is necessary to have a balanced measurement. In the context of the CTT, task duration is a result of a more or less arbitrary decision, as there are no subtasks involved. Our analysis shows that CTT performance in the dual task conditions declined steadily over time, whereas the difference between the conditions stayed more or less the same. Baseline performance (which was only included for completeness) increased slightly, though not substantially. Given the fact that the 90 s time segments were highly correlated with overall performance, a shorter duration of trials appears plausible, provided there is sufficient training on the CTT before measuring the baseline. For Experiment III, we took this into account.

EXPERIMENT II – DISPLAY POSITION

In our second experiment, we wanted to complement Experiment I by gathering further evidence for the CTT's capability to reflect distraction-level differences in aspects relevant to human–machine interface (HMI) design. Once again we referred to the European statement of principles (CEC, 2008) to identify display position as another important factor. As one of its installation principles, the document states that "Visual displays should be positioned as close as practicable to the driver's normal line of sight." (p. 11). An explanation has been given by Stevens et al. (2002), who pointed out that "visual displays positioned close to the driver's normal line of sight reduce the total eyes-off-the-road time relative to those that are positioned further away" (p. 23). Since driving is primarily a visual task, it appears obvious that eyes-off-the-road time is directly linked to crashes (Green, 1999). As a potential tool for assessing driver distraction, the CTT should be able to distinguish between displays that are placed in different positions with regard to the line of sight. We hypothesised that a secondary task display located closer to the CTT display will result in lower CTT deviation values than a display further away.

Method

Participants. Twenty students (all licensed drivers) of Chemnitz University of Technology took part in this experiment; 15 female and 5 male students, with a mean age of 22.0 years (SD = 2.1). None of the participants had taken part in the previous experiment. Each participant completed all available experimental conditions.

Material.

CTT. The CTT employed in this experiment, as well as the overall setup (screen size and position, operation of task, etc.), were identical to Experiment I.

Visual task. The task that had to be dealt with on the secondary task screen was the so-called surrogate reference task (SuRT; Mattes & Hallén, 2009), which required participants to scan stimulus displays for the one stimulus that differed from others surrounding it. Target and distracters were white circles in front of a black background (Figure 5). Participants responded by moving a vertical grey indicator bar to the position of the identified target and pressing the enter key for confirmation, followed by the next display. For this experiment, we used a rather difficult version of the task. Distracters were 8 mm in diameter (visual angle approx. 0.92°), and the target was 9 mm (visual angle approx. 1.03°). The task was presented on an 8.37″ screen to the right of the participant. The indicator bar was controlled by using a standard keyboard. The position of the screen was varied between the two experimental conditions: in the “close” condition, the screen was set up directly to the right of the CTT screen (approx. 30° angle between the centre of the CTT screen and the centre of the SuRT screen), and in the “distant” condition, it was moved about 60 cm further to the right (approx. 75° angle).

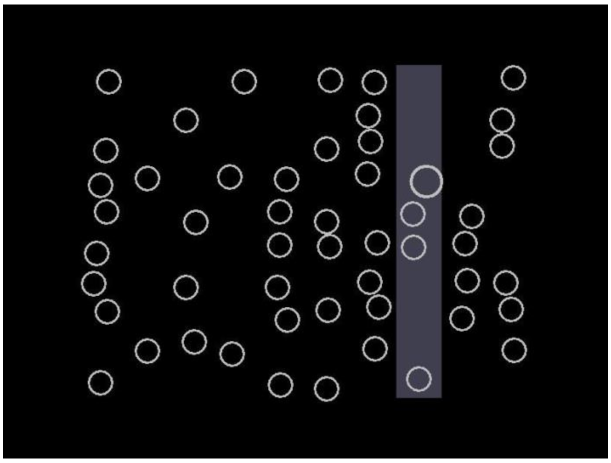

Figure 5. Surrogate reference task (SuRT); example screen.

Procedure. Participants received instructions on the CTT and the SuRT, followed by a short CTT training. A baseline trial with the CTT as a single task was recorded. Two trials (one for each condition) of 3 min duration each followed, in which participants operated the CTT and the SuRT in parallel. The order of the two different conditions was balanced over subjects. Overall, the experiment took less than 20 min.

Results

In Figure 6, CTT standard deviation values are displayed for the baseline and the two experimental conditions. It is obvious that the closer position of the SuRT screen to the CTT screen corresponded to lower deviation values in the CTT than the distant position. A repeated measures ANOVA revealed significant differences between the conditions, $F(2, 38) = 90.556$, $p < .001$, $\eta_p^2 = .83$. More importantly, post-hoc testing (Bonferroni-corrected) confirmed that the close screen position went with significantly better performance than the distant position ($p < .004$). Also, both experimental conditions differed significantly from the baseline (both $p < .001$).

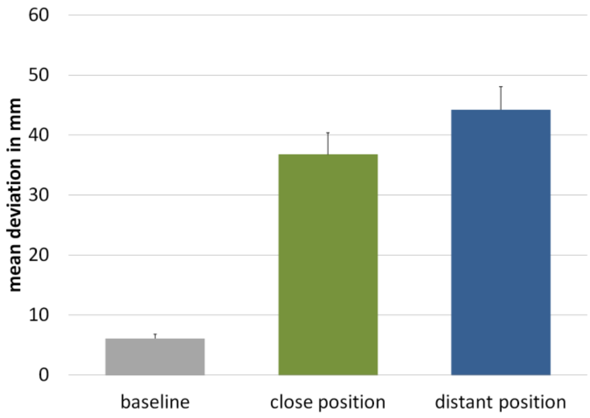

Figure 6. Mean deviation from the centre position.

Secondary task performance was measured as the number of SuRT trials completed. We initially chose this measure over the number of correctly completed trials, as this provides a more realistic picture of how much time was devoted to the secondary task. In any case, the number of mistakes was extremely low, so the two measures did not differ substantially. For the closer position, participants completed 23.3 trials on average (*SE* = 1.6), only slightly more than for the more distant position (*M* = 22.0, *SE* = 1.6). As expected, a t-test for dependent measures was not significant, *t*(19) = .57, *p* = .575, *d* = .13.

