Investigating the Effect of Cognitive Load on UX: A Driving Study

Authors:
Jessica Jung
Andreas Maier
Anne Gross
Natalie Ruiz
Fang Chen
Bo Yin

IESE-Report No. 103.11/E
Version 1.0
December 2011

A publication by Fraunhofer IESE

Fraunhofer IESE is an institute of the Fraunhofer Gesellschaft.
The institute transfers innovative software development techniques, methods and tools into industrial practice, assists companies in building software competencies customized to their needs, and helps them to establish a competitive market position.

Fraunhofer IESE is directed by
Prof. Dr. Dieter Rombach
(Executive Director)
Prof. Dr.-Ing. Peter Liggesmeyer
(Scientific Director)
Fraunhofer-Platz 1
67663 Kaiserslautern
Germany

Abstract

Cognitive load (CL) and user experience (UX) are understood to be critical in the automotive domain, little evaluation has taken place regarding the investigation of the effect of CL on UX in this domain. This position paper introduces a study plan and research goals that aim to investigate whether and how different levels of CL influence UX, using a driving simulator. Besides a discussion of theoretical background and related work, we present initial results regarding an appropriate experiment design and how to manipulate and measure these variables in our specific study context. The paper concludes with a summary and outlook on next steps.

Keywords: automotive, cognitive load, user experience, driving simulator, experimentation, case study

Table of Contents

1 Introduction

The concepts of mental workload, i.e., cognitive load (CL), and user experience (UX) are widely recognized to be critical in the automotive domain. CL in this domain is often tied to attentional resources and safety issues, whereas UX is mainly related to design issues. It has been previously established that high CL influences driving performance [3], but is the driver's perception of their experience with using in-car devices (UX) while driving also influenced by CL? To our knowledge, little evaluation has taken place regarding the investigation of such an effect of CL on UX, especially in the automotive domain.

This position paper introduces our current research aimed at conducting a controlled experiment in a driving simulator to investigate the effect of CL on UX. The main research questions that underlie our work include: (1) is there a relationship between CL and UX (as exemplarily illustrated in Figure 1) (2) If so, how strong is this relationship? (3) Does this relationship evolve or change over time, especially regarding different driving situations (easy, moderate, hard), different user interfaces (UIs) (graphical or tangible Uis), or different input methods (speech, touch) to interact with a device in a car (e.g. navigation system). However, in order to investigate all these research questions, the first challenge to be solved is related to the design of such an experiment, in particular: how to manipulate and measure these "human-oriented" variables CL and UX in a driving simulator.

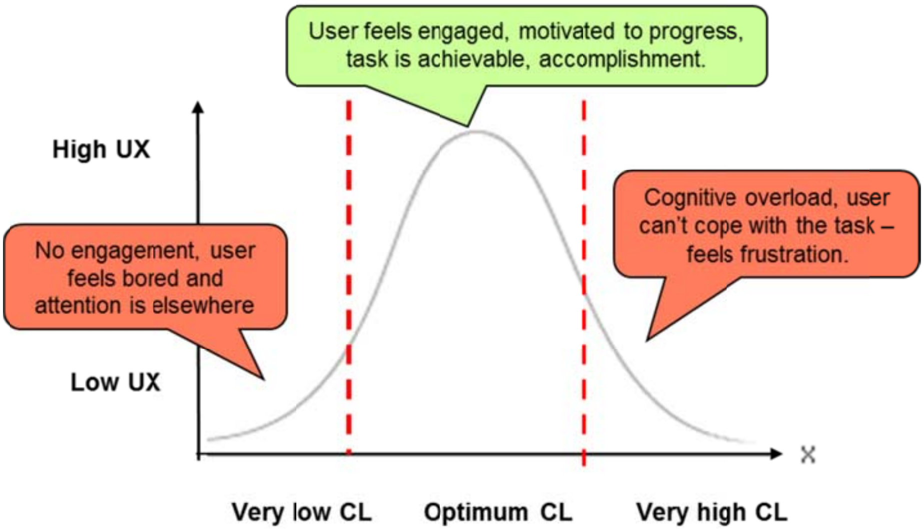

Figure 1: Possible relation between CL and UX

Gaining knowledge about the relationship between CL and UX and suitable measurement and recording techniques for CL and UX could be beneficial in several ways: (1) knowledge about the relationship could be integrated into software engineering approaches, e.g. in forms of guidelines or design decisions for devices which adapt best to a user's CL and still provide a reasonable or high UX. (2) suitable measurements could be integrated into requirements engineering and/or interaction design approaches to derive solutions to lower CL and thus increase UX, (3) the identification of the most appropriate input method, perception modality, and output media according to the level of CL, which influences the UI design could be supported, and (4) CL could be adapted to current driving situations by providing adequate UI designs in order to optimize UX in that situations.

