ALCT - A METHODICAL APPROACH TOWARD EVALUATING THE INFLUENCE OF SECONDARY TASKS DURING AUTOMATED DRIVING

Wolfgang Spiessl, Marion Mangold
BMW Group Research and Technology, Human Machine Interaction
Hanauer Str. 46, 80992 München
Wolfgang.Spiessl@bmw.de, Marion.Y.Mangold@partner.bmw.de

ABSTRACT: In the automotive industry we are seeing an increasing degree of automation. Technically future cars could be able to take over control completely in certain situations if the driver desires. With no immediate control task drivers possibly shift their attention toward secondary activities in the car, e.g. phoning. But naturally automated control systems can fail and leave it to humans to take over and correct. So far there is no method to assess the influence of secondary activities on error recognition and reaction ability in the context of automated driving in such situations. We propose a new methodology to assess this. In an experiment we found promising results in favour of our approach.

1 INTRODUCTION

Automation in today's vehicles has progressed dramatically in the past decade. Advanced driver assistance systems (ADAS) are gradually able to take over more and more driving tasks up to complete control over the vehicle in specific (mostly unpleasant) situations, as driving at low speed in congested stop and go traffic [9]. Despite these fully automated systems still being researched, it becomes clear that the vision of automated or partially autonomous driving is within sight. Surveys show that drivers want to use gained resources for secondary activities in the car, for example phoning or using information services (e.g. [5]), which is permissible during well-performing automation. However, as sophisticated and reliable automated systems might get in the foreseeable future, they will not be 100% error-free, even with expensive high-end technology. There will always be a chance of sensor or actuation failure or unforeseen situations that may result in a potential incorrect action taken by the automated system. In such cases the driver must still be able to intervene and correct the erroneous behaviour of the car. Therefore it is essential to examine the influence of secondary activities during automated driving situations. With no standard tool available to answer this question we created a new method that we called Autonomous Lane Change Test (ALCT).

2 METHOD DESCRIPTION

2.1 Lane Change Test

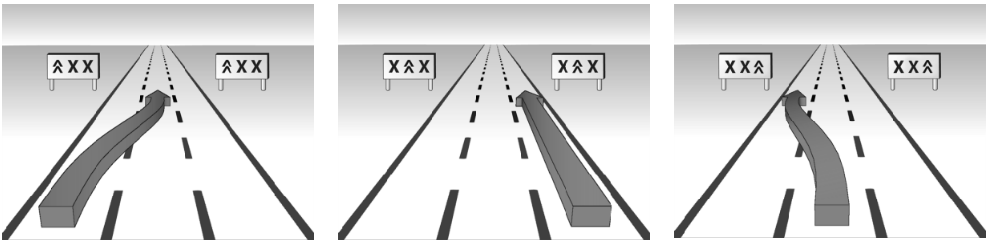
The ALCT is based on the Lane Change Test (LCT, [6], [7]) which is in process of standardisation by the International Organization of Standardization (ISO, [4]). The LCT is designed to evaluate driver distraction through secondary activities during a simulated driving task. The original setup comprises a standard desktop PC and monitor, as well as a gaming steering wheel along with accelerator and brake pedals. A participant drives at 60 km/h on a straight

3-lane-road without traffic and has to change lanes according to traffic signs that are located on both sides of the road. The signs show an arrow next to two X's in order to indicate which of the three lanes the participant should drive in. Signs are always visible; sign contents are invisible at first and become visible when the car is within a certain distance. The driver is instructed to change lanes as fast as possible when the sign contents become visible and also to accurately stay inside the lane. In addition to the driving task the driver is asked to perform an arbitrary secondary task, usually from the automotive context, such as entering a destination into the navigational system, reading texts presented on a display, or phoning. The mean deviation from a normative model indicates the distraction level. According to the dual task paradigm the higher the commitment is in a secondary task the higher the deviation from the normative model [7].

2.2 Autonomous Lane Change Test

The ALCT method is designed as an instrument to assess the influence of secondary tasks during an automated driving scenario. In our approach we adapted the original LCT to the paradigm of automated driving. The low complexity and highly structured virtual driving scene of the LCT provides an excellent basis for eliminating confounding variables.