As in Experiment I, we assessed whether a short CTT trial duration produced the same results as a full trial. In Figure 7, CTT performance is again displayed segmented in ten 18s portions. After an initial increase, dual task performance appears to be relatively stable over time. There is a minimal increase in baseline performance (difference between first half and second half of trial 0.9mm). We again found highly significant correlations between the first and second half of the measurement and overall performance (first and second half: *r* = .90; first half and overall: *r* = .98; second half and overall: *r* = .97).

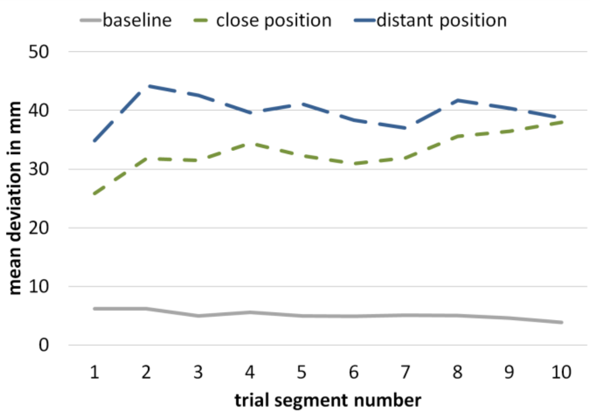

Figure 7. Mean deviation from the centre position over time. Each trial segment covers 18 s.

Discussion

In our second experiment, we found further evidence for the assumption that the CTT is a promising candidate for measuring driver distraction elicited by IVIS. The task could differentiate between different levels of distraction caused by different positions of the SuRT screen. Secondary task performance indicated that participants did not compensate for the distant location of the screen by solving less SuRT trials. Performance decrements were only to be observed for the CTT.

With regard to performance over time, we found a relatively stable operation of the CTT in both the dual task conditions and the baseline. At the same time, the correlations between overall performance and first / second half of the measurement support the assumption that a shorter duration would lead to the same general results. We consider the results of this and the previous experiment sufficient evidence to shorten the length of experimental trials to 90 s in further studies.

EXPERIMENT III – VISUAL VERSUS COGNITIVE DISTRACTION

In our third experiment, we tried to use a more standardised approach to uncover specific capabilities of the task. Visual distraction ("Tasks that require the driver to look away from the roadway to visually obtain information"; Department of Transport, 2010) has been mentioned previously as a critical factor in causing crashes. However, visual distraction is not the only influential form of inattention. Cognitive distraction ("Tasks that are defined as the mental workload associated with a task that involves thinking about something other than the driving task"; Department of Transport, 2010) has been shown to influence visual behaviour and vehicle control (e.g. Harbluk, Noy, Trbovich, & Eizenman, 2007), as well. The assessment of the CTT's ability to distinguish between different levels of visual and cognitive distraction appeared to be the logical next step towards an appropriate judgement of the task's potential. Since the operation of the CTT requires substantial visual attention resources, we predicted that a higher demand in visual attention in a secondary task would be reflected in degraded CTT performance. The CTT's cognitive demand, however, is rather low. Controlling the task is fairly easy, as no complex cognitive processes are involved. Still, handling a dual-task situation, regardless of the respective task's nature, is a cognitively demanding process in itself. Therefore, we hypothesised that variations in cognitive demand in a secondary task would influence CTT performance as well, though most certainly less pronounced than for variations in visual demand. In our approach, we followed Mattes and Hallén (2009), who tried to validate the lane change task (LCT) in a similar way.

Method

Participants. Twenty-four students of Chemnitz University of Technology took part in this experiment. All participants possessed a valid driving licence. Three datasets had to be excluded from the final analysis (one dataset was corrupted, and two participants were statistical outliers on various measures). The mean age of the remaining participants was 21.9 years (*SD* = 3.2), with 16 female and 5 male. None of the participants in Experiment III had taken part in one of the previous experiments. Each participant completed all available experimental conditions.

Material.

CTT. The CTT employed in this experiment, as well as the overall setup (screen size and position, operation of task, etc.), were identical to the previous experiments.

Cognitive task. The cognitive task used in the experiment closely resembled the one that was employed by Mattes and Hallén (2009) for validating the LCT. We defined cognitive easy and cognitive difficult conditions. For the cognitive-easy task, participants had to count forwards in steps of two from 212 on, or in steps of five from 45 on. In the cognitive difficult version, the task was to count backwards in steps of 6 from 831 on, or in steps of 7 from 581 on. Participants were told to keep a constant pace while counting, however without specific instruction on the actual pace.

Visual task. The visual task we employed in the experiment was again the SuRT (see Figure 5). In contrast to Experiment II, we defined two different conditions (hereafter referred to as visual easy and visual difficult conditions) that were identical to the one used by Mattes and Hallén (2009), by varying distracter size and number of indicator bar sections. For the visual easy condition, distracters were 4 mm in diameter (visual angle approx. 0.46°), the target was 8 mm (visual angle approx. 0.92°), with only two indicator bar sections. In the visual difficult condition, distracters were 7 mm in diameter (visual angle approx. 0.80°), the target was again 8 mm (visual angle approx. 0.92°), with eight indicator bar sections. The task was presented on an 8.37″ screen to the right of the participant. The indicator bar was controlled by using a standard keyboard. Position of screen and keyboard matched the requirements of the LCT ISO (ISO 26022, 2010).

Instrument for subjective ratings. To confirm our a-priori assumptions about the cognitive and visual tasks, we administered a questionnaire to obtain subjective assessments of the demands the different tasks represent. We used a scale that we derived from the classical NASA-TLX (Hart & Staveland, 1988), asking participants to use a continuous scale to indicate (i) the mental demand, (ii) the visual demand and (iii) the temporal demand represented by the task, (iv) the effort necessary, and (v) the frustration experienced. In their rating, participants were requested to assess the dual task situation (CTT + respective task).

Procedure. First, participants received instructions on how to perform the CTT, followed by a short CTT training session. Then, we recorded two baseline trials with the CTT as a single task. Participants were then informed about the first of the secondary tasks (i.e. either cognitive or visual). They completed four experimental trials of 90 s; two for each difficulty level. The trials were blocked for difficulty, with the blocks in randomised order, each block preceded by a short familiarisation phase.

The same procedure was then used for the remaining secondary task. The subjective rating scale was filled at the end of each respective task condition. The whole experiment was completed in approximately 30 min.

