Learning effects in the lane change task (LCT)—Evidence from two experimental studies

Tibor Petzoldt, Nina Bär, Claudia Ihle and Josef F. Krems

Technische Universität Chemnitz, 09107 Chemnitz, Germany

E-mail: tibor.petzoldt@psychologie.tu-chemnitz.de

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

Abstract

Given the ever-growing distribution of new in-vehicle information systems, the assessment of their distraction potential becomes an important issue. An accurate estimation of their impact on driver behaviour should be made in the early stages of product development. Several easy-to-use methods can be used to make this early estimate, one of them being the lane change task (LCT). As this task is being considered as an ISO standard, questions about factors that might influence or even distort the results obtained through this procedure arise. One problem, which is the focus of this paper, is the possible occurrence of learning effects. We report the results of two experiments that show that participants' performance improves significantly after just one LCT encounter, and that this improvement is rather stable.

**Keywords:** Lane change task; Driving; Distraction; Learning

1. Introduction

As new in-vehicle information systems have become increasingly popular (Starry, 2001), they have also become more and more the subject of thorough investigation, as it is often argued that such systems lead to an increase in driver distraction ("technology-based distraction", Young, Regan, & Hammer, 2003). As field research (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006), official crash statistics (Kuratorium für Verkehrssicherheit, 2008; New Zealand Ministry of Transport, 2007), and in-depth crash analyses (Stutts, Reinfurt, Staplin, & Rodgman, 2001) have shown, distraction is a major issue in traffic accidents. Also, field and laboratory studies have highlighted certain aspects of the impact that driver distraction has on driving performance. Critical deterioration of driving performance has been observed concerning lateral position, speed maintenance, reaction times, and gap acceptance (see Young et al., 2003 for an overview), often as a result of changes in glance behaviour (Lansdown, 2001).

Industry leaders and government authorities are well aware of this problem. The assessment of the potential distraction caused by new in-vehicle devices is an issue that has been (and is still being) addressed in several projects, such as the Driver Workload Metrics Project from the Crash Avoidance Metrics Partnership (CAMP; Angell, Auflick, Austria, Kochhar, Tijerina, Biever, et al., 2006) and the Advanced Driver Attention Metrics (ADAM; Breuer, Bengler, Heinrich, & Reichelt, 2003) project. Within ADAM, a set of easy-to-use methods has been developed and evaluated to assess the extent of distraction imposed by performing secondary tasks while driving. 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 already become an ISO-standardised procedure (ISO 16673, 2007). 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). A third method, the lane change test/task (LCT; Mattes, 2003), has also been developed to address the issue of visual distraction. As both occlusion and PDT are rather artificial procedures, they lack face validity, as neither bears any obvious resemblance to activities connected to actual driving. The LCT, however, employs the look of a driving simulator (see Section 1.1), and therefore, aside from its scientific grounding, it has a certain appeal when it comes to communicating results obtained in studies on driver distraction. Consequently, the LCT is now under investigation as a potential ISO standard. The major objective of this paper is to investigate whether the LCT delivers results stable enough for it to be considered a good, reliable measure of distraction.

1.1. The lane change task (LCT)

The LCT is a simple, inexpensive dual-task method intended to estimate secondary task demand on a driver as a result of the operation of an in-vehicle device in a laboratory setting. Participants have to control a simulated vehicle on a three-lane road, with no other traffic present, and are instructed to change lanes according to signs appearing on both sides of this road (Figure 1). Participants are required to maintain a constant speed of 60 kph. Exceeding this limit is not possible. The signs appear around every 150 m; duration between lane changes is therefore around 9 s.
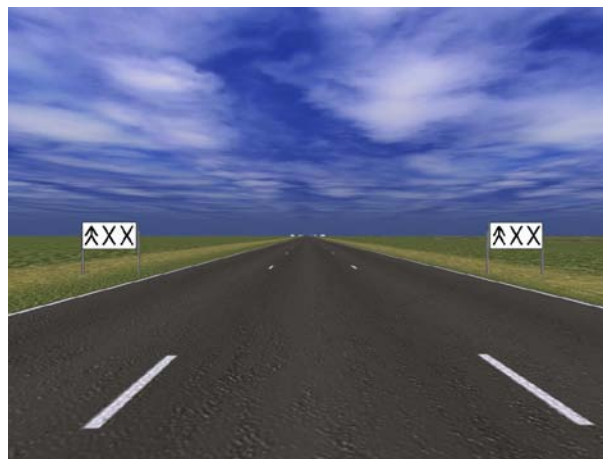
Figure 1. Lane change task (LCT) example screen.

Main performance measures are the mean deviation (MDEV) from a nominal lane change model, or the MDEV from a participant's own baseline (adaptive model). Additional measures as discussed in the ISO draft might be a modified standard deviation of lateral position, the proportion of missed or erroneous lane changes, or the mean delay in lane change initiation.

Recent studies employing the LCT have so far mainly focused on methodological aspects such as the connection between secondary task time and LCT performance (Burns, Trbovich, McCurdie, & Harbluk, 2005; Harbluk, Mitroi, & Burns, 2009), the introduction of new performance metrics (Engström & Markkula, 2007; Harbluk, Burns, Lochner, & Trbovich, 2007), or the susceptibility to different experimental contexts (Bruyas, Brusque, Auriault, Tattegrain, Aillerie, & Duraz, 2008; Rognin, Alidra, Val, Lescaut, & Chalandon, 2007). However, regardless of the studies available, at the time of this writing some issues remain unresolved. Baseline values vary considerably between different test sites. For instance, Weir, Kwok, and Peak (2007) reported MDEV values of around 0.64, Rognin et al. (2007) values of 1.6. Bearing in mind that the LCT is expected to become a standardised procedure, that is, is supposed to produce comparable results under comparable circumstances, this

variance in performance metrics is quite worrisome. Possible factors that might influence the metrics in such a dramatic manner have to be assessed. One first step has been taken by Petzoldt, Bär, and Krems (2009), who analysed the influence of the participant sample's composition, and more specifically, the effect of gender on LCT performance. They reported substantial differences between male and female participants in terms of LCT and secondary task performance. However, this effect cannot account for all the variance that has been observed in the aforementioned studies. Another possible factor of influence might be experience with the LCT, which is the focus of this paper.

