

Dialogue behaviour under high cognitive load

Jessica Villing

Graduate School of Language Technology and
Department of Philosophy, Linguistics and Theory of Science
jessica@ling.gu.se

Abstract

For safety reasons, in-vehicle dialogue systems should be able to take the cognitive load of the driver into consideration. However, it is important to distinguish between two types of cognitive load, namely if the cognitive load is affecting the driving behaviour or not. We will present findings from a data collection carried out in the DICO project¹, where the dialogue behaviour under high cognitive load is analysed, and present a novel theory of how to distinguish between different types of workload.

1 Introduction

In-vehicle dialogue systems demand dialogue management that takes the cognitive workload of the driver into consideration. The primary task is the driving, and therefore it is necessary to develop a dialogue system that interferes as little as possible with the driving task. However, the driver's cognitive workload might increase for various reasons, and it is important to distinguish between workload that is driving-induced (i.e. due to, for example, a heavy traffic situation) and workload that is dialogue-induced (i.e. due to a complicated dialogue). If the workload is driving-induced it is probably necessary to pause the dialogue to enable for the driver to concentrate on the driving task, whereas if the workload is dialogue-induced it is instead necessary to facilitate the dialogue task, for example by reformulating a question.

¹www.dicoproject.org

2 Data collection

DICO is a project that aims to develop a proof-of-concept demo system, showing how a spoken dialogue system can be an aid for drivers. To study how an additional distraction or increase in the cognitive load would affect a driver's dialogue behaviour, a data collection has been made. The goal was to elicit a natural dialogue (as opposed to giving the driver a constructed task such as for example a math task) and make the participants engage in the conversation.

The participants (two female and six male) between the ages of 25 and 36, drove a car in pairs while interviewing each other. The interview questions and the driving instructions were given to the passenger, hence the driver knew neither what questions to discuss nor the route in advance. Therefore, the driver had to signal, implicit or explicit, when she wanted driving instructions and when she wanted a new question to discuss. The passenger too had to have a strategy for when to change topic. The reasons for this setup was to elicit a natural and fairly intense dialogue and to force the participants to change topic and/or domain (e.g. to get driving instructions).

The participants changed roles after 30 minutes, which meant that each participant acted both as driver and as passenger. The cognitive load of the driver was measured in two ways. The driver performed a Tactile Detection Task (TDT (van Winsum et al., 1999))², and workload was also measured by using an IDIS system³.

²When using a TDT, a summer is attached to the driver's wrist. The driver is told to push a button each time the summer is activated. Cognitive load is determined by measuring hit-rate and reaction time.

³IDIS determines workload based on the driver's behaviour (for example steering wheel

The participants were audio- and videotaped, and then transcribed with the transcription tool ELAN⁴, using an orthographic transcription. The annotation scheme was designed to enable analysis of utterances with respect to topic change for each domain.

Domain and topic was defined as:

- *interview* domain: discussions about the interview questions where each interview question was defined as a topic
- *navigation* domain: navigation-related discussions where each navigation instruction belonging to the same row in the given route was defined as a topic
- *traffic* domain: discussions about the traffic situation and fellow road-users where each comment not belonging to a previous event was defined as a topic
- *other* domain: anything that does not fit within the above domains where each comment not belonging to a previous event was defined as a topic

Topics change has been coded as follows:

- *begin-topic*: whatever → topic A (new)
- *end-topic*: topic A (finished) → whatever
- *interrupt-topic*: topic A (unfinished) → whatever
- *resume-topic*: whatever → topic A (unfinished)
- *reraise-topic*: whatever → topic A (finished)

Cognitive load has been annotated as:

- **reliable workload**: annotated when workload is reliably high according to the TDT (reliability was low if response button was pressed more than 2 times after the event).
- **high**: high workload according to IDIS
- **low**: low workload according to IDIS

movements or driver applying the brake). See <http://www.roadsafe.com/news/article.aspx?article=210>

⁴<http://www.lat-mpi.eu/tools/elan/>

Silence, regardless of length, has been coded as a pause.

The annotation schema has not been tested for inter-coder reliability. While full reliability testing would have further strengthened the results, we believe that our results are still useful as a basis for future implementation and experimental work.

2.1 High workload

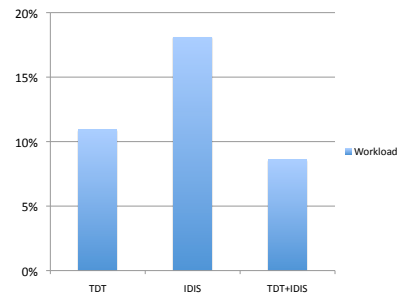


Figure 1: Percentage of annotated workload time.

