Towards a Cognitive Load Ready Multimodal Dialogue System for In-Vehicle Human-Machine Interaction

Robert Neßelrath
German Research Center for Artificial Intelligence (DFKI)
66123 Saarbrücken, Germany
robert.nesselrath@dfki.de

Michael Feld
German Research Center for Artificial Intelligence (DFKI)
66123 Saarbrücken, Germany
michael.feld@dfki.de

ABSTRACT

This position paper approaches one of the critical topics in the development of multimodal HMI for the automotive domain: keeping the driver's distraction low. However, the estimation of the cognitive load (CL), of which distraction is one symptom, is difficult and inaccurate. Instead our research indicates that an approach to predict the effect of dialogue and presentation strategies on this is more promising. In this paper we discuss CL in theory and related work, and identify dialogue system components that play a role for monitoring and reducing driver distraction. Subsequently we introduce a dialogue system framework architecture that supports CL prediction and situation-dependent decision making & manipulation of the HMI.

1. INTRODUCTION

Due to the enhancement of on-board electronics in modern cars during recent years, the amount of information the driver receives has been steadily increasing. Traditional car displays and controls like speedometer, light, wiper settings or radio add to the information load that is produced by the actual traffic and environmental context. Nowadays, many modern cars offer much additional information, services and assistance systems for driving, navigation, "infotainment", entertainment and comfort. However, although it is generally accepted that some of this information can be beneficial to increase safety and the driver's comfort, it cannot be denied that the flood of cognitive stimuli harbours the risk of distracting the driver from his primary task, namely to steer the car. Automobile manufacturers face the challenge by developing user interfaces that reduce the effect on CL. This can be achieved by using different interface modalities or adapting the provided dialogue strategies in order to reach a certain goal. Unfortunately there exist only a few patterns and guidelines that support the HMI development process or give an a-priori prediction of the influence of dialogue and information presentation strategies on the driver's workload and the distraction from his primary task. Moreover, the effective load depends on numerous situational parameters that cannot be foreseen, including the driver's mental model and the interplay of stimuli. In this paper, we explore models and strategies for supporting development and evaluation of cognitive load aware multimodal user interfaces.

AutomotiveUI'13, October 27-30, Eindhoven, The Netherlands.
Adjunct Proceedings.

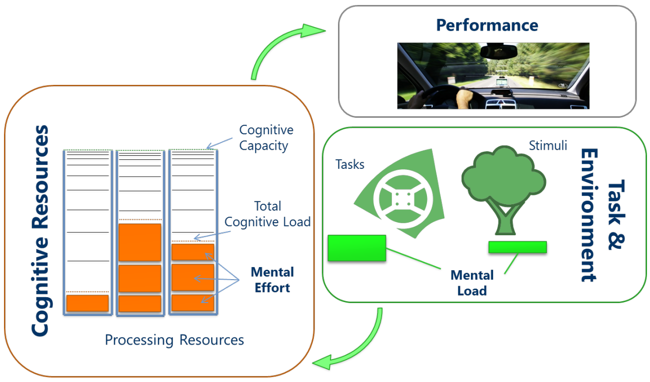

Figure 1: Interplay between mental load, mental effort, performance and CL. Mental load is imposed by stimuli and tasks. The mental effort is the actual allocated amount of CL, that is individual for every user and distributed over different resources. There is a greater interference between two tasks when they share the resources of one category. The overall mental effort for one resource should not exceed the cognitive capacity, since it directly influences the driving performance [19].

2. COGNITIVE LOAD THEORY

In psychology, CL theory addresses the cognitive effort required when learning new tasks. The theory maintains that it is easier to acquire new knowledge and expertise if the kind of learning instruction keeps the CL, and therefore the demand on a user's working memory, low [4][21]. The theory differentiates between three types of CL: *intrinsic load*, *germane load*, and *extraneous load*. The intrinsic load results from an interaction between the amount and type of the material being learned and the expertise of the learner. Extraneous load relates to the manner in which the material being learned is presented. The germane load is needed for processing the learned content and organize it into new schemata or activating existing ones. The three types are additive; together they build the overall load that should not exceed the cognitive capacity limit [18].

Paas and Van Merriënboer [19] describe assessment factors on CL. Figure 1 depicts the simplified interplay between them. The *mental load* is imposed by the task or environmental demands and is constant for a given task in a given environment, independent of a particular user's characteristics. The mental capacity actually allocated is represented by the *mental effort*. It is the outcome of the interaction between the task and the subject's characteristics. Thus, this represents the actual CL on the individual. The quality of the task solution is a third measure, the *performance*. It is influenced by the suspected mental load, the effectively invested mental effort and the individual prior knowledge and experience of the subject.