Results

Prior to analysis, we tested whether it was justifiable to merge the two separate trials for each condition into one single score. We found no significant differences for any of the combinations, so we computed one single value for baseline, cognitive easy, cognitive difficult, visual easy and visual difficult conditions. Furthermore, we assessed the correlation between the two trials for each condition. With r = .70 for the two baselines, and r > .85 for the other pairs, we found highly significant correlations that confirm a high stability of the measurement.

CTT and secondary tasks. In Figure 8, CTT standard deviation values are displayed for the five conditions. The baseline condition is again included for purposes of completeness. A two-factor ANOVA for repeated measures revealed that task type (cognitive vs. visual), $F(1, 20) = 86.27$, $p < .001$, $\eta_p^2 = .81$, as well as task difficulty, $F(1, 20) = 28.47$, $p < .001$, $\eta_p^2 = .59$, had a significant influence on CTT performance. Visual tasks, as well as difficult tasks, resulted in higher deviations than cognitive and easy tasks. We also found a significant interaction between task type and task difficulty, $F(1, 20) = 10.44$, $p = .004$, $\eta_p^2 = .34$. As Figure 9 shows, the effect of task difficulty was mainly driven by the visual task, whereas raising task difficulty for the cognitive task had hardly an impact on CTT. For completeness, we also compared each experimental condition to the baseline with a paired t-test (Bonferroni-corrected), resulting in highly significant differences between the baseline and the two visual tasks (each $p < .001$). For the cognitive tasks, however, no significant differences were found ($p = .788$ for cognitive easy, $p = .112$ for cognitive difficult).

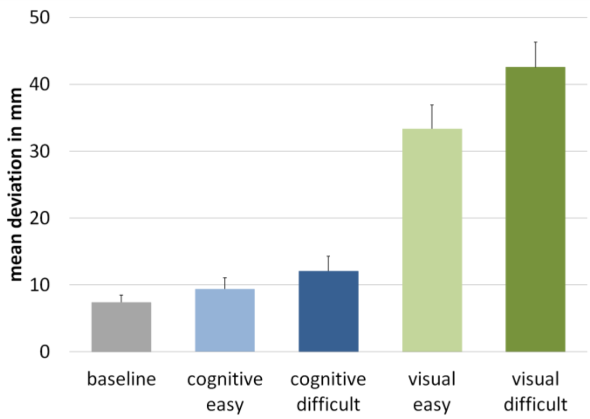

Figure 8. Mean deviation from the centre position.

Secondary task performance in the visual task was again measured as the number of SuRT trials finished. In the easy condition, participants completed 57.1 trials (SE = 4.6), in the difficult condition 11.4 (SE = 0.8). For the cognitive task, the number of counts made during the run of a trial was used as a performance measure. On average, 72.6 counts (SE = 4.1) were made in the easy and 26.0 (SE = 0.8) in the difficult condition. A two-factor repeated measures ANOVA confirmed a significant effect of task type, $F(1, 20) = 29.95$, $p < .001$, $\eta_p^2 = .60$, as well as task difficulty, $F(1, 20) = 229.42$, $p < .001$, $\eta_p^2 = .92$. No significant interaction was found, $F(1, 20) = 0.04$, $p = .853$, $\eta_p^2 = .00$.

Subjective ratings. In Figure 9, the results of the subjective ratings are displayed. It is clearly visible that for nearly all dimensions, the difficult version of a task resulted in higher ratings (with the exception of "visual demand" for the cognitive tasks). A two-factor ANOVA for repeated measures showed a significant effect of task difficulty on all five rating dimensions (see Table 1). Of special interest are also the significant interactions for visual demand, "mental demand" and "effort". Figure 9 shows that an increased task difficulty in the visual task had, as expected, a much higher impact on the rating of visual demand compared to the cognitive task. In contrast, an increased difficulty in the cognitive tasks had stronger impact on mental demand and effort in comparison to the visual task.

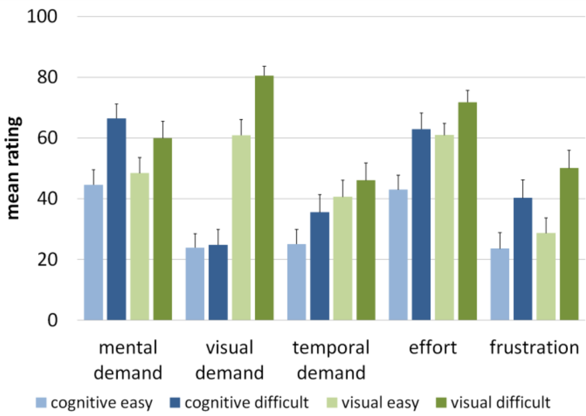

Figure 9. Subjective assessment of secondary tasks.

Table 1. ANOVA results for subjective assessment of secondary tasks (main effects and interaction).

	task type		task difficulty		interaction	
	F (1,20)	*P*	*F* (1,20)	*p*	*F* (1,20)	*p*
mental demand	.10	.760	37.29	<.001	4.71	.042
visual demand	82.68	<.001	21.91	<.001	14.40	.001
temporal demand	9.55	.006	6.82	.017	.88	.359
effort	9.87	.005	29.72	<.001	5.76	.026
frustration	3.64	.071	31.13	<.001	.60	.447

Discussion

The results of Experiment III further strengthen the claim that the CTT might be a useful method to assess driver distraction. It was able to capture variations in visual distraction, just as differences between cognitive and visual distraction. Distraction as measured by the CTT was generally much higher for visual than for cognitive tasks. This is in line with Mattes and Hallén's (2009) results with the same secondary tasks. For cognitive distraction, variations of task difficulty did not have an impact. The CTT appears to be especially suitable to assess visual distraction, whereas especially smaller variations in cognitive distraction would most certainly not be reflected by changes in performance.

The analysis of secondary task performance proved useful again. Although the differences in CTT performance between the two cognitive tasks were minimal, secondary task duration differed substantially. As many authors emphasise, distraction is not only magnitude, but also duration of distraction demand (e.g. Burns et al., 2010). However, since CTT performance with the cognitive tasks also hardly differed from the baseline, task duration could be neglected in this specific case, as, at least according to the measurement, no distraction occurred at all. For the visual tasks, it took participants much longer to complete the difficult task, which strengthens the impression that especially difficult visual tasks might have highly detrimental effects in terms of driver distraction.