Figure 1 shows workload measured uniquely by IDIS, uniquely by the TDT and annotations made by IDIS and TDT jointly.

The difference in annotation time can be explained by the fact that IDIS analyses driving behaviour while the TDT measures the driver's reaction time. IDIS is developed to decide when it is suitable to show alarms that are non-critical (such as the indicator for low level of wind screen washer fluid). Since showing the alarm is not time critical, IDIS does not measure the individual driver's workload directly. Taking this into consideration, IDIS measurements alone might be too general and approximate when it comes to adapting a dialogue system to the driver's cognitive load. However, neither IDIS nor TDT in isolation say anything about *what* is causing the high cognitive load, only that something makes the driver unable to pay full attention to the task at hand. These differences can be used to decide what type of workload the driver is experiencing, which will be explained next.

3 Workload management

To determine *type* of workload, the dialogue manager could be extended with a *Cognitive Load Manager (CLM)* which has access to two workload detectors, a *Speech Analyser (SA)* and a *Vehicle State Analyser (VSA)*, see figure 2.

Since the driver is talking to a dialogue system the most convenient method for determining

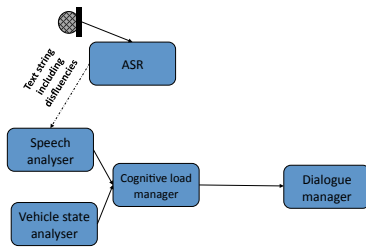


Figure 2: Architecture of the cognitive load manager.

workload level would be to analyse the speech. Studies have for example shown that an increased number of disfluencies such as deletions can indicate increased workload (Shriberg, 2001; Lindstrom et al., 2008). The driver might also make sudden changes of domain, e.g. talk as if addressing fellow road-users, to indicate that she is busy sorting out a difficult traffic situation (Villing et al., 2008). There are no commercial SA systems present today, however research has shown that it is possible to detect workload by analysing the speech signal (Yin et al., 2008).

The VSA analyses the driving behaviour to find signs of increased workload. Variants of VSA-like modules are a reality in the vehicle industry today. For example, if the driver puts the brake on, makes a left turn or manages the radio or the fan, it is assumed that the workload is high.

The CLM collects data from the detectors and determines type of workload based on the combined signals from the SA and the VSA. Type of workload can be set to *driving-induced* (workload that is affecting the driving performance, detected by the VSA) or *dialogue-induced* (workload that is not affecting the driving performance, detected by the SA) based on four assumptions, shown in Table 1.

4 Results

4.1 High workload annotations

Figure 3 shows the number of instances of high workload detected by IDIS alone (possibly driving-induced), by TDT alone (dialogue-induced) and by both IDIS and TDT jointly (driving-induced) for each domain. The TDT makes most annotations in the *other* and *interview* domains and fewest in the *traffic* domain, while the TDT and IDIS jointly makes most annotations in the *traffic* and *other* domains and fewest in the

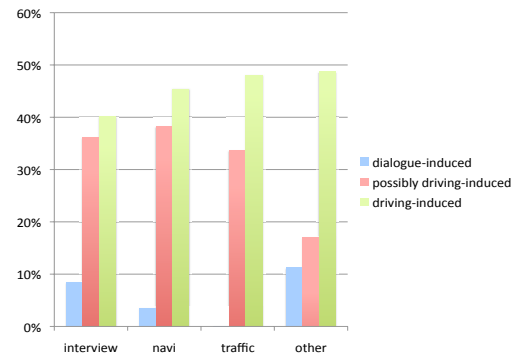


Figure 3: High workload measured for each domain.

interview domain.

To make the SA more powerful, we wanted to investigate if an analysis of dialogue behaviour might improve the possibility to determine workload level. The most frequent topic changes are shown in Figure 4. Most interview related topics are discussed during dialogue-induced workload, while traffic related topics are discussed during driving-induced workload. During possibly driving-induced workload the topics are fairly equally spread. These results are further discussed in (Villing, 2009).

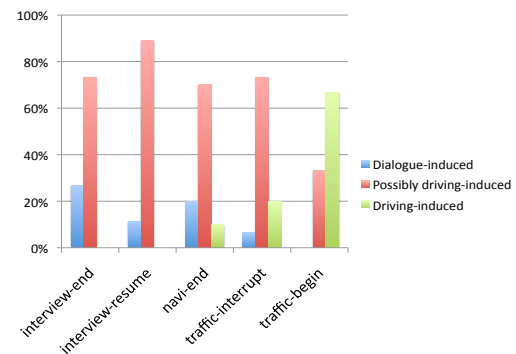


Figure 4: Topic shift during high workload.