Current theories for working memory are based on models which consist of multiple independent processors associated with different modes. Baddeley [2] [3] describes the two independent components *visio-spatial sketchpad* and *phonological loop* that are coordinated by a central executive module. The first processes visual input and spatial information, the second stores auditory-verbal information. The *four-dimensional multiple resource model* [22] divides resources into four categories/dimensions, postulating that there is a greater interference between two tasks when they share the resources of one category. The categories are *stages* (perceptual/cognitive vs response), *sensory modalities* (auditory vs visual), *codes* (visual vs spatial) and *channels of visual information* (focal vs ambient) [22].

To sum up, the influence of tasks and cognitive stimuli on the CL is dependent on various factors. These are the task difficulty, the individual experience of the user and the distribution of load among different working memory resources. Finally, also the individual subject can have an active influence on the CL by ignoring information and focusing on a specific task. Furthermore, the working memory theory suggests that a distribution of information presentation on different modalities and the opportunity to solve tasks in a cross-modal way can help to reduce the load on single resources.

3. MMDS COMPONENTS AFFECTING CL

How can the knowledge from theory be exploited for CL awareness in multimodal dialogue systems (MMDS)? We can state that a precise prediction of CL is nearly impossible, since it is dependent on many uncertain factors like the situation, personal experience and even the amount of concentration the driver is willing to invest into a situation. But, theory says that the *mental load* is constant on a given task and independent from the user's characteristics. We want to use this observation as a starting point for CL estimation by finding concepts for the evaluation of dialogue and presentation strategies.
It is not the goal of HMI researchers to explain human cognition in detail. In fact, their research focuses on how presentation and interaction design affect the CL of a user, especially in scenarios in which he controls safety-critical systems like flying an aeroplane, crisis management or steering a vehicle. Some projects treat this question and test strategies for manipulating the CL with changes in interaction design for a multimodal system [15][17]. Related work helped us to identify three components of a multimodal dialogue application that potentially have an influence on the cognitive load:

Multimodal Input & Presentation
The realization of unimodal presentation and the way in which information is presented directly influences the user's attention. [10] analyzed the impact of presentation features like font size and contrast on glance time for a visual display. Presentation complexity on the basis of presentation layout models is predicted in [7]. [6] propose a system design for in-vehicle spoken dialogue complexity management. Other related cognitive research showed that multimodality has a great effect on the CL [16][17].
Hence we assume that the presentation planner is a CL relevant component. Besides the realization of unimodal presentation it coordinates the combination of several modalities (multimodal fusion/fission). Considering the working memory theory postulating that there is a greater interference of tasks if they share the same resource category, presentation planning can keep CL low by selecting modalities with less impact or distributing content on different modalities.

Dialogue Management:
The strategy how to solve tasks in collaboration with the user affects the CL of the user. We demonstrate this by the example of a cinema seat reservation task. In order to successfully reserve a seat, the reservation system needs some relevant information like the movie name, day and time. In a dialogue system the dialogue management is responsible for providing a dialogue strategy that requests this information from the user. The strategies can differ in the amount of information the system collects in a single dialogue turn. One approach is to collect all information at once: A GUI modality would provide a single screen with input elements for all values required ; to use speech dialogue, the system would allow more complex and content-rich utterances. A different approach is to collect the needed information step by step by asking the user in a question-answer-based speech dialogue or by providing multiple GUI windows with lower information density.

Discourse Processing & Context Resolution:
In natural conversations, speakers use referring expressions like anaphora in order to avoid the superfluous effort of rearticulating already established entities. In his *informational load hypothesis*, Almore [1] claims that the noun phrase anaphoric processing optimizes the cost of activating semantic information. Like in the Gricean maxim of quantity [8] a speaker makes a dialogue contribution only as informative as is minimally required. [16] adapts this idea and found out that users communicate more likely multimodally when establishing new content. Following this idea, a component that is responsible for the context resolution of referring expressions and that allows dialogue applications to support multiple forms of referring expressions (e.g. anaphora or deictic expressions) can optimize the CL.

4. MEASURING EFFECTS OF CL

Several measures have been used in psychology and HMI research to estimate the amount of CL. Generally methods can be classified in four categories.

Subjective Measures
A traditional way to assess the subjective workload of a user is *introspection*. The results are acquired by a questionnaire e.g. with the NASA Task Load Index (NASA-TLX) [9]. Because this method is an intrusive procedure and would add an additional task to the CL, it can only be done after the experiment. Beside other scales also in-depth interviews should help to gain more detailed information.

Physiological Measures
One possibility for real-time assessment is to use physiological measures based on the assumption that the subject's cognitive stress is reflected in the human physiology [11]. Physiological indicators that have been used in previous research are heart rate, brain activity, galvanic skin response and eye activity [12, 20] (e.g. blinking or saccadic eye movements).