It has to be acknowledged, though, that the comparison of tasks and task difficulties is somewhat problematic. There is no independent assessment of task difficulty, so it is not entirely clear how, for example, the cognitive difficult task relates to the visual difficult task in terms of absolute task difficulty. Nor can it easily be assumed that the difference between the two levels of difficulty is the same for both visual and cognitive tasks. A helpful indicator is the subjective assessment of the demand that the different task combinations induced, which in general supported our a-priori assumptions about the tasks' properties. But even although this assessment indicates that the difference between cognitive and visual tasks cannot just be ascribed to a substantial disparity in overall difficulty between task types, and that also the distances between easy and difficult are in general quite similar, the results still have to be handled with caution.

As for the experimental procedure, the change from longer to shorter test trials did not appear to distort results. At the same time, participants' complaints about the heavy demand of the experimental situation decreased. We conclude that if CTT duration is to be set at a fixed value, 90-s trials provide sufficient information for meaningful calculations, without overloading participants.

EXPERIMENT IV – NAVIGATION SYSTEM

After confirming the general usefulness of the CTT as a method to assess driver distraction through the differentiation between artificial secondary tasks of varying distraction-related demand, we aimed for a more realistic, i.e. externally valid assessment of the CTT. As the purpose of the method would be to assess distraction demand, especially of driver information systems, we regarded the CTT's validation on a navigation system as the next important step. The basic approach was to develop tasks that were objectively more or less distracting, and test whether the CTT is able to reflect this difference in distraction. Research has focused on the visual load of IVIS as a critical issue for driver distraction. Wierwille (1993) described the need for visual sampling of the traffic environment for safe driving, which is influenced by the operation of IVIS. As the European statement of principles (CEC, 2008) states: "Increasing the frequency and/or duration of glances required to detect and acquire visually displayed information may increase the risk of potentially dangerous traffic situations caused by driver preoccupation with non-primary driving-related tasks" (p. 13). Since interactions of drivers with IVIS often extend beyond visual acquisition of information by requiring manual interaction as well, the document also recommends minimising the number and length of these interactions, argues for interruptability and states that the pace of interaction should be user-controlled. We therefore decided to design IVIS tasks that have an influence on visual sampling and manual control of the CTT by varying the need to attend to the IVIS visually and manually to complete the specified task. We hypothesised that IVIS tasks that are lower in visual and motor demand result in better CTT performance compared to tasks that are higher in visual and motor demand.

Method

Participants. Twenty-five students of Chemnitz University of Technology, again all licensed drivers, took part in this experiment, 18 female and 7 male, with a mean age of 21.6 years (SD = 1.5). Experiment IV participants did not take part in any of the previous experiments. Each participant completed all available experimental conditions.

Material.

CTT. The CTT employed in this experiment, as well as the overall setup (screen size & position, operation of task etc.), were identical to the previous experiments.

Easy destination entry task. All tasks were performed on a nomadic navigation system (TomTom Go 710) that was placed to the right of the CTT, in a position where an aftermarket navigation system would typically be located (again following the LCT ISO). Participants entered destination letter by letter on a touchscreen display. To complete the task, participants selected the desired destination from a drop-down appearing above the letter block (or to proceed to the next step). Participants also learned how to navigate through menus and practiced this thoroughly before the actual experiment took place.

The easy destination entry task required participants to enter a single city name. An example would be “Verona, Italy”, with the system’s current position (and starting point for route calculation, set by default) set as “Chemnitz, Germany”. Once a route was calculated, participants were instructed to choose the option “avoid toll roads”, which resulted in a recalculation of the route. The destinations were chosen so that the system needed considerable time for the computation of the route and the recalculation. This allowed for a substantial total task time, with only few interactions with the system necessary to complete the task. Overall, the easy destination entry task was designed to take approximately 60 s as a single task (based on pre-tests).

Difficult destination entry task. The difficult destination entry task required participants to use the function “extended route planning”. Here, participants entered a start location (city and street), such as “Dresden, Waldstraße 1”, as well as a destination (city and street), such as “Berlin, Hauptstraße 1”. Locations were chosen such that they were easy to learn and recall (e.g. well-known German cities, simple street names, and always the same street number). Several actions, such as button or virtual button presses, were necessary to complete the task, whereas the calculation of the route was very fast. Again, the task was designed to take approximately 60 s as a single task.

Mobile phone synchronisation task. The third task assessed with the CTT was the navigation system’s mobile phone synchronisation function, which allows the system to use the mobile phone to connect to the internet. The participants’ task was to navigate through the different menus on the navigation system and a mobile phone (as instructed) to synchronise them. While both of the devices had to be operated repeatedly to complete the task, it was not required (nor possible) to do so in parallel. Again, the task was designed to take approximately 60 s as a single task. In contrast to the easy and difficult destination entry tasks, we had no specific a-priori assumptions about the task’s distraction demand in relation to the other two tasks. While the coordination of two different additional devices without doubt created additional load, there were also phases during the synchronisation process in which no action or attention was required. Therefore, we had to rely on the subjective rating of the task’s demand, especially in relation to the easy and difficult destination entry tasks, as a criterion for validation.

Instrument for subjective ratings. We tried to confirm our a-priori assumptions about the easy and difficult destination entry tasks through subjective ratings. Also, we tried to create some criterion against which to validate the CTT performance in the mobile phone synchronisation task. We used the questionnaire employed in Experiment III. Again, participants were requested to assess the dual task situation (CTT + destination entry) with their rating.

Procedure. First, participants received instructions on the CTT, followed by a short CTT training session. Then, we recorded two baseline trials with the CTT as a single task. Participants then received instructions on the first of the destination entry tasks (order-balanced). They first completed one destination entry as a single task, then an additional one in parallel to the CTT as practice trials. Then, five destinations had to be entered in parallel to operating the CTT in the experimental phase. Each entry was completed as an individual trial (i.e. not five entries while continuously operating the CTT), resulting in five separate measurements. The same procedure was followed for the second of the destination entry tasks (i.e. single-task training, followed by dual-task training, followed by five experimental trials). Participants then received instructions on the mobile phone synchronisation procedure, and completed one experimental trial in this task. As we had absolutely no a-priori assumptions about the task's distraction demand in relation to the other two, we decided to always present this task last, so to not interfere with the comparison of the other two. The subjective rating scale was filled in at the end of each respective task condition. Overall, the experiment took approx. 45 min.