In the following, we describe the theoretical background of both CL and UX in the context of automotive research and related studies in this field. This is followed by a description of current efforts to design such an experiment and select appropriate measures. We conclude this paper with a short summary and outlook on next steps.

2 Background and Related Work

2.1 Cognitive Load

CL is variously described as the level of perceived effort associated with thinking and reasoning by a user, to complete the task successfully [2]. It is well-established that the two main limitations of working memory resources are its capacity and duration [2]. Only a limited number of item “chunks” can be held in working memory at any one time and then only for a short amount of time [5]. These limitations are never more evident than when users undertake complex tasks, or when in the process of learning – resulting in extremely high demands being placed on working memory. The construct of CL refers to the working memory demand induced by a complex task in a particular instance where novel information or novel processing is required [15]. The CL construct comprises at least two separate load sources: intrinsic load and extraneous load. Intrinsic load refers to the inherent complexity of the task; whereas extraneous load refers to the representational complexity – that is, complexity that varies depending on the way the task is presented. In an automotive system, the task complexity would be dictated by the domain. For example, a sample task might be driving on a rainy day at night, while listening to the radio or following GPS instructions. The tools and interfaces the driver employs to complete the task, e.g. a steering wheel with cruise control, a paper based directory, or even Google maps on their iPhone, contribute to extraneous load. Situations that induce high levels of CL can impede efficient performance on designated tasks [15].

Since the capacity of working memory is limited, the user’s CL should be maintained below a safe threshold to prevent task failure [16]. A number of methods have been used, both in HCI and other domains, to estimate experienced levels of CL. Existing methods for measuring CL, used separately and in combination, include (1) subjective measures, such as self-rating scales, (2) physiological techniques, such as heart rate [13] and pupil dilation [4], (3) Task or performance based measures, such as critical error rates; and (4) Behavioral measures, such as linguistic indices [8].

Unfortunately, many of these data acquisition and analysis methods are highly intrusive and can disrupt the normal flow of the task, or are physically uncomfortable for the user. Historically, the most consistent results for CL assessments have been achieved through self-rating subjective measures; however, the assessments can only be made available offline after completion of the task. Among the other types of methods, behavioral methods appear to be the most suitable for practical CL monitoring systems which need accurate, non-intrusive, objective and real-time measures. Speech features can be a particularly good choice within behavioral methods, since speech data already exists in many real-life tasks (e.g. telephone conversation, voice control) and can be easily collected in a non-intrusive and inexpensive way [16]. However, while most of these types of measures are suitable for research purposes, many are unfeasible for automotive applications.

2.2 User Experience in Cars

UX as a concept is well known and widespread in the field of UI design, ranging from software design, website design, mobile devices etc. The definition of UX itself is rather controversial, but it can generally be described as the experiences a user has before, while or after interacting with a product, system, service or object [10]. These experiences include all kinds of subjective perceptions such as emotions and cognitions as well as behaviors. Furthermore, the context in which the product is used influences how the user perceives experiences with the product. To conclude, UX is a multidimensional and complex construct which cannot be assessed with a single item questionnaire. Recently, UX became increasingly important especially in the automotive domain: A comprehensive description of input and output devices in a car with regard to placement and modality was created in [7]. That design space provides a basis for analyzing and discussing different UI arrangements in cars, enables the comparison of alternative UI setups and the identification of new opportunities for interaction and the placement of controls. The newly formed design space comprises considerations about input and output modalities, recommendations for instantiations the usage of different interface objects, and guidelines for the positioning of input and output devices.