Performance Measures
Supposing that the performance of task solution is influenced by the CL, conclusions about the latter can be drawn from performance measures. Two performance types can be observed. One is the dialogue task processing performance by considering the amount of time required for solving a task, error rate or type of errors. The other one is the driving performance since the response or reaction time to a stimulus event provides information about the actual CL. An example for this is the Lane Change Test[13], that predicts the level of user distraction by measuring the reaction time of the driver to commands to change lane.

Behavioural Measures
Under high CL users tend to change their interaction behaviour. [5] define *response-based behavioural features* as those that can be extracted from any user activity that is predominantly related to deliberate/voluntary task completion, for example, eye-gaze tracking, mouse pointing and clicking, keyboard usage, use of application, gesture input or any other kind of interactive input used to issue commands to the system. Characteristics of speech, such as pitch, prosody, speech rate and speech energy, can change under high CL. Further features in speech which may indicate cognitive stress are high level of disfluencies, fillers, breaks or mispronunciations.

The different measurement categories involve advantages and disadvantages for the use in a multimodal dialogue system. While subjective measures are not practicable for real-time assessment, physiological sensors are often integrated in cumbersome equipment and it must be guaranteed that the methods are non-intrusive. Furthermore we need concrete models and heuristics in order to map sensor data on concrete CL describing values, to make matters worse a measuring unit for CL does not exist, yet (similar questions arise for behavioural measures). A promising approach is to start with subjective and performance measures that give more concrete conclusions about the driver's distraction and use these findings as evidence for the development and validation of models for the analysis of the two other measures.

5. DESIGN CONSIDERATIONS FOR A CL-AWARE DIALOGUE PLATFORM
Our goal is to create a multimodal dialogue platform that supports state-of-the-art functionalities like multimodal and context fusion, discourse processing and multimodal fission. However, we want to extend this dialogue platform to support research on the estimation of CL and to make it CL-aware. The platform we are building together with an associated development toolkit allows the rapid and flexible creation of new dialogue applications [14]. A great focus is therefore placed on a carefully considered model-based approach and a modular platform architecture with respect to strategies for CL evaluation, estimation and prediction, that allows the easy adaptation or replacement of components. With an adequate development toolkit, the validation of theories and models from cognitive science with live experiments can thus be improved.

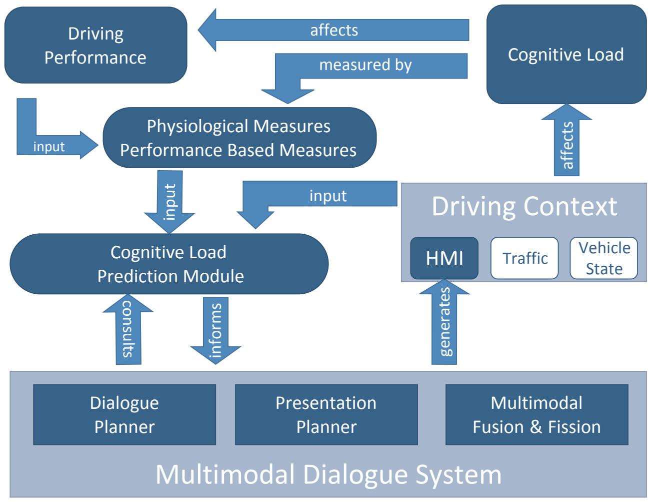

Figure 2: Relevant dialogue system components have a direct influence on the CL and the driving performance. These can be measured and estimated in order to provide situation adapted dialogue and presentation strategies.

Figure 2 shows the concept for our cognitive load aware multimodal dialogue system architecture. The three components mentioned in section 3 (Dialogue Planner, Presentation Planner, Multimodal Fusion/Fission) to a large extent define and generate the human-machine interface (HMI) of the dialogue system that is part of the driving context. This context directly affects the CL of the driver and may have influence on his driving performance. It is possible to integrate arbitrary components for performance and physical measurement. Combined with the driving context they form the source data for a CL prediction module. This module and its algorithms will be adjustable and replaceable for different use cases, theories and measurement methods. Thus, the system will be able to support on the one hand more pragmatic heuristic estimation approaches for use in live applications and on the other hand the evaluation of more complex models from cognitive science.