Results

CTT and secondary tasks. In Figure 10, CTT standard deviation values are displayed for the four conditions. For the two destination entry tasks, an average deviation over the five test trials was calculated. As can be seen in the figure, the difficult destination entry task resulted in higher deviations than the two other tasks. A repeated measures ANOVA revealed significant differences between the conditions, $F(3, 72) = 79.54$, $p < .001$, $\eta_p^2 = .77$. Post-hoc comparisons (Bonferroni-corrected) confirmed a significant difference between difficult destination entry and the other two conditions (both $p < .001$), whereas there was no difference between easy destination entry and the mobile phone synchronisation ($p = .254$). All experimental conditions differed significantly from the baseline (all $p < .001$).

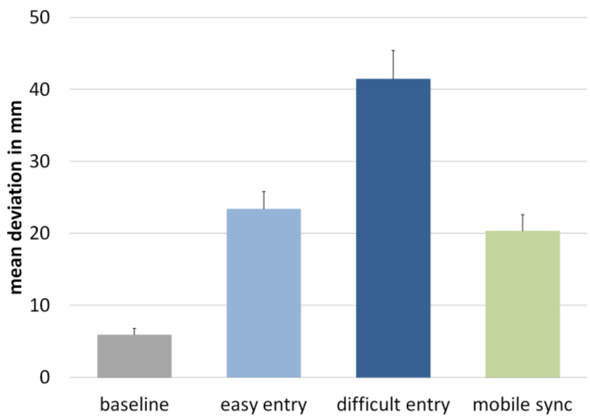

Figure 10. Mean deviation from the centre position.

Secondary task performance was measured as task duration. The easy destination entry task was completed fastest (M = 89.5 s; SE = 2.4 s), while it took much longer to complete the mobile phone task (*M* = 133.5 s; *SE* = 4.3 s), with the difficult destination entry task in between the two tasks (*M* = 108.7 s; *SE* = 4.9 s). Again, a repeated measures ANOVA revealed significant differences between the dual-task conditions, $F(2, 48) = 46.02$, $p < .001$, $\eta_p^2 = .66$. Post-hoc comparisons (corrected for multiple comparisons) confirmed significant differences between all three conditions (all $p < .001$).

Subjective ratings. In Figure 11, the results of the subjective ratings are displayed. The general pattern is the same for all five scales – difficult destination entry is scored highest, and mobile phone synchronisation lowest. ANOVA for repeated measures and post-hoc analysis (Bonferroni corrected) confirmed this picture, as nearly all comparisons reached statistical significance (Table 2).

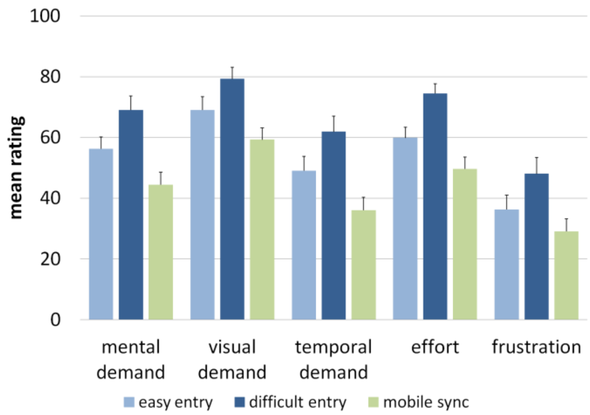

Figure 11. Subjective assessment of secondary tasks.

Table 2. ANOVA results and post-hoc pairwise comparison for subjective assessment of secondary tasks (easy = easy destination entry, diff. = difficult destination entry, mobile = mobile phone synchronisation).

	ANOVA		easy vs. diff.	easy vs. mobile	diff. vs. mobile
	F (2,48)	*p*	*p*	*p*	*p*
mental demand	21.98	<.001	.002	.026	<.001
visual demand	16.83	<.001	.004	.084	<.001
temporal demand	23.46	<.001	.001	.015	<.001
effort	33.30	<.001	<.001	.005	<.001
frustration	9.70	<.001	.063	.235	.001

As the subjective ratings should serve as an indicator of the CTT's accuracy in assessing the mobile phone synchronisation task's distraction potential, we calculated the correlation between CTT performance (i.e. CTT standard deviation) and subjective rating (all scales) over all conditions, to determine whether the overall pattern of results is similar. We found significant correlations between CTT standard deviation and all five scales (from $r = .24$ to $r = .52$).

Discussion

In our fourth experiment, we found evidence for the CTT's capabilities as a measure for distraction that can serve beyond purely artificial tasks. It was able to reliably differentiate between a simple and difficult navigation system task. Subjective measures complemented these results. As for the mobile phone synchronisation task, the pattern of results found in the subjective ratings appears to be comparable to that for the CTT's standard deviation, which can be interpreted as an indicator of the CTT's accuracy in the tasks assessment. The visual inspection of the resulting figures strengthens this claim. Task duration, on the other hand, seems to find no reflection in the subjective ratings. It would appear that participants did not consider this as a relevant factor when reporting subjective assessments of task demand.

GENERAL DISCUSSION AND CONCLUSIONS

We conducted four experiments in order to find out whether the CTT has the potential to be a useful measure to assess driver distraction elicited by IVIS. The results of all four experiments confirm this assumption. The task was able to reflect simple manipulations in distraction demand of secondary tasks (Experiments I and II), and to assess differences between visual and cognitive distraction, just as different degrees of visual distraction (Experiment III). Moreover, it succeeded in differentiating between actual IVIS tasks of varying demand (Experiment IV). And finally, the differences that the CTT found corresponded with subjective assessments of workload in most cases. It appears that the CTT can serve as a method to assess driver distraction. The results show that especially visual distraction and variations thereof were easily detected. For cognitive distraction, the task proved hardly sensitive. Although it might have increased the value of the task if it would be able to detect cognitive distraction as well, the fact that mainly visual distraction impacted on CTT performance underscores the similarities between the CTT and the operational level of driving. When comparing visual and cognitive distraction, Kaber, Liang, Zhang, Rogers and Gangakhedkar (2012) found that the presence of visual distraction led to larger steering errors. Accordingly, Liang and Lee (2010) reported that visual distraction went with abrupt steering control, large lane variance and delayed reactions to lead vehicle brake events, results they did not find for cognitive distraction. Interestingly, they also found that steering control and lane variation were associated with eyes-off-road time. It appeared that glance duration accounted for most of the performance decrements in the different distraction conditions. An assessment of glance behaviour while operating a CTT / secondary task combination would be a valuable addition to better understand what drives CTT performance. Likewise, a comparison of this glance behaviour to glance behaviour in a realistic driving situation with the same secondary tasks could help validate the CTT.