1.2. LCT and training

One important question that needs to be answered is whether the repeated assignment of the same participants might influence LCT results. When evaluating in-vehicle information systems, manufacturers usually rely on a predefined group of possible participants who are quite regularly deployed for studies of this kind. It can be argued that this practice leads to training effects that might distort the results obtained. As long ago as the 1800s Ebbinghaus (1885/1971) described the "learning curve," which refers to the relation between the amount of learning and the time it takes to learn. Especially in the early stages of skill acquisition, a tremendous increase in performance quality can be observed. It is doubtful if this early stage of rapid learning is already completed after the short LCT familiarisation phase that precedes the actual test. For novices, learning, and therefore performance advancement, might still occur in experimental trials. Experts would be expected to perform at a high level right from the beginning. Shinar, Tractinsky, and Compton (2005) investigated the effect of practice on interference from a phone task while driving in a simulator. They found that after five sessions of driving and using the phone, there was a learning effect on most of the driving measures, and that this training was even sufficient to eliminate driving impairment caused by the phone task altogether in the group of more experienced drivers.

Not only for the LCT, but also for the secondary task variations in performance can be expected. Jahn, Krems, and Gelau (2009) described the course of skill acquisition in operating navigation systems, pointing out that training effects might occur for the secondary tasks, as well. Also, training effects do not have to be limited to single tasks only. It has been demonstrated that dual tasks are more than just the sum of their component tasks (Bahrick, Noble, & Fitts, 1954; Bahrick & Shelly, 1958). It appears that dual-task training not only leads to an increased automation of the respective tasks, but also helps develop the skill to optimally allocate resources between them. Damos and Wickens (1980) identified such time-sharing skills and their development in dual-task training. They also found evidence for the transfer of those skills to other task combinations. It can be argued that previous encounters with the LCT in conjunction with a secondary task might facilitate time sharing

in subsequent experimental instances and might do so even with different secondary tasks. Again, considering the learning curve, LCT novices might still be acquiring the necessary skill in experimental trials, while experts start at their best and show an overall superior performance. So the first question to be answered is whether a first encounter with the LCT in an experimental setting serves as LCT training for subsequent encounters. Given the very simple nature of the initial training for the LCT and secondary task, and the practical absence of dual-task training before the actual test, we hypothesise that training effects will occur.

The second question to be examined is whether such a training effect is stable. In this context, the "forgetting curve" is a relevant concept, as it describes the course of forgetting acquired knowledge over time (Ebbinghaus, 1885/1971). The largest portion of forgetting appears to take place in the time directly after learning. With increasing time, forgetting happens less and less. For the LCT, it can be argued that even if training takes place, the effects will disappear very quickly. Taking into account that participants are not usually tested on a day-to-day basis, but rather with intervals of weeks and months, it is possible that the effects of training become negligible after some time, making any experimental control for training or experience redundant.

However, Ebbinghaus also provided evidence that could suggest the opposite. First, there is the effect of overlearning. Overlearning occurs when knowledge already acquired and understood perfectly is still being learned. The same applies to motor skills. In such cases, the forgetting curve does not appear. A proper and very valid example is driving a car. After learning how to drive, people of course just go on driving, making every ride another learning trial. Even if not driving for some time, an experienced driver will not forget how to drive. It is questionable whether such an effect occurs after solving the LCT once. However, regular LCT driving might indeed lead to overlearning. This would result in better performance in subsequent LCT trials even with larger inter-experiment intervals. A second aspect of relevance is the concept of savings. Ebbinghaus (1885/1971) stated that, even if forgetting occurs, people do not require the same amount of time to reacquire knowledge or skills once learned and then forgotten as they did to learn it the first time. Thus, although during the first LCT encounter learning is still happening in the experimental trials, this is not necessarily the case in subsequent studies. There, the simple act of training for the LCT and secondary task before starting the actual experiment might just be enough to reach the performance plateau that was reached the first time only during the experiment, and not before. This effect, in contrast to overlearning, can be expected even after only one previous LCT experience.

In this paper, we present two experiments that try to shed light on the issue of the LCT and training. The first experiment assesses whether training effects occur at all. The second experiment more

closely resembles the practical use of the LCT in terms of inter-experiment interval, addressing the question of whether any training effect found earlier is of practical relevance.

2. Experiment I

In this experiment, we tried to uncover any effect of experience that might occur. The LCT was used in conjunction with two different secondary tasks that resemble different aspects of in-vehicle tasks. The surrogate reference task (SuRT) is a task that is chunkable and allows for interruption. The critical tracking task (CTT), in contrast, requires continuous attention. We hypothesized that any training with the LCT would result in improved dual-task performance, either through direct learning effects or indirectly through the freeing of resources. More specifically, we expected dual-task training (LCT + secondary) to produce better performance than just LCT training.

2.1. Method

2.1.1. Participants

Fifty-two participants took part in this study; 5 had to be removed from the data set for being statistical outliers in terms of LCT performance (mean values more than 2 *SD*s different from group average). All of the remaining 47 (age *M* = 29.1 years, *SD* = 8.5 years, 25 male, 22 female) were in possession of a valid driving license and drove a mean of 16,500 km a year, *SD* = 2,250 km (outlier with 350,000 km a year excluded here). None had previous LCT experience. All of the student participants (32) received course credit; the remainder received monetary compensation. Students and nonstudents were distributed equally over the different experimental conditions.

2.1.2. Material

2.1.2.1. Lane change task (LCT)

For presentation of the LCT, the desktop setup as described in the ISO draft was employed. A standard PC system with a 19" flat screen was used. To control the vehicle, a MOMO force-feedback game steering wheel with foot pedals was connected to the PC. The length of a single LCT trial corresponded to the length of one LCT track, 1,800 m, which should take roughly 3 min, provided the participants follow instructions. Any secondary task was terminated as soon as the end of the track was reached.