Figure 5 shows the average duration of the driver's pauses.

Figure 6 shows that the majority of driver utterances are produced during low workload.

5 Discussion

Figure 5 and 6 shows that an analysis of the speech can give clues about workload level. The duration of the pauses is increasing during high workload, and especially during driving-induced workload. This supports our hypothesis that the dialogue system should pause when the driver needs to concentrate on the driving task. This trend can also be derived from Figure 6, since the number

	SA + VSA	SA	VSA
driver speaking	driving-induced	dialogue-induced	false alarm
driver not speaking	-	-	possibly driving-induced

Table 1: CLM output based on information from the SA and the VSA.

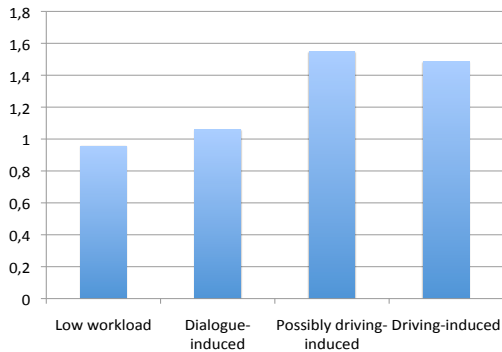


Figure 5: Average pause duration for the driver (in seconds).

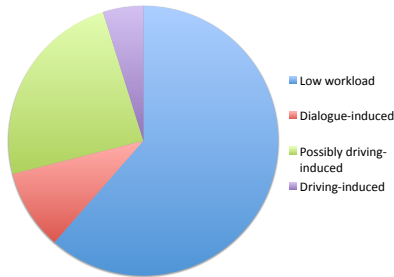


Figure 6: Distribution of driver utterances during low and high workload.

of utterances are decreasing dramatically under high workload when comparing with low workload. The driver seems to make fewest utterances during driving-induced workload.

Looking at Figure 3, it seems like the VSA-like systems present today would benefit from cooperating with a system that is able to make a deeper analyse of the cognitive load of the driver. For example, the *traffic* domain holds almost no dialogue-induced workload annotations but second most driving-induced, supporting the theory that people often make comments about the traffic situation to signal that they have to concentrate on the driving task. The results, although tentative, can be seen as an indication that it is possible to distinguish between different types of cognitive load by analysing both driving behaviour and speech, and that different types of workload demand different dialogue strategies.

6 Future work

Next we will analyse the DICO material regarding interruptions, to find a relevant *interruption place* in the dialogue, i.e. a place where it is most suitable to pause in order to disturb the driver as little as possible.

The resumption behaviour will also be analysed to see who takes the initiative to resume the dialogue and how it is done. The findings will form a basis for a theory of in-vehicle dialogue management.

References

- Anders Lindstrom, Jessica Villing, Staffan Larsson, Alexander Seward, Nina Aberg, and Cecilia Holtelius. 2008. The effect of cognitive load on disfluencies during in-vehicle spoken dialogue. In *Proceedings of Interspeech 2008*.
- Elisabeth Shriberg. 2001. To "errrr" is human: ecology and acoustics of speech and disfluencies. *Journal of the International Phonetic Association*, 31:153–169.
- W van Winsum, M Martens, and L Herland. 1999. The effect of speech versus tactile driver support messages on workload, driver behaviour and user acceptance. tno-report tm-99-c043. Technical report, Soesterberg, Netherlands.
- Jessica Villing, Cecilia Holtelius, Staffan Larsson, Anders Lindstrom, Alexander Seward, and Nina Aberg. 2008. Interruption, resumption and domain switching in in-vehicle dialogue. In Bengt Nordstrom and Aarne Ranta, editors, *Proceedings of GoTAL, 6th International Conference of Advances in Natural Language Processing*, volume 5221, pages 488–499, August.
- Jessica Villing. 2009. In-vehicle dialogue management - towards distinguishing between different types of workload. In *Proceedings of SiMPE, Fourth Workshop on Speech in Mobile and Pervasive Environments (to appear)*.
- Bo Yin, N. Ruiz, Fang Chen, and E. Embikairajah. 2008. Investigating speech features and automatic measurement of cognitive load. In *Proceedings of 2008 IEEE 10th Workshop on Multimedia Signal Processing*, pages 988–993.