This aspect touches on a broader issue - the link between CTT performance and actual driving performance. Of course, the CTT, just like other easy-to-use methods, will not be able to give a direct estimation of accident risk. However, at least the relative pattern of results should be similar in all potential settings, whether simulation, real world or CTT (see e.g. ISO 26022, 2010). The tasks and task variations used in our four experiments were chosen based on research literature and design guidelines. While the assumptions and outcome expectations for the experiments were certainly valid from a theoretical point of view, what is still lacking is an actual validation of the results on simulator or real world data. As a logical next step in the assessment of the CTT is to test it with more realistic secondary tasks, this problem becomes even more apparent, since for such tasks often no clear a priori assumptions about the pattern of results that would be expected can be made. In that case, the external validation on additional data is indeed crucial. Although the theoretical basis for our choice of tasks appears to justify the inferences drawn from the results of our experiments, they would certainly benefit from a further investigation of the secondary tasks and their effects on simulator or real world driving performance.

Further questions remain. The λ-value, which defines the level of difficulty in the CTT, was chosen on the basis that this value appeared to represent a medium level of difficulty. It is unclear if the pattern of results would have been identical if the task would have been easier or more difficult. The level of difficulty that most appropriately reflects the load experienced in “normal” or “emergency” driving situations still has to be assessed. Also, choosing to change the bar’s colour once it deviates too far from the centre of the screen is a matter of debate. It is certainly true that for some driving situations, additional stimuli help drivers to recover from making driving errors (e.g. rumble strips when crossing lanes). However, there are probably even more situations where the opposite is true, such as situations in which no distinct stimulus assists to correct such errors. Whether or not changing the bar’s colour would produce a different patterns of results needs further investigation. The decision to have baseline trials always and only at the beginning of an experiment has to be reconsidered as well. Initially, we assumed that the CTT as a single task is too easy to benefit from practice or experience. However, it appeared that when segmenting the baseline data in 20s portions, performance increased slightly over time. An appropriate strategy might be to follow LCT ISO recommendations (ISO 26022, 2010) again, and assess baseline performance at the beginning and end of experiments.

Of course there are excellent driving simulators available that might provide researchers with a much more realistic assessment of secondary tasks. The value of the task lies in the fact that it can provide a very cost-effective solution, as one among a battery of tests, for early testing of new in-vehicle devices. Thus, it falls into the same category of tools such as the PDT, the occlusion method or the LCT. It is very easy to employ, requires only minimum practice, and is very easy to analyse. At the same time, it is very flexible. Compared to the PDT (and similar to the others), the CTT does not require an additional driving task (simulator or real world), which increases flexibility and ease of use. The PDT has its advantage in the fact that it is able to capture cognitive distraction, which we could not confirm for the CTT. Occlusion cannot quantify performance decrements, just as it is not suitable for the assessment of the effects of display eccentricity. On the other hand, it can measure interruptability of a secondary task, something which cannot be achieved with the CTT. Compared to the LCT (to which it bears some resemblance as both are tracking tasks, and the dependent measures for both are the deviations from some ideal position), it appears that the CTT has advantages in several aspects. The hardware necessary for the CTT's use (standard PC) is even simpler than for the LCT (standard PC + game steering wheel). Also, there is less need for training, as the definition of the bar's ideal position in the CTT is much more straightforward than the normative path defined in the LCT, which requires some explanation and respective training. In addition, the possibility to use secondary tasks of different length, without the need to repeat the tasks or to stop them halfway through allows for presumably more accurate assessments of real-world tasks. Finally, although the CTT does not include actual unexpected events, the fact that any lengthy withdrawal of visual attention is inevitably reflected in task performance is, in our opinion, grounds for measurement of the potential risk associated with visual driver distraction. The LCT, however, has been shown to be more sensitive to cognitive distraction (although visual distraction is its primary focus). Whether or not the CTT's advantages outweigh the specific shortcomings in comparison to other methods certainly varies from case to case. Further assessment of the CTT, especially with more realistic secondary tasks, is required to draw a complete picture of the CTTs capabilities. Overall, however, it appears that the simplicity of the task and the results obtained in the reported experiments make the CTT an interesting candidate for the assessment of visual distraction caused by in-vehicle devices.

REFERENCES

Alliance of Automobile Manufacturers (2006). *Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems*. Washington, DC: Alliance of Automobile Manufacturers.

Angell, L. S., Auflick, J., Austria, P. A., Kochar, D., Tijerina, L., Biever, W. J., . . . Kiger, S. (2006). *Driver Workload Metrics Task 2 Final Report* (DOT HS 810 635). Washington, DC: NHTSA.

Ascone, D.S. (2009). *An examination of driver distraction as recorded in NHTSA databases* (DOT HS 811 216). Washington, DC: NHTSA.

Barnett, G., Licko, V., & Thompson, T. (1985). Behavioral pharmacokinetics of marijuana. *Psychopharmacology*, *85*, 51-56.

Baumann, M., Keinath, A., Krems, J.F., & Bengler, K. (2004). Evaluation of in-vehicle HMI using occlusion techniques: Experimental results and practical implications. *Applied Ergonomics, 35,* 197-205.

Buettner, A. (2009, December 2*). Global automotive telematics shipments to expand by factor of four by 2016* [Press release]. Retrieved from http://www.isuppli.com/Automotive-Infotainment-and-Telematics/News/Pages/Global-Automotive-Telematics-Shipments-to-Expand-by-Factor-of-Four-by-2016.aspx.

Burke, M., Gilson, R., & Jagacinski, R. (1980). Multi-modal information processing for visual workload relief. *Ergonomics*, *23*, 961-975.

Burns, M., & Moskowitz, H. (1980). Effects of diphenhydramine and alcohol on skills performance. *European Journal of Clinical Pharmacology*, *17*, 259-266.

Burns, P., Harbluk, J., Foley, J.P., & Angell, L. (2010). The importance of task duration and related measures in assessing the distraction potential of in-vehicle tasks. In P*roceedings of the Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2010)*, Pittsburgh, PA, pp. 12-19.