2.1.2.2. Surrogate reference task (SuRT)

The SuRT we employed in the experiment required participants to scan stimulus displays for the one stimulus that differed from the others surrounding it (Figure 2). Target and distracters were white circles on a black background. Distracter size could be varied to create different levels of difficulty (later referred to as SuRT1, SuRT2, and SuRT3). Participants gave their response by moving a grey indicator bar to the position of the identified target and pressing the enter key for confirmation, after which the next display appeared. As a performance metric, we analysed the number of displays correctly solved per LCT trial, as this might best reflect any attentional strategy. The task was presented on an 8.37" screen to the right of the participant. The indicator bar was controlled using a standard keyboard. Position of screen and keyboard matched the requirements of the ISO draft.
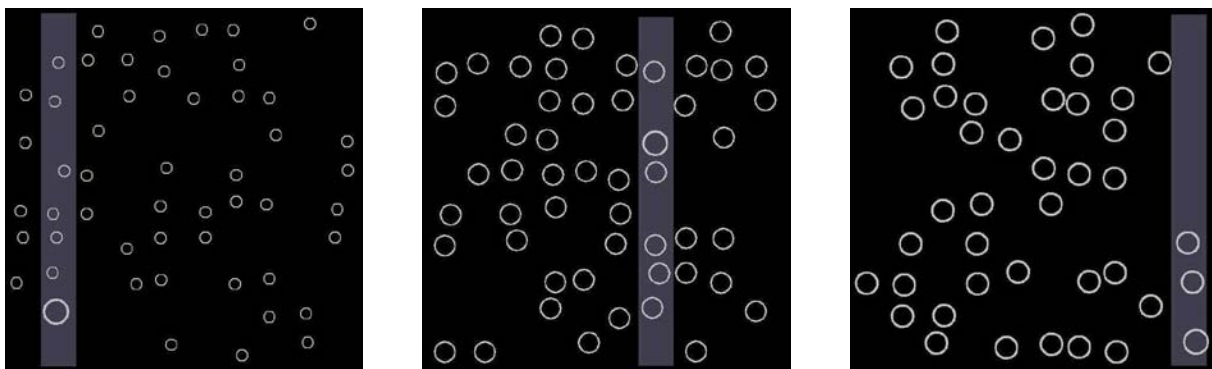

Figure 2. Surrogate reference task (SuRT) examples for low (SuRT1, left), moderate (SuRT2, centre) and high (SuRT3, right) difficulty.

2.1.2.3. Critical tracking task (CTT)

The main goal in the CTT (Figure 3) is the manual control of a dynamic unstable element. This element is a simple horizontal bar that tends to leave the proposed target position at the centre of the screen. While the bar moves up or down continuously, participants try to control this deviation by using the up and down keys on a keyboard, with the ultimate goal of bringing the bar back to the centre. If the bar gets too far away from the middle of the screen, its colour changes to red to alert the participant and capture his/her attention (see Figure 3, right). The task allows for the variation of difficulty by letting the experimenter choose the level of instability. Three different levels of difficulty were used in the experiment (later referred to as CTT1, CTT2, and CTT3). The standard deviation from the central position is used as a performance indicator. Display and keyboard were the same as for the SuRT.
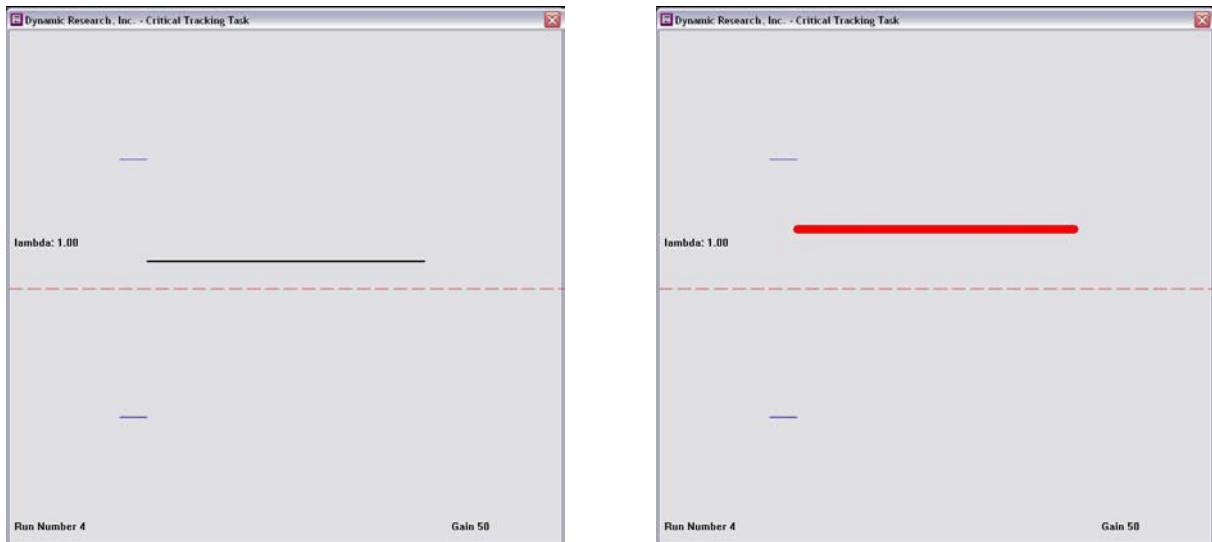
Figure 3. Critical tracking task (CTT) example screens.

2.1.2.4. Driving activity load index (DALI)

For a subjective rating of workload, we employed the DALI questionnaire (Pauzié & Pachiaudi, 1997). The questionnaire is derived from the NASA-TLX (Hart & Staveland, 1988) and is intended for the assessment of workload experienced while driving with an additional task. We used five of the questionnaire's seven scales that were suitable for the experimental setup: global attention demand, visual demand, stress, temporal demand, and interference (the interference subscale was not used in baseline drives, as its purpose is to capture the interference caused by a concurrent secondary task). The auditory demand and tactile demand subscales (specific constraints induced by vibrations during the test) were omitted.

2.1.3. Procedure

Participants were divided into three groups. The control group did not receive any training and therefore had to show up only for one testing session. The second group received an "LCT only" training. In the training session, they drove the LCT with varying instructions (e.g., "drive only with your right hand") to avoid boredom and fatigue for about 20–30 min. The third group received full LCT and secondary task training. They started with a short phase of LCT driving to get familiar with the task. After that, a baseline was recorded. Then, the different secondary task conditions (blocked for task type, random for difficulty within task type) were administered in a balanced fashion. After each single trial, participants had to fill in the DALI questionnaire.