6. CONCLUSION & OUTLOOK
We designed a multimodal dialogue system platform that allows the rapid development of multimodal applications. We propose to extend and adapt the platform in so far that it is able to support the estimation of the user's current CL. Besides supporting the monitoring of CL the developer should also be able to react to it accordingly by changes in interaction design, e.g. in order to reduce it in subsequent interaction. The following two goals are additionally in focus of our research:

Support for application developers - Results from our studies can be used to find patterns and propose guidelines that help to develop interfaces with a low effect on the CL. Since not every application designer will have adequate experience to apply these in practice, a system that predicts the complexity of an interaction design and supports the application developer in his work will provide a valuable benefit. Thus, during the design process, dialogue platform tools can advise the developer with CL predictions for dialogue and presentation strategies.

Support for situation-adaptive systems - A future goal is to build systems that adapt their communication behaviour with respect to the current context and CL of the driver. For this purpose, our architecture allows the cooperation between the dialogue system and the prediction module in order to plan situation-aware behaviour of the HMI.

7. ACKNOWLEDGMENTS

This work was funded by the German Ministry of Education and Research (project SiAM: Situation-Adaptive Multimodal Interaction for Innovative Mobility Concepts for the Future; grant number 01IW11004).

8. REFERENCES

[1] A. Almor. Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, 106(4):748 – 765, October 1999.

[2] A. Baddeley. Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, 4:829–837, 2003.

[3] A. Baddeley. Working memory: theories, models, and controversies. *Annual review of psychology*, 63:1–29, 2012.

[4] P. Chandler and J. Sweller. Cognitive load theory and the format of instruction. In L. E. Associates, editor, *Cognition and Instruction*, pages 292–332. Taylor & Francis, 1991.

[5] F. Chen, N. Ruiz, E. Choi, J. Epps, M. A. Khawaja, R. Taib, B. Yin, and Y. Wang. Multimodal behavior and interaction as indicators of cognitive load. *ACM Trans. Interact. Intell. Syst.*, 2(4):22:1–22:36, Jan. 2013.

[6] V. Demberg and A. Sayeed. Linguistic cognitive load: implications for automotive UIs. In *Adjunct Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotive UI 2011)*, 2011.

[7] C. Endres, M. Feld, and C. Müller. A layout-based estimation of presentation complexity. In *Adjunct Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2012)*, pages 19–20, Portsmouth, New Hampshire, USA, October 2012.

[8] H. Grice. Logic and conversation. In *Syntax and Semantics*. Academic Press, 1975.

[9] S. G. Hart and L. E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Human mental workload*, 1(3):139–183, 1988.

[10] D. Imbeau, W. W. Wierwille, and Y. Beauchamp. Age, display design and driving performance. *Automotive ergonomics*, pages 339–357, 1993.

[11] A. E. Kramer. *Physiological Metrics of Mental Workload: A Review of Recent Progress*. Defense Technical Information Center, 1990.

[12] S. P. Marshall. The index of cognitive activity: Measuring cognitive workload. In *In Proceedings of the 7th IEEE Conference on Human Factors and Power Plants*, 2002.

[13] S. Mattes and A. Hallén. *Driver distraction: Theory, Effects and Mitigation*, chapter Surrogate Distraction Measurement Techniques: The Lane Change Test, pages 107–122. CRC Press, 2009.

[14] R. Nesselrath and D. Porta. Rapid development of multimodal dialogue applications with semantic models. In *Proceedings of the 7th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems (KRPD-11). Twenty-Second International Joint Conference On Artificial Intelligence (IJCAI -11)*, Barcelona, Spain, July 2011.

[15] A. Niculescu, Y. Cao, and A. Nijholt. Manipulating stress and cognitive load in conversational interactions with a multimodal system for crisis management support. In A. Esposito, N. Campbell, C. Vogel, A. Hussain, and A. Nijholt, editors, *Development of Multimodal Interfaces: Active Listening and Synchrony*, Lecture Notes in Computer Science 5967, pages 134–147, Heidelberg, 2010.

[16] S. Oviatt, R. Coulston, and R. Lunsford. When do we interact multimodally?: cognitive load and multimodal communication patterns. In *Proceedings of the 6th international conference on Multimodal interfaces*, ICMI '04, pages 129–136, New York, NY, USA, 2004. ACM.

[17] S. Oviatt, R. Coulston, S. Tomko, B. Xiao, R. Lunsford, M. Wesson, and L. Carmichael. Toward a theory of organized multimodal integration patterns during human-computer interaction. In *Proceedings of the 5th international conference on Multimodal interfaces*, ICMI '03, pages 44–51, New York, NY, USA, 2003. ACM.

[18] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. V. Gerven. Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1):63–71, 2003.

[19] F. G. Paas and J. J. Merriënboer. Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6(4):351–371, 1994.

[20] M. Schwalm, A. Keinath, and H. D. Zimmer. Pupillometry as a method for measuring mental workload within a simulated driving task. *Human Factors for assistance and automation*, pages 1–13, 2008.

[21] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285, 1988.

[22] C. D. Wickens. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2):159–177, 2002.