Carsten, O.M.J., Merat, N., Janssen, W.H., Johansson, E., Fowkes, M., & Brookhuis, K.A. (2005). *HASTE Final Report.* Leeds, UK: Institute for Transportation Studies, University of Leeds.

Commission of the European Communities (2008). Commission recommendation of 26 May 2008 on safe and efficient in-vehicle information and communication systems: Update of the European statement of principles on human-machine interface. *Official Journal of the European Communities, L216*, 1-42.

Damos, D.L., Bittner, A.C., Kennedy, R.S., Harbeson, M.M., & Krause, M. (1984). Performance Evaluation Tests for Environmental Research (PETER): Critical tracking test. *Perceptual and Motor Skills, 58*, 567-573.

Department of Transport (2010). *Overview of the National Highway Traffic Safety Administration's Driver Distraction Program* (DOT HS 811 299).

Derrick, W. (1988). Dimensions of operator workload. *Human Factors*, *30*, 95-110.

Distracted-driving campaign requires attitude adjustment. (2010, January 26). *USA Today,* p. 8A.

Elmenhorst, D., Elmenhorst, E., Luks, N., Maass, H., Mueller, E., Vejvoda, M., . . . Samel, A. (2009). Performance impairment during four days partial sleep deprivation compared with the acute effects of alcohol and hypoxia. *Sleep Medicine*, *10*, 189-197. doi:10.1016/j.sleep.2007.12.003.

FESTA consortium (2011). *FESTA Handbook. Version 3 (draft) revised by FOT-NET*.

Godthelp, H. (1984). *Studies on human vehicle control* (Dissertation). Soesterberg, NL: Institute for Perception, TNO.

Godthelp, H. (1986). Vehicle control during curve driving. *Human Factors, 28*, 211–221

Greco, M., Stucchi, N., Zavagno, D., & Marino, B. (2008). On the portability of computer-generated presentations: The effect of text-background color combinations on text legibility. *Human Factors*, *50*, 821-833. doi:10.1518/001872008X354156.

Green, P. (1999). *Visual and task demands of driver information systems* (UMTRI-98-16), Ann Arbor, MI: University of Michigan Transportation Research Institute.

Harbluk, J.L., Mitroi, J.S., & Burns, P.C. (2009). Three navigation systems with three tasks: Using the lane-change test (LCT) to assess distraction demand. In *Proceedings of the 5th international driving symposium on human factors in driver assessment, training, and vehicle design,* Big Sky, MT, pp. 24-30.

Harbluk, J.L., Noy, Y.I., Trbovich, P.L., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis and Prevention, 39*, 372-379.

Hart, S.G., & Staveland, L.E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: Elsevier.

Hurwitz, J.B., & Wheatley, D.J. (2002). Using driver performance measures to estimate workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *46*, 1804-1808.

ISO 15008. (2009). *Road vehicles - Ergonomic aspects of transport information and control systems - Specifications and test procedures for in-vehicle visual presentation*. Geneva, Switzerland: ISO.

ISO 16673. (2007). *Road vehicles - Ergonomic aspects of transport information and control systems - Occlusion method to assess visual demand due to the use of in-vehicle systems*. Geneva, Switzerland: ISO.

ISO 26022. (2010). *Road vehicles - Ergonomic aspects of transport information and control systems - Simulated lane change test to assess in-vehicle secondary task demand*. Geneva, Switzerland: ISO.

Jahn, G., Oehme, A., Krems, J.F., & Gelau, C. (2005). Peripheral detection as a workload measure in driving: Effects of traffic complexity and route guidance system use in a driving study. *Transportation Research Part F, 8*, 255-275.

Jamson, A.H., & Merat, N. (2005). Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. *Transportation Research Part F, 8*, 79-96.

Jex, H.R. (1967). Two applications of a critical-instability task to secondary work load research. *IEEE Transactions on Human Factors in Electronics, 8*, 279-282.

Jex, H.R., McDonnell, J.D., & Phatak, A.V. (1966). A "critical" tracking task for manual control research. *IEEE Transactions on Human Factors in Electronics, 7*, 138-145.

Kaber, D.B., Liang, Y., Zhang, Y., Rogers, M.L., & Gangakhedkar, S. (2012). Driver performance effects of simultaneous visual and cognitive distraction and adaptation behaviour. *Transportation Research Part F, (15)*, 491-501. doi:10.1016/j.trf.2012.05.004.

Karlsson, I.C.M., Rämä, P., Alonso, M., Engelbrektsson, P., Franzén, S., Henar Vega, M., . . . Welsh, R. (2009). *TeleFOT D.2.2.1 Testing and Evaluation Strategy*.

Keinath, A., Baumann, M., Gelau, C., Bengler, K., & Krems, J.F. (2001). Occlusion as a technique for evaluating in-car displays. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics - Aerospace and transportation systems* (pp. 391-397). Aldershot, UK: Ashgate.

Kimura, K., Marunaka, K., & Sugiura, S. (1997). Human factors considerations for automotive navigation systems—Legibility, comprehension, and voice guidance. In Y.I. Noy (Ed.), *Ergonomics and safety of intelligent driver interfaces* (pp. 153-167). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Klein, R., & Jex, H. (1975). Effects of alcohol on a critical tracking task. *Journal of Studies on Alcohol*, *36*, 11-20.

Kuratorium für Verkehrssicherheit. (2008). Road accident statistics 2007. In *Reihe Verkehr in Österreich. Heft 40*, Wien.

Lamble, D., Laakso, M., & Summala, H. (1999). Detection thresholds in car following situations and peripheral vision: Implications for positioning of visually demanding in-car displays. *Ergonomics, 6*, 807-815.

Liang, Y., & Lee, J.D. (2010). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis and Prevention, (42)*, 881-890. doi:10.1016/j.aap.2009.05.001.

Martens, M., & van Winsum, W. (2000). *Measuring distraction - the peripheral detection task*. Soesterberg, NL: TNO Human Factors Research Institute.

Mattes, S. (2003). The lane-change-task as a tool for driver distraction evaluation. In H. Strasser et al. (Eds.), *Quality of work and products in enterprises of the future* (pp. 57-60). Stuttgart: Ergonomia.

Mattes, S., & Hallén, A., (2009). Surrogate distraction measurement techniques: The lane change test. In M.A. Regan, J.D. Lee, & K.L. Young (Eds.), *Driver distraction. Theory, effects, and mitigation* (pp. 107-122). Boca Raton, FL: CRC Press.

Michon, J.A. (1979). *Dealing with danger: Report of the European Commission MRC Workshop on physiological and psychological factors in performance under hazardous conditions, Gieten, The Netherlands, 23-25 May, 1978* (VK 79-01). Haren, NL: Traffic Research Center, University of Groningen.

Miura, T. (1986). Coping with situational demands: A study of eye movements and peripheral vision performance. In A.G. Gale et al. (Eds.), *Vision in vehicles-II* (pp. 126-137). Amsterdam: Elsevier.

National Highway Traffic Safety Administration (2012). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices* (NHTSA-2010-0053).

New Zealand Ministry of Transport. (2010). *Reported injury crashes 2009.*

O'Donnell, R., Moise, S., & Schmidt, R. (2005). Generating performance test batteries relevant to specific operational tasks. *Aviation, Space, and Environmental Medicine*, *76*(7 suppl.), C24-C30.

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

Patten, C.J.D., Kircher, A., Östlund, J., Nilsson, L., & Svenson, O. (2006). Driver experience and cognitive workload in different traffic environments. *Accident Analysis and Prevention, 38*, 887-894.

Pauzié, A. (2002). In-vehicle communication systems: The safety aspect. *Injury Prevention*, *8*(4 suppl.), 26-29.

Petzoldt, T., Bär, N., Ihle, C., & Krems, J.F. (2011). Learning effects in the lane change task (LCT) - Evidence from two experimental studies. *Transportation Research Part F, 14*, 1-12. doi:10.1016/j.trf.2010.09.001

Pickrell, T.M., & Ye, T.J. (2011). *Driver electronic device use in 2009* (DOT HS 811 517). Washington, DC: NHTSA.

Ramaekers, J., Kauert, G., van Ruitenbeek, P., Theunissen, E., Schneider, E., & Moeller, M. (2006a). High-potency marijuana impairs executive function and inhibitory motor control. *Neuropsychopharmacology*, *31*, 2296-2303. doi:10.1038/sj.npp.1301068.

Ramaekers, J., Moeller, M., van Ruitenbeek, P., Theunissen, E., Schneider, E., & Kauert, G. (2006b). Cognition and motor control as a function of Delta9-THC concentration in serum and oral fluid: Limits of impairment. *Drug and Alcohol Dependence*, *85*, 114-122.

Ramaekers, J., Muntjewerff, N., van Veggel, L., Uiterwijk, M., & O'Hanlon, J. (1998). Effects of nocturnal doses of mirtazapine and mianserin on sleep and on daytime psychomotor and driving performance in young, healthy volunteers. *Human Psychopharmacology: Clinical and Experimental*, *13*(0 suppl.), S87-S97.

Ramaekers, J., Uiterwijk, M., & O'Hanlon, J. (1992). Effects of loratadine and cetirizine on actual driving and psychometric test performance, and EEG during driving. *European Journal of Clinical Pharmacology*, *42*, 363-369.

Regan, M.A., Lee, J.D., & Young, K.L. (Eds.) (2008). *Driver distraction: Theory, effects and mitigation*. Boca Raton, Florida: CRC Press.

Rehmann, J., Stein, E., & Rosenberg, B. (1983). Subjective pilot workload assessment. *Human Factors*, *25*, 297-301.

Rosenberg, B.L., Rehmann J., & Stein, E.S. (1982). *The relationship between effort rating and performance in a critical tracking task* (DOT/FAA/CT-82/66). Washington, DC: Federal Aviation Administration Office of Systems Engineering Management.

Sanchez, D., Garcia, E., Saez, M., Benmimoun, M., Pütz, A., Aust, M.L., . . . Obojski, M.-A. (2012). *euroFOT D.6.3 Final results: User acceptance and user-related aspects*.

Senders, J.W., Kristofferson, A.B., Levison, W.H., Dietrich, D.W., & Ward, J.L. (1967). The attentional demand of automative driving. *Highway Research Record, 195*, 15-33.

Sheridan, T. (2004). Driver distraction from a control theory perspective. *Human Factors, 46*, 587-599.

Shieh, K.-K., & Lin, C.-C. (2000). Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference. *International Journal of Industrial Ergonomics, 26*, 527-536.

Short, M., Pont, M.J., & Huang, Q. (2004). *Safety and reliability of distributed embedded systems: Simulation of motorway traffic flows* (ESL04/02). Leicester, UK: Embedded Systems Laboratory, University of Leicester.

Society of Automotive Engineers (2004). *Navigation and route guidance function accessibility while driving* (SAE recommended practice 2364). Warrendale, PA: Society of Automotive Engineers.

Starry, C. (2001). The emerging in-vehicle intelligent transportation systems market. *Business Economics, 32*, 49–55.

Stevens, A., Quimby, A., Board, A., Kersloot, T., & Burns, P. (2002). *Design guidelines for safety of in-vehicle information systems*. Project Report PA3721/01. Crowthorne, UK: Transport Research Laboratory TRL.

Stutts, J.C., Reinfurt, D.W., Staplin, L., & Rodgman, E.A. (2001). *The role of driver distraction in traffic crashes*. Washington, DC: AAA Foundation for Traffic Safety.

Törnros, J., & Bolling, A. (2006). Mobile phone use - effects of conversation on mental workload and driving speed in rural and urban environments. *Transportation Research Part F, 9*, 298-306.

US Department of Transportation (2009, October 1). *U.S. Transportation Secretary Ray LaHood announces administration wide effort to combat distracted driving* [Press release]. Retrieved from http://distraction.gov/files/for-media/10.01.09.pdf.

Wagner, R.C., & Weir, D.C. (2006). *Software users guide for the critical tracking task (CTT)* (DRI-TM-06-140-2).

Wickens, C., Braune, R., & Stokes, A. (1987). Age differences in the speed and capacity of information processing: 1. A dual-task approach. *Psychology and Aging*, *2*, 70-78.

Wickens, C., & Kessel, C. (1980). Processing resource demands of failure detection in dynamic systems. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 564-577.

Wierwille, W.W. (1993). Demands on driver resources associated with introducing advanced technology into the vehicle. *Transportation Research Part C*, 1, 133-142.

Young, K., Regan, M., & Hammer, M. (2003). *Driver distraction: A review of the literature*. Melbourne, Australia: Monash University, Accident Research Centre.