The testing session was the same for all three groups and followed the procedure of the training session for the dual-task-trained group. First, there was a familiarisation phase, then the recording of a baseline drive, and afterwards the balanced application of the secondary task conditions, each directly followed by the DALI. Training and testing session were a maximum of 1 week apart.

2.2. Results

2.2.1. LCT + SuRT, between-groups comparison

To assess possible differences between the three training groups, we analysed the mean deviation of the driven course from the normative model (MDEV) by calculating a two-way analysis of variance (ANOVA) for mixed designs. The analysis revealed a main effect of secondary task condition, $F(3, 132) = 46.61$, $p < .001$. Pairwise comparison of conditions, however, showed differences only between the baseline and all the SuRT conditions (all $p < .001$). We found no main effect of training, $F(2, 44) = 3.09$, $p = .056$. Also, there was no interaction between the secondary task condition and the type of training. As can be seen in Figure 4 (left), baseline values do not differ greatly, but the fully trained group outperformed the other two in the SuRT conditions. The control group showed the largest performance decrement with a concurrent secondary task.
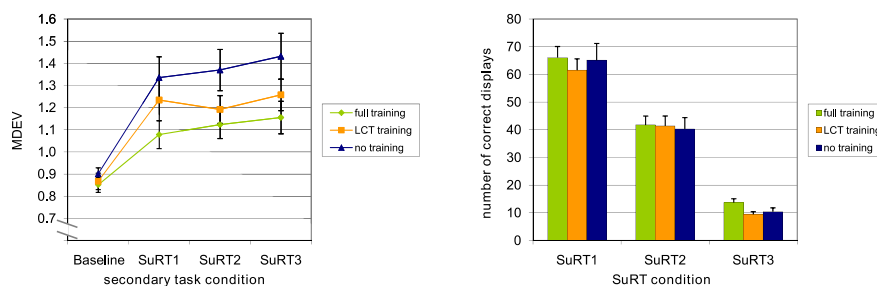
Figure 4. Mean deviation (MDEV) of the driven course from the normative model in the LCT (left) and number of correct displays in the SuRT (right) for different training groups (error bars indicate standard error).

To control for possible effects of variations in attention allocation, we analysed secondary task performance, measured as the number of displays correctly solved (Figure 4, right). A two-way ANOVA for mixed designs produced the anticipated main effect of secondary task difficulty, $F(2, 88) = 285.33$, $p < .001$. No effect of training, $F(2, 44) = .28$, $p = .759$, and no interaction between secondary task and training were found.

We analysed subjective measures as obtained with the DALI (for an overview, see Table 1) by calculating a two-way ANOVA for mixed designs for each subscale. Every subscale produced a

significant main effect of secondary task condition (all $p < .001$). In addition, the scales global attention demand, $F(2,44) = 4.86$, $p = .012$, and visual demand, $F(2,44) = 7.07$, $p = .002$, showed significant main effects of training. Post hoc testing revealed a difference in global attention demand between the fully trained group and the control group ($p = .018$), as well as differences in visual demand between the fully trained group and both other groups ($p = .009$ and $p = .008$). Other scales showed no effect of training, and no interactions were found.

Table 1. Subjective assessment of the lane change task (LCT) plus the surrogate reference task (SuRT) as obtained by the driving activity load index (DALI); for between-groups comparison see rows "no training" vs. "LCT only training" vs. "full training (testing session)"; for comparison of training sessions see "full training (testing session)" vs. "full training (training session)".

scale	group	baseline		LCT + SuRT1		LCT + SuRT2		LCT + SuRT3	
		M	SD	M	SD	M	SD	M	SD
global attention demand	no training	3.24	1.48	4.71	1.05	5.41	0.94	5.88	0.33
	LCT only training	3.21	1.48	4.79	1.12	5.00	1.30	5.86	0.36
	full training (testing session)	2.50	1.15	3.75	1.39	4.25	1.39	5.44	0.81
	full training (training session)	3.69	1.58	4.38	1.26	4.69	1.01	5.75	0.58
visual demand	no training	3.59	1.46	4.35	1.37	5.18	0.88	5.82	0.53
	LCT only training	3.21	1.53	4.64	1.15	5.14	0.86	5.93	0.27
	full training (testing)	2.38	0.96	3.38	1.09	4.06	1.24	5.69	0.60
	full training (training)	3.69	1.62	4.25	1.29	4.88	1.09	5.94	0.25
stress	no training	1.94	1.09	3.76	1.25	3.88	1.05	5.06	0.97
	LCT only training	2.57	1.50	3.64	1.78	4.00	1.62	4.57	1.74
	full training (testing session)	1.81	1.17	3.00	1.26	3.44	1.67	4.31	1.58
	full training (training session)	2.94	1.48	3.81	1.22	3.69	1.25	4.94	1.18
temporal demand	no training	1.71	1.40	3.59	1.54	3.82	1.59	4.47	1.62
	LCT only training	2.00	1.36	3.21	1.53	3.93	1.77	3.93	1.44
	full training (testing session)	1.81	1.22	2.94	1.57	3.19	1.42	4.06	1.73
	full training (training session)	2.19	1.42	3.88	1.09	3.88	1.20	4.75	1.18

2.2.2. LCT + SuRT, comparison of training sessions

Since there are two complete data sets (training session and testing session) available for the fully trained group, we calculated a two-way repeated measures ANOVA on the MDEV to assess possible learning effects within this group. Again, we found a main effect of secondary task condition, $F(3, 45) = 23.46$, $p < .001$. Also, we found a main effect of training, $F(1, 15) = 5.42$, $p = .034$. There was no interaction between the secondary task condition and training level. As can be seen in Figure 5 (left), in the second session the lane keeping/changing quality increased for all secondary task conditions.
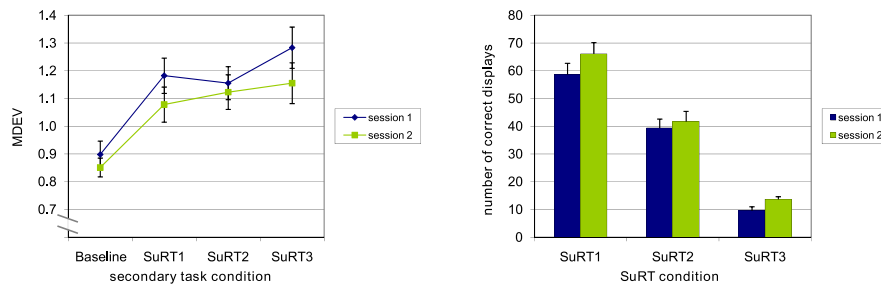
Figure 5. LCT (left) and SuRT (right) performance for different testing sessions (error bars indicate standard error).

When analysing secondary task performance (Figure 5, right), we found a main effect of secondary task difficulty, $F(2, 30) = 164.21$, $p < .001$, but no significant effect of training level, $F(1, 15) = 1.70$, $p = .213$. There was, however, a significant interaction between secondary task difficulty and training level, $F(2, 30) = 5.59$, $p = .009$.
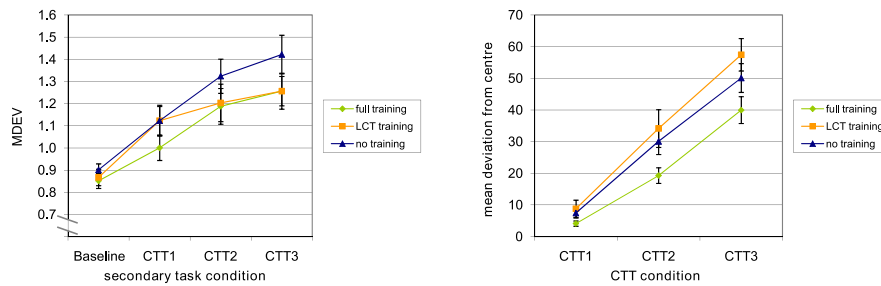
The analysis of the DALI's (Table 1) subscales again revealed a significant main effect of secondary task condition on each scale (all $p < .001$). In addition, the scales global attention demand, $F(1,15) = 14.60$, $p = .002$; visual demand, $F(1,15) = 32.09$, $p < .001$; stress, $F(1,15) = 13.33$, $p = .002$; and temporal demand, $F(1,15) = 9.77$, $p = .007$, showed significant main effects of training. No interactions were found.

2.2.3. LCT + CTT, between-groups comparison

The analysis for assessing training effects concerning the LCT/CTT combination follows the same procedures as for the SuRT. For differences between the training groups, we calculated a two-way ANOVA for mixed designs on the MDEV values (Figure 6, left). The analysis revealed a main effect of secondary task condition, $F(3, 132) = 76.57$, $p < .001$. Pairwise comparison of conditions showed significant differences between all of them (five out of six $p < .001$). However, we found no effect of training, $F(2, 44) = 1.13$, $p = .332$. There was no interaction between the secondary task condition and the type of training.

Analysing secondary task performance, we found a main effect of secondary task difficulty, $F(2, 88) = 182.00$, $p < .001$, as well as of training, $F(2, 44) = 4.10$, $p = .023$. There was no interaction between the two. As can be seen in Figure 6 (right), the "full training" group produced the smallest deviations, whereas the "LCT training" group performed even worse than the control group.

Figure 6. LCT (left) and CTT (right) performance for different training groups (error bars indicate standard error).

Subjective measures as obtained with the DALI (Table 2) showed a pattern similar to that of the SuRT. There was again a significant main effect of secondary task condition for each subscale (all $p < .001$). In addition, the scales global attention demand, $F(2,44) = 3.32$, $p = .045$, and visual demand, $F(2,44) = 3.72$, $p = .032$, showed significant main effects of training. Post hoc testing, however, showed no significant differences between the groups. Other scales showed no effect of training, and no interactions were found.

Table 2. Subjective assessment of LCT + CTT (critical tracking task) as obtained by the DALI; for between-groups comparison see rows "no training" vs. "LCT only training" vs. "full training (testing session)"; for comparison of training sessions see "full training (testing session)" vs. "full training (training session)"

scale	group	baseline		LCT + CTT1		LCT + CTT2		LCT + CTT3	
		M	SD	M	SD	M	SD	M	SD
global attention demand	no training	3.24	1.48	4.24	1.25	5.29	0.77	5.76	0.56
	LCT only training	3.21	1.48	4.36	1.08	5.21	0.80	5.93	0.27
	full training (testing session)	2.50	1.15	3.88	1.20	4.38	1.02	5.31	0.87
	full training (training session)	3.69	1.58	4.06	1.06	5.38	0.72	5.69	0.60
visual demand	no training	3.59	1.46	3.94	1.56	5.29	1.05	5.29	0.99
	LCT only training	3.21	1.53	4.21	1.12	5.07	1.00	5.64	0.63
	full training (testing)	2.38	0.96	3.38	1.31	4.25	1.44	4.88	1.54
	full training (training)	3.69	1.62	4.06	1.24	5.06	1.00	5.31	1.08
stress	no training	1.94	1.09	3.18	1.33	4.47	1.50	4.88	1.17
	LCT only training	2.57	1.50	3.29	1.64	4.14	1.61	4.36	1.34
	full training (testing session)	1.81	1.17	2.94	1.34	4.00	1.63	4.38	1.75
	full training (training session)	2.94	1.48	3.50	1.32	4.44	1.46	5.19	1.11
temporal demand	no training	1.71	1.40	3.18	1.59	4.47	1.74	4.82	1.59
	LCT only training	2.00	1.36	3.00	1.57	4.14	1.46	4.14	1.66
	full training (testing session)	1.81	1.22	2.88	1.36	3.81	1.38	4.31	1.54
	full training (training session)	2.19	1.42	3.06	1.48	4.19	1.83	4.56	1.71