A deeper understanding of how drivers concretely interact with today's entertainment, communication, navigation, and information systems in their cars and how they balance those interactions with the primary driving task is gained by General Motors' UX design team's contextual design project "Journey" [6]. One important conclusion drawn from the project was that too often driving was the least important thing in the car. Drivers are distracted by conversations, navigation and entertainment adjustments, addressing needs of passengers, and getting environmental information. Furthermore, the authors claim that UX depends on the brand of the car: expensive vehicles are expected to provide easy to use functionalities and a driver should experience the higher-quality aesthetics compared to those of lower tier vehicles from the same manufacturer.

Another approach for designing experiences while interacting with an advanced driving assistance system is introduced in [9]. It aims at creating positive UX by a story-based design workflow. Hereby, UX focusses on the fulfillment of psychological needs, so psychological needs have to be identified which can be important when interacting with advanced driving assistance systems and systems are designed based on that needs. The workflow can be described as follows: based on results of collecting positive driving experiences that emerged from critical situations (by interviews and PANAS-based questionnaire), UX schemes are identified that integrate several actual experiences and their specific psychological needs. Afterwards, a fictional UX story of a positive experience is created for each UX scheme and a corresponding UX storyboard is created. Finally, design ideas are implemented in a simple driving simulation based on the traffic scenarios described in the UX stories.

Until now, the majority of driving simulator studies have been related to safety issues, e.g., how driving performance is influenced by performing a secondary task, such as using a mobile phone while driving [3], or what kinds of psycho-physiological data could be used as an input for driving assistance devices, e.g. face recognition to detect different states of emotion or brain waves to predict driver intentions etc. However, studies investigating UX in such a driving simulator context are still hard to find. In order to plan and conduct an experimental study which could answer our research questions, we are primarily interested in how experimental variations of CL and data collection for CL as well as UX has been done by other researchers and which measurement methods are recommended.

One study which relates to our study objectives was performed by Liu [11]. The author examined driver performance and subjective ratings on workload for young and elderly drivers under high and low-load driving conditions. Additionally the drivers had to respond with either a navigation or a push button task to a simple or complex advanced traveler information system (ATI), with which information was presented in three conditions: only visually, only aurally or multimodal (both visually and aurally). The driving load conditions resulted by varying driving load factors, e.g. lane width, road type, and traffic density, while information complexity was varied by the number of information chunks presented (little vs. lot) in the ATI. Methodologically, objective measures (e.g. response time, variance in steering wheel position) and subjective measures (system preference and SWAT workload assessment: time stress, visual effort, psychological stress) were assessed. The subjective data collection took place at the midpoint and the end of each scenario (combination of driving condition, task and information system), which means that the driving performance in the driving simulator was abruptly interrupted with filling out a questionnaire for several times, whereas the objective measurement was collected automatically by the driving simulator system itself. The author showed that there were significant differences in the performance measures (e.g. response time, subjective workload) for the different display types: auditory and multimodal displays produced a better driving performance. Also the visual display was a distraction for the driver's attention, which resulted in a less safe driving behavior. The multimodal display of information was preferred by the participants over the two other possibilities.

A general measurement recommendation is given by [12] which can be applied to the automotive domain when considering secondary tasks, which generate distraction and additional CL. They compared concurrent thinking aloud (CTA) with retrospective thinking aloud (RTA) in combination with eye-tracking to analyze a website's usability. They concluded that the use of CTA is reactive in terms of significantly influencing relevant eyetracking parameters, e.g. fixation duration or number of fixations, whereas this does not account for RTA. Also the authors claim, that the use of CTA causes a less natural study situation compared to RTA. RTA on the other side is a more time and resource-consuming technique, but should be used, when natural user behavior is of main research interest.

3 Initial Results

3.1 Experiment Design

Initial discussions towards a suitable design revealed that there were many variables that could be manipulated (like driving task levels, UI designs, driving scenarios, user groups, etc.) which quickly resulted in a design requiring more than 300 participants to discern whether any relationship existed. To handle this complexity we came up with a simpler design in which we first

			Set 1	Set 2	Set 3
A	Task	easy	X	X	X
		medium			
		hard			
B	Driving scenario	baseline (no driving)			X
		driving (normal condition)	X	X	
		driving (hard condition)			
C	UI Design	DesignA	X		X
		DesignB		X	
D	User group	normal	X	X	X

Figure 2: Exemplary set of conditions

define sets of conditions as groups of independent variables (see Fig.2). Furthermore, we add an additional independent variable for CL (such as a secondary task), which will uniformly increase the CL manipulation when present. Based on these variables we will define several designs in which the set is fixed but the CL variable is manipulated and run repeated-measures (within-subjects) studies on each design and measure the UX on the outcomes, i.e., whether there is a change in UX (see Fig.3). Based on such a design we will be able to answer interesting questions as for instance whether the UX varies as CL varies when interacting with a given secondary system (e.g. navigation system).