In a driving simulator setup a car drives within the identical virtual driving scene as in the LCT. When a road sign is approached and the sign contents become visible the car starts to change to the indicated lane autonomously. However, the car may make errors (cf. Fig. 1): it changes when it is not supposed to (error 1), it doesn't change when it is supposed to (error 2), or it changes, but to the incorrect lane (error 3). In these cases the driver must intervene as fast as possible to correct the vehicle's faulty behaviour by turning the steering wheel quickly by at least 90 degrees in the direction of the correct lane, as indicated by the road signs. Errors can always be detected directly at sign appearance as the car will immediately start changing lanes. Changing across two lanes – as is common in the LCT – is not allowed, since then there would not necessarily be a clearly determinable point for when an error can be recognized as such.



a) b) c)

Fig. 1: Error types: a) Lane change when not supposed to (error 1); b) no lane change when supposed to (error 2), and c) change to the wrong lane (error 3)

Buld et al. [2] recommend an overall system reliability of at least 90% in order to establish a sufficient level of automation trust. According to this, 9 out of 10 lane changes in the ALCT setup are performed correctly and 1 out of 10 is incorrect. Lower reliability and higher error rates respectively can lead to less trust in automation and therefore lower attendance to secondary activities along with lower system acceptance. Errors are randomly distributed and can occur with any sign except the first two and the last sign. Each error type occurs exactly three times.

At occurrence of an error the time between sign appearance and error recognition and appropriate reaction is measured. Furthermore the number of errors not corrected within the time span until the next sign appears and the number of times drivers turned the steering wheel without a faulty reaction by the car, is also recorded. Thus response time, missed errors and false reactions are the objective measures indicating the distraction imposed by secondary tasks.

Although the ALCT is systematically based on the LCT there are some clear differences we would like to point out and explain the reasons for their design. The most obvious difference is the overall environment. The LCT is designed as a PC based system; the ALCT is set in a more realistic environment with a car mock-up, a genuine steering wheel and large display of the driving scene presented on three plasma screens (cf. Fig. 2). This is done in order to create deeper involvement with the task and potentially contribute to establishing more trust in the automation. This is also the cause for a longer run time with more signs (and lane changes respectively). With an error rate of 10% there is the need of a certain number of critical events in order to extract sufficient reaction data. Another clear difference is the way of interaction. The LCT is a permanent control and manoeuvre task whereas the ALCT requires automation supervision and only situational correction as it would be the case with potential automated control systems.

Fig. 2: ALCT setup in a driving simulation environment

In the original design road signs are always visible (although in the far distance) and the content becomes visible before passing. We decided to show only the next upcoming sign when reaching a certain distance and display the contents immediately to avoid drivers being able to expect and prepare for an error. Further we introduced the possibility of a straight driving manoeuvre. Here the car simply stays on the previous lane indicated by a corresponding road sign. Thus not changing the lane does not automatically imply an error and must be verified by checking the traffic sign. Major differences are shown in Table. 1:

Table. 1: Comparison of Lane Change Test and Autonomous Lane Change Test

	LCT	ALCT
Environment	PC	driving simulator
Type of interaction	manoeuvre and control	supervision and correction
Duration of interaction	permanent	situational
Track length	3000m	11000m
Test duration	3 min	11 min
Total number of signs	18	90
Signs visible	always	60m before passing
Sign contents visible	40m before passing	60m before passing
Sign distance	144 – 188m	100 – 140m
Possible manoeuvres	6	7
Change across 2 lanes	yes	no
Straight driving	no	yes
Measurements	mean lane deviation, standard lane deviation	mean response time, errors missed, false reactions

The described configuration was found after several explorative pre-studies with different parameter sets varying distance between road signs, point of sign emergence, point of sign visibility, as well as number and distribution of errors. Experimenting with showing only one road sign on a random side of the road –

in order to avoid a fixed area to check for an emerging sign – resulted in strong confusion of the participants. Some participants implicitly assumed that the road side the sign appeared indicated the direction the car had to change to without noticing the content of the sign. Therefore this idea was not pursued any further.

3 EXPERIMENT

The feasibility and validity of the ALCT were examined in a first experiment. The main questions were if the ALCT can discriminate between different secondary tasks and if there is a verifiable main effect of secondary tasks on reaction performance.

We conducted a study with 28 participants (19 male, 9 female). They were 19 to 47 years in age with an average age of 29.36 years (SD = 6.03). All participants were in possession of a driving license for 1 to 21 years with an average of 10.75 years (SD = 5.73).

3.1 Secondary Tasks

For a representative selection of relevant secondary tasks we defined a categorization scheme based on Wickens' model of multiple resources [12]. The theory states that performance decreases when two tasks require similar resources. We included the dimensions 'modality' and 'encoding', and added 'degree of interaction'. Modality means the main channel of information transport (visual vs. auditory), encoding is defined as the way the information is represented (verbal vs. spatial) and the degree of interaction indicates if an active interaction is necessary or information is rather passively consumed. Finally we chose a set of six tasks differing in dimensions that are likely to be performed during driving and exist in today's in-car information systems: entering a destination into the navigational system (visual, verbal, active), selecting a target directly on an interactive map (visual, spatial, active), making a phone call (auditory, verbal, active), listening to an audio book (auditory, verbal, passive), reading a long, unstructured text on a display as well as reading a single line of text on a display (visual, verbal, passive).

3.2 Experimental Design

Each participant performed each of the six tasks described above while driving in the ALCT setup (within-design). A baseline drive without a secondary task was also performed. In total seven conditions were permuted in a way that no condition had the same predecessor nor successor more than once. We instructed participants to keep at least one hand on the steering wheel at all times. During a run they were occupied permanently with a task, e.g. after finishing the input of a navigation destination they were told the next target right away. In order to avoid effects of learning and previous knowledge each task was practised before a run until the test person felt comfortably familiar with it. We presented all visual tasks on a display mounted on the right side next to the steering wheel (cf. Fig. 2). Interactive tasks – except the phone call task – required manual operation of a multifunctional control element (push, shift, rotate) mounted near the arm rest in the centre console. Gazes were recorded using a head-mounted camera-based eye tracking system.

The dependent variable "influence of secondary tasks" was operationalized as the reaction of the participant at the occurrence of an error (mean response time, missed errors, incorrect reactions), subjective measurements (driver activity load index (DALI [8]), distraction ranking, secondary task performance) and gaze behavior (duration of gaze at the display or driving scene, error detection time). Each drive lasted 11mins and comprised 90 signs with an error probability of 10% (as described above). Response time is defined as the interval between the instant of sign appearance and the moment the participant turns the steering wheel over the threshold of 90 degrees, where it includes all steps of the information processing chain (cf. [1], for example).

The phone call was realized by orally asking the participants general knowledge questions. The questions were multiple-choice with four possible answers. Participants were asked to explain why they had chosen a certain answer and not the others. After a run with a 'passive' task (texts and audio book), participants had to answer content-related questions (cf. 3.3.4).

3.3 Results

3.3.1 Mean response time

An analysis of variance (ANOVA) for repeated measures showed a significant main effect of performing secondary tasks on mean response time with $F(3.644, 98.389)=9.545$; $p<.000$ (cf. Fig. 3). Driving without secondary tasks showed the shortest mean response time (1.71 sec.), finding locations on an interactive map resulted in the longest mean response time (2.06 sec.) which means an increase by 20.5%. Performing a pairwise comparison of all conditions we found all tasks except the audio book task differ significantly from the baseline regarding the mean response time. These results indicate that the ALCT is principally able to discriminate regarding different secondary tasks.

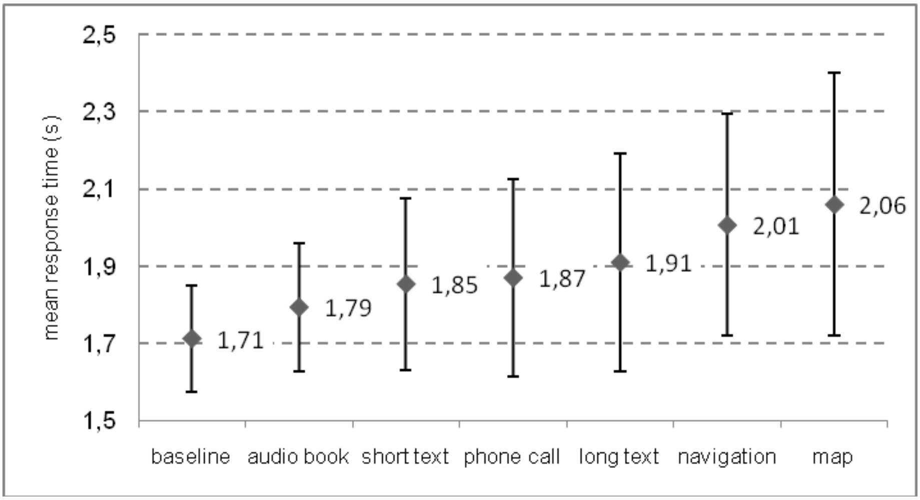


Fig. 3: Mean response times for each condition

We also found interesting results when grouping tasks to the corresponding dimensions (cf. 3.1). A repeated measures ANOVA revealed significantly longer response times for ‘active’ tasks than for ‘passive’ tasks, as well as for both conditions compared with the baseline (F[1.645, 44.417]=26.220; p=.000). Furthermore we found a significant difference between ‘visual’ and ‘auditory’ tasks with t[27]=-3.113; p<.004. ‘Verbal’ and ‘spatial’ tasks had no significant difference (t[27]=-.881; p=.386).

There was no significant difference between male and female participants regarding mean response time (F[1, 26]=4.059; p=.054). Response times grouped by error types (cf. Fig. 1: error 1: M=1.93; SD=0.20, error 2: M=2.09; SD=0.26, error 3: M=1.73; SD=0.19) showed a significant main effect using a repeated measures ANOVA (F[2, 54]=41.263; p= .000). Post-hoc analysis also showed significant differences between all error types regarding response times.

3.3.2 Error data

During the experiment a total number of 64 (M=2.29) errors were missed (i.e. not corrected) with zero missed errors in the audio book condition up to 26 errors in the interactive map (cf. Table. 2) We found a significant main effect of secondary tasks on missed errors (X2[6]=73.414; p=.000). Comparing different dimensions there is a significant summed-up difference between ‘active’ and ‘passive’ tasks (X2[1]=50.104; p=.000) as well as between ‘verbal’ and ‘spatial’ tasks (X2[1]=28.651; p=.000): ‘active’ and ‘spatial’ tasks lead to more missed errors than ‘passive’ and ‘verbal’ tasks. ‘Visual’ and ‘auditory’ tasks did not show a significant difference (X2[1]=1.193; p=.339);

Table. 2: Absolute number of missed errors in the experiment

Condition	audio book	baseline	long text	short text	navigatio n	phone call	map
Missed errors	0	1	1	3	16	17	26

We recorded a total number of 44 (M=1.57) false reactions in the experiment ranging from one in the baseline condition to 19 during the phone call condition (cf. Table. 3). This also resulted in a significant main effect (X2[6]=38.413; p=.000). When comparing different dimensions we found analogously to above that ‘visual’ tasks result in significantly more false reactions than ‘passive’ tasks (X2[1]= 16.555; p=.000). The dimensions ‘modality’ (X2[1]= 3.967; p=.066) and ‘encoding’ (X2[1]=.17; p=.835) did not show significant differences.

Table. 3: Absolute number of false reactions in the experiment

Condition	baseline	long text	audio book	short text	map	navigatio n	phone call
False reactions	1	2	2	5	6	9	19

The results of the examination of the missed errors and false reactions indicate, above all, ‘active’ tasks are cognitively the most demanding ones and interfere most with the supervision of automated driving task.

3.3.3 Gaze data

We recorded gaze data for all ‘visual’ tasks. We divided the field of sight into two areas of interest (AOIs): central information display (CID) and driving scene

(FRONT). An ANOVA showed a main effect regarding total gaze time for both areas ($F_{FRONT}[3, 81]=44.210$; $p=.000$ and $F_{CID}[2.283, 61.650]=53.455$; $p=.000$). Mean gaze times (cf. Fig. 4) on the CID ranged from 2.13s (SD=0.62) in the map condition to 4.24s (SD=1.31) in the long text condition, on FRONT from 0.90s (SD=0.35) in the short text condition to 0.71s (SD=0.19) in the long text condition.

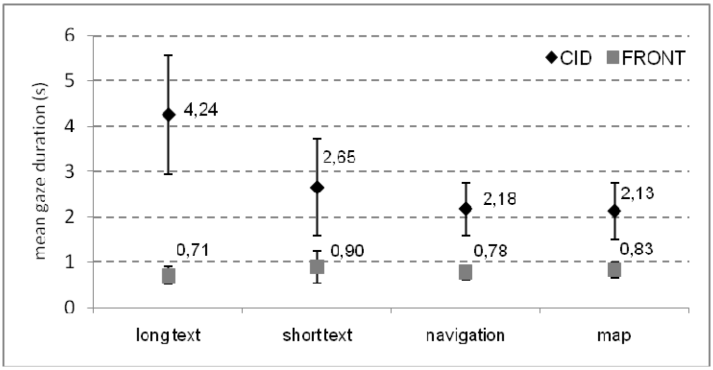


Fig. 4: Mean gaze times on the central information display and the frontal driving scene

In the collected gaze data we were able to identify missed errors seen by the participants (defined by looking at the road signs for over 200ms) but not reacted to. So in the map condition participants missed 26 errors in total (cf. Table. 2) which contained eight errors actually overseen by not watching the road a single time while the road signs were visible and 18 errors that showed at least one gaze to the road signs but resulted in no correcting action. The navigation condition showed 14 of 16 errors that were seen but not reacted to opposing two errors overseen. Both text reading conditions had all errors seen but not reacted to. These results are clear evidence of the so-called out-of-the-loop phenomenon (e.g. [11]) which simply put means both temporal loss of control abilities and lack of situational awareness [13]. This is a common problem with automated systems.

3.3.4 Subjective data

After each task we had participants rate their experienced workload using the DALI questionnaire. In addition, participants had to create a ranking of all six task conditions regarding the experienced difficulty, i.e. ranging from the easiest to the most demanding task, following the experiment. Interestingly we found the exact same order of tasks for both the DALI rating and the overall ranking as we did for response times. This match of objective and subjective results supports the discriminatory power of the presented method and the choice and categorization of tasks.

We also measured the participants' performance on secondary tasks. This is the number of completed navigation and map destinations, as well as read pages of text and number of correctly answered questions in the text and phone condition. We could not find any significant correlations regarding response time, missed errors or false reactions.

4 CONCLUSION AND DISCUSSION

The ALCT incorporates signal detection, supervision and choice reaction rather than control and reaction as in the LCT. In case of an automation error it is crucial to detect the error and react appropriately as fast as possible, rather than maintaining permanent accurate action and control. The chosen setting and variables show plausible results with respect to the investigated subject, i.e. measuring the influence of secondary tasks on detecting and responding to automation errors. It is likely that different methods assessing driving performance (e.g. LCT, Peripheral Detection Task – PDT [10], etc.) would have shown a similar tendency of overall workload imposed by most of the assessed secondary tasks. However, since they do not consider the context of automated driving and use different measurements the results cannot be compared directly. One example may be the phone call task which shows average response times but high numbers of missed errors and false reactions (cf. Table. 2 and Table. 3). These are measurements that are not considered within the original LCT, amongst others. The ALCT can provide detailed insight into how automation supervision and normative-actual value comparison are affected by tasks that are distracting in various aspects.

The ALCT covers three potential automation errors: commission errors (action when not necessary), omission errors (no action when necessary) [3] and incorrect action (action when necessary but incorrectly executed). This is mapped to lane change errors as depicted in Fig. 1. We deliberately included only lateral control errors for the purpose of simplicity. In theory it would also be possible to require longitudinal action in the ALCT, e.g. by introducing a ‘brake sign’. This could be examined in a future experiment.

It could be argued that the driving scene is too simple and that the visual cue of showing road signs which can be anticipated to a certain degree, is too obvious. The results of the experiment do not confirm this: although the driving scene is simple and the road signs are easy to spot, we found strong effects in response times, missed lane changes as well as false reactions. Despite quite clear results, it is not possible to make an absolute statement about the criticality of single tasks during automated driving, as there is no reference to compare to thus far. In order to do this a specific experiment with an implementation of an automated car control system in a real-life driving scenario would be necessary.

Currently, there is a version of the ALCT under development with active haptic feedback through the steering wheel that will move according to the automatic change of lanes, which was not the case in the experiment described above. It will also be possible to intervene directly and take over the task of steering for the time of the lane change in order to correct an error. This will help to understand what effect haptic feedback and temporary direct control lends to error recognition compared to non-haptic feedback and indirect control.

5 REFERENCES

[1] Broadbent, D.: ‘Perception and Communication‘ (Pergamon Press, London, 1958)

[2] Buld, S., Krüger, H.-P., Hoffmann, S., Kaussner, A., Tietze, H., Totzke, I.:

‘Wirkungen von Assistenz und Automation auf Fahrerzustand und Fahrsicherheit‘, Veröffentlichter Abschlussbericht Projekt EMPHASIS: Effort-Management und Performance-Handling in sicherheitsrelevanten Situationen (Würzburg, 2002)

[3] Hollnagel, E.: ‘Looking for errors of omission and commission or The Hunting of the Snark revisited’, ‘Reliability Engineering and System Safety’, 68, 2000, pp. 135–145

[4] ISO/DIS 26022.2 40.20: ‘Road vehicles -- Ergonomic aspects of transport information and control systems -- Simulated lane change test to assess in-vehicle secondary task demand’, 2009

[5] Kompaß, K.; ‘Fahrerassistenzsysteme der Zukunft – auf dem Weg zum autonomen Pkw?‘ In: Schindler, V., Sievers, I. (eds): ‘Forschung für das Auto von Morgen: Aus Tradition entsteht Zukunft‘ (Springer Verlag, 2007), pp. 261–285

[6] Mattes, S., Hallén, A.: ‘Surrogate Distraction Measurement Techniques: The Lane Change Test’ in: Regan, M. A., Lee,J. D., Young, K. L. (eds.): ‘Driver Distraction: Theory, Effects and Mitigation’ (CRC Press, 2009) pp. 107–122

[7] Mattes, S.: ‘The Lane-Change-Task as a Tool for Driver Distraction Evaluation’, in: Strasser, H., Bubb, H. (eds.): ‘Quality of Work and Products in Enterprises of the Future’ (Ergonomia Verlag, 2003), pp. 57–60

[8] Pauzié, A., Pachiaudi, G.: ‘Subjective evaluation of the mental workload in the driving context’ in: Rothengatter, T., Carbonell Vaya, E. (eds.): ‘Traffic & Transport Psychology: Theory and Application’ (Pergamon, 1997), pp. 173–182

[9] Schaller. T., Schiehlen, J., Gradenegger, B.: ‘Stauassistenz – Unterstützung des Fahrers in der Quer- und Längsführung: Systementwicklung und Kundenakzeptanz‘, 3. Tagung Aktive Sicherheit durch Fahrerassistenz, Garching, Germany, April 2008

[10] van Winsum, W., Martens, M. H., Herland, L.: ‘The effects of speech versus tactile driver support messages on workload, driver behaviour and user acceptance’, Report No. TM-99-C043, Soesterberg, Netherlands, TNO Human Factors, 1999

[11] Wickens, C. D., Hollands, J. G.: ‘Signal Detection, Information Theory and Absolute Judgment‘ in: Wickens, C. D., Hollands, J. G. (eds): ‘Engineering Psychology and Human Performance’, (Prentice Hall, London, 2000)

[12] Wickens, C. D.: ‘Attention, Time-Sharing, and Workload’ in: Wickens, C. D. (ed.) ‘Engineering Psychology and Human Performance’ (Harper-Collins Publishers Inc, New York, 1992), pp. 439–479

[13] Wiener, E. L., Curry, R. E.: ‘Flight deck automation: promises and problems’ in: ‘Ergonomics’, 23 (10), 1980, pp. 995–1011