2.2.4. LCT + CTT, comparison of training sessions

For the fully trained group, we also assessed possible learning effects within this group from session one to session two by calculating a two-way repeated measures ANOVA on the MDEV values. Again, we found a main effect of secondary task condition, $F(3, 45) = 32.81$, $p < .001$. Also, we found a main effect of training, $F(1, 15) = 6.70$, $p = .021$. There was no interaction between the secondary task condition and training level. As can be seen in Figure 7 (left), the lane keeping/changing quality increased in all secondary task conditions.
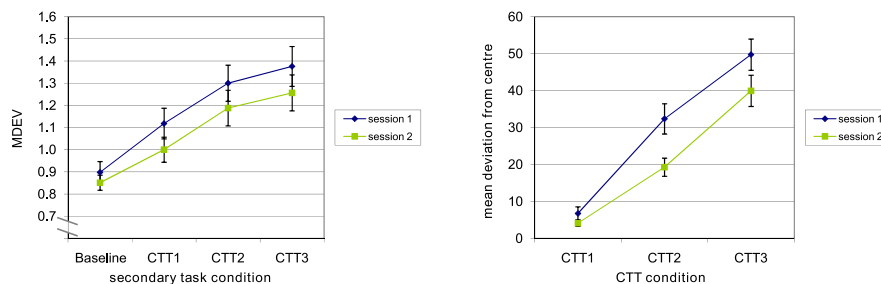
Figure 7. LCT (left) and CTT (right) performance for different testing sessions (error bars indicate standard error).

Regarding secondary task performance (Figure 7, right), we again found substantial effects of both secondary task difficulty, $F(2, 30) = 87.02$, $p < .001$, and training, $F(2, 15) = 51.87$, $p < .001$, as well as a significant interaction between the two, $F(2, 30) = 3.69$, $p = .037$.

The analysis of the DALI's subscales again produced a significant main effect of secondary task condition (all $p < .001$) on each scale. In addition, the scales global attention demand, $F(1,15) = 27.92$, $p < .001$; visual demand, $F(1,15) = 19.96$, $p < .001$; stress, $F(1,15) = 39.12$, $p < .001$; and interference, $F(1,15) = 11.04$, $p = .005$, showed significant main effects of training. For global attention demand, there also was a significant interaction, $F(1,15) = 4.13$, $p = .027$.

2.3. Discussion

We conducted this first experiment to find out whether any training effect occurs when repeatedly employing the same participants for LCT testing. The results clearly show that this effect exists, and that it is substantial. When comparing the different variations of training, it becomes obvious that especially dual-task training has a tremendous impact on performance. This impact, however, is not uniform. When looking at the SuRT as the secondary task, we find relevant differences between the groups in LCT performance, whereas SuRT performance does not seem to improve. In contrast, we

find no distinction between the three training groups in LCT performance when the CTT is the secondary task, whereas CTT performance improves with the level of training. The comparison of the two sessions for the fully trained group, however, strongly supports the view that both primary and secondary task performance improve with the amount of training. This impression is further strengthened by subjective assessments of the participants' workload. Between groups as well as between sessions comparisons show on various scales that training lowers the level of experienced load significantly. These findings are somewhat disturbing, given that as yet there are no specifics in the ISO draft on how to deal with this issue. Although the amount of training within an experiment is allowed to vary to give each participant the chance to reach some sort of optimum performance, our results show that this just might not be enough. At the same time, it has to be acknowledged that the short training–testing interval chosen for this experiment is rather artificial. It remains to be proven that the effect of experience we found is stable over a period of time that makes it practically relevant.

3. Experiment II

After we were able to confirm short-term learning effects in LCT and secondary task performance, we tried to assess the durability of those effects. As LCT testing is usually not done on a daily basis, the assessment of long-term effects bears much more practical relevance. Therefore, we tested another group of participants a minimum of 4 months after their initial (full) training, and again compared them to a control group with no previous LCT experience. As the LCT itself was not sensitive to variations of SuRT difficulty in our first experiment, we chose to use only the CTT as the secondary task.

3.1. Method

3.1.1. Participants

Forty-eight participants took part in this experiment; three outliers had to be removed. All of the remaining 45 (age M = 24.8 years, SD = 4.2 years, 14 male, 31 female) were in possession of a valid driving license and drove a mean 10,950 km a year, SD 8,550 km. None had previous LCT experience. All of the student participants (39) received course credit; the remainder received monetary compensation.

3.1.2. Material

The material used was identical to the material in the first study. We again used the LCT, employing the same setup (desktop according to the ISO draft). As a secondary task we used the CTT again, with settings identical to those in the first experiment. The DALI was applied as well, although only after the testing session.

3.1.3. Procedure

Two groups of participants were employed in the experiment. One group served as a control, similar to in the first experiment. They completed one testing session with a familiarisation phase, baseline driving, and then combined LCT + CTT driving, with CTT conditions in randomised order. Each LCT drive was concluded with filling in the DALI. The second group consisted of participants who had already taken part in a small preceding study that required them to drive the LCT in conjunction with the CTT, with the respective prior familiarisation and baseline drives. In this sense, this group was comparable to the "full training" group of Experiment I. Data on these drives were available. Half of the participants were tested around 4 months, the other half approximately 7 months after their initial "training" (the reason being two different points of time for the preceding study). Testing was done in the same way as for the control group.

3.2. Results

3.2.1. LCT + CTT, between-groups comparison

To assess possible long-term learning effects, we first computed a two-way ANOVA for mixed designs on the MDEV values (Figure 8, left). The analysis revealed a main effect of secondary task condition, $F(3, 129) = 41.69$, $p < .001$, which is coherent with the results of our previous experiment. There was no significant effect of training, $F(1, 43) = 3.71$, $p = .061$, and also no interaction between the secondary task condition and the type of training.
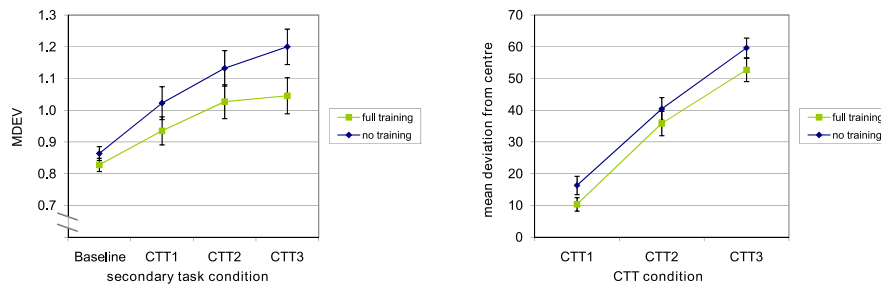
Figure 8. LCT (left) and CTT (right) performance for different training groups (error bars indicate standard error).

We also calculated an ANOVA on the adaptive MDEV values. The results were close to those obtained for the normative MDEV, with a significant influence of secondary task condition, $F(3, 129) = 5.04$, $p < .001$, no significant effect of training, $F(1, 43) = 4.04$, $p = .051$, and no interaction between the secondary task condition and training.

Analysing secondary task performance (Figure 8, right), we found the anticipated main effect of secondary task difficulty, $F(2, 86) = 198.59$, $p < .001$. However, no significant effect of training was found, $F(1, 43) = 2.26$, $p = .140$. There was no interaction between the two.

The analysis of the DALI's subscales (Table 3) produced a significant main effect of secondary task condition (all $p < .001$) on each scale. In addition, for the scale global attention demand, we found a significant influence of training, $F(1,41) = 5.79$, $p = .021$. Other scales showed no significant effect of training, and no interactions were found.

Table 3. Subjective assessment of LCT + CTT as obtained by the DALI (between-groups comparison).

		baseline		LCT + CTT1		LCT + CTT2		LCT + CTT3	
scale	group	M	SD	M	SD	M	SD	M	SD
global attention demand	no training	4.86	1.04	4.95	1.17	5.68	0.48	5.95	0.21
	full training (testing session)	4.00	1.10	4.62	0.97	5.24	0.70	5.76	0.44
visual demand	no training	4.26	1.42	4.83	1.19	5.48	0.73	5.78	0.52
	full training (testing session)	3.77	1.23	4.82	0.91	5.23	0.81	5.68	0.57
stress	no training	2.61	1.41	3.74	1.32	4.52	1.27	4.96	1.22
	full training (testing session)	2.50	1.10	3.45	1.06	4.50	1.10	4.68	1.04
temporal demand	no training	2.26	1.45	3.17	1.59	4.00	1.62	4.30	1.55
	full training (testing session)	2.86	1.61	3.86	1.42	4.09	1.34	4.64	1.18

3.2.2. LCT + CTT, comparison of training sessions

Since we had two full data sets (initial training and testing) available for the fully trained group again, we compared this group's performance in the two sessions. The two-way repeated measures ANOVA on the MDEV showed a significant effect of secondary task condition, $F(3, 63) = 25.60$, $p < .001$. Also, there was a significant influence of training, $F(1, 21) = 14.84$, $p = .001$, on LCT performance (Figure 9, left). We also found a significant interaction between the secondary task condition and training, $F(3, 63) = 3.76$, $p = .015$. A similar picture emerged for the adaptive MDEV values. Again, there was a significant effect of secondary task condition, $F(3, 63) = 35.65$, $p < .001$, and a significant effect of training, $F(1, 21) = 16.21$, $p = .001$. However, there was no interaction between the secondary task condition and the type of training for the adaptive MDEV values.

Analysing secondary task performance (Figure 9, right), we found the anticipated main effect of secondary task difficulty, $F(2, 42) = 182.72$, $p < .001$. Also, a significant effect of training was found, $F(1, 21) = 20.34$, $p < .001$. There was no interaction between the two.
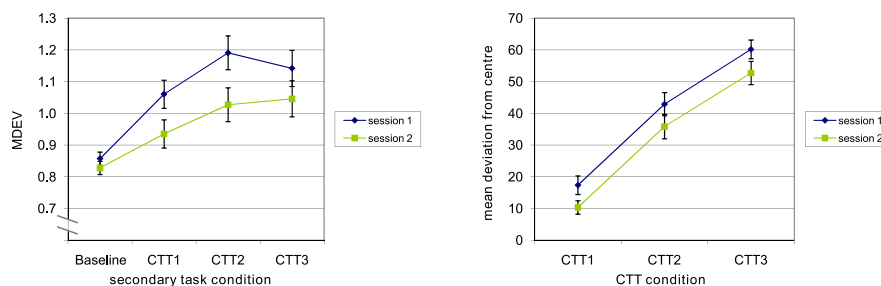
Figure 9. LCT (left) and CTT (right) performance for different testing sessions (error bars indicate standard error).

3.3. Discussion

With this second experiment, we tried to find out whether the effect of training found in our previous experiment is stable over a longer period of time. The results obtained clearly point in this direction. When comparing the two groups in Experiment II, we found small differences in LCT performance, using different metrics. Also the ratings in the DALI seem to support this view. Even more impressive are the results of the within-group comparison for the dual-task training group. Although first and second testing were at least 4 months apart, we still found significant improvements in primary and secondary task performance.

4. General discussion and conclusions

We carried out two experiments to find out if experience has an impact on LCT performance in the short and long run. Our first experiment shows that some form of training might indeed facilitate performance in the LCT as well as a given secondary task. Subjective ratings support this view, as trained participants reported lower levels of demand. This finding is not too surprising, as the purpose of a familiarisation phase for any given task is to provide some sort of training to the participants so they can perform at their best in the experimental phase. What is disturbing is that the familiarisation phase in LCT studies is obviously not sufficient to reach the aspired optimum performance level. Still, this would be a minor problem if this was the case every time a participant took part in an LCT study, so comparability of these studies in general would be ensured. However, it seems that training effects are rather stable, so this optimum performance level is reached much faster in subsequent LCT encounters. Performance is better in the LCT and the secondary task, and reported demand is lower, even after longer periods of time without any exposure to those tasks. Given these results, the occurrence of this training effect might threaten the validity of LCT results.

The implications, however, are not straightforward. The findings could be used to support an "only novice participants" or "only expert participants" experimental design, regardless of any practical limitations that go with either of the two options. Using novice participants reflects a "first encounter" situation of a driver with a certain system. One might think of a rental car scenario, with a driver having to use an unfamiliar navigation system. There are no well-known, automated procedures available, and there is no a priori strategy of attention allocation that optimally fits the task at hand. An LCT assessment with novice participants would be just that, therefore providing insight into the maximum distraction or load a certain system can cause. Using expert participants, on the other hand, reflects the "everyday user"—a driver who knows how to complete a given task in a very efficient fashion, in terms of secondary task completion as well as attention allocation. Such a measurement captures more of an average distraction or load that the system will cause. Both approaches seem useful, as they reflect different use cases. Also, it is quite easy to imagine a system that is very distracting for a novice user but quite easy to handle for an expert, and another system that is easy or hard to operate for both. The difference between novice and expert would therefore be not just a matter of generally higher levels of distraction. One would rather expect some sort of interaction between the level of experience and distraction, making the assessment of both novice and expert drivers an appealing approach.

Some final remarks have to be made regarding the two different secondary tasks that were used in the experiments. It has to be acknowledged that the SuRT, which allows for interruption and easy

resumption, more closely resembles possible in-vehicle tasks such as the operation of a navigation system, than the CTT does. There is hardly a task to be found that requires continuous attention like the CTT does. However, the CTT is the task that the LCT is better able to assess. One might be inclined to conclude that the LCT is not useful in assessing distraction caused by available systems, and that any learning effect found with the LCT + CTT combination probably would not apply to a LCT + regular in-vehicle task setup. Still, it is possible to argue otherwise. Although many in-vehicle tasks do not necessarily require the devotion of continuous attention, it is nevertheless possible to do so. Even though most systems allow for a strategy of interruption and resumption, there is no reason to assume a priori that users will follow such a strategy. Just as experienced and inexperienced users are equally relevant in the assessment of distraction, it is equally relevant and necessary to assess both "optimal" and "suboptimal" use of a system. If there is interest in the maximum amount of distraction a system might elicit, it would even be necessary to instruct the participants to follow a "continuous attention devotion" strategy, as there might be users in the real world doing just that. So, although there might be no system currently available that requires strategies of attention allocation comparable to the CTT, it can still be assumed that there are a substantial number of drivers on the roads who employ such deficient strategies. In this regard, the CTT is a useful secondary task to emulate this scenario, and the LCT is an important tool to assess its distractive impact.

References

Angell, L., Auflick, J., Austria, P. A., Kochhar, D., Tijerina, L., Biever, W., et al. (2006). *Driver workload metrics project: Final report (DOT HS 810 635).*

Bahrick, H., Noble, M., & Fitts, P. (1954). Extra-task performance as a measure of learning a primary task. *Journal of Experimental Psychology*, *48*, 298–302.

Bahrick, H., & Shelly, C. (1959). Time sharing as an index of automatization. *Journal of Experimental Psychology*, *56*, 288–293.

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.

Breuer, J., Bengler, K., Heinrich, C., & Reichelt, W. (2003). Development of advanced driver attention metrics (ADAM). In H. Strasser et al. (Eds.), *Quality of work and products in enterprises of the future* (pp. 37–39). Stuttgart: Ergonomia.

Bruyas, M.-P., Brusque, C., Auriault, A., Tattegrain, H., Aillerie, I., & Duraz, M. (2008). Impairment of lane change performance due to distraction: Effect of experimental contexts. In *Proceedings*

of the European conference on human centred design for intelligent transport systems, Lyon, France, pp. 89–100.

Burns, P. C., Trbovich, P. L., McCurdie, T., & Harbluk, J. L. (2005). Measuring distraction: Task duration and the lane-change test (LCT). In *Proceedings of the Human Factors and Ergonomics Society 49th annual meeting,* Orlando, FL, pp. 1980–1983.

Damos, D. L., & Wickens, C. D. (1980). The identification and transfer of timesharing skills. *Acta Psychologica*, *46*, 15–39.

Ebbinghaus, H. (1971). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie* [Memory: A contribution to experimental psychology]. Darmstadt: Wissenschaftliche Buchgesellschaft. (Original work published 1885)

Engström, J., & Markkula, G. (2007). Effects of visual and cognitive distraction on lane change test performance. In *Proceedings of the 4th international driving symposium on human factors in driver assessment, training, and vehicle design,* Stevenson, WA, pp. 199–205.

Harbluk, J. L., Burns, P. C., Lochner, M., & Trbovich, P. L. (2007). Using the lane-change test (LCT) to assess distraction: Tests of visual-manual and speech-based operation of navigation system interfaces. In *Proceedings of the 4th international driving symposium on human factors in driver assessment, training, and vehicle design,* Stevenson, WA, pp. 16–22.

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.

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.

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.

Jahn, G., Krems, J. F., & Gelau, C. (2009). Skill acquisition while operating in-vehicle information systems: Interface design determines the level of safety-relevant distractions. *Human Factors, 51*, 136–151.

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.

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.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data* (HS 810 59).

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

Lansdown, T. C. (2001). Causes, measures and effects of driver visual workload. In P. A. Hancock & P. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 351–369). Mahwah, NJ: Erlbaum.

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.

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.

New Zealand Ministry of Transport. (2007). *Reported injury crashes 2006.*

Pauzié, A., & Pachiaudi, G. (1997). Subjective evaluation of the mental workload in the driving context. In T. Rothengatter & E. Carbonell (Eds.), *Traffic & transport psychology: Theory and application* (pp. 173–182). Oxford: Pergamon.

Petzoldt, T., Bär, N., & Krems, J. F. (2009). Gender effects on lane change test (LCT) performance. In *Proceedings of the 5th international driving symposium on human factors in driver assessment, training, and vehicle design,* Big Sky, MT, pp. 90–96.

Rognin, L., Alidra, S., Val, C., Lescaut, A., & Chalandon, X. (2007). Driver strategies when interacting with information and comfort systems. In *Proceedings of the 20th international technical conference on the enhanced safety of vehicles,* Lyon, France, Paper number 07-0272.

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.

Shinar, D., Tractinsky, N., & Compton, R. (2005). Effects of practice, age, and task demands on interference from a phone task while driving. *Accident Analysis and Prevention, 37*, 315–326.

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

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

Weir, D. H., Kwok, A. M., & Peak, J. L. (2007). Exploratory lane change test experiments in support of ISO WG8 LCT task force. *Presentation given at the meeting of the Task Force Calibration Task*, Milano, Italy.

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