			Design 1	Design 2	Design 3
A	Set	Set 1	X		
		Set 2		X	
		Set 3			X
B	Cognitive Load (secondary task)	no (Baseline)	X		
		Low	X	X	X
		Medium	X	X	X
		High	X	X	X

Figure 3: Experiment Design

3.2 CL Measurements

Speech is an implicit, non-intrusive data stream that reflects changes in cognitive load – it does not require any overt action by the user (e.g. self-report scales) that may interrupt task flow – hence is suitable for driving tasks. An automatic CL measure from speech in real-time, previously developed by [16] will be used to assess whether the CL manipulation is effective. This means that speech will be required in all conditions.

This measure of CL relies on the classification of speech features using pattern recognition techniques after splitting speech data into discrete levels (e.g. high and low). This strategy correlates each load level to an individual class of speech. Given sufficient training instances, the designed CL levels are modeled by corresponding speech class models, as illustrated in Figure 4. We will also be able to compare the load levels induced by each "UI design set" and the relative differences in CL induced between design sets.

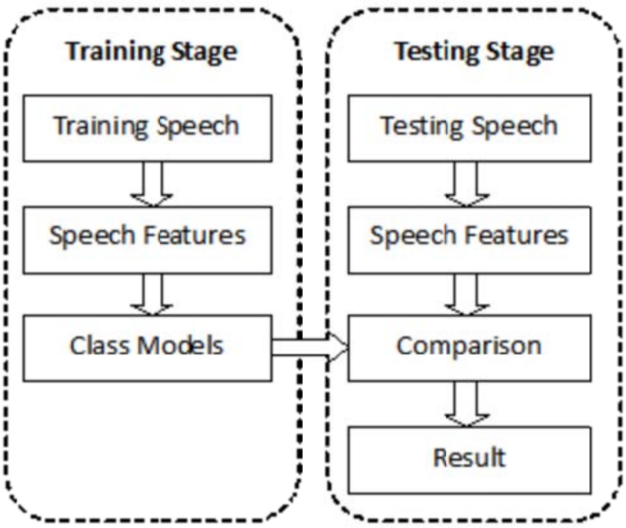

Figure 4: Typical statistical classification system diagram

The classifier works by evaluating each target speech segment such that the likelihood between input speech and models is maximized, so as to determine its CL class [16]. The speech classifier has a number of advantages, such as: (1) automatic CL measurement derived from speech, that can be generated in real-time, (2) automatic model creation from pre-classified speech without manual labelling or analysis, (3) speaker-independent measurement, there is no need to create a model for each individual subject, and (4) novel use of the background speech model to support the solution. Further details of this system can be found in [16].

3.3 UX Measurements

The driving simulator poses some constraints on measurement possibilities. It would be inconvenient for the experiment flow, if a study participant had to interrupt his driving behavior to fill out a questionnaire about the momentary emotional state and impressions right after using a technical device. Also, methods which rely on speech output by the participants such as CTA are not suitable here, because the speak channel is occupied by controlling the inducement of different levels of CL (manipulation check). We are aware of the fact that every driving situation creates CL, which is increased with each secondary task performed in parallel. To minimize the CL created by secondary tasks that are not objects of our inquiry we are not going to let participants think aloud or press any buttons for the conveyance of their current state of UX when driving or let them fill in a questionnaire while or shortly after a single driving task.

A better alternative would be a combination of multiple non-obtrusive measurement options with state-of-the art methods to gather data regarding UX. Hence we basically follow a threefold approach: (1) use of a combination of objective psychophysiological measurements during driving: by measuring electrical indicators and non-electrical signals, conclusions can be drawn on (un-)conscious psychological processes. Examples include variables of the cardiovascular system (e.g. heart rate), emotions (e.g., via skin conductance, face detection), muscular or respiration activity, and eye movements [13], (2) use of standard questionnaires (e.g., SAM, PANAS [1]) right after driving, to measure the overall emotional state and (3) RTA while watching videos of the participant's driving performance combined with displaying the data from psychophysiological measurement after driving as a trigger.

Furthermore, we would like to adapt and use the story-based design workflow ([9] and Section 2.2) for both the selection of a driving situation which may create a positive UX (as reference value) and the RTA of the UX. That is, we will interview participants after the driving tasks, let them fill in PANAS-based questionnaires, and let them describe if their psychological needs have been fulfilled when they had been in the simulated driving situation. At this point, the thinking-aloud method can be applied: when they retrospect the UX story in their minds, the subjects comment every situation they remember. It will not matter if they remember situations that really happened or if situations are observed to be more negative or more positive than the subjects estimate afterwards. UX is an ongoing estimation of stimuli in particular situations and thus can be influenced by many other factors like surroundings, mental states, or similar experiences. What matters with UX is the subjective estimation of situations, not the objective, empirical observation. That is why our approach of evaluating the UX retrospectively can lead to comparative results of our experiment.

4 Summary and next steps

This paper introduced theoretical background and initial results of our ongoing research aiming to investigate whether and how different levels of CL influence UX in a driving simulator. Next steps include to test the suitability of the measurements in smaller case studies (e.g., driving tasks with wii) and prepare experimental material. The experiment itself is planned to be conducted in January 2012.

5 Acknowledgements

This work is done within the collaboration "Fraunhofer Project Centre on Transport and Logistics at NICTA". NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

References

[1] All About UX Blog: UX evaluation methods. Retrieved November 6, 2011: http://www.allaboutux.org/all-methods.

[2] Baddeley, A.D. (1992). Working Memory. Science, 255, 556-559.

[3] Barón, A., and Green, P. (2006). Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review. Technical Report UMTRI 2006-5. Ann Arbor, MI: University of Michigan Transportation Research Institute.

[4] Beatty, J. (1982). Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. Psychological Bulletin. 91(2): p. 276-92.

[5] Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences, 24, 1, 87-114.

[6] Gellatly, A.W., Hansen, C., Highstrom, M., and Weiss, J.P. (2010). Journey: General Motors' move to incorporate contextual design into its next generation of automotive HMI designs. In Proceedings of the 2nd Int'l Conf. on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '10). ACM, New York, NY, USA, 156-161.

[7] Kern, D., and Schmidt, A. (2009). Design space for driver-based automotive user interfaces. In Proceedings of the 1st Int'l Conf. on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '09), 3-10.

[8] Kettebekov, S. (2004). Exploiting prosodic structuring of coverbal gesticulation. In Proceedings of the ICMI'04. PA, USA, pp 105-112.

[9] Knobel, M., Schwarz, F., Palleis, H., and Schumann, J. Towards Designing Experiences for Advanced Driving Assistance Systems. Workshop User Experience in Cars. In conjunction with 13th IFIP TC13 Conf. on HCI (INTERACT '11). ACM, New York, NY, USA, 3-10.

[10] International Standards Organization (2010). ISO 9241-210: Ergonomics of Human-System Interaction – Part 210: Human-Centered Design for Interactive Systems.

[11] Liu, Y.-C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. Ergonomics, 44:4, 425-442.

[12] Jo, Y.-M.M., and Stautmeister, A. (2011). "Don't make me think aloud" – Lautes Denken mit Eyetracking auf dem Prüfstand. In M. Eibl (Ed.) Mensch & Computer 2011. München: Oldenbourg-Verlag. 172-177.

[13] Paas, F., and vanMerrienboer, J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. Educational Psychology Review, 6, 51–71.

[14] Stemmler, G. (2004). Physiological processes during emotion. In P. Philippot, & R.S. Feldman (Eds.). The regulation of emotion. Mahwah, NJ: Erlbaum. 33-70.

[15] Sweller, J., vanMerrienboer, J., and Paas, F. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10, pp. 251-296.

[16] Yin, B., Chen, F., Ruiz, N., and Ambikairajah, E. (2008). Speech-based Cognitive Load Monitoring System. In IEEE Int'l Conf. on Acoustic Speech and Signal Processing (ICASSP'08) Las Vegas: IEEE, pp. 2041-204.

Document Information

Title:	Investigating the Effect of Cognitive Load on UX: A Driving Study
Date:	Dezember, 2011
Report:	IESE-103.11/E
Status:	Final
Classifikation:	Public